# Mapping atherogenesis mechanisms in smooth muscle cells by targeting genes linked to coronary artery disease

Julián Albarrán-Juárez[1*], Anton Markov[1], Anne Louise Jensen[2], Peter Loof Møller[2,3], Anna K. Uryga[4], Djordje Djordjevic[5], Jakob Hansen[1], Lise Filt Jensen[1], Diana Sharysh[1], Charles Pyke[5], Jaime Moreno[5], Giulia Borghetti[4], Julian Bachmann[4], Kate Herum[4], Lisa Maria Røge[1], Matthew Traylor[6], Michael Nyberg[4], Mette Nyegaard[3], Jacob Fog Bentzon[1,7,8].

**Affiliations:**

1. Department of Clinical Medicine, Aarhus University, Aarhus, Denmark.
2. Department of Biomedicine, Aarhus University, Aarhus, Denmark.
3. Department of Health Science and Technology, Aalborg University, Aalborg, Denmark.
4. Department of Cardiovascular Biology, Research & Early Development, Novo Nordisk A/S, Copenhagen, Denmark.
5. Department of Computational Biology, Novo Nordisk A/S, Copenhagen, Denmark.
6. Department of Genetics, Novo Nordisk A/S, Research Centre Oxford, UK.
7. Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain.
8. Steno Diabetes Center, Aarhus, Denmark.

***Address for correspondence**

Julián Albarrán-Juárez
Atherosclerosis Research Unit,
Department of Clinical Medicine, Aarhus University
A503, Palle Juul-Jensens Boulevard 11,
8200 Aarhus N
Denmark
Email: jalbarran@clin.au.dk

**Abstract**

Recent genome-wide association studies (GWAS) have identified multiple vascular cell-expressed genes linked to coronary artery disease (CAD), suggesting that smooth muscle cells (SMCs) and SMC-derived metaplastic cells are promising targets for novel antiatherosclerosis therapies. However, the disease-promoting pathways of most GWAS-identified genes are unknown, hindering their translation into therapeutic targets. This study integrated public GWAS data for CAD and single-cell RNA sequencing (scRNA-seq) analyses of human atherosclerotic plaques to identify 20 GWAS risk genes with a putative mechanism of action in SMCs or SMC-derived cells.

Gene perturbation experiments in SMCs coaxed to plaque-relevant phenotypes revealed that the selected risk genes, despite encoding very different types of proteins, regulated shared sets of genes associated with contractile functions, cell cycle pathways, NFκB, and type I interferon signaling. By integrating information about GWAS gene effect direction and a deep analysis of cholesterol- and stretch-induced gene modules in SMCs, we find evidence that cholesterol-induced signaling is a pro-atherogenic disease mechanism in SMCs that is upregulated by detrimental and downregulated by protective GWAS genes.

Overall, our study identifies a set of candidate disease mechanisms in SMCs that are regulated by multiple GWAS genes across several SMC assays. Furthermore, it provides proof-of-concept for using GWAS gene effect directionality to predict the pathogenic effect of candidate disease mechanisms that can be extended to other GWAS genes and cell types in the future.

**Keywords**

Smooth muscle cells, coronary artery disease, atherosclerosis, genome-wide association studies, single-cell RNA sequencing, gene networks, signaling pathways.

## Background

Atherosclerotic cardiovascular disease encompasses myocardial infarction (MI), ischemic stroke, and peripheral artery disease and is the leading cause of death and disability in the world [1,2]. Atherosclerosis is caused by high levels of low-density lipoproteins (LDLs) and other risk factors (e.g., hypertension and diabetes), which induce inflammation, activation of local arterial smooth muscle cells (SMCs), and the buildup of fibrous, calcified, and necrotic tissue in the arterial intima. After decades of silent development, plaques may suddenly precipitate thrombosis, leading to clinical disease [3]. Lifestyle changes and pharmacotherapies that halt disease progression by lowering the causal risk factors for atherosclerosis are the central therapies. Despite the availability of these effective interventions, a significant residual risk remains, thus highlighting the need for new approaches that target novel mechanisms.

Vascular SMCs are promising cell targets for new therapies that can work orthogonally to risk factor reduction. Recent murine lineage tracing and human single-cell RNA sequencing (scRNA-seq) studies have shown that local SMCs lose their quiescent contractile phenotype, undergo proliferation and modulate to abundant fibroblast-like and osteochondrocyte-like cells during atherogenesis [4–8]. Many genes linked to coronary artery disease (CAD) and MI in genome-wide association studies (GWAS) are expressed in SMCs or SMC-derived cells [9,10]. This indicates that the expansion, modulation, and function of these cells are critical determinants of the outcome of human atherosclerosis [11–13]. The disease-promoting pathways in SMCs are, however, poorly understood, and this impedes their use as targets for drug development.

In this study, we combined public GWAS for CAD and scRNA-seq plaque data to identify risk genes with a putative mechanism of action in SMCs or SMC-derived cells. By systematically knocking down 20 risk genes that encode very different types of proteins, we performed an unbiased analysis of SMC pathways that are genetically linked to human atherosclerosis, and therefore, they represent potential targets for SMC-directed therapies.

## Results

### Identification of GWAS genes with a potential mechanism of action in SMCs

A crude list of 242 genomic variants associated with CAD independently of LDL cholesterol was drawn from the Cardiovascular Disease Knowledge Portal (CVDKP) and FinnGen study (Freeze 5) **(Figure 1A** and **Supplementary Table 1).** To identify candidate GWAS genes with a potential function in plaque SMCs or SMC-derived cells, we looked up the expression of 368 genes predicted to be regulated by the variants in human plaque scRNA-seq data, assembled by integrating published data sets of human coronary [5] and carotid [6,7] atherosclerosis **(Figure 1B)**. The integrated dataset comprised 50,390 cells after quality control and filtering, including a large mesenchymal supercluster (15,018 cells) of cells with SMC, pericyte, transitional, fibromyocyte, osteochondrogenic, fibroblast, or undefined phenotype based on marker gene expression **(Figure 1C** and **Supplementary Table 2).** Of the 368 examined genes, 92 protein-coding genes showed enriched expression in the mesenchymal supercluster compared with other cell types **(Figure 1D and Supplementary Table 1)**. From this pool, we selected 17 genes (*ADAMTS7*, *C1S*, *COL4A2*, *CRISPLD2*, *CTTN*, *DSTN*, *FHL1*, *GEM*, *HHIPL1*, *LMOD1*, *LOXL1*, *LRP1*, *MFGE8*, *PDGFD*, *PLPP3*, *TIMP2*, and *TNS1*), chosen to represent a wide range of putative functions **(highlighted in Figure 1D)**. The list included genes with strong expression in mesenchymal subclusters (e.g., *C1S, DSTN, MFGE8*), genes with experimental evidence of functions in SMCs (e.g., *LMOD1, PDGFD*)[14–16], and genes encoding both extracellular (*PDGFD, C1S, COL4A2, HHIPL1, LOXL1, ADAMTS7, MFGE8, CRISPLD2, TIMP2*) and intracellular (*DSTN, CTTN, FHL1, ALDH2, LMOD1, GEM, RBPMS2, PLPP3, TNS1*, *LRP1)* proteins, according to the human protein atlas [17]. We added three CAD GWAS genes (*ALDH2, TCF21, RBPMS2*) with published evidence for a function in SMCs [5,18,19]. Lead genomic variants and plaque single-cell expression profiles for the 20 selected genes are shown in **Supplementary Table 3** and **Supplementary Figure 1**, respectively.

Expression of the selected target genes was also investigated in a spatial transcriptomics data set of human coronary plaques [20,21]. *MFGE8, TIMP2, TNS1, C1S, COL4A2, DSTN, LMOD1,* and *LRP1* were abundantly expressed in coronary artery plaque and underlying media. *ALDH2, CTTN, GEM, LOXL1, PLPP3*, and *RBPMS2* expression were restricted mainly to the media, while *CRISPLD2, TCF21*, *HHIPL1,* and *ADAMTS7* did not have sufficient detectable expression to determine the localization (**Supplementary Figure 2**).

### Establishment of cellular assays to evaluate human smooth muscle cell function

To test gene function in several SMC phenotypes that are relevant in atherosclerotic plaques, we coaxed human aortic SMCs into three different cell states. The underlying rationale was that target genes that might drive important functions in one context may be inactive in

another. The baseline condition was a standard SMC culture, where cells spontaneously modulate, losing contractile proteins, increase proliferation, and display a less elongated morphology [22,23]. Secondly, SMCs were overloaded with cholesterol to drive cholesteryl ester droplet formation in SMCs [24], which leads to a foam cell-like state characterized by increased inflammatory signaling [25,26]. Furthermore, mechanical stretching (10% elongation) was applied to mimic the physiological mechanical forces of the arterial wall and potentially drive SMCs toward a more contractile phenotype [27]. RNA-seq analysis confirmed that 72h of cholesterol overloading in SMCs upregulated inflammatory genes (e.g., *CCL20, CXCL2, CXCL3, CXCL5, IL1B*), cholesterol efflux genes (*ABCA1* and *ABCG1*) and downregulated contractile genes (e.g., *ACTA2, CNN1, LMOD1, MYOCD*, and *TAGLN*) compared with cells cultured under baseline conditions **(Figure 2A)**. In contrast, 6h of physiological mechanical stretching downregulated several inflammatory genes (e.g., *CCL2, CXCL1, CXCL3, IL1B, IL6*) and upregulated contractile genes (e.g., *CNN1, ACTA2, TAGLN, ACTG2*) compared with SMCs cultured similarly under static conditions **(Figure 2B)**.

**Knockdown of genes associated with coronary artery disease in human SMCs**

To analyze the function of the selected CAD risk genes in human SMCs, we knocked down each target gene by custom-designed small interfering RNAs (siRNAs) and subjected cells to baseline, cholesterol overload, or mechanical stretch conditions. Effects on gene expression were analyzed by RNA sequencing using a nested design where 2-3 target gene knockdowns were run concurrently with control siRNA-treated cells to mitigate batch effects **(Figure 2C)**. Overall, we analyzed the transcriptome of 345 samples (3 to 4 biological replicates per target gene and condition). The numbers of differentially expressed genes (DEGs) compared with controls were higher under baseline conditions (from 159 to 3266 DEGs, 884 at average) and cholesterol loading (from 107 to 3122 DEGs, 1078 at average) than after mechanical stretch (from 12 to 1612 DEGs, 592 at average). Principal component analysis (PCA) of total RNA sequencing results confirmed the transcriptional changes across the baseline, cholesterol overloading, and mechanical stretch conditions **(Figure 2D).** It also revealed considerable variation within each condition among the control cells, underscoring the importance of the nested design with matched controls.

The knockdown efficiency of the target genes was analyzed by real-time qPCR (qPCR), as shown in **Supplementary Figure 3.** Gene knockdown analysis was efficient for most target genes, decreasing the level of target gene expression by 75 to 90% compared to control samples. However, for *TCF21* and *PLPP3,* the designed siRNAs reduced the expression by less than 50%. Therefore, *TCF21* and *PLPP3* were not further analyzed in the study. RNA-seq data confirmed the knockdown efficiency and interestingly showed mutual regulation among the target genes (**Supplementary Figure 4**). For example, the knockdown of *LMOD1*

significantly upregulated the expression of *GEM, CTTN, and TNS1* and downregulated the expression of *ALDH2, COL4A2, CRISPLD2, FHL1, PDGFD, and TIMP2* in baseline and cholesterol overloading conditions.

**Regulation of smooth muscle cell pathways after target gene knockdown**

Gene set enrichment analysis (GSEA) for all comparisons in this study showed that cell cycle, interferon signaling, cytokines, and Rho GTPases were among the top gene ontology categories and biological pathways regulated after target gene knockdowns (**Supplementary Table 4)**. For a more direct analysis of pathways and genes relevant to SMC functions, we analyzed tailored gene signatures representing SMC contraction, cytoskeleton, cell adhesion, extracellular matrix organization, cell cycle, and inflammatory/immune responses, collected from several sources, including KEGG [28], Reactome [29], and WikiPathways [30] **(Supplementary Table 5).** Gene sets involved in SMC contraction, extracellular matrix synthesis, and cell cycle were regulated by multiple target genes, and we also identified a strong regulation of the nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) and interferon pathways **(Figure 3A-C)**.

For a subset of target genes (e.g., *ADAMTS7, CTTN, GEM, LMOD1, PDGFD, TIMP2*), the regulation of specific signaling pathways remained consistent across all examined cellular states, but often, it was only detected in one or two of the examined cell states and, in some cases (e.g., *LRP1, COL4A2, and TNS1*), in opposite directions, as exemplified for contraction, NFκB, and interferon signaling in **Figure 3D**. This may indicate that certain mechanisms gain or wane in importance in different cell states. For example, genes involved in extracellular matrix organization were generally downregulated by knockdowns in the baseline conditions, presumably reflecting the high activity of these genes in this cell state, while up- and downregulation were seen in other states (**Figure 3A-C**). Basal expression of the target genes, although in several cases up- or downregulated by cholesterol or stretch (**Supplementary Figure 5**), did not predict knockdown outcomes.

To validate the observed regulation of NFκB- and type I interferon-induced genes, we performed two types of experiments. First, we tested whether the siRNA transfections themselves could cause the observed inflammatory signaling. Depending on specific recognition motifs in the sequence, some siRNAs can directly activate type I interferon responses [31,32]. The siRNAs used in our experiments were designed not to contain these motifs, and we confirmed experimentally that neither the control siRNA nor lipofectamine transfection alone altered the expression of *IL6*, *IFI6, IFITM1,* or the contractile marker *MYOCD* compared with untreated controls **(Supplementary Figure 6)**.

Second, since genes induced by NFκB and type I interferon signaling may vary by cell type, we defined the NFκB and type I interferon gene signatures in human SMCs by stimulating with the prototypical inducers tumor necrosis factor (TNF) and interferon alpha (IFN-alpha) followed by RNA sequencing **(Supplementary Figure 7A-B** and **Supplementary Table 6)**. We found that the experimentally defined gene sets were partly overlapping and confirmed that they were regulated by multiple target genes with *C1S* and *HHIPL1* knockdown, producing the strongest up- and down-regulation across the three cell states **(Supplementary Figure 7C)**.

**Widespread alterations in inflammation- and cell cycle-related regulons**

To identify the transcription factors (TFs) that drive the shared transcriptomic changes among different knockdowns, we detected regulons (a set of genes jointly regulated by the same transcription factor) among up and downregulated genes. To focus on regulons with in vivo relevance, we restricted to regulons that were detected to be active in at least one mesenchymal cell subcluster of human plaques by scRNA-seq (SCENIC) analysis **(Figure 4A)**. Corroborating our analysis on the pathway and marker gene levels, we found that regulons putatively controlled by cell cycle/proliferation (E2F3, E2F4, ETS1, NFATC1, SP1, MEIS1, and TAF1) and NFκB/interferon (IRFs, STAT, RELA, RELB, NFKB1) transcription factors were significantly enriched. Members of the AP-1 complex (FOS, JUN), SOX, and Kruppel-like factors (KLFs) known as regulators of SMC phenotype were also identified **(Figure 4B)**. A detailed list of the regulons identified in every targeted gene knockdown and cellular assay is shown in **Supplementary Table 7.**

**Independent regulation of common SMC marker genes and functional readouts**

The growth of SMC-derived cells in atherosclerosis involves loss of SMC contractile gene expression, increased proliferation, and switching to modulated cell phenotypes characterized by extracellular matrix production and inflammatory signaling [8,33,34]. This metaplastic process explains the inverse correlation between contractile genes and markers of modulated cells and inflammatory signaling in plaque scRNA-seq data [8]. It is, however, less clear if reciprocal regulation of these programs already exists within a particular SMC cell state. To analyze this, we retrieved the expression levels of multiple marker genes used in the literature to define SMC phenotype, including contractile gene markers (*ACTA2, ACTG2*, *CNN1, LMOD1, MYOCD, TAGLN*), modulated SMC markers (*FN1, KLF4, LUM, SPP1*), cell proliferation markers (*MKI67, PCNA*), NFκB activation markers (*IL1B, IL6, PTGS2, VCAM1*) and type I interferon-induced genes (*IFIT1, IFITM1, ISG15, MX1, OAS1*).

Contractile marker gene expression was widely regulated under baseline conditions, being stimulated by *C1S, GEM, LOXL1, MFGE8, RBPMS2,* and *TIMP2* knockdown and repressed by *ADAMTS7, ALDH2, COL4A2, CRISPLD2, CTTN, DSTN, FHL1*, *HHIPL1, LMOD1,* and *PDGFD* knockdown **(Figure 5A)**. Changes in cholesterol-overloaded and stretched cells were similar albeit with some differences in the magnitude of regulation. The selected contractile genes were mostly regulated as a group, consistent with their common transcriptional regulation [35], with *MYOCD* showing the most independent regulation.

Interestingly, the selected markers for modulated SMCs (*FN1, KLF4, LUM*) were often independently regulated and not necessarily upregulated when contractile genes were repressed and vice versa. **(Figure 5B)**. Neither did we observe a consistent reciprocal regulation between contractile genes and marker genes for proliferation (*MKI67, PCNA)*, which were downregulated after all target gene knockdowns except for *TNS1* **(Figure 5B)**. Moreover, while markers for NFκB signaling and type 1 interferon signaling were regulated by multiple target genes (**Figure 5C**), those knockdowns that dampened inflammatory signaling, e.g., *DSTN* and *CTTN*, were not necessarily the same that upregulated modulated SMC markers and vice versa.

Finally, we examined proliferation and migration, which are functional readouts often used to gauge SMC phenotype. Proliferation measured by EdU labeling was downregulated in almost all target gene knockdowns (except for *COL4A2* and *LOXL1*, which showed no difference), consistent with the cell cycle marker gene analysis **(Figure 5D).** In contrast, a significant difference in cell migration was only detected for *ADAMTS7* knockdown **(Figure 5E)**.

Overall, this analysis, made possible by having a large set of independent cell perturbations, indicated that contractile gene expression and markers/functional readouts of proliferation, modulation, and inflammatory signaling are not tightly co-dependent in cultured SMCs. This is important because such markers are often used as markers for SMC phenotypic switching in cell culture under the assumption that they are linked and report cell identity changes similar to those occurring during atherosclerosis, which appears not to be justified.

**Proatherogenic effect of cholesterol overload indicated by GWAS gene effect directions**

The genetic variation identified in GWASs mostly influences CAD risk by altering gene expression rather than through alterations in the encoded protein, and genes can be detrimental or protective depending on whether lower expression decreases or increases disease risk, respectively. To classify our selected target genes, we searched for co-localizing expression or protein quantitative trait loci (e/pQTLs) using HyPrColoc [36] with a posterior probability of alignment threshold of 0.7. We determined the impact of the e/pQTL on CAD risk by Mendelian randomization [37]. We used eQTL data from GTEx V8 [38] and eQTLgen [39], and pQTL data from UK Biobank [40] and in a large Icelandic cohort [41].

The analysis predicted *ALDH2, C1S, CTTN, LOXL1, MFGE8, PDGFD, and TIMP2* to be detrimental and *FHL1, HHIPL1, LMOD1, LRP1, RBPMS2* to be protective genes for CAD development **(Supplementary Table 8)**. The analysis was inconclusive for the residual genes. We then used this classification as a tool to estimate which of the many regulated pathways identified in our SMC analysis were likely pro- versus anti-atherogenic. The underlying rationale is that a pro-atherogenic mechanism will tend to be upregulated by the knockdown of protective genes, while an anti-atherogenic mechanism will tend to be upregulated by the knockdown of detrimental genes.

Analyzing the broader gene signatures induced by cholesterol overloading and stretch revealed that cholesterol overloading in SMCs had the characteristics of a pro-inflammatory mechanism (**Figure 6A**): knockdown of protective genes, such as *LMOD1, FHL1, or LRP1,* upregulated the cholesterol overload-induced gene signature, while knockdown of detrimental genes, such as *C1S* and *TIMP2*, downregulated it. Interestingly, this pattern of opposite regulation by protective and detrimental genes was not seen among gene knockdowns that simultaneously upregulated stretch-induced genes, suggesting some level of interaction (e.g., *HHIPL1, MFGE8*).

Individual pathways (e.g., SMC contractility, NFκB signaling, type I interferon) were not consistently regulated by protective versus detrimental genes (analyses not shown). This may reflect the complex and mutually independent regulation of these pathways across the gene knockdowns, necessitating a multivariate analysis to disentangle that was not possible with our limited number of informative genes.

**Specific SMC gene modules regulated by protective and detrimental GWAS genes**

To evaluate which parts of the cholesterol- and stretch-induced responses may be pro- or anti-atherogenic, we defined gene modules in SMCs subjected to baseline, cholesterol overload, and mechanical stretch conditions by analyzing gene co-expression in the 87 samples transfected with control siRNA. We obtained 24 modules (M1 to M24), ranging in size from 1570 genes in M1 to 30 genes in M24, with evidence of differential regulation across the SMC states (**Figure 6B).** Details of genes and their connectivity are provided in **Supplementary Table 9-10,** and a graphical representation of the analysis can be found in **Supplementary Figure 8.**

We then explored which of these 24 gene modules meets the criteria of a pro- or anti-atherogenic mechanism, being regulated oppositely (if at all) by protective and detrimental genes. Gene module 2 (M2) and module 21 (M21) were the top modules identified with these characteristics (**Figure 6C**). Module 2 (795 genes) was characterized by genes with high activity in the stretch condition but low in baseline and cholesterol-overloading conditions. In comparison, module 21 (36 genes) contained genes with high activity during cholesterol-

overloading conditions but not at baseline or stretch. Enrichment analysis revealed that module 2 contained genes involved in RNA processing, while module 21 comprised genes associated with the wingless-type MMTV integration site family (Wnt) signaling pathway (**Supplementary Table 10**). Expression of the two modules across the target knockdowns showed a positive correlation, and both had the signature of a pro-atherogenic mechanism being upregulated by protective gene knockdowns (e.g., *LMOD1, HHIPL1*) and downregulated by detrimental gene knockdowns (e.g., *ALDH2, C1S*) (**Figure 6D**). These elements of SMC responses to metabolic and mechanical stress may, therefore, be particularly interesting for further exploration. An expression-based gene scoring analysis for all the modules is shown in **Supplementary Figure 9.**

## Discussion

Many CAD susceptibility genes in GWAS studies do not exert their effects through classical risk factors such as circulating cholesterol levels or hypertension [42]. Therefore, unidentified disease mechanisms contributing to residual risk, which are not addressed by current therapies, may exist. During atherosclerosis, a subset of SMCs de-differentiate, proliferate, and modulate to other mesenchymal cell phenotypes, many of which abundantly accumulate cholesteryl ester lipid droplets [8,43,44]. The accumulation of the SMC-derived cells and the extracellular matrix proteins they secrete is a central mechanism by which atherosclerotic plaque grows and a main determinant of plaque stability [45]. Cells with SMC-like phenotype build the fibrous cap, which protects against plaque rupture and thrombosis, but the modulated SMCs in the plaque interior may play detrimental roles by contributing to the formation and expansion of the necrotic core [46,47]. The phenotypic transition of SMCs and the important functions of the modulated cell types in plaques make them interesting therapeutic targets, but little is known about the molecular pathways that could be targeted.

Recent studies have searched for genes regulating a fixed set of atherosclerosis-relevant SMC functions (e.g., proliferation, inflammatory gene expression) using the genetic variation across a large set of patient-cultured SMCs [48,49]. Here, we followed the complementary strategy of introducing gene perturbations experimentally and performing an open search for alterations in SMC gene expression and function. This has previously been done for individual genes, but as exemplified by our findings, many patterns of atherosclerosis-relevant SMC regulation can only be revealed by looking at multiple gene perturbations in the same context. We systematically performed a knockdown screen of genes that are expressed in SMCs or their modulated progeny in coronary and carotid plaque scRNA-seq datasets and have genetic evidence for a causal role in CAD that is not explained by plasma lipids. The goal was to use these knockdowns as an instrument to understand what gene signatures and pathways are relevant for the pathophysiological function of SMCs in CAD. While it is not yet possible to

recreate the SMC phenotypic transitioning process in vitro, we steered human SMCs to several plaque-relevant states to increase the likelihood that central pro- or anti-atherogenic mechanisms of SMCs were featured in our assays. Finally, by integrating what is known about the effect on CAD of gene variants controlling the expression of our GWAS target genes, we attempted to determine which of the observed regulated pathways are pro-atherogenic and anti-atherogenic.

Most of the genes targeted in our study have not previously been investigated for their role in human SMC function or modulation, but some have been examined in independent studies, which can be used for cross-validation of our data. PDGFD was shown previously to promote the modulation of cultured SMCs and increase inflammatory signaling [14], while LMOD1 was required for the maintenance of the contractile phenotype [15,16]. These findings were replicated in our data. The knockdown of *PDGFD* decreased the expression of genes related to NFκB/interferon signaling and collagen synthesis, while the knockdown of *LMOD1* reduced genes related to SMC contractile gene expression.

Overall, our analysis produced four key findings. Firstly, we observed widespread regulation of cytoskeletal and contractile genes, NFκB, and type I interferon-induced genes, as well as cell cycle genes by the target gene knockdowns at both the pathway, marker gene, and regulon analysis levels. Importantly, there is experimental evidence from mouse models that several of these programs in SMCs are causal mediators of atherogenesis. Heterozygous deletion of MYOCD, the transcriptional co-activator of most contractile genes, results in accelerated atherosclerosis with more inflammatory cells in mice [50]. NFκB signaling in cultured SMCs interferes with MYOCD function and drives proliferation [51], and knockout of the gene encoding the required signaling factor IKKβ in SMCs protects against the development of murine atherosclerosis [52]. Furthermore, we recently found that the depletion of modulated SMCs in regressing atherosclerosis is associated with the loss of NFκB signaling in those cells [8]. In contrast, the role of type I interferon signaling in SMCs and atherosclerosis is less understood [53], but the strong regulation observed here, which could not be attributed to artifacts of siRNA transfection, warrants further studies of the potential causal role. Finally, the observed almost uniform downregulation of cell cycle genes and proliferation rates in cell culture assays was striking. SMC proliferation is a key driver of plaque development in atherosclerosis, but it is unlikely that almost all the GWAS genes, including both protective and detrimental ones, regulate proliferation in the same direction in vivo. We speculate that our findings can be explained by cultured SMCs being selected for their ability to proliferate - those that proliferate better overgrow others. If the state of the cell is optimized for the fastest possible proliferation, it makes sense that most perturbations to the cell would result in

reductions in proliferation. This inherent quality of SMC culture can easily be overlooked when gene knockdowns are examined individually.

A second main finding was that several knockdowns produced different effects in cells subjected to baseline, cholesterol overload, and mechanical stretch conditions. In most cases, the difference was a lack of significant effect in one condition when that effect was present in (an)other condition(s), but some effects were even opposite in different conditions. This is a testament to the importance of the context within which gene function is analyzed, and it touches on a major limitation in the field, further discussed below, that it is not yet known which SMC culture assays are predictive for relevant SMC functions in vivo.

Thirdly, we found that commonly used markers for SMC phenotypes and functional readouts like migration and proliferation were independently regulated. This pattern has been reported before [54]. This contrasts with the clear inverse associations between contractile gene markers (e.g., ACTA2, MYH11) and markers of modulated SMCs (e.g., LUM) and inflammatory signaling in plaque scRNA-seq data [5,8], and the fact that healthy contractile SMCs are non-proliferative and non-migrating, whereas activated, non-contractile SMCs in disease expand and move. The likely explanation is that studies in cultured SMCs are confined to examining cellular mechanisms in a semi-modulated, proliferative state, whereas SMCs in vivo undergo much more pervasive changes, transitioning from contractile SMCs characterized by low NFκB signaling and minimal synthetic gene expression to non-contractile mesenchymal cell types exhibiting active NFκB signaling and robust synthetic gene expression. These cellular transitioning processes cannot yet be fully modeled in vitro, and none of the knockdowns were consequential enough to transform the overall cell phenotype. This is evidenced by the PCA analysis in **Figure 2D**, which shows that no knockdown shifted gene expression from one cell state cluster to another. The key takeaway is to acknowledge this limitation of SMC culture studies, and not use SMC markers, proliferation, and migration in vitro as if they reported on the full trajectory that SMCs take in atherosclerotic plaques in vivo.

The final key finding was that the cholesterol-induced gene response had the characteristics of a pro-atherogenic mechanism. Among the gene knockdowns that regulated cholesterol-induced genes, upregulation was seen after the knockdown of predicted protective genes and downregulation after the knockdown of predicted detrimental genes **(Figure 6)**. This type of analysis is only hypothesis-generating, but, importantly, the causal role of cholesterol overloading in SMCs is supported by experimental evidence. Cholesterol overload of cultured SMCs with cyclodextrin-complexed cholesterol, as performed in this study, leads to foam cell formation with cholesteryl ester droplet accumulation [24]. While this method may not generate cells that fully acquire the transcriptional profile of SMC-derived foam cells in vivo [26], it induces gene expression that overlaps with gene expression of modulated SMCs in progressing

mouse lesions [8]. Part of this regulation appears to be mediated by cholesterol overload-induced activation of NFκB signaling [25], possibly through eliciting endoplasmic reticulum stress [55]. It will be important to investigate the upstream and downstream mechanisms of cholesterol overload in SMCs to identify nodes that can be targeted in experimental models to understand causal mechanisms and, potentially, as future targets for therapy. Further explorations of *C1S, LMOD1*, *FHL1, LRP1, and TIMP2*, identified here as strong regulators of the cholesterol-induced gene response, and the specific gene modules M2 and M21 implicated in the mechanism, can be starting points.

In a general perspective, the approach taken here, combining an in vitro screen with genetic effect predictions, can serve as proof-of-concept for much larger-scale studies in SMCs or other atherosclerosis-relevant cell types to predict the effect of blocking versus stimulating specific mechanisms in atherosclerosis.

*Limitations*

There are several important limitations in our study to consider. First, cultured SMCs, even if subjected to cholesterol overload and stretch, do not mimic the full diversity of SMC-derived cell phenotypes in atherosclerosis. Second, the target genes were identified using a relaxed criterion for genome-wide significant association with CAD. Combined with the uncertainty involved in predicting the causal gene at the associated locus, it is possible that some of our target genes are not de facto causal human CAD genes. Third, the number of informative GWAS genes in our screen was not sufficient for a full exploration of the relationships between the effect direction of GWAS genes on human CAD and the mechanisms they regulate in SMCs.

*Conclusions*

In summary, our study maps major mechanisms regulated by a set of CAD risk genes with a likely mechanism of action in plaque SMCs. By integrating knowledge of the effect direction of a subset of the analyzed GWAS genes, we identify cholesterol overload of SMCs, as well as specific cholesterol- and stretch-regulated gene modules, as putative pro-atherogenic mechanisms regulated by human genetic predisposition to CAD. The data released with this report include 345 RNA-seq data sets that can be used to explore the mechanisms downstream of 18 SMC-relevant GWAS genes, the effect of cholesterol overload and mechanical stretching on SMC cell state, and interactions between GWAS gene manipulation and SMC state.

## Methods

### Mining of genome-wide association studies (GWAS) datasets

We focused on GWAS from the Cardiovascular Disease Knowledge Portal (CVDKP) [56] and Finngen r5 [57]. Datasets were extracted from CVDKP using Signal Sifter (https://cvd.hugeamp.org/signalsifter.html). We searched for coronary artery disease (CAD) traits, excluding those with lipoproteins (LDL) associations. A ratio was calculated between log10(P value) of CAD/LDL. Hits were included if ratio > 2 and LDL P value > 5e-8. For Finngen datasets, credible sets and top hits were extracted using coronary atherosclerosis ("I9_CORATHER") as a phenotype of interest, choosing the "Traditional" (with a gene nearest to genomic variant) or "Credible Sets" (where genomic variant is fine mapped to a leading variant and related nearest gene) data table. Genes were removed if already identified in the previous method, LDL P value ≤ 5e-8 or co-association to lipoprotein traits were previously discovered.

### Gene annotation

Each genomic variant was subsequently checked using the results of "Variant-to-Gene" (V2G) and "Locus-to-Gene" (L2G) fine-mapping pipelines of Open Targets Genetics platform (https://genetics.opentargets.org/) (Ghoussaini et al, 2020), and if any additional gene was identified, it was added to the list.

### Integration of single-cell RNA sequencing studies from public datasets

To estimate the transcriptomic patterns of target genes and gene networks in different cell types, we collected single-cell RNA sequencing (scRNA-seq) data of human atherosclerotic lesions and healthy arteries published previously [5–7]. Sample information was gathered from the Gene Expression Omnibus database (accession GSE159677, GSE155512 and GSE131778). Raw FASTQ files were downloaded from NIH Sequence Read Archive (SRA) and processed via CellRanger v6.1.2 with 10X Genomics-derived human reference hg38 (2020-A July 7, 2020). Single-cell data filtering, integration, and clustering were done in the Seurat R package [58]. Cells were filtered as follows: number of UMIs > 500, number of detected genes between 200 and 4500, percent of mitochondrial genes < 10% and hemoglobin genes < 1%, ratio of mitochondrial to ribosomal genes < 0.5 [59]. Additionally, we discarded the barcodes suspected as doublets by any of the three methods used: DoubletFinder [60], scDblFinder [61], and Scrublet [62].

We normalized data using the standard "LogNorm" method in Seurat, and integrated datasets using Harmony [63]. Cell clusters were identified using a shared nearest neighbor modularity optimization-based algorithm (implemented in Seurat by default) with 2000 most variable genes, 30 principal components, and a resolution of 1. Marker genes of the identified cell

clusters were found with Wilcoxon–Mann–Whitney test implemented in Seurat (**Supplementary Table 11**). Automatic cell type annotation was performed using popularVote pipeline (https://tabula-sapiens-portal.ds.czbiohub.org/annotateuserdata) with the Tabula Sapiens "Vasculature" dataset as a reference [64]. To validate the SMC-specific origin of identified clusters, we integrated human data with a scRNA-seq dataset of SMC-derived cell population from Myh11-CreERT2 x TdTomato (TdT) reporter mice with induced atherosclerosis and high-fat diet as described elsewhere (Carramolino et al, 2024). Cell clusters within SMC-specific supercluster were further re-annotated based on marker genes (differentially expressed in a specific cell cluster compared to other cell clusters).

**Selection of smooth muscle cell population-specific genes**

According to scRNA-seq results, several clusters of SMCs (contractile, transitional, osteochondrogenic, undefined SMCs, fibromyocytes), fibroblasts, and pericytes were included in a unified supercluster of "mesenchymal cells." Since a clear delineation between the spectrum of differentiated and modulated smooth muscle cells, fibroblasts, and other transcriptionally close cell types remains challenging, we consider a whole mesenchymal supercluster to determine the SMC specificity of each gene. If a gene had the highest average expression in this supercluster or a gene is expressed in a larger percentage of mesenchymal cells than any other type, or the average expression of a gene in smooth muscle cells was 1.5 standard deviations above the average for all other cell types, we call that gene enriched in mesenchymal supercluster and considered as a candidate gene.

**Spatial transcriptomics**

Spatial transcriptomics samples were generated from formalin-fixed and paraffin-embedded (FFPE) human coronary arteries at different stages of the disease, isolated from explanted hearts, as previously described [20,21]. Briefly, 5µm thickness sections were mounted on Visium slides and processed for spatial transcriptomics according to the 10X Genomics Visium FFPE Version 1 protocol. Samples were deparaffinized, stained with hematoxylin and eosin (H&E), and imaged using VS200 Slide Scanner (Olympus Life Science) before decrosslinking, destaining and overnight probe hybridization with the 10X Visium Human version 1 probe set. The following day, hybridized probes were released from the tissue and ligated to spatially barcoded oligonucleotides on the Visium Gene expression slide. Barcoded ligation products were then amplified and used for the construction of libraries, which consequently were sequenced on a NovaSeq 6000 sequencing platform (Illumina) using a NovaSeq 6000 S2 Reagent Kit v1.5 (Illumina) according to the manufacturer's instructions. Afterwards, reads were aligned with their corresponding probe sequences, mapped to the Visium spot where each probe was captured, and finally aligned with the original H&E-stained image of the tissue

section using the software SpaceRanger version 1.3.0 (10X Genomics). Samples ranged from 568 to 1,746 spots.

**Human aortic smooth muscle cells**

We used healthy human aortic smooth muscle cells (haSMCs) from a single donor (ATCC, PCS-100-012; Lot #80323179). SMCs were expanded at 37 °C, 5% CO2 in vascular cell basal medium with cell growth kit from ATCC (PCS100 030 and PCS 100 042) from passage P0 to passage P3. During cell expansion, plates were coated with type I collagen from rat tail (Corning, 354236) diluted in 0.1% acetic acid [15 µg/mL] for >20min at room temperature (RT) and washed once with phosphate-buffered saline (PBS). For experiments, SMCs of passage P4 and P5 were cultured in smooth muscle cell growth medium containing 5% fetal calf serum and supplements (Provitro, 211-0601).

**Small interference RNA (siRNA) transfection**

The expression of each candidate gene was suppressed by two different small interfering RNAs (siRNAs) per target gene (QIAGEN, 20 μM); the sequences are shown in **Supplementary Table 12**. Following the guidelines provided by the manufacturer, the siRNAs utilized in this study were screened for various sequence motifs known to trigger an interferon or cytokine response. Consequently, siRNAs containing these motifs were excluded from our study. Transfection was performed with Opti-MEM (Thermo Fisher, Cat. No. 31985062) and Lipofectamine RNAiMAX (Invitrogen, Cat. No. 13778150) on two consecutive days according to the manufacturer's instructions. Control SMCs were transfected with control siRNA (QIAGEN, AllStars Negative S103650318, 20 μM).

**Baseline modulation of human smooth muscle cells**

Human SMCs were cultured on type I collagen-coated plastic 12-well plates with growth medium as described above. Cells cultured under these conditions spontaneously modulated their phenotype and thus were used as a “baseline.” When the cells reached 70% confluency, they were transfected with siRNAs for each gene on two consecutive days, as described above. Every RNA sample was a pool of three wells, and we used four biological replicates per condition.

**Cholesterol overload**

Human SMCs were cultured on type I collagen-coated plastic 12-well plates with growth medium. At 70% of confluency, cells were either untreated or treated for 72 hours (h) with water-soluble cholesterol (Sigma-Aldrich, C4951, final concentration 10 μg/mL). For initial

RNA-seq analysis, we compared expression levels of SMCs treated with cholesterol overload for 72 h versus untreated cells. For target gene characterization, SMCs were treated with cholesterol and transfected in parallel with siRNAs for each gene on two consecutive days while continuing cholesterol treatment for another day to reach 72 h in total. Every RNA sample was a pool of three wells, and we used four biological replicates per condition.

**Mechanical stretch**

Human SMCs were seeded onto flexible silicon 6-well plates coated with ProNectin (Bioflex, Cat. No. BF 3001P) and cultured in growth medium until 70% confluency. Cells were then transfected with siRNAs for each gene on two consecutive days, as described above. Then, SMCs were subjected to 10% cyclic stretch with a sinusoidal waveform at a frequency of 1 Hz for 6 h with the Flexcell tension system (Flexcell International Corp., FX-5000 T) or left under static conditions. For initial RNA-seq analysis, we compared expression levels of SMCs subjected to stretch versus untreated cells. For target gene characterization, SMCs were transfected with siRNAs for each gene on two consecutive days. A day later, control and target transfected cells were subjected to stretch. Every RNA sample was a pool of two wells and we used four biological replicates per condition. Data was evaluated by RNA-seq as described above.

**TNF and IFN treatment**

Stimulation with tumor necrosis factor-alpha (TNF) or interferon-alpha (IFN) was performed for 24h when SMCs reached a confluency of ~80-90% (approximately 48 h after seeding). SMCs were cultured on type I collagen-coated plastic 12-well plates and stimulated with TNF (Abcam, ab259410, 10 ng/mL), IFN (Abcam, ab48750, 100 ng/mL), or left untreated. Every RNA sample was a pool of three wells (12-well plates), and we used four biological replicates per condition.

**Cell proliferation**

A total of 50,000 human SMCs were seeded on glass coverslips coated with type I collagen in a 24-well plate. After siRNA transfection, cells were then cultured for an additional 24 h in growth medium containing 5-ethynyl-2'-deoxyuridine (EdU,10 μM). Cells were fixed in 3.7% paraformaldehyde in PBS for 15 min. The EdU-positive cells were labeled according to the manufacturer's instructions (Sigma-Aldrich, BCK-EDU488), followed by nuclei labeling with DAPI (Thermo Fisher, D1306). Coverslips were mounted on slides and imaged using a Nikon Eclipse Ti2 inverted microscope. The NIS-Elements software was used for image acquisition, and ImageJ [65] was used for quantification. Three to four biological replicates were evaluated per gene.

**Scratch assay for migration**

Human SMCs were seeded in 24-well plates coated with type I collagen and transfected with siRNAs as described above. Then, the confluent cell monolayer was scratched in the center of the well using a 10 µl pipette tip before washing once with PBS to clear cell debris and floating cells. Cells received fresh media, and time-lapse images were captured hourly for 24 h using a Nikon Eclipse Ti2 inverted microscope with an integrated chamber ensuring proper cell culture conditions of 37 °C and 5% $CO_2$. The migration ability was measured by calculating the percentage of gap closure at different time points compared to time 0 (baseline) using the NIS-Elements software. Three to four biological replicates were evaluated per gene.

**Total mRNA sequencing**

RNA was isolated from wells using the NucleoSpin RNA Plus Kit (Macherey-Nagel, 790984) and sequenced by BGI (Copenhagen, Denmark). All RNA samples passed quality control (Agilent 4200 electrophoresis system). Non-stranded cDNA libraries were prepared with polyA-selected mRNA, and 100 bp paired-end sequencing was conducted with at least 20M pair reads per sample using DNBSEQ. RNA sequencing data were processed with the Nextflow-based pipeline "nf-core/rnaseq" (v3.5). Sequenced reads were aligned to the GRCh38 human reference genome using STAR aligner (v2.6.1d), and gene-level quantification was done with Salmon (v1.5.2). The reference genome sequence and gene annotation were downloaded from AWS iGenome using the Ensembl version as a source. Further analysis was performed in R (v4.1.1) with Bioconductor (v3.14) packages. Gene counts were imported using tximport. Gene filtering was performed as previously described, and only genes with ≥10 counts per million (CPM) in at least 3 samples were analysed. DESeq2 [66] was used for data normalization, variance stabilizing transformation, principal component analysis, and estimation of differential gene expression between compared groups. Estimated P values (Padj) were adjusted for false discovery rate by Benjamini and Hochberg procedure [67]. Genes with Padj <0.05 and absolute log2 fold change (LFC) ≥0.5 were considered differentially expressed, or significantly dysregulated. To identify involved pathways, the lists of differentially expressed genes were used for gene-set enrichment analysis using a Web-based gene set analysis toolkit (WebGestalt) [68] with categories from GeneOntology [69], Reactome [29], WikiPathways [30], and KEGG PATHWAY [28] databases.

**Real time-qPCR**

The relative gene expression was measured by real-time qPCR (RT-qPCR) using *HPRT1* as housekeeping gene. Intron-spanning primers amplifying 50<150bp amplicons were designed using the Primer-BLAST design tool from NCBI [70]. Primer efficiency was tested by plotting the sample quantity of the dilution series to the average ct values of each dilution on a logarithmic

scale. The primer sequences are shown in **Supplementary Table 13**. The Maxima SYBR Green qPCR master mix kit [1:2] including ROX [0.18μM] (Thermo Fisher Scientific, K0252) was mixed with forward and reverse primers [268 nM] and cDNA [12.5 ng] in a final volume of 14 μL. On a CFX Opus Biorad machine, the RT-qPCR program was run with an initial denaturation step of 95°C for 10 min followed by 40 cycles of 95°C for 30 sec, 60°C for 1 min, and 72°C for 1 min, and one cycle of 95°C for 1 min, 60°C for 30 sec, and 95°C for 30 sec.

**Regulon enrichment analysis**

The lists of unidirectionally dysregulated genes were screened for the enrichment of transcription factor binding motifs using R/Bioconductor package *RcisTarget* [71]. A set of identified motifs and related transcription factors from TRANSFAC [72] database collection was used for further regulon analysis. Regulon activity analysis in the integrated scRNA-seq dataset data of human atherosclerotic lesions and healthy arteries [5–7] was performed using the SCENIC pipeline [71] with default parameterization and human genome reference (hg38). Area under the curve (AUC) values estimated by the pipeline were averaged for every cell cluster and z-scaled by regulon. Regulons with higher relative activity (positive z-score) in the cluster of interest were considered active.

**Genetic direction of selected CAD-associated genes**

For each of the selected genes, we determined if there was colocalization for the latest CAD GWAS loci [73] with any eQTLs from arteries, liver, and whole blood from GTEx or with pQTLs from UK Biobank, deCODE genetics or recent studies as described in the results. If there was significant colocalization between the QTL and the CAD GWAS locus, we determined the direction of effect using Mendelian Randomization with the QTL as the instrumental variable and the CAD GWAS as the outcome. The second analysis stage was to determine whether missense or pLOF variants in any of the genes overlapped with any significant associations from the CAD GWAS.

**Gene co-expression network inference**

We inferred the control gene co-expression network (GCN) using the results of total mRNA sequencing for SMCs transfected with control siRNA in every cellular assay. We inferred the control gene co-expression network (GCN) using the results of total mRNA sequencing for human SMCs transfected with control siRNA in every cellular assay. Briefly, we employed a gene expression matrix after variance stabilizing transformation. We selected the top 5000 most variable genes for weighted gene co-expression network analysis (Langfelder and Horvath 2008) in BioNERO R package (Almeida-Silva and Venancio 2022). Seven of the 94 control samples were identified as outliers by the standardized connectivity method (Zk) [74] and

discarded for further analysis. We used Spearman's correlation to build a signed network with the power soft threshold 7, achieving scale-free topology $R^2$=0.843 (scale-free topology and mean connectivity for different power thresholds are shown in **Supplementary Figure 8**). Correlation analysis between module eigengene (the first principal component of the gene expression matrix of the corresponding module) and cellular state was used to identify gene modules responsive to baseline modulation, cholesterol overload, or mechanical stretch. Noteworthy, we did not adjust RNA sequencing data to correct the difference between experimental batches before GCN inference since they are compromised by study design with a cellular state. Instead, we considered the modules with clear patterns of higher differences between cellular states than within-state variation. Over-representation analysis of genes assigned to different GCN modules in Reactome, KEGG Pathway, WikiPathways, GeneOntology, and Broad Institute Hallmark50 gene categories was performed with WebGestalt, as previously described (**Supplementary Table 10**).

**Gene signature scoring**

We performed gene expression-based scoring of regulation for those gene signatures which, to our knowledge, are related to SMC contractility, adhesion, and cell cycle, as well as with synthesis of components of extracellular matrix, and inflammatory signaling. Specific gene sets were collected from pathway databases: KEGG Pathway [28], Reactome [29], and WikiPathways [30] (**Supplementary Table 5**). Additionally, we generated sets of top 100 up- and downregulated genes after cholesterol loading, stretch, TNF, or IFN stimulation of SMCs (Padj <0.01, absolute LFC ≥ 1, arranged by LFC) and used them as gene signatures. We also used the gene lists for every module of the SMC-specific GCN as gene signatures (for large modules, only 100 genes with the highest intramodular connectivity were chosen). R/Bioconductor package *singscore* [75] was employed for independent sample-wise scoring with DESeq2-produced Wald z-statistic used for gene ranking. Relative z-scores for selected gene signatures were estimated for every comparison of target knockdown versus related control in all studied cellular assays. Thus, the estimated z-scores reflect the change in total transcriptional activity for every gene signature.

We assessed the ability of GCN module-specific gene signatures to predict the effect of target genes with identified genetic direction on CAD. Activity score of each GCN module was calculated for every target knockdown and assay, as described above. Random forest model (implemented in the *randomForest* package [76]) was trained with gene module scores as explanatory variables and the effect of target gene knockdown on CAD as instrumental variable. The embedded measure of node impurity, mean decrease in Gini index, was used to assess how important the contribution of each variable was in the final model prediction.

**Data availability**

All data generated via RNA-seq in siRNA-guided target gene knockdown experiments are deposited at Zenodo and are free to access. All other data supporting the findings in this study are included in the main article and associated Supplementary Tables.

**Sources of funding**

This study was supported by grants from the Novo Nordisk Foundation (nos. NNF17OC0030688 and NNF20SA0061466 to J.F.B.), the European Research Council under the European Union's Horizon 2020 research and innovation program (grant no. 866240 to J.F.B.), and the Aarhus University Research Foundation (Starting Grant, AUFF-E-201 9-7-23 to J.A-J.). The results of this manuscript are part of the Project THOR (Targeting smooth muscle cells in atherosclerosis therapy) of the Open Discovery Innovation Network (ODIN) initiative.

**Author contributions**

J.A-J., participated in the design of the experiments, performed in vitro experiments, analysis, interpretation of data, and generation of figures, A.M., performed the re-integration of scRNA-seq datasets and RNA-seq bioinformatics analysis, and contributed to the generation of figures, A.L.J., performed in vitro experiments and analysis, P.L.M., performed the initial analysis on GWAS datasets and potential candidate genes, A.K.U., interpreted the data, D.D., contributed to the scRNA-seq analysis and selection of target genes, J.H., performed in vitro experiments and analysis, L.F.J., performed in vitro experiments and analysis, D.S., performed bioinformatics analysis in SCENIC, C.P., and J.M., performed the spatial transcriptomics analysis, G.B., interpreted the data, J.B., interpreted the data, K.H., interpreted the data, L.M.R., provided experimental assistance for in vitro experiments, M.T., performed the genetic analysis of directionality, M.N., contributed in the conceptualization of the project and interpreted the data, Mette.N., conceptualized the project and interpreted the data, J.F.B., conceptualized the project, design the experiments, and interpreted the data. J.A-J., A.M., and J.F.B. drafted the manuscript and figures with important contributions from A.L.J., P.L.M., A.K.U., J.H., L.F.J., D.S., C.P., G.B., J.B., K.H., J.M., M.T., M.N., and Mette.N. All authors read and approved the manuscript.

**Disclosures:**

A.K.U., D.D., C.P., J.M., G.B., J.B., K.H., M.T., and M.N. are employed at Novo Nordisk A/S.

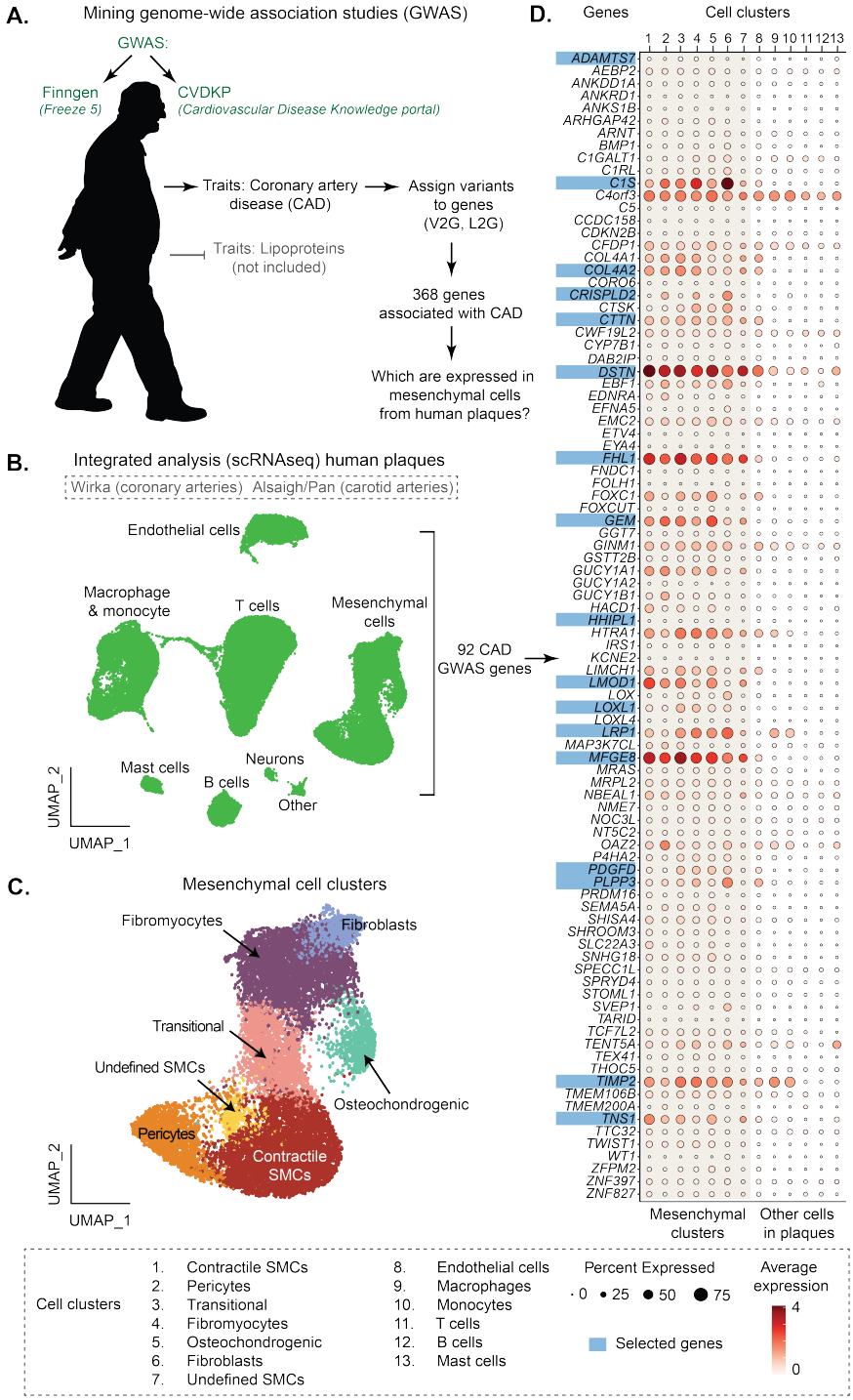



**Figure 1: From GWAS and scRNA-seq data to the selection of genes. A.** GWAS results from Finngen and CVDKP were utilized to identify variants associated with coronary artery disease (CAD), excluding those associated with lipid traits. After gene fine-mapping, we identified 368 genes. **B.** We then tested the expression of these genes on integrated single-cell transcriptomic data from publicly available studies of human atherosclerotic lesions. This led to identifying 92 protein-coding genes preferentially expressed in mesenchymal clusters. **C.** Shows a supercluster of mesenchymal cells was subdivided into 7 clusters that were identified by their marker genes as contractile smooth muscle cells (SMCs), transitional, osteochondrogenic, and undefined SMCs, pericytes, fibromyocytes, and fibroblasts in integrated data. **D.** Dot plot shows the expression of genes enriched in the mesenchymal supercluster. Normalized and log2-transformed gene expression values are averaged for every cell cluster.

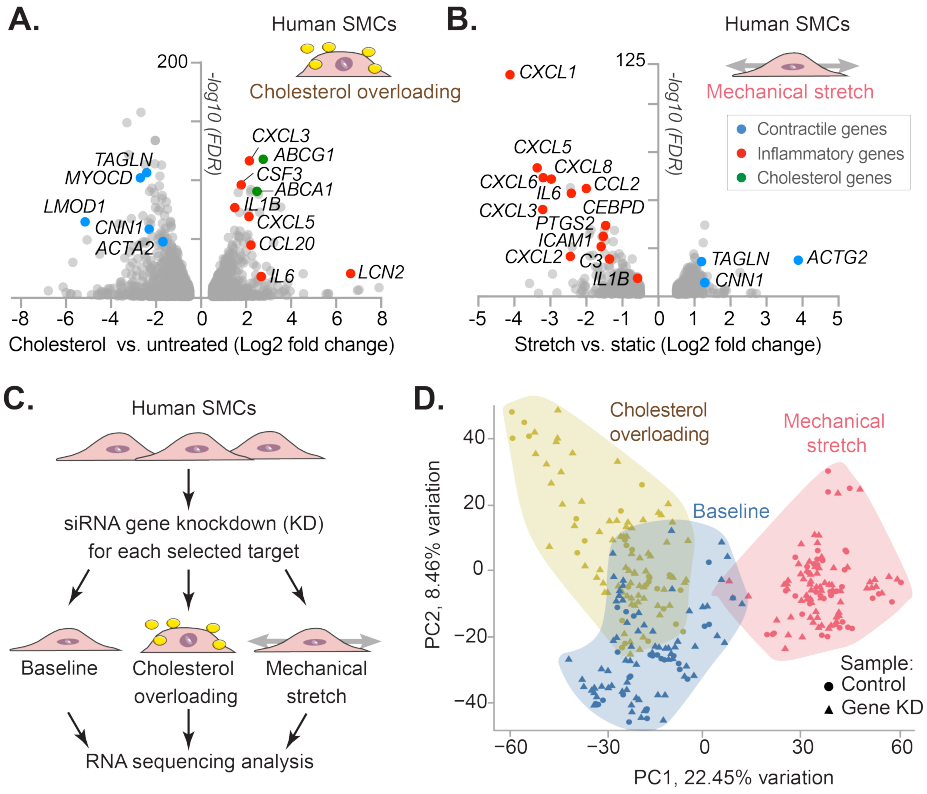



**Figure 2: *In vitro* models of vascular disease in human smooth muscle cells. A.** Gene expression in human smooth muscle cells after cholesterol overload (volcano plot). **B.** Gene expression in SMCs after physiological (10%, 1hz, for 6 h) mechanical stretch (volcano plot). A and B show significantly regulated genes (*P value* < 0.05 and absolute log2 fold change ≥ 0.5) compared to their control condition. FDR. False discovery rate-adjusted values. **C.** Overview of the workflow. SMCs were transfected with siRNAs for each gene, subjected to different cellular assays, and analyzed by RNA sequencing. **D.** Principal component analysis (PCA) of transcriptome profiles of studied SMC samples (*n*=345) revealed the most difference between mechanical stretch and other cellular assays.

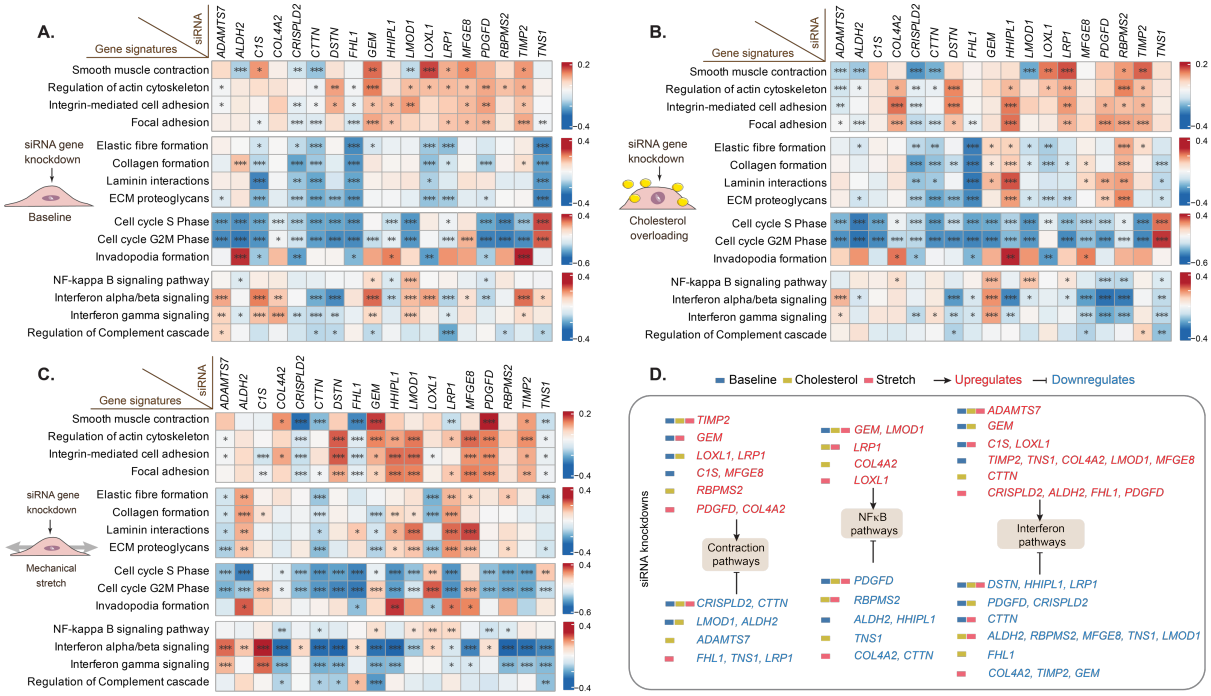



**Figure 3: Gene signatures disturbed after target gene knockdown.** Expression-based scoring of gene signatures implicated in SMC contractility, synthesis of extracellular matrix proteins, cell cycle, and inflammatory response in the different cellular assays: **A.** baseline, **B.** cholesterol overload, and **C.** mechanical stretch. The color scale represents a z-score of deviation in the transcriptional activity of relevant pathway genes in target knockdown compared with a control group. Asterisks highlight statistical significance based on false discovery rate (FDR)-adjusted P values (Padj): * Padj <0.1, ** Padj <0.01; *** Padj <0.001. Extracellular matrix (ECM) **D.** Examples of pathways regulated by differentially expressed genes induced by siRNA knockdowns of target genes.

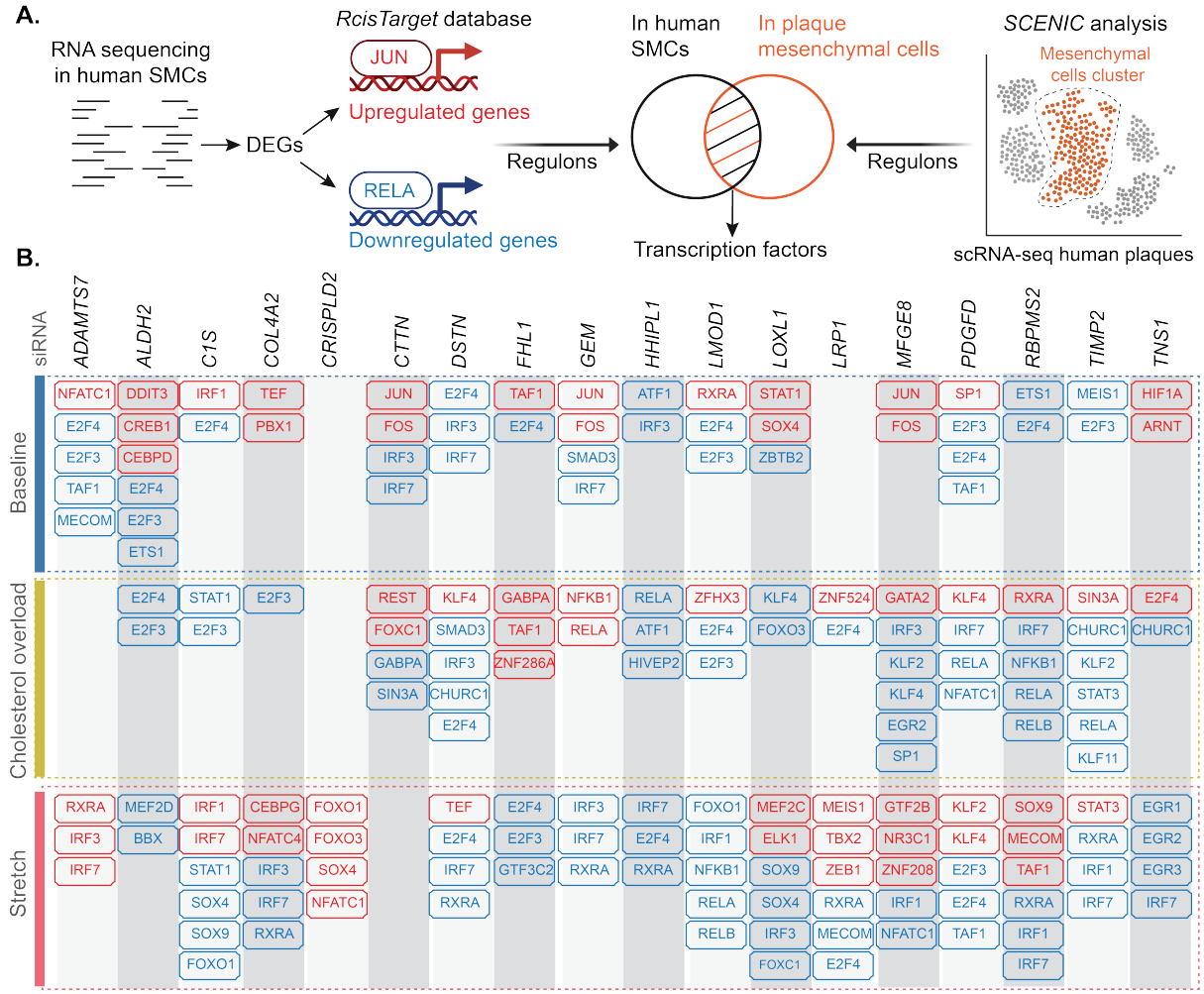



**Figure 4: Regulons with altered activity following target knockdown in different cellular assays**. **A.** Regulon is a collection of genes with a shared transcription factor binding site suggesting potential co-regulatory mechanisms by identical transcription factors (TFs). For each targeted gene knockdown and corresponding cellular assay in human SMCs, we identified regulons with a significant presence within the differentially expressed genes (DEGs), separately for up- and downregulated gene sets (JUN and RELA are shown as examples). Identified regulons were then overlapped with regulons detected to be active in at least one mesenchymal subcluster by SCENIC analysis of the integrated single-cell RNA sequencing dataset from human plaques. Drawing created with Biorender.com. **B.** An overview of shared regulons across various mesenchymal cell clusters and altered by targeted gene knockdown in cellular assays are named by putatively binding TFs. Regulons with up-and down-regulated genes are depicted as red and blue boxes.

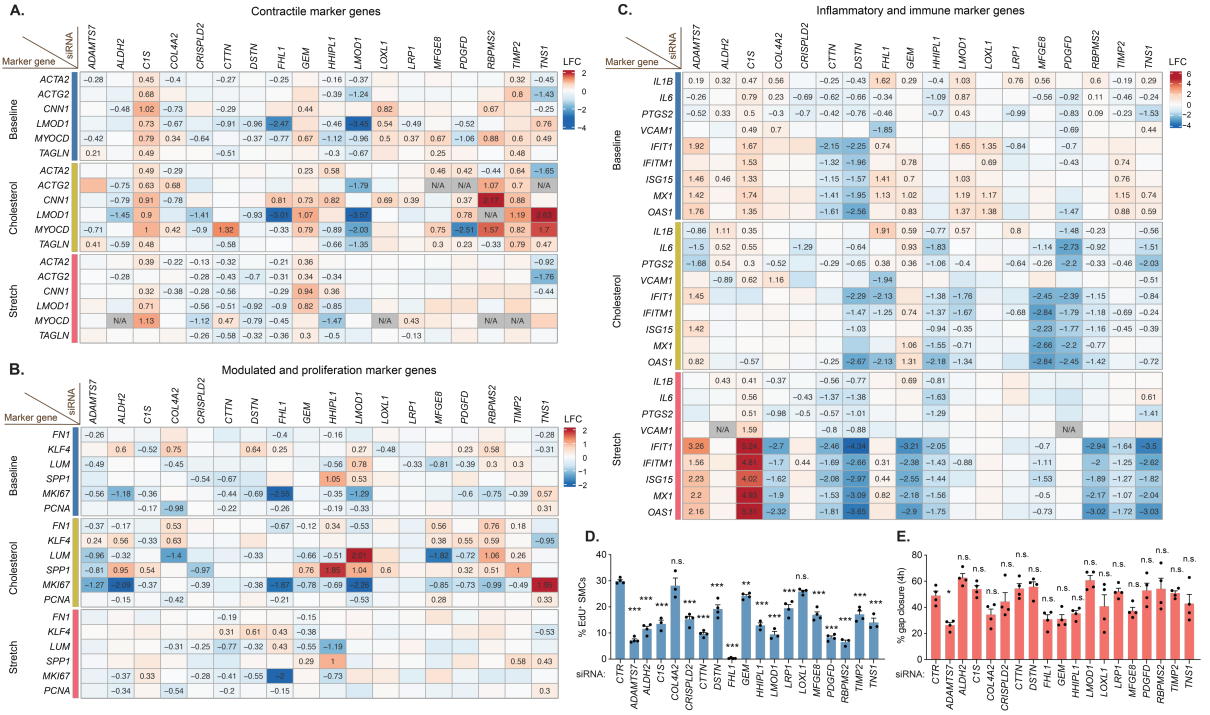



**Figure 5: Regulation of conventional gene markers of smooth muscle cell function.** Differences in the expression of a subset of marker genes important for smooth muscle cell function in target gene knockdowns compared to related control groups on each cellular assay. Results are shown for different cellular pathways: **A.** contractile, **B.** modulated and proliferation, and **C.** inflammatory and immune marker genes. The number in heatmaps indicates log2-transformed fold changes (LFC). Only statistically significant LFC numbers are shown. N/A indicates that gene expression was not detected. **D.** SMCs were transfected with negative control siRNA (control) or siRNAs targeting each gene and labeled with EdU (5-ethynyl-2'-deoxyuridine) for 24 h to evaluate their proliferation. Shown is the percentage of EdU+ SMCs per view field (3-4 biological replicates were evaluated per gene). **E.** SMCs were transfected with the different siRNAs, and a scratch was made in the center of the well. Cells were incubated at 37°C and 5% $CO_2$ in a time-lapse microscope (Nikon, Eclipse Ti2), and cell migration was monitored for 24 h in total. The percentage of gap area closure is shown 4 h after the scratch (3 to 4 biological replicates were evaluated per gene). Samples were compared to the control cell using one-way ANOVA followed by a Dunnett test. A * represents an adjusted P value < 0.05, ** P value < 0.1; and *** P value < 0.01, or n.s (not significant, P value > 0.05) compared to control.

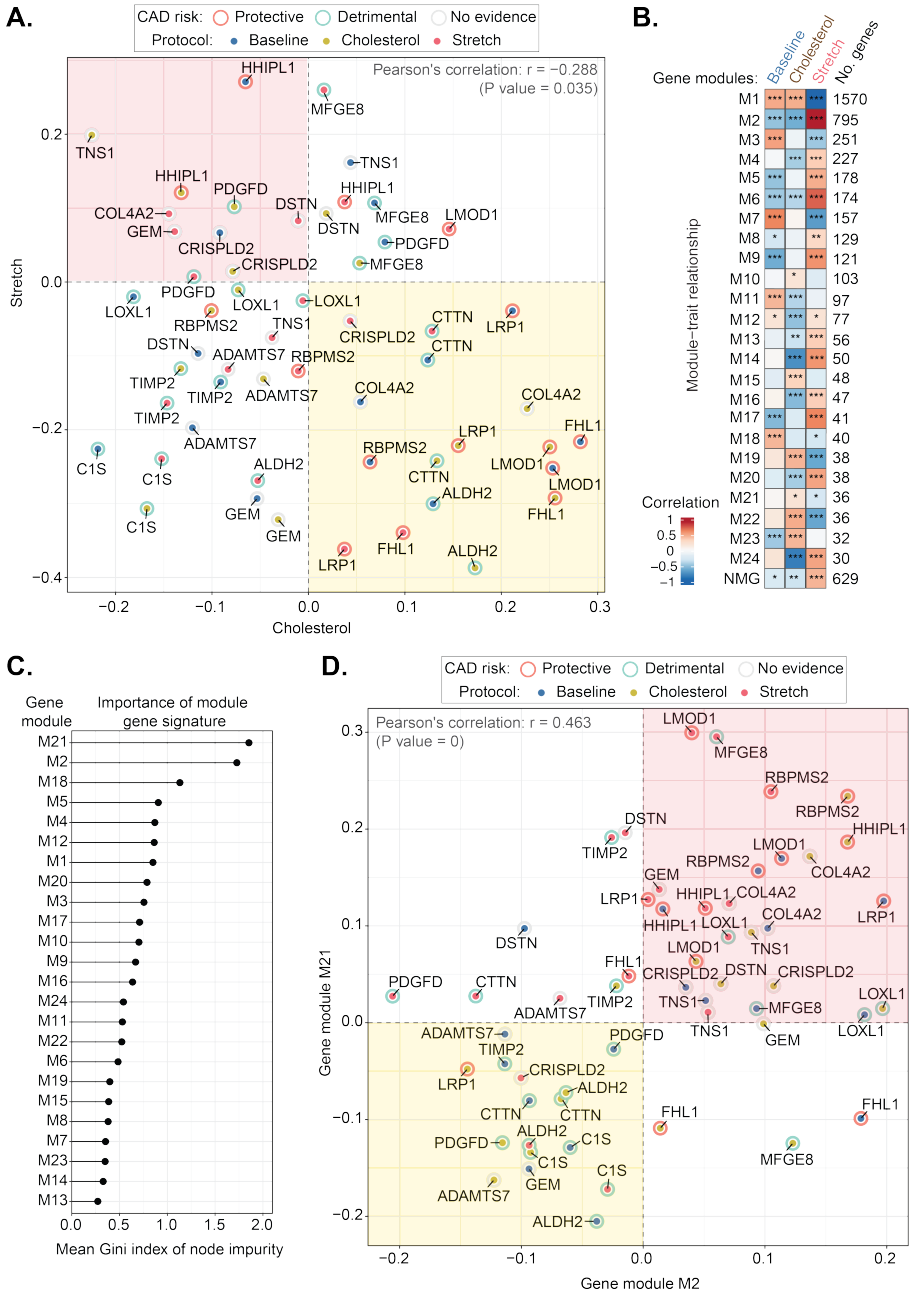
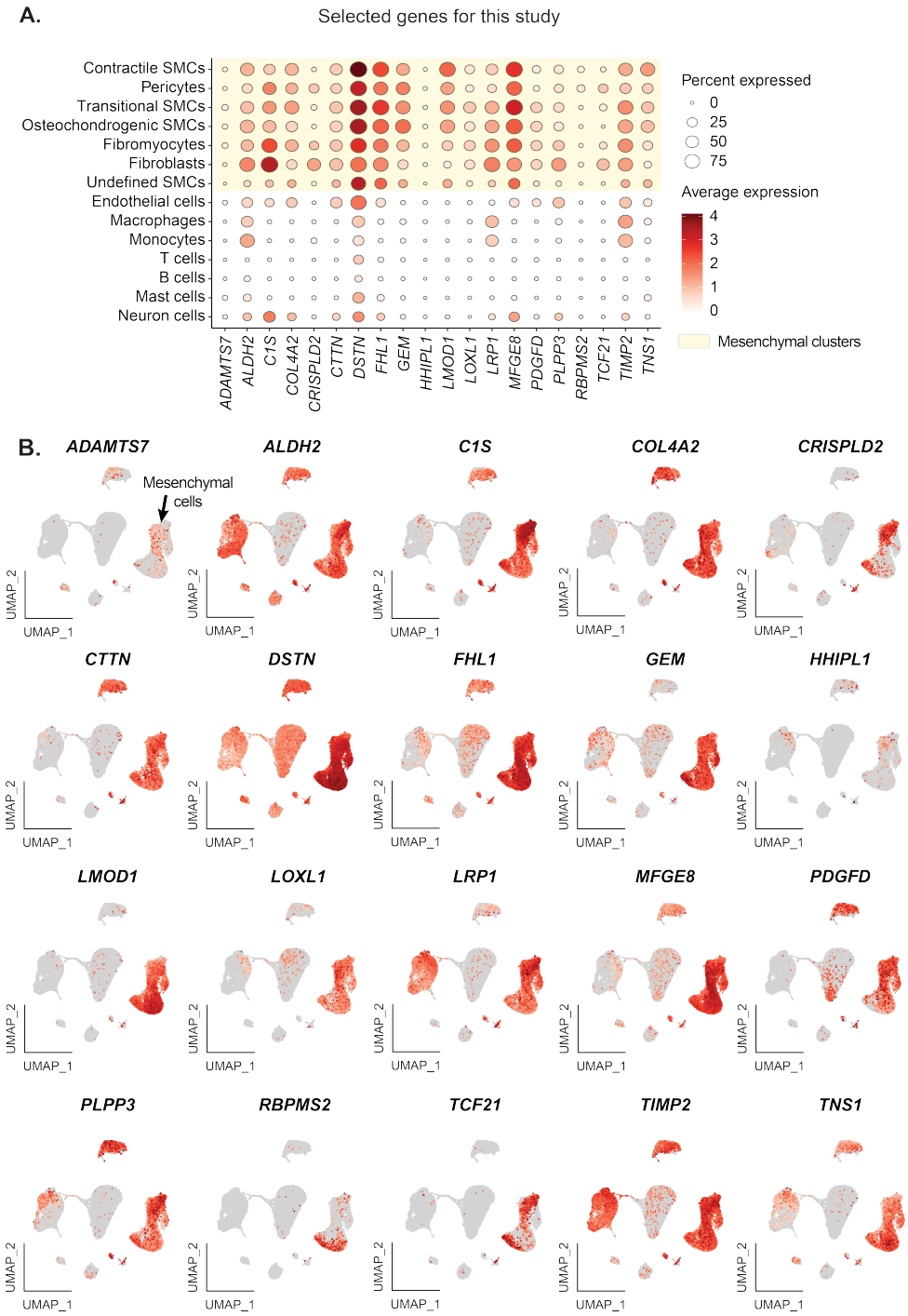
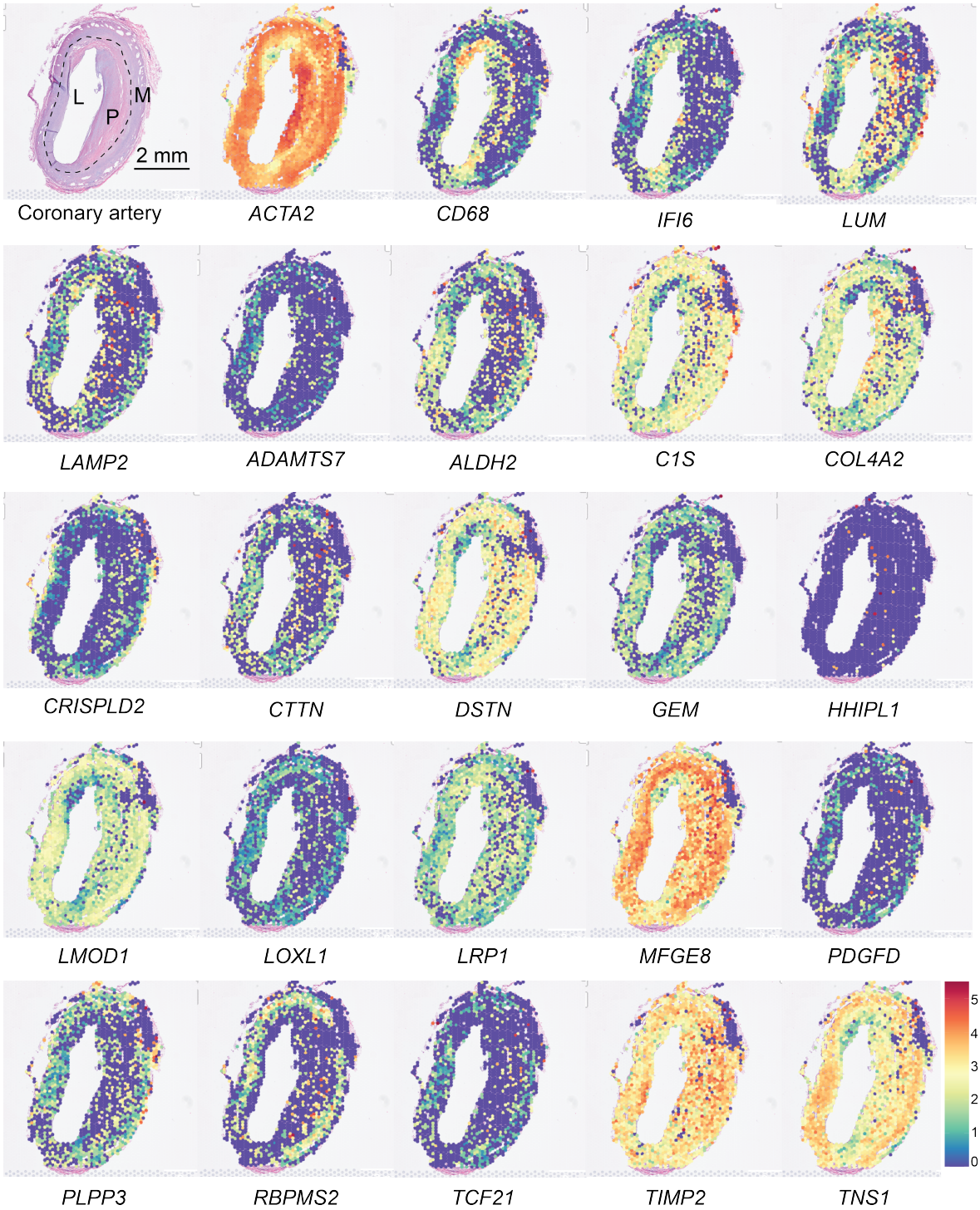



**Figure 6. Gene signatures in human SMCs. A.** Regulation of cholesterol and stretch-specific gene signatures (top 50 uniquely up-regulated genes after either cholesterol overload or stretch) scored in target gene knockdowns. The type of cellular assay (baseline, cholesterol overload, or stretch) is marked by dots and the estimated effect of the target gene on human CAD (protective, detrimental, or unknown) by circles. **B.** Correlation of co-expressed gene modules with the cellular assays. The color scale on the heatmap represents the Pearson's correlation coefficient between module eigengene (summarized expression profile) and coaxed cellular state. Asterisks highlight significant correlation: * P value <0.05, ** P value <0.01; *** P value <0.001. NMG, Non-module genes not clustered in any co-expressed module. **C.** Random Forest model and Gini index-based estimate of feature importance were used to prioritize identified module gene signature scores by their ability to discriminate between target genes with detrimental and protective effects on CAD risk. The higher the estimate, the better the gene module score predicts the opposite effects of target genes. **D.** Relationship between the top gene modules (M02 and M21) gene signature dynamics in target gene knockdowns, genetic effect direction of target genes, and cellular assay. The top 100 highly connected genes used in the analysis for the large M2 module.

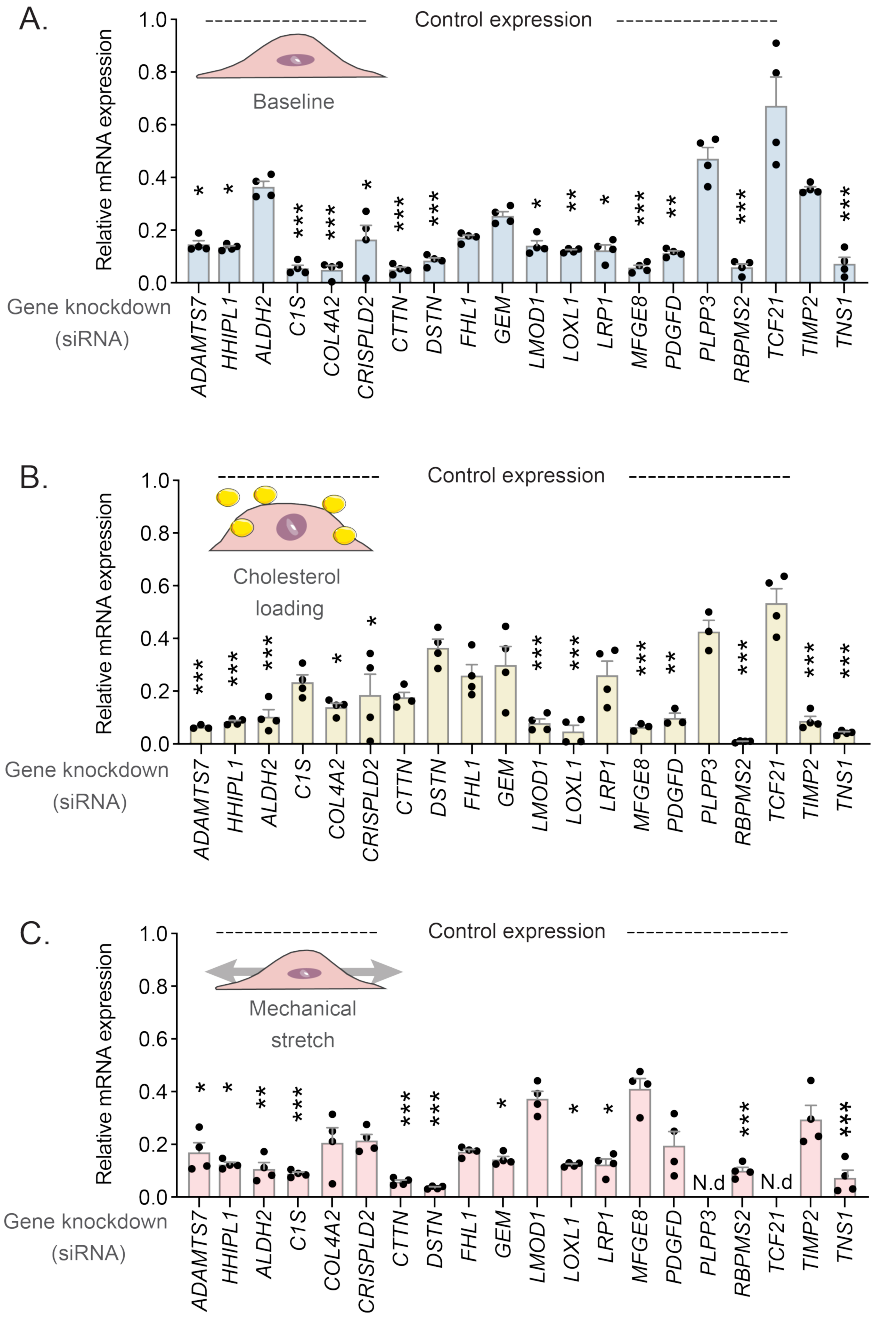
## Supplementary material:



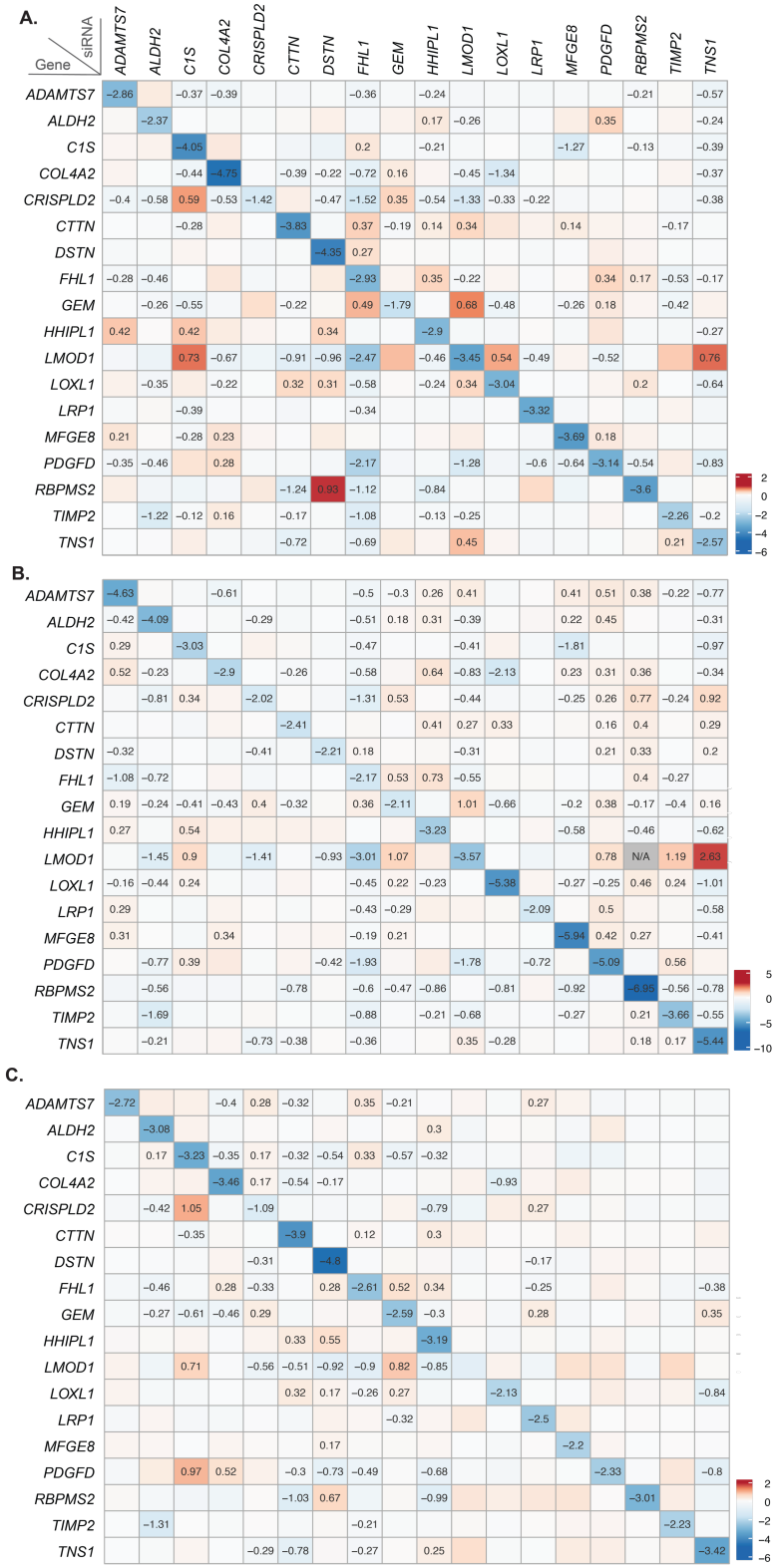

**Supplementary Figure 1: Expression of genes selected for the study in human atherosclerotic plaques. A.** Dot plot shows the expression of genes enriched in the mesenchymal supercluster and selected for further study. Normalized and log2-transformed gene expression values are averaged for every cell cluster. **B.** UMAP plots of the selected target genes as analyzed by integrated public scRNA-seq data on human atherosclerotic lesions. Scale is based on log2-transformed gene expression.



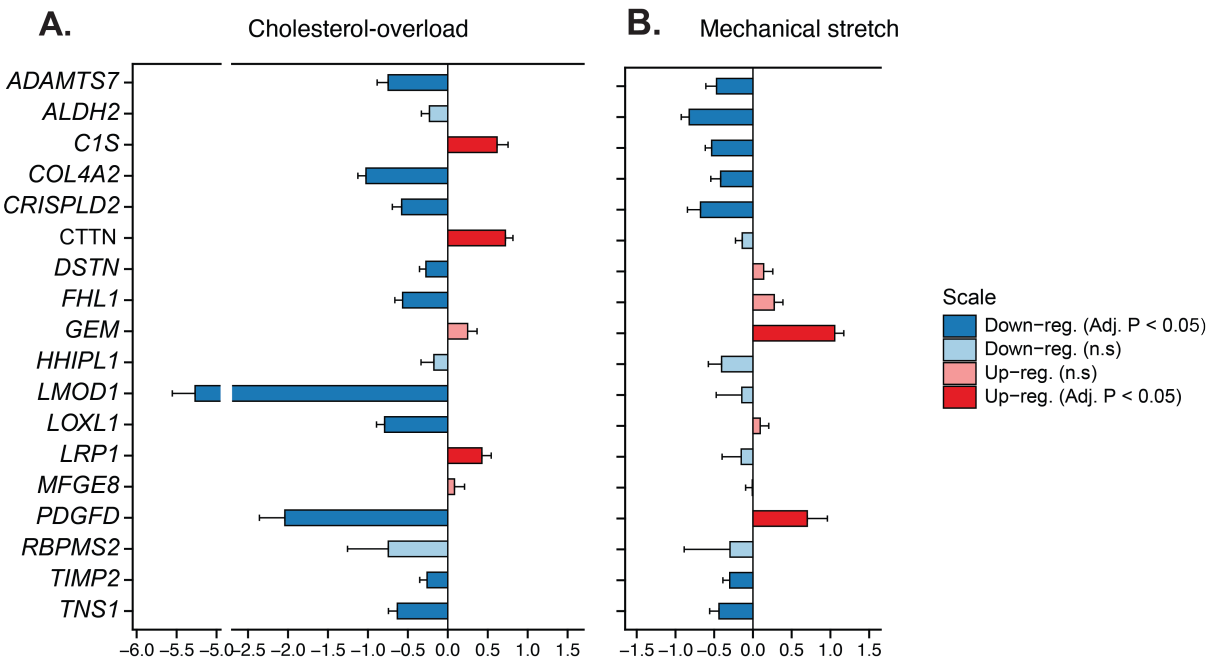

**Supplementary Figure 2: Spatial transcriptomics of selected genes in the study of human coronary arteries with atherosclerotic plaque.** Coronary artery plaque is shown in H&E staining *top left* M (media), P (plaque), and L (lumen). Selected markers are shown to identify cell populations (smooth muscle cells; *ACTA2*, macrophages; *CD68*, inflammatory; *IFI6*, modulated smooth muscle cells; *LUM*, and macrophage-like *LAMP2*). Barcoded spots represent a mixture between 5 and 10 cells. Shown is a representative section that contains 1746 spots. Expression intensity is designated by the scale color.



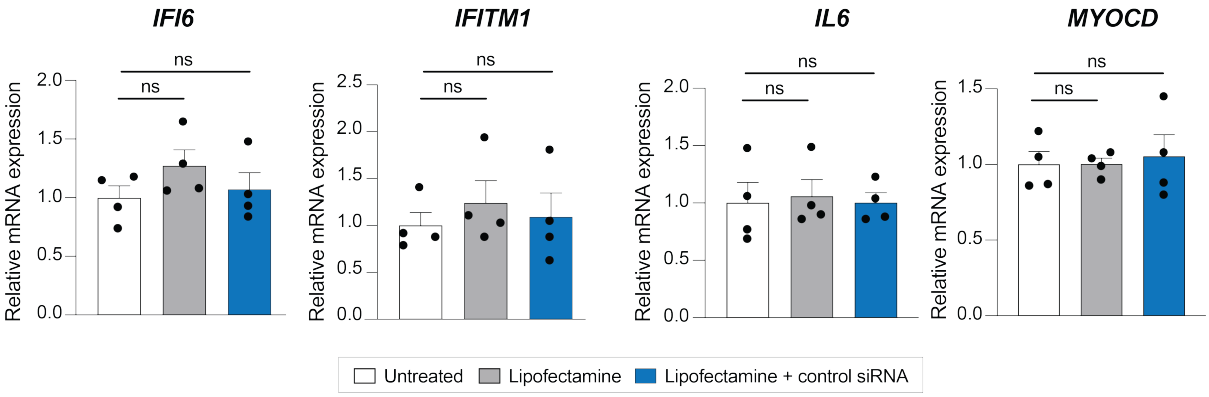

**Supplementary Figure 3: Validation of gene knockdown by real-time qPCR**. The knockdown efficiency of target genes in SMCs was evaluated by real-time qPCR under baseline (**A**), cholesterol loading (**B**), and mechanical stretch (**C**) conditions. Gene expression was normalized to the *HPRT1* housekeeping gene (control expression). N.d (gene expression was not determined). Four biological replicates were tested for each gene. * Adjusted P value < 0.05, ** P value < 0.01; and *** P value < 0.001 compared to control. Comparisons were analyzed by the Kruskal-Wallis test followed by Dunn's posthoc test with a control of false discovery rate by Benjamini and Hochberg method.



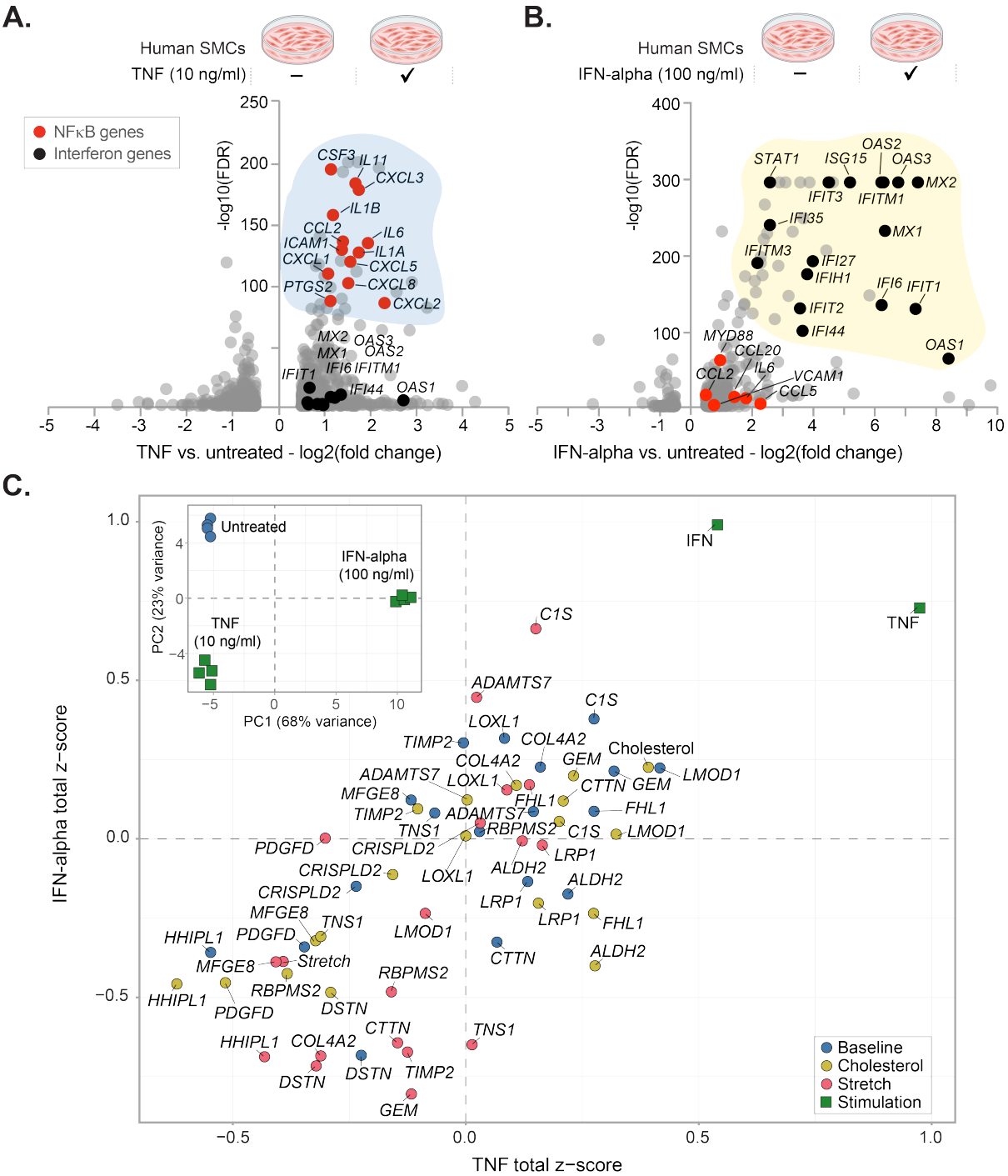

**Supplementary Figure 4. Knockdown efficiency of target genes by RNA-seq.** The knockdown efficiency of target genes was evaluated by RNA-seq analysis after baseline (**A**), cholesterol overloading (**B**), and mechanical stretch (**C**) conditions. Genes that are knocked down with siRNAs are located on the X-axis. The measured gene expression is on the Y-axis. Scale is based on log2 transformed fold-change values, and numbers are shown for only significant results with an adjusted P value < 0.05. N/A, the values are not available.



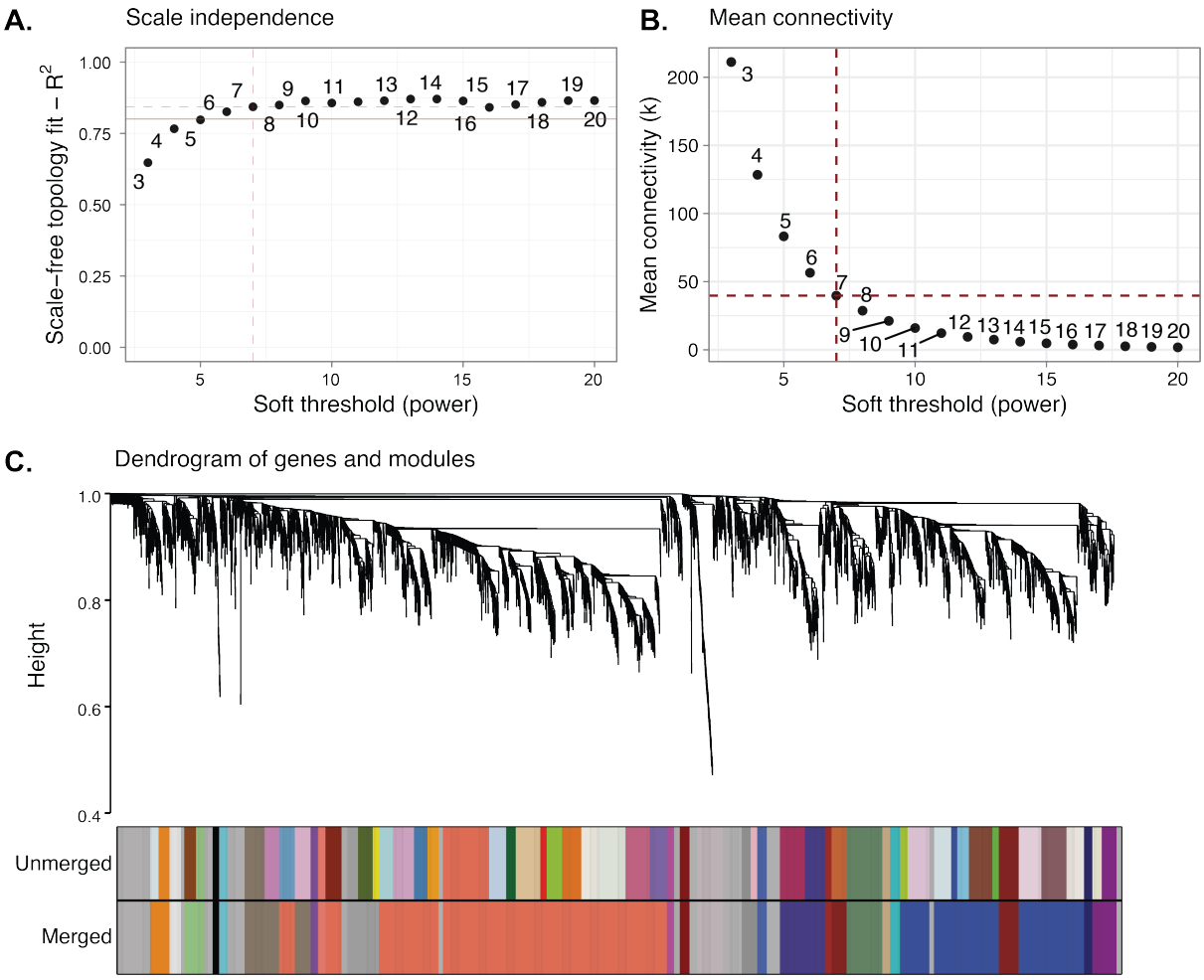

**Supplementary Figure 5: Regulation of target genes in the different cellular assays.** Differences in target gene expression in cholesterol overload (**A**) and mechanical stretch (**B**) assays compared to baseline. Bars represent log2-transformed fold change (LFC) of average gene expression, error bars are estimated standard error values of LFC. The color scale indicates the direction of fold change (up- and down-regulated genes), significant results (P value adjusted for false discovery rate < 0.05) are highlighted with a saturated color. n.s indicates not significant values.



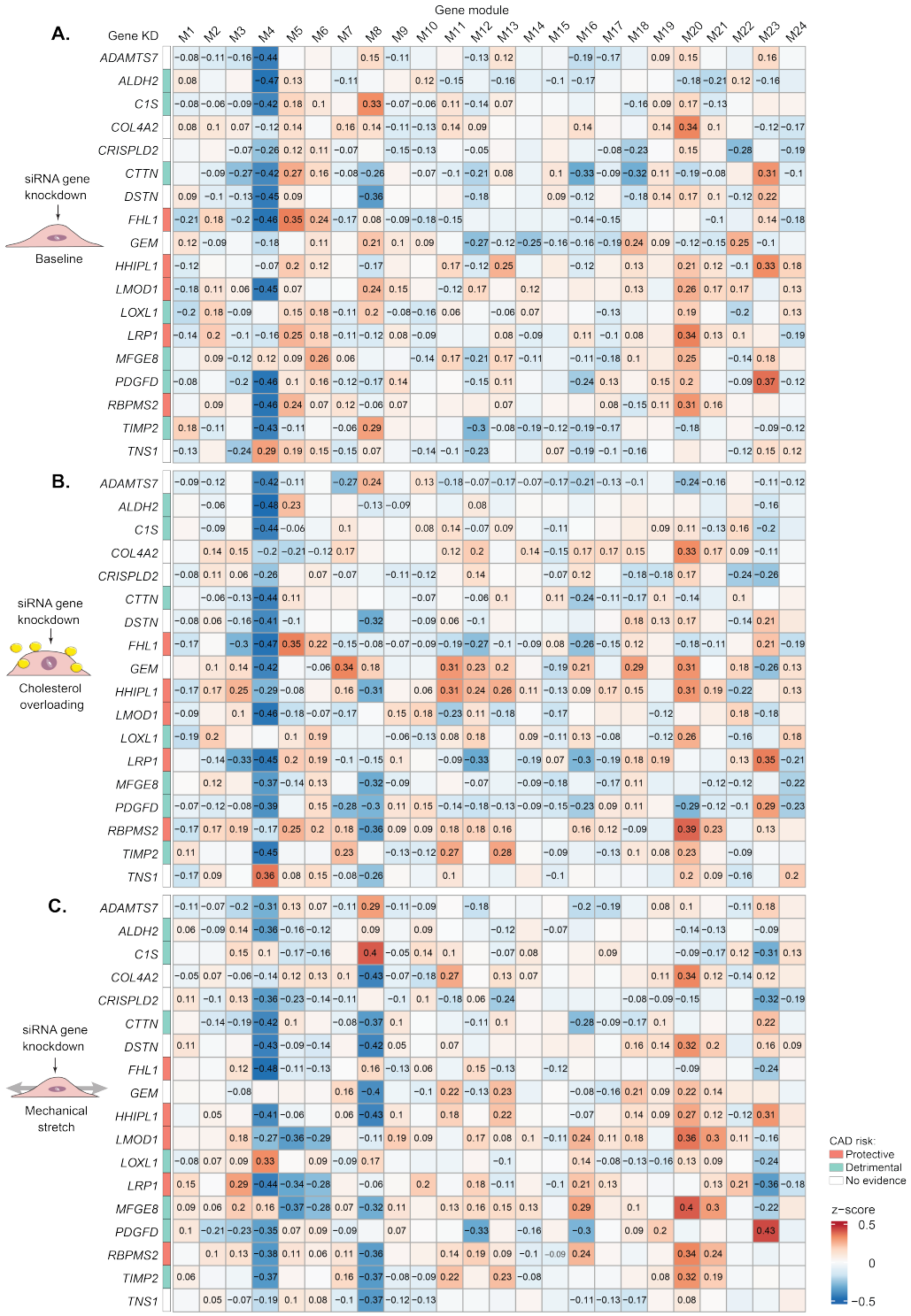

**Supplementary Figure 6: Effect of lipofectamine and control siRNA *per se* in knockdown experiments.** Shown is the expression of interferon-regulated genes (*IFI6, IFITM1*), TNF-regulated (*IL6*), or a prototypical contractile gene marker (*MYOCD*) in SMCs treated with lipofectamine, control siRNA or left untreated. Gene expression was normalized to the housekeeping gene *HPRT1*. Four biological replicates were tested for each gene. Statistical significance was tested by one-way ANOVA. Non-significant results (P > 0.05) are shown as ns.




**Supplementary Figure 7: NFκB and type I interferon gene signatures in human SMCs.** Transcriptomic changes in human SMCs treated with **(A)** tumor necrosis factor (TNF, 10 ng/ml) and with **(B)** interferon alpha (IFN-alpha, 100 ng/ml) for 24h compared to untreated cells. Shown in A and B are DEGs with a P value < 0.05 and absolute log2 fold-change ≥ 0.5. **C.** To understand how NFκB and type I interferon gene networks are regulated by selected CAD genes, we scored the alterations in the expression of TNF and IFN-alpha gene signatures in every target gene knockdown and cellular assay (baseline, cholesterol overload, or stretch). The scores of the TNF and IFN-alpha stimulated cells compared to unstimulated control are added for reference (green square symbols). The inserted plot in **C** shows the principal component analysis (PCA) of cells stimulated with TNF, IFN-alpha, or left untreated. Some elements in A and B were created with BioRender.com




**Supplementary Figure 8: Gene co-expression network analysis in control human aortic smooth muscle cells. A** and **B** show the soft threshold power analysis results to ensure an optimal scale-free model fit index and mean connectivity, respectively. The red solid line demarcates a lower cutoff for the fit index (R^2=0.8). Dashed lines highlight the selected soft threshold**. C.** Dendrogram plot of 5000 genes in a co-expression network clustered based on a dissimilarity measure and their assignment to different modules (shown as different colors).




**Supplementary Figure 9. Gene modules co-expressed in human SMCs and disturbed after target gene knockdown.** Expression-based scoring of gene signatures for SMC-specific modules of co-expressed genes in the different cellular assays: **A.** baseline, **B.** cholesterol overload, and **C.** mechanical stretch. For gene modules 1 to 10, only the top 100 genes with the highest intramodular connectivity are considered for this figure. The color scale represents a z-score of deviation in the transcriptional activity of the relevant module's gene signature in target knockdown compared with a control group. Only statistically significant z-score values are shown. Genetic direction is shown for genes associated with a decreased risk for coronary artery disease (CAD, in salmon color), increased (emerald color), or unknown (white color). Gene knockdown (KD).

**References:**


1. Lindstrom, M. *et al.* Global Burden of Cardiovascular Diseases and Risks Collaboration, 1990-2021. *J. Am. Coll. Cardiol.* **80**, 2372–2425 (2022).
2. Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V. & Roth, G. A. The Global Burden of Cardiovascular Diseases and Risk. *J. Am. Coll. Cardiol.* **80**, 2361–2371 (2022).
3. Bentzon, J. F., Otsuka, F., Virmani, R. & Falk, E. Mechanisms of plaque formation and rupture. *Circ. Res.* **114**, 1852–1866 (2014).
4. Jacobsen, K. *et al.* Diverse cellular architecture of atherosclerotic plaque derives from clonal expansion of a few medial SMCs. *JCI Insight* **2**, 95890 (2017).
5. Wirka, R. C. *et al.* Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat. Med.* **25**, 1280–1289 (2019).
6. Pan, H. *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* **142**, 2060–2075 (2020).
7. Alsaigh, T., Evans, D., Frankel, D. & Torkamani, A. Decoding the transcriptome of calcified atherosclerotic plaque at single-cell resolution. *Commun. Biol.* **5**, 1–17 (2022).
8. Carramolino, L. *et al.* Cholesterol lowering depletes atherosclerotic lesions of smooth muscle cell-derived fibromyocytes and chondromyocytes. *Nat. Cardiovasc. Res.* (2024) doi:10.1038/s44161-023-00412-w.
9. Örd, T. *et al.* Single-Cell Epigenomics and Functional Fine-Mapping of Atherosclerosis GWAS Loci. *Circ. Res.* **129**, 240–258 (2021).
10. Turner, A. W. *et al.* Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat. Genet.* **54**, 804–816 (2022).
11. Howson, J. M. M. *et al.* Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat. Genet.* **49**, 1113–1119 (2017).
12. Örd, T. *et al.* Dissecting the polygenic basis of atherosclerosis via disease-associated cell state signatures. *Am. J. Hum. Genet.* **110**, 722–740 (2023).
13. Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. 2021.05.24.21257377 Preprint at https://doi.org/10.1101/2021.05.24.21257377 (2021).
14. Kim, H.-J. *et al.* Molecular mechanisms of coronary artery disease risk at the PDGFD locus. *Nat. Commun.* **14**, 847 (2023).
15. Perisic Matic, L. *et al.* Phenotypic Modulation of Smooth Muscle Cells in Atherosclerosis Is Associated With Downregulation of LMOD1, SYNPO2, PDLIM7, PLN, and SYNM. *Arterioscler. Thromb. Vasc. Biol.* **36**, 1947–1961 (2016).
16. Nanda, V. *et al.* Functional regulatory mechanism of smooth muscle cell-restricted LMOD1 coronary artery disease locus. *PLoS Genet.* **14**, e1007755 (2018).
17. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
18. Zhu, H. *et al.* Aldehyde dehydrogenase 2 deficiency promotes atherosclerotic plaque instability through accelerating mitochondrial ROS-mediated vascular smooth muscle cell senescence. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1865**, 1782–1792 (2019).
19. Nakagaki-Silva, E. E. *et al.* Identification of RBPMS as a mammalian smooth muscle master splicing regulator via proximity of its gene with super-enhancers. *eLife* **8**, e46327 (2019).

20. Slabowska, A. O. *et al.* A systematic evaluation of state-of-the-art deconvolution methods in spatial transcriptomics: insights from cardiovascular disease and chronic kidney disease. *Front. Bioinforma.* **4**, (2024).
21. Bleckwehl, T. *et al.* Encompassing view of spatial and single-cell RNA-seq renews the role of the microvasculature in human atherosclerosis. 2023.12.15.571796 Preprint at https://doi.org/10.1101/2023.12.15.571796 (2023).
22. Campbell, J. H. & Campbell, G. R. Smooth Muscle Phenotypic Modulation—A Personal Experience. *Arterioscler. Thromb. Vasc. Biol.* **32**, 1784–1789 (2012).
23. Chamley-Campbell, J., Campbell, G. R. & Ross, R. The smooth muscle cell in culture. *Physiol. Rev.* **59**, 1–61 (1979).
24. Rong, J. X., Shapiro, M., Trogan, E. & Fisher, E. A. Transdifferentiation of mouse aortic smooth muscle cells to a macrophage-like state after cholesterol loading. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13531–13536 (2003).
25. Vengrenyuk, Y. *et al.* Cholesterol loading reprograms the microRNA-143/145-myocardin axis to convert aortic smooth muscle cells to a dysfunctional macrophage-like phenotype. *Arterioscler. Thromb. Vasc. Biol.* **35**, 535–546 (2015).
26. Conklin, A. C. *et al.* Meta-Analysis of Smooth Muscle Lineage Transcriptomes in Atherosclerosis and Their Relationships to In Vitro Models. *Immunometabolism* **3**, e210022 (2021).
27. Jensen, L. F., Bentzon, J. F. & Albarrán-Juárez, J. The Phenotypic Responses of Vascular Smooth Muscle Cells Exposed to Mechanical Cues. *Cells* **10**, 2209 (2021).
28. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
29. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2022).
30. Agrawal, A. *et al.* WikiPathways 2024: next generation pathway database. *Nucleic Acids Res.* gkad960 (2023) doi:10.1093/nar/gkad960.
31. Guo, X. *et al.* Transfection reagent Lipofectamine triggers type I interferon signaling activation in macrophages. *Immunol. Cell Biol.* **98**, 88–88 (2020).
32. Echeverri, C. J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat. Rev. Genet.* **7**, 373–384 (2006).
33. Dobnikar, L. *et al.* Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. *Nat. Commun.* **9**, 4567 (2018).
34. Shankman, L. S. *et al.* KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nat. Med.* **21**, 628–637 (2015).
35. Mack, C. P. Signaling Mechanisms That Regulate Smooth Muscle Cell Differentiation. *Arterioscler. Thromb. Vasc. Biol.* **31**, 1495 (2011).
36. Foley, C. N. *et al.* A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat. Commun.* **12**, 764 (2021).
37. Benn, M. & Nordestgaard, B. G. From genome-wide association studies to Mendelian randomization: novel opportunities for understanding cardiovascular disease causality, pathogenesis, prevention, and treatment. *Cardiovasc. Res.* **114**, 1192–1208 (2018).
38. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
39. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).

40. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
41. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
42. Webb, T. R. *et al.* Systematic Evaluation of Pleiotropy Identifies 6 Further Loci Associated With Coronary Artery Disease. *J. Am. Coll. Cardiol.* **69**, 823–836 (2017).
43. Allahverdian, S., Chehroudi, A. C., McManus, B. M., Abraham, T. & Francis, G. A. Contribution of intimal smooth muscle cells to cholesterol accumulation and macrophage-like cells in human atherosclerosis. *Circulation* **129**, 1551–1559 (2014).
44. Basatemur, G. L., Jørgensen, H. F., Clarke, M. C. H., Bennett, M. R. & Mallat, Z. Vascular smooth muscle cells in atherosclerosis. *Nat. Rev. Cardiol.* **16**, 727–744 (2019).
45. Francis, G. A. The Greatly Under-Represented Role of Smooth Muscle Cells in Atherosclerosis. *Curr. Atheroscler. Rep.* **25**, 741–749 (2023).
46. Allahverdian, S., Chaabane, C., Boukais, K., Francis, G. A. & Bochaton-Piallat, M.-L. Smooth muscle cell fate and plasticity in atherosclerosis. *Cardiovasc. Res.* **114**, 540–550 (2018).
47. Alonso-Herranz, L., Albarrán-Juárez, J. & Bentzon, J. F. Mechanisms of fibrous cap formation in atherosclerosis. *Front. Cardiovasc. Med.* **10**, 1254114 (2023).
48. Aherrahrou, R. *et al.* Genetic Regulation of SMC Gene Expression and Splicing Predict Causal CAD Genes. *Circ. Res.* **132**, 323–338 (2023).
49. Aherrahrou, R. *et al.* Secreted Protein Profiling of Human Aortic Smooth Muscle Cells Identifies Vascular Disease Associations. *Arterioscler. Thromb. Vasc. Biol.* **44**, 898–914 (2024).
50. Ackers-Johnson, M. *et al.* Myocardin Regulates Vascular Smooth Muscle Cell Inflammatory Activation and Disease. *Arterioscler. Thromb. Vasc. Biol.* **35**, 817–828 (2015).
51. Tang, R.-H. *et al.* Myocardin inhibits cellular proliferation by inhibiting NF-kappaB(p65)-dependent cell cycle progression. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3362–3367 (2008).
52. Sui, Y. *et al.* IKKβ links vascular inflammation to obesity and atherosclerosis. *J. Exp. Med.* **211**, 869–886 (2014).
53. Kong, P. *et al.* Inflammation and atherosclerosis: signaling pathways and therapeutic intervention. *Signal Transduct. Target. Ther.* **7**, 1–24 (2022).
54. Althoff, T. F. & Offermanns, S. G-protein-mediated signaling in vascular smooth muscle cells - implications for vascular disease. *J. Mol. Med. Berl. Ger.* **93**, 973–981 (2015).
55. Chattopadhyay, A. *et al.* Cholesterol-Induced Phenotypic Modulation of Smooth Muscle Cells to Macrophage/Fibroblast–like Cells Is Driven by an Unfolded Protein Response. *Arterioscler. Thromb. Vasc. Biol.* **41**, 302 (2020).
56. 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.* **122**, 433–443 (2018).
57. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
58. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
59. Sun, B. *et al.* Double-jeopardy: scRNA-seq doublet/multiplet detection using multi-omic profiling. *Cell Rep. Methods* **1**, None (2021).

60. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329-337.e4 (2019).
61. Germain, P.-L., Lun, A., Meixide, C. G., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using *scDblFinder*. Preprint at https://doi.org/10.12688/f1000research.73600.2 (2022).
62. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281-291.e9 (2019).
63. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
64. THE TABULA SAPIENS CONSORTIUM. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
65. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
66. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
67. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
68. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).
69. The Gene Ontology Consortium *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
70. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).
71. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
72. Wingender, E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief. Bioinform.* **9**, 326–332 (2008).
73. Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815 (2022).
74. Oldham, M. C., Langfelder, P. & Horvath, S. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Syst. Biol.* **6**, 63 (2012).
75. Foroutan, M. *et al.* Single sample scoring of molecular phenotypes. *BMC Bioinformatics* **19**, 404 (2018).
76. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **2**, 18–22 (2002).