The 9p21.3 coronary artery disease risk locus drives vascular smooth muscle cells to an osteochondrogenic state

Elsa Salido[1], Carolina de Medeiros Vieira[1], José Verdezoto Mosquera[2], Rohan Zade[1], Clint L. Miller[2] and Valentina Lo Sardo[1†]

[1]Department of Cell and Regenerative Biology; University of Wisconsin-Madison; Madison, WI 53705 USA
[2]Center for Public Health Genomics, Department of Public Health Sciences, Department of Biochemistry and Molecular Genetics; University of Virginia; Charlottesville, VA 22908 USA

[†]Corresponding Author:

Dr. Valentina Lo Sardo, PhD

Department of Cell & Regenerative Biology
University of Wisconsin-Madison
School of Medicine and Public Health
Madison, WI 53705
Email: losardo@wisc.edu

Abstract

Genome-wide association studies have identified common genetic variants at ~400 human genomic loci linked to coronary artery disease (CAD) susceptibility. Among these genomic regions, the most impactful is the 9p21.3 CAD risk locus, which spans a 60 kb gene desert and encompasses ~80 SNPs in high linkage disequilibrium. Despite nearly two decades since its discovery, the functional mechanism of this genomic region remains incompletely resolved. To investigate the transcriptional gene programs mediated by 9p21.3 risk locus, we applied a model of induced pluripotent stem cells (iPSCs) from risk and non-risk donors at 9p21.3, as well as isogenic lines with a full haplotype deletion. Upon differentiation to vascular smooth muscle cells (VSMC), single-cell transcriptomic profiling demonstrated iPSC-VSMC phenotypes resembling those from native human coronary arteries, confirming the robustness of this model. Remarkably, our analyses revealed that VSMCs harboring the 9p21.3 risk haplotype preferentially adopt an osteochondrogenic state. Importantly, we identified LIMCH1 and CRABP1 as signature genes critical for defining this transcriptional program. Our study provides new insights into the mechanism at the 9p21.3 risk locus and defines its role in driving a disease-prone transcriptional state in VSMCs.

Introduction

In the past two decades, GWASs have made significant advancements in identifying genetic variants associated with CAD[1–4]. However, translating this knowledge into functional and mechanistic insights remains a challenging task. One of the key bottlenecks is the lack of an appropriate model system to evaluate the causal link between human-specific genetic traits and the biology of the tissues and cell types implicated in the disease. As the majority of genetic variants associated with CAD reside in noncoding regions, they are predicted to impact gene regulatory mechanisms[5]. Furthermore, the lack of conservation across species makes it difficult to use standard animal models to dissect human-specific CAD risk loci[6]. How noncoding genetic risk factors for CAD influence the ability of vascular cells to acquire disease-prone states is not well defined.

The 9p21.3 CAD risk locus was discovered in 2007 through three independent CAD GWASs[7–9]. This 60 kb genomic region is unique to humans and contains approximately 80 SNPs in strong linkage disequilibrium in most populations, presenting a challenge to pinpoint causal CAD risk variants[7–9]. The locus resides in a gene desert region and only partially overlaps with the extensively spliced long noncoding RNA (lncRNA) gene *CDKN2B-AS1* (ANRIL)[10]. The lack of conservation of this region across species has posed challenges for conducting *in vivo* studies; the 9p21.3 locus is poorly syntenic in rodents, and no study has identified a murine homolog of ANRIL[11]. A previous study reported the deletion of a putative syntenic region in the noncoding region downstream of the CDKN2A/B genes on murine chromosome 4. These mice exhibited increased tumor incidence and increased aortic smooth muscle cell proliferation but not atherosclerosis[12]. Given the lack of conservation of this genomic region across most mammals, except

primates, newer approaches that maintain the complexity of the human genome are urgently needed to study the function of the human 9p21.3 locus.

Vascular smooth muscle cells (VSMCs) are the most abundant cells in arteries and exhibit remarkable plasticity, characterized by their ability to transition to diverse cell states under physiological conditions and during CAD pathogenesis[13,14]. VSMCs are traditionally separated into two major cell states: 1) “contractile” quiescent cells that regulate vascular tone and 2) “synthetic” highly proliferative and highly migratory cells that produce extracellular matrix (ECM) and have reduced contractile marker expression[13–15]. Recent studies using single-cell RNA and ATAC sequencing of healthy and atherosclerotic coronary arteries have revealed more granular VSMC phenotypic states[16–20]. For instance, VSMCs can transition to macrophage-like and foam-like cells[21], fibromyocytes[19], SEM (Stem cell, Endothelial cell, and Monocyte intermediary state)[16], fibrochondrocytes[20,22,23] and osteoblast-like cells[22]. However, to date, the impact of the 9p21.3 CAD locus on VSMC states and transitions has not been described.

To determine the role of the 9p21.3 risk haplotype in influencing cell state changes in VSMCs, we used a system that we have previously generated and validated, based on induced pluripotent stem cells (iPSCs), vascular smooth muscle cell differentiation and haplotype editing[24]. Using this system, we conducted comprehensive single-cell transcriptional profiling to identify VSMC states and potential state trajectories mediated by the risk haplotype at the 9p21.3 locus. Here we show for the first time that, in an isogenic background, the risk haplotype at 9p21.3 drives VSMCs to acquire an osteochondrogenic-like state. Risk iPSC-VSMCs show increased expression of SOX9 and COL2A1, despite maintaining VSMC identity. Our pipeline revealed two signature genes, accompanying SOX9 and COL2A1, LIMCH1 and CRABP1, defining the 9p21.3-risk-driven VSMC state. Moreover, our analysis revealed that the risk haplotype at 9p21.3 alters the expression of genes at other CAD loci, suggesting that an interconnected CAD gene network may increase risk susceptibility in concert.

RESULTS

iPSC-VSMCs are reflective of ex vivo human SMCs and reveal a distinct transcriptomic landscape caused by the CAD risk haplotype at 9p21.3

To explore the impact of the 9p21.3 CAD risk locus on cell state transition and VSMCs plasticity, we leveraged a system based on induced pluripotent stem cells (iPSCs), we previously generated and characterized[24]. Briefly, iPSC lines were derived from both risk and nonrisk homozygous backgrounds at 9p21.3. TALE-mediated genome editing was used to delete the 60 kb CAD region at 9p21.3[24]. These isogenic iPSC lines share identical genetic backgrounds and differ exclusively in the 9p21.3 CAD region, allowing strict and controlled evaluation of the function of 9p21.3 in determining cellular phenotypes. We established a collection of lines composed of four genotypes: risk cells with an intact haplotype (RR), isogenic risk knockouts (RRKO), nonrisk cells with an intact haplotype (NN), and nonrisk knockouts (NNKO). To control for potential variability across iPSC lines, we used two clones per genotype for parallel

vascular smooth muscle cell (VSMC) differentiation. A total of 8 lines (carrying the 4 genotypes) were differentiated to VSMCs over the course of 17 days as we previously described[24]. At D17, the iPSC-VSMCs were harvested, and a total of ~27,000 cells were used for 10X Chromium 3' GEM single-cell sequencing (Figure 1A). After QC and removal of low-quality cells (detailed in Methods), we visualized our data by uniform manifold approximation and projection (UMAP) embeddings to evaluate gene expression patterns across our samples.

To establish the robustness of this *in vitro* system in mimicking endogenous cellular phenotypes, we performed gene expression comparison with publicly available datasets of human heart arteries[17], batch integration and co-embedding of data from iPSC-VSMCs and *ex vivo* native VSMCs from human coronary or carotid arteries[20]. We evaluated the expression levels of established SMC markers, namely, *ACTA2*, *CNN1*, and *TAGLN* (Figure 1B), in parallel with publicly available scRNA sequencing data from healthy human heart arteries[17]. After *de novo* clustering of the human artery data and cell type annotation through cell type-specific markers, we confirmed that iPSC-VSMCs represent an overall homogenous population expressing VSMC markers (Figure 1B and Extended data 1). To further support the validity of iPSC-VSMC model, we leveraged the newly established MetaPlaq atlas, a multidataset atlas of human coronary and carotid artery cells[20]. Co-embedding of the iPSC-VSMCs and the human artery scRNA-seq atlas in a low dimensional space showed strong qualitative overlap between iPSC-VSMCs and the *ex vivo* SMC cluster. However, we detected minimal overlap only of a few cells with the fibroblast population (Fibs) and a small subset of iPSC-VSMCs projected on the endothelial cell (EC) cluster, suggesting similarities between iPSC-VSMC and *ex vivo* human SMC transcriptomic profiles (Figure 1C).

We performed an unsupervised clustering analysis including all genotypes of iPSC-VSMCs and detected 10 transcriptionally distinct cellular clusters (Figure 2A). Although iPSC-VSMCs exhibit overall homogeneous expression of cell type-specific markers, this analysis showed that our system can detect different cell subpopulations, potentially representing different cellular states. To establish how different 9p21.3 haplotypes may influence the transcriptional landscape of VSMCs, we analyzed and quantified the cluster composition of our iPSC derivatives sorted by genotype (Figure 2A). Remarkably, cluster composition analysis revealed that clusters 0 and 1 account for ~50% of the NN, NNKO, and RRKO samples, while the RR cells are mostly devoid of these clusters. Clusters 2 and 5 are equally represented across all genotypes. Cluster 8 is mostly represented in RRKO and NNKO cells, and cluster 7 is mainly represented in NNKO cells. While NN, NNKO, and RRKO cells share similar cluster patterns, 72% of RR VSMCs comprise clusters 3, 4, and 6 (Figure 2A). Grouping the cells by genotype highlights the presence of four distinct UMAP areas (Figure 2B). RR cells tend to group together and away from other genotypes (NN, NNKO, RRKO), which lack the risk haplotype at 9p21.3. The remaining genotypes exhibit intermixing in three regions of the UMAP. RR cells mostly populated three clusters, 3, 4, and 6, each comprising 80% or more of the risk cells (Figure 2B). NN, NNKO, and RRKO are intermixed in multiple clusters, suggesting that the deletion of the 9p21.3 CAD risk region mostly resembles a NN-like status. These data highlight

the presence of transcriptionally independent cellular clusters in the overall population of iPSC-VSMCs, with three clusters populated exclusively by VSMCs carrying the risk haplotype at 9p21.3, indicating potential independent cellular states. Qualitative analysis of cell cultures across the genotypes, (RR, RRKO, NN, and NNKO) showed comparable cell differentiation features and density at this differentiation stage (Figure 2C).

Since the 9p21.3 CAD region resides close to the cyclin-dependent kinase inhibitors CDKN2A and CDKN2B, we evaluated the expression of these genes in the distinct clusters in VSMCs, which could possibly explain the divergent state (Figure 3A). Interestingly, *CDKN2A* and *CDKN2B* did not localize to any specific clusters and were not enriched in RR VSMCs, excluding the possibility that the difference in the RR VSMC transcriptomic profile is due to genotype-specific expression of these genes (Figure 3B). Previously, we reported that *CDKN2B-AS1* (*ANRIL)*, the lncRNA that partially overlaps the CAD region at 9p21.3, has genotype-specific isoform expression in iPSC-VSMCs harboring different genotypes at 9p21.3. Specifically, total *ANRIL* levels are more highly expressed in RR VSMCs compared to NN VSMCs. While long *ANRIL* isoforms, which terminate in exon 20, are equally expressed in RR and NN VSMCs, short *ANRIL* isoforms are expressed exclusively in RR VSMCs[24,25]. Here, we confirmed increased *ANRIL* expression, attributable to short *ANRIL* isoforms, as the long isoforms did not differ by genotype (Figure 3C).

The 9p21.3 risk haplotype reduces VSMC plasticity

To identify cluster-specific gene expression signatures, we conducted differential expression analysis (DEA) and identified the top 10% of cluster-specific genes across all clusters (Figure 4A). We highlighted a set of genes differentially expressed specifically in risk clusters 3-4-6 (Figure 4A). Gene Ontology (GO) enrichment analysis on all significant (Bonferroni corrected p value < 0.05) genes driving each cluster revealed two main patterns of biological processes in the NN, NNKO, and RRKO cells: (1) clusters 5 and 7 are enriched for cell cycle-related processes and concomitantly depleted of extracellular matrix organization processes. (2) clusters 0, 1, 2, and 8 are enriched in muscle development, wound healing and adhesion processes and depleted of cell cycle-related processes (Extended data 2). Remarkably, this analysis revealed a different trend in RR cells. Several terms related to healthy SMC function, such as cell adhesion and wound healing, were depleted in two of the three risk clusters (4-6). This trend was accompanied by enriched cell cycle genes in cluster 4 and non-SMC lineages, including genes related to cartilage development and axonogenesis in clusters 3 and 6, respectively (Figure 4B). Interestingly, Cluster 6 exhibited strong enrichment for neuron-related processes.

To support these findings we performed a closer examination using the GTEx human scRNA-seq atlas. We analyzed the expression pattern of the genes in the neural-linked GO terms across different tissues, including arteries and several tissues with high SMC content (Ovary, Bladder, Lung, etc.) (Extended data 3A). We detected global expression of these genes across many tissues, suggesting that the enrichment

in neuron-like GO terms is a result of biased functional annotation of most of these genes, rather than an alternative differentiation of RR VSMCs into neuron-like derivatives (Extended data 3A). Further, we did not observe significant enrichment of neuronal-specific genes in the risk cells (Extended data 3B). Thus, the presence of gene signatures related to axon guidance and neuronal projections may instead reflect morphological changes occurring in RR VSMCs.

VSMCs are known to exhibit high plasticity and have been shown to acquire numerous transitional states[14,26,27]. Given the distinct gene expression signatures in RR VSMCs, we next investigated whether different genotypes at 9p21.3 could drive divergent transcriptional trajectories. We independently evaluated the dynamics of cell cluster transitions across each genotype using pseudotime analysis in Monocle3 (Figure 5A). We hypothesized that RR cells would follow a divergent cell trajectory, while the other three genotypes would follow a similar trajectory. We selected cluster 2 as an anchor for the trajectory analysis, given the balanced cell compositions by all four genotypes and we evaluated the trajectory of gene expression separately by genotype. The NNKO and RRKO lines converged on the same trajectory, predicting that the absence of the 9p21.3 haplotype has an equal effect on transcriptional states, independent of the genetic background. The *KO trajectory* is characterized by a transition from cluster 2 through 0 and terminating at 7. Gene ontology analysis of these clusters indicated a transition from a quiescent (clusters 2 and 0) to a more proliferative state (cluster 7). The *nonrisk trajectory* mirrors the KO transition except for the last stage, terminating in cluster 0 instead of 7, suggesting that fewer NN cells transition to a proliferative (synthetic-like) state than NNKO or RRKO cells (Figure 5A-B). Analysis of the *risk trajectory* substantiated our hypothesis of divergent state acquisition for RR VSMCs. The risk states populate the upper right portion of the UMAP, moving from cluster 2 through clusters 3, 4, and 6. Gene Ontology showed increasing non-VSMC signatures in these clusters, suggesting that the risk haplotype drives VSMCs to acquire a divergent state (Figure 5A-B).

To date, VSMC phenotypic modulation has been described mostly in murine models of atherosclerosis[14,26,27]. Through rigorous lineage tracing studies in mice and, more recently, through single-cell transcriptomic analysis of human artery samples, numerous VSMC state nuances have been identified. To investigate whether the cell state transition of our *in vitro* system recapitulates previously identified VSMC states or phenotypic switches, we analyzed the expression of known master regulators of VSMC phenotypic modulation. The traditional modulators of VSMC phenotypic switch, MYOCD and SRF[28,29], did not localize to any SMC cluster by genotype. Moreover, the stem cell marker KLF4[30–32] was not detected in our dataset (Extended data 4A). Overall, the expression patterns of these transcriptional regulators, which were previously identified to play a role in VSMC modulation, do not differ by genotype or SMC cluster, excluding the possibility that they drive the RR cell state. Previous literature has reported a variety of disease states for VSMC in CAD wherein they transdifferentiate to other cell types or take on aberrant cell states[14,16,19,20,26,32]. We then sought to investigate markers of CAD-related VSMC states in RR VSMCs from our dataset. We used a single-cell dataset from healthy human arteries (Hu *et al*. 2021) together with

two additional published datasets using endogenous atherosclerotic human arteries (Wirka *et al*. 2019; Pan *et al*. 2020) and looked for evidence of either transition to fibromyocytes and SEM (Stem cell, Endothelial cell, and Monocyte intermediary state), or transdifferentiation to other cell types, including macrophages or endothelial cells. However we did not detect expression of known fibromyocyte markers (*TNFRSF11B or VCAM1*) (Extendend data 4B) in iPSC-VSMCs overall, precluding the possibility that RR VSMCs acquire one of these known VSMC states (Extended data 4B). We observed a similar lack of expression of macrophage (*CD68* and *AIF1*) or endothelial cell (*PECAM1* and *VWF*) (Extended data 4C) markers, suggesting that the RR VSMCs are unlikely to originate from transdifferentiation or alternative differentiation to other cell types. Overall, this analysis shows no significant resemblance between the RR cells and any of the cell types and states we surveyed (Extended data 4).

In mouse models of atherosclerosis, *Lgals3* was previously linked to macrophage-like VSMCs during atherosclerosis progression[33]. More recently, *Lgals3* has been described as a marker of VSMC plasticity and an indicator of a transitional state for subpopulations of VSMCs that undergo phenotypic modulation[31]. Interestingly, *LGALS3* was uniformly expressed in all clusters populated by RRKO, NN and NNKO cells, while the expression of *LGALS3* was significantly decreased in RR cells (Figure 6A). This pattern suggests that NN, NNKO and RRKO cells exhibit a dynamic transitional phenotype, while RR cells likely lack plasticity and do not undergo known phenotypic modulation.

To assess whether other genes within known CAD genomic loci are affected in the iPSC-VSMC RR population, we analyzed the expression of 31 genes residing in known CAD loci[1] and found 13 genes that were homogeneously expressed in the Risk population and significantly differentially expressed between the RR and other genotypes (Figure 6B). This may suggest that the 9p21.3 locus influences multiple other genes linked to CAD and that the risk status of the cells may arise from a cumulative effect of gene expression alterations at other genomic loci.

Risk iPSC-VSMCs exhibit an osteochondrogenic-like state, and novel CAD signature genes

Informed by the GO analysis, we explored the possibility that RR VSMCs acquire an osteo/chondrogenic state. *SOX9* and *RUNX2* are master regulators of bone and cartilage differentiation and have been shown to be upregulated, in CAD vessels, in VSMCs transitioning toward osteogenic- and chondrogenic-like cells[22,23,26]. Recently, *SOX9* has been linked to vascular stiffening and aging through its modulation of collagen expression[34]. Evaluation of key markers for osteo- and chondrogenic lineages between RR cells and their isogenic RRKO cells revealed overall increased expression of both lineage markers in the presence of 9p21.3 risk (Figure 7A). Remarkably, *SOX9* was uniformly expressed at higher levels in RR cells compared to RRKO cells (Figure 7B-C). Additionally, *SOX9* upregulation was accompanied by significant upregulation of *COL2A1* and concomitant downregulation of *COL1A1* and *COL1A2*, suggesting that the cells acquire a chondrogenic fate and alter ECM production (Figure 7B-C).

Using the most comprehensive human artery single-cell atlas available, MetaPlaq[20], we examined whether iPSC-VSMCs resemble VSMC states in *ex vivo* arterial samples. Integration of SMCs and fibroblasts from the atlas with iPSC-VSMCs revealed a remarkable overlap of iPSC-VSMCs and *ex vivo* arterial SMCs (Figure 7D and Extended data 5). Using fine-grained cell annotations in this atlas, we demonstrated that iPSC-VSMCs mostly overlap with Transitional SMCs, Fibromyocytes, Foam-like SMCs and Fibrochondrocytes (Figure 7E). Pseudotime trajectory analysis incorporating SMCs and fibroblasts from human arteries and iPSC-VSMCs from this study showed that the nonrisk iPSC-VSMCs (NN, NNKO, RRKO) mostly overlap with transitional populations of SMCs, corroborating the findings obtained using *LGALS3* as a marker of phenotypic modulation. Interestingly, RR cells resemble fibrochondrogenic SMCs and partially foam-like cells. Additionally, we identified a divergent cell population, mostly derived from RR VSMCs (Figure 7F), which may represent a less abundant transitional cell population in the current atlas.

To better define the transitional state of the RR VSMC population and identify new signature genes of 9p21.3 risk-specific phenotypes, we conducted a differential expression analysis between RR and all other genotypes and selected the most differentially expressed genes in the risk cells by p-value and fold change. Closer observation of the top 50 DEGs revealed only 11 that were uniformly and uniquely expressed in the entire RR cell population. We highlight two genes with the highest fold enrichment and p-value for expression in RR cells: *LIMCH1* and *CRABP1* (Figure 8A).

Immunocytochemistry confirmed that the expression of transcripts resulted in a high abundance of the LIMCH1 protein across the entire RR population and highlighted the expression of CRABP1 in ~20% of RR VSMCs. The expression of these genes was not detected in NN cells, as predicted by the scRNAseq analysis (Figure 8C). To confirm the unique expression of these signature genes across multiple samples, we generated iPSC-VSMCs from 2 additional RR donors and 2 NN donors. Gene expression analysis by qPCR confirmed that the expression of *LIMCH1* and *CRABP1* was greater in the VSMCs harboring the risk haplotype than in those harboring the NN haplotype (Fig 8B). Time course expression analysis demonstrated early expression of both genes around day 9 of the differentiation protocol which increased over the next 7-9 days. The expression of these genes remained negligible in the other genotypes (Extended data 6). Next, we analyze the coexpression pattern of *LIMCH1* and *CRABP1* across the different genotypes. *LIMCH1*$^+$ cells account for 64.7% of the RR VSMC population. Analysis of coexpression with the other signature genes revealed that 19.2% are exclusively expressing *LIMCH1*, 20.3% are *LIMCH1*$^+$-*SOX9*$^+$, 16.4% *LIMCH1*$^+$-*SOX9*$^+$-*CRABP1*$^+$ and 8.8% are *LIMCH1*$^+$-*CRABP1*$^+$. RRKO, NN, and NNKO cells have 7-10% of cells expressing negligible levels of *LIMCH1* alone or *LIMCH1*+-*SOX9*+, indicating that this signature predominates with RR VSMCs (Figure 8D). Finally, we evaluated the expression of these genes in MetaPlaq and found that *LIMCH1* and *CRABP1* were enriched in a few cell populations. *LIMCH1* can be found throughout the SMC population, with slight enrichment in the fibromyocyte population and a subpopulation of endothelial cells previously predicted to undergo EndoMT (Figure 8E). Interestingly, *CRABP1* was mostly found in fibroblasts but was found in a subpopulation of cells predicted to resemble

myofibroblasts. Together, our analyses reveal signature genes in iPSC-VSMC resembling novel unidentified cellular states in human arterial tissue, which warrant further investigation to assess their role in atherosclerosis.

DISCUSSION

Genome-wide association studies (GWASs) have identified ~400 genomic loci linked to CAD susceptibility[1,35,36]. However, the functions of these genomic regions, target genes, and relevant cell types remain to be fully determined. Moreover, how CAD-associated genetic variants and haplotypes impact cell state maintenance dynamics and cell plasticity in response to pathophysiologic stimuli is also unclear. Recent efforts using single-cell omics have identified CAD-associated cellular states in human coronary and carotid arteries[16,18–20]. Most studies, performed using human samples, capture a spatial and temporal snapshot of the vascular tissue, bearing the limitations of a steady-state influenced by the original location of the tissue analyzed, potential donor-specific epigenetic marks and the effect of the surrounding niche.

The 9p21.3 locus was the first CAD risk locus identified by GWAS almost two decades ago[7,8,37], but its function remains to be fully defined. The risk haplotype frequency (~50% of the population of non-African descent) and its strong additive effect for each risk allele account for ~10-15% of the incidence of CAD in the US[38], affecting ~3 million people. Numerous studies have explored the function of the 9p21.3 locus using various approaches, including primary cultures[39–43], cell lines[42,44] and primary tissues[45,46]. However, the discoveries have been limited by a few important factors: i) they have mostly described correlations between the risk haplotype and cellular phenotypes without robust evidence of causality[47,48]; ii) most studies using primary cultures have predominantly employed cell types not primarily associated with atherosclerosis (whole blood, T cells, PMBCs)[39–42]; and iii) functional studies on the 9p21.3 CAD locus employing genetic manipulation have been based on immortalized cell lines (HEK cells)[42] or mouse models that lack synteny with the human 9p21.3 region[12]. In recent years, our group has provided evidence that the 9p21.3 CAD risk haplotype affects vascular smooth muscle cells (VSMCs) both transcriptionally and functionally[24,25]. However, many questions remain unanswered on how the 9p21.3 locus affects VSMC heterogeneity, state changes, and disease trajectories.

In this study, we leveraged the power of induced pluripotent stem cells (iPSCs) and their vascular derivatives to determine the role of the 9p21.3 risk haplotype in VSMC state transition. iPSCs accurately maintain the genetic background of the donor without carrying the epigenetic marks of the somatic tissue from which they originate, erased during the reprogramming process. Paired with fine genome editing, they provide a unique opportunity to unravel genotype-to-phenotype causal effects at specific genomic regions. Moreover, since they grow in an isolated and controlled environment, they are excellent tools for evaluating cell-autonomous mechanisms in specific cell types based on their genetic repertoire.

In this study, we leverage a human-cell type-specific approach to annotate CAD risk loci and identify novel VSMC populations, ultimately revealing causal effects of the 9p21.3 CAD risk locus in vascular wall cells.

First, we show that iPSC-derived VSMCs reflect the overall transcriptional profile of *ex vivo* SMCs and recapitulate the heterogeneity of native tissue cells. This highlights the power of this model to explore cellular state dynamics in the vascular wall. Using isogenic knockout lines in both risk and nonrisk backgrounds at 9p21.3, we provide evidence that VSMCs carrying the risk haplotype at 9p21.3 show a transcriptional signature with minimal overlap with other genotypes and low heterogeneity. Remarkably, the isogenic knockout cells restore a signature similar to non-risk cells, demonstrating that the deletion of the 9p21.3 locus is sufficient to revert their phenotype.

Second, we provide direct evidence that the 9p21.3 locus drives the VSMC state toward an *osteochondromyocyte* phenotype. Risk cells maintain SMC identity but show strong upregulation of *SOX9* and *COL2A1*, along with other markers of osteogenic and chondrogenic lineages. Osteochondrogenic SMC states, characterized by loss of contractility and increased expression of *SOX9*, *RUNX2* and *BGLAP*, have been previously reported to be associated with CAD[22]. We previously showed that RR VSMCs have reduced contraction force[24], and now provide supporting data in through upregulated of osteogenic markers. However, RR VSMCs are not enriched in *RUNX2* cells, suggesting that the 9p21.3 risk haplotype may prompt VSMCs toward a novel state that does not fully recapitulate previously described VSMC states. Deletion of the 9p21.3 risk haplotype in the same genomic background does not produce a similar signature, providing strong evidence for a genotype-to-phenotype causal effect of the risk allele. Recently, SOX9 has been linked to vascular aging and the regulation of extracellular matrix composition, leading to arterial stiffness[34]. Moreover, LGALS3 has been described as a marker of SMC plasticity[31]. In our model, RR VSMCs showed increased expression of *SOX9* and decreased expression of *LGALS3*. We speculate that the 9p21.3 CAD risk locus alters the vascular wall by reducing VSMC plasticity and increasing stiffness. This phenotype could ultimately affect the adaptability of cells to external stimuli and potential damage, increasing vessel sensitivity to recurrent insults and the inability to transition to different phenotypes essential for maintaining healthy vessel function.

Third, we demonstrate the role of 9p21.3 CAD risk locus in driving VSMCs states in a cell-autonomous manner, independent of cues from the surrounding tissue niche. Most studies exploring VSMCs phenotypic modulation have leveraged lineage tracing mouse models or human tissues, hence, it has been hard to decipher whether the surrounding tissue or other vascular layers play a role in triggering VSMCs phenotypic switch. Here, we demonstrate that VSMCs carrying the risk haplotype at 9p21.3 have an intrinsic potential to acquire an osteochondromyocyte-like state. Its deletion restores a VSMC trajectory that resembles known transitional VSMCs. Moreover, we show that the presence of the risk allele is sufficient to guide this cell transition independent of other vascular cells.

Fourth, in this study, we also identified two new signature genes, *LIMCH1* and *CRABP1*, that, combined with *SOX9* and *COL2A1* expression, define the 9p21.3 risk status. These genes are uniquely expressed in risk VSMCs and are minimally expressed in other genotypes, suggesting that they are controlled by the

9p21.3 risk haplotype. Although *LIMCH1* and *CRABP1* have been previously detected during heart development[49,50], their roles in VSMC differentiation and state transition have not been described. Despite identifying new roles in mediating SMC cellular states, further studies are needed to fully characterize their vascular-specific functions. Importantly, we provide evidence that these genes are expressed in *ex vivo* human coronary samples, limiting potential artifacts of our system. Moreover, *LIMCH1* and *CRABP1* expression levels were consistently increased in iPSC-VSMCs from additional donors, strengthening the link between the 9p21.3 risk haplotype and upregulation of these signature genes.

Our integrative transcriptome analysis, which included iPSC-VSMCs, *ex vivo* SMCs and fibroblasts, showed a strong similarity of iPSC-VSMCs not carrying the risk haplotype (NN, NNKO, RRKO) with *ex vivo* transitional SMCs, thus rationalizing this model to study VSMC phenotypic modulation. Interestingly, the RR VSMCs also composed a unique cluster that resembled a continuous and advanced stage of *osteochondromyocyte* in the pseudotime trajectory. We thus speculate that the atlas may not include the full representations of this extreme phenotype due to technical challenges in extracting these cells from calcified human plaques. Alternatively, it is possible that our iPSC-based culture system enriches this phenotype to a degree that is not captured *in vivo*. We cannot exclude the possibility that this extreme phenotype *in vivo* is buffered by other signals within the complex tissue environment.

Overall, our study provides new evidence on the role of the 9p21.3 CAD risk locus in altering VSMC state transitions. Moreover, we highlight the power of the iPSC-based system to reveal the biology of VSMCs and potentially other disease-relevant cell types to understand the dynamics of cellular trajectories driven by the most impactful CAD genomic locus.

Methods

Ethical compliance and cell lines

The authors have complied with all ethical regulations. The human subjects were enrolled, and informed consent was obtained from a study approved by the Scripps IRB (#115676). The cells were transferred and maintained under a study approved by the University of Wisconsin Madison through Material Transfer Agreement (MSN246500). Unedited lines from the same donor are available at WiCell-Inc. The lines included in this study were obtained from a total of five donors. Lines included in the scRNA sequencing analysis were obtained from two donors: C512 (RR clones 1-5 and 2-3 and isogenic RR KO clones 1-9 and WB46) and HE463 (NN clones E14 and E17 and isogenic NN KO clones E50 and E56). Isogenic KO lines were derived by TALEN-mediated genome editing as previously described by Lo Sardo *et al*. (Cell 2018). Additional iPSC-VSMCs, including unedited clones from RR donors HE787 and C021 and C939 and NBD5433, were used to confirm the main findings.

iPSC derivation and differentiation

Induced pluripotent stem cells were previously derived and characterized (Lo Sardo *et al*. 2017). Peripheral blood mononuclear cells (PBMCs) were reprogrammed into induced pluripotent stem cells (iPSCs) through episomal plasmid transfection of reprogramming factors (Lo Sardo *et al*. 2018). iPSCs were cultured on Matrigel-coated (Corning) cell culture plates and mTeSR (Stem Cell Technologies) and cultured at 37°C under normoxic conditions (20% O_2) with 5% CO_2. Plates were coated according to manufacturer's guidelines, diluting Matrigel in cold DMEM/F12 at 0.5mg/mL. Cells The cells were cultured in mTeSR and fed daily with splits performed approximately every 4-5 days or when the culture reached ~75% confluency. EDTA (0.5 mM) in PBS was used for the detachment of iPSCs for regular splitting.
iPSCs were differentiated into VSMCs as previously described (Lo Sardo *et al*. 2018) adapted from Cheung *et al*. 2014. Briefly, iPSCs were cultured in mTeSR until approximately 50% confluency. The media were changed to CDM-BSA (250 mL of IMDM (GIBCO, Cat. 12440-053), 250 mL of F12 (GIBCO, Cat. 11765-054), 5 mL of pen/strep (GIBCO, Cat. 15140122), 5 mL of chemically defined lipid concentrate (Life Tech, Cat. 11905-031), 20 µL of monothioglycerol (Sigma-Aldritch, Cat. 96-27-5), 0.5 g of BSA (Sigma-Aldritch, Cat. A9418), 250 µL of transferrin (GIBCO, Cat. 51500-056), and 350 µL of human recombinant insulin (Sigma-Aldritch, Cat. 11376497001)) supplemented with bFGF (10 ng/mL) and activin A (10 ng/mL) for 24 hours before seeding for differentiation. Cells were seeded on gelatin-coated plates in CDM-PVA (the same as CDM-BSA, in which BSA was replaced with 0.5 g of PVA (Sigma–Aldrich, 9002-89-50)) and differentiated into the lateral mesoderm lineage for five days. On day 5, the cells were detached with Triple Express (Gibco) and reseeded on gelatin-coated plates in CDM-PVA supplemented with PDGF-BB (10 ng/mL) and TGFb1 (2 ng/mL) at a density of 20,000-30,000 cells/cm^2. Twelve days after seeding, the cells were harvested for single-cell sequencing, maintained in DMEM (GIBCO, Cat. 10566-016) supplemented with 10% FBS, Pen/Strep (GIBCO, Cat. 15140122), GlutaMAX (GIBCO, Cat. 35050061), and NEAA (GIBCO, Cat. 11140050) and passaged every 5-6 days.

Single-cell sequencing of cell suspensions and library preparation

On day 17, terminally differentiated VSMC were detached using Tryple E (Cat. 12604-013) by incubation for 3-5 minutes at room temperature. Tryple E was diluted with CDM-PVA, and the cell suspension was centrifuged at 200 × g for 5 minutes. The cells were gently resuspended in 0.04% BSA in PBS, passed through a 30 µm cell strainer and then counted using Trypan Blue (Sigma–Aldrich, Cat. 72-57-1) solution and an Invitrogen Countess II cell counter. The cell suspension was centrifuged at 200 × g for 5 minutes, after which the supernatant was removed. One milliliter of PBS supplemented with 0.04% BSA was added, and the cells were counted again. After centrifugation, the samples were gently resuspended in 0.04% BSA in PBS at a concentration of 1000 cells/µl. The cells were maintained on ice and subjected to library preparation at the UW-Madison Biotechnology Center.

RNA Preparation and Sequencing

The authors utilized the University of Wisconsin–Madison Biotechnology Center's Gene Expression Center Core Facility (Research Resource Identifier - RRID:SCR_017757) for single-cell RNA library preparation and the DNA Sequencing Facility (RRID:SCR_017759) for sequencing. GEM and library preparation were conducted through the Core using a 10x Genomics Chromium Next GEM Single Cell 3' Reagent Kit v3.1 and a single-cell G Chip Kit (Lot# 161273). All protocols were followed according to the manufacturer's instructions. The target cell recovery (TCR) values were determined according to the quality and viability of the samples and calculated for each sample to obtain ~3000 cells/sample. GEM generation and post-GEM clean-up procedures were completed according to the manufacturer's instructions.

The number of cycles for the thermocycler protocol during the cDNA amplification step was set to 12 based on targeted cell capture. Quality control for cDNA amplification was then performed using an Agilent HS DNA chip, and 1 µL of sample was diluted to 10 µL with nuclease-free water to calculate the total cDNA yield. Fragmentation end repair and A-tailing, adaptor ligation and cleanup, sample index PCR, postsample index PCR double-sided size selection, and post-library construction quality control were performed according to the manufacturer's instructions. The SPRI reagent lot number was 18198500, and the Dual Index TT Set A lot number was 162043. MiSeq was used for QC, and balance of the samples was performed using calculations based on UMI tools (Smith *et al*, 2017). Sample libraries were balanced for the number of estimated reads per cell and run on an Illumina NovaSeq system on a NovaSeq S1 flow cell at the University of Wisconsin–Madison Biotechnology Center's Gene Expression Center Core Facility.

Single-cell RNA-seq data were first analyzed by the UW Bioinformatics Resource Center using Cell Ranger pipelines from 10x Genomics. The data were demultiplexed with the Cell Ranger mkfastq command wrapped around Illumina's bcl2fastq. Cell Ranger software was then used to perform alignment, filtering, barcode counting, UMI counting, and gene expression estimation for each sample according to the 10x Genomics documentation. The gene expression estimates from each sample were then aggregated using CellRanger (cellranger aggr) to compare experimental groups with normalized sequencing depth and expression data. This produced a dataset of 27,088 cells with a median of 3,169 genes per cell detected and an average read depth of 30,186 confidently mapped reads per cell per sample.

Bioinformatic analysis of single-cell sequencing of iPSC-VSMCs

Filtering. The data were analyzed in R studio (v. 4.2.1) using Seurat V4.3.0. Genes expressed in fewer than 3 cells and cells with fewer than 200 genes expressed were omitted. The data were filtered to remove dying cells, doublets, and empty GEMs using UMI count (5,000-120,000), feature count (2,500-8,500), and percent mitochondrial RNA (>20%).

Normalization, PCA, and clustering. Normalization, principal component analysis (PCA), and clustering were conducted using Seurat V4.3.0. After filtering, counts were normalized via SC transformation, analyzed via PCA, and clustered via the Louvain method. Variation in the percentage of mitochondrial RNA

decreased during SC transformation. Five principal components and a resolution of 0.4 were used in the final clustering.

scRNA data figures. All UMAPS, heatmaps, and violin plots were generated using the Seurat V4.3.0 commands RunUMAP and dimPlot for UMAPs and Doheatmap and VlnPlt for the heatmaps and violin plots, respectively. Dot plots and bar charts were generated with ggplot2 (version 3.4.2) DotPlot and ggplot commands.

Gene Ontology. Gene Ontology (GO) analysis was conducted using clusterProfiler (version 3.0.4), which revealed genes that were significantly differentially expressed according to the differential gene expression analyses. The organism was set to 'Human' using organism = "org.Hs.eg.db", gene names were mapped to ENTREZID identifiers using the mapIds command, and enrichment analysis was conducted using the enrichGO command with subontology = BP, p value cutoff = 0.05, and p value adjustment = BH.

Trajectory Analysis. Trajectory analysis was conducted using Monocle3 (version 1.3.1.). The data were made into a Monocle readable object using the as.cell_data_set command, the cells were clustered using the cluster_cells command, the trajectories were calculated using learn_graph, the starting nodes were chosen using order_cells, and the final trajectories were plotted using plot_cells with coloring by pseudotime.

Bioinformatic analysis of publicly available datasets

Comparison to published datasets was conducted by downloading the original count matrices and filtering/clustering using the metrics and filters below (details in the table). Thresholds for filtering were set to omit excessively high % mitochondrial RNA and unusually high or low UMI and feature counts relative to the bulk of the cells in the dataset to capture the representative majority of cells and omit dead cells, doublets, or empty GEMs. Once the downloaded datasets were clustered and a UMAP was generated, clusters were annotated for cell types using cell type-specific markers provided by the publication. The resolution for published datasets was empirically estimated to approximate the same number of cell type clusters as in the original publication as closely as possible.

Dataset	UMI Count filters	Feature Count filters	% mitochondrial RNA filter	Number of Principle components	Resolution
Hu *et al*, 2021	100-18000	1000-4000	<8%	5	0.5
Wirka *et al*, 2019	100-20000	1000-4000	<6%	5	0.5
Pan *et al*, 2020	5000-15000	1800-4000	<7.5%	5	0.4

Immunocytochemistry

Cells were first fixed using 4% paraformaldehyde (15710, Electron Microscopy Science) in PBS. The fixed samples were incubated with 0.5% Triton X-100 (T8787) in PBS for cytoplasmic and nuclear staining. The cells were then blocked with 5% FBS (GeminiBio 100-106) in PBS for 1 hour at room temperature. The blocking solution was aspirated and replaced with primary antibodies against LIMCH1 (NBP2-39006, Novus Biological) and/or CRABP1 (sc-166897, Santa Cruz) in blocking buffer and incubated for approximately 2 hours at room temperature or overnight at 4°C. The primary antibody solution was then aspirated, followed by three PBS washes. The secondary antibodies Donkey anti-mouse 488 (Alexa-fluor A32766, Life Technology), Donkey anti-rabbit 555 (Alexa-fluor A32794, Life Technology) and DAPI (MBD0015, Sigma–Aldrich) in blocking buffer were added to the samples, which were then incubated for 1 hour at room temperature. Three additional PBS washes were performed after secondary antibody incubation, and the samples were imaged via fluorescence microscopy.

RNA extraction, cDNA synthesis, and qPCR

The cell pellets were dissociated using TRIzol (15596018). RNA was then extracted using a Zymo Direct-zol RNA Miniprep kit (R2052) according to the manufacturer's protocol. The elution volume was between 25 and 50 µl for each sample, according to the number of cells. The samples were then quantified using a Nanodrop spectrophotometer and stored at -80°C until needed.

To make cDNA, 200 ng of RNA sample was diluted in nuclease-free water to a total volume of 15 µl. A master mixture of 1 µl of iScript Reverse Transcription and 4 µl of iScript Reaction Mix (1708891, Bio-Rad) per sample was used for a total of 20 µl of reaction. The samples were placed in a thermocycler following the manufacturer's protocol. This consisted of 5 minutes of priming at 25°C, reverse transcription for 20 minutes at 46°C, and reverse transcription inactivation for 1 minute at 95°C. The samples were then directly subjected to qPCR or stored at -80°C.

For qPCR, each well of a 384-well plate was filled with iTaq Universal SYBR Green Supermix (1725121) and nuclease-free water, and cDNA samples from each line were used for detection on a Bio-Rad CFX OPUS 384 Touch Real-Time PCR machine. For ANRIL qPCR, SsoAdvanced Universal SYBR Green Supermix (17252721) was used instead. Samples were loaded in triplicate to assess technical variability. Each reaction was performed using cDNA template equivalent to 2-8 ng of total RNA, varying based on the expression level of each transcript of interest. A complete list of primers used for qPCR is provided in the table below.

Gene	**Primer**	**Sequence**
PPIA	hPPIA Fwd	5'- CCGAGGAAAACCGTGTACTATTAG -3'
	hPPIA Rev	5'- TGCTGTCTTTGGGACCTTG -3'
ANRIL Total	ANRIL-5 Fwd	5'- ACACATCAAAGGAGAATTTTCTTGG -3'

	ANRIL-6 Rev	5’- GTACTGACTCGGGAAAGGATTC -3’
ANRIL Long	ANRIL 18-19 Fwd	5'- GCACACTGGTCAGCTGTTTT -3'
	ANRIL 18-19 Rev	5'- ACTTTTCATCTTTCTGTATTTCAA -3'
LIMCH1	hLIMCH1 Fwd	5'- ATGATTCCTTCGACAGCCTG -3'
	hLIMCH1 Rev	5'- TCTTCACATCTGGCAGCTTTC -3'
CRABP1	CRABP1 Fwd	5'- TGAGAACAAGATCCACTGCAC-3'
	CRABP1 Rev	5'- TCACTCTCGGACATAAATTCTGG -3'

MetaPlaq atlas integration

MetaPlaq data integration, co-embedding and trajectory analyses were conducted as previously described in Mosquera *et al*. 2023[20]. Briefly, we subjected the iPSC-VSMC scRNA-seq datasets to an automated data processing pipeline (scRNAutils) that was also used to process the datasets in MetaPlaq, including doublet and ambient RNA removal, filtering for high quality cells, and normalization to account for variability in sequencing depth across cells. PCA was used for dimensionality reduction of the normalized counts matrix and Shared-Nearest-Neighbors (SNN and Louvain community detection was used for clustering in Seurat v4. Clustering results were visualized using UMAP embeddings. Seurat objects were batch integrated using reciprocal PCA, which was shown to optimize biological preservation and removal of unwanted technical batch effects. Clusters were annotated using the original labels in MetaPlaq[20]. Finally, similarity between iPSC-VSMC and MetaPlaq clusters was calculated using the Euclidean distance proximity metric as previously described[51].

Acknowledgments

The authors would like to thank Dr. Ali Torkamani, Dr. Amitabh Pandey and Doug Evans at the Scripps Research Translational Institute, Kathryn Hubert in Dr. Deneen Wellik’s laboratory and Huan Yang in Dr. Bo Liu’s laboratory at UW-Madison SMPH for initial guidance on sequencing data analysis. The authors utilized the University of Wisconsin – Madison Biotechnology Center Gene Expression Center (Research Resource Identifier – RRID:SCR_017757). The authors acknowledge the following funding. E.S. was supported by the NIH T32 Genetics Training grant (T32GM007133) and Advanced Opportunity Fellowship through SciMed Graduate Research Scholars, both at the University of Wisconsin Madison. C.D.M.V. received a fellowship supported by the American Heart Association grant 23IAUST1027463 /Board of Regents University of Wisconsin System and the University of Wisconsin Cardiovascular Research Center. V.L.S. received funding for this study from the University of Wisconsin Madison School of Medicine and

Public Health, the UW-Madison Human Genomics and Precision Medicine Center and the UW-Madison office of the Vice Chancellor for Research and Graduate Education.

Author Contributions

E.S. and V.L.S. conceived the project and designed, performed and supervised all the experiments and analyses. C.D.M.V. performed the qPCR analysis. R.Z. performed immunocytochemistry and fluorescence microscopy. E.S. and V.L.S. performed and analyzed the scRNA-seq data and interpreted the data. J.V.M. and C.L.M. performed the data integration with the Metaplaq dataset. C.L.M. provided input on the data interpretation. The manuscript was written by E.S. and V.L.S. with input from the other authors.

Competing Interests

The authors declare no competing financial interests.

REFERENCES

1. Nikpay, M. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics* **47**, 1121–1130 (2015).

2. Tcheandjieu, C. *et al.* High heritability of ascending aortic diameter and trans-ancestry prediction of thoracic aortic disease. *Nat Genet* **54**, 772–782 (2022).

3. Vries, P. S. de *et al.* Whole-genome sequencing uncovers two loci for coronary artery calcification and identifies ARSE as a regulator of vascular calcification. *Nat. Cardiovasc. Res.* **2**, 1159–1172 (2023).

4. Koyama, S. *et al.* Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet* **52**, 1169–1177 (2020).

5. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005–D1012 (2019).

6. Leypold, N. A. & Speicher, M. R. Evolutionary conservation in noncoding genomic regions. *Trends Genet* **37**, 903–918 (2021).

7. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).

8. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).

9. Samani, N. J. *et al.* Genomewide Association Analysis of Coronary Artery Disease. *New England Journal of Medicine* 1–11 (2007) doi:10.1056/nejmoa072366).

10. Holdt, L. M., Kohlmaier, A. & Teupser, D. Circular RNAs as Therapeutic Agents and Targets. *Frontiers in physiology* **9**, 1262 (2018).

11. He, S., Gu, W., Li, Y. & Zhu, H. ANRIL/CDKN2B-AS shows two-stage clade-specific evolution and becomes conserved after transposon insertions in simians. *BMC evolutionary biology* **13**, 247 (2013).

12. Visel, A. *et al.* Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature* **464**, 409–412 (2010).

13. Owens, G. K., Kumar, M. S. & Wamhoff, B. R. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiological reviews* **84**, 767–801 (2004).

14. Alexander, M. R. & Owens, G. K. Epigenetic control of smooth muscle cell differentiation and phenotypic switching in vascular development and disease. *Annual review of physiology* **74**, 13–40 (2012).

15. Basatemur, G. L., Jørgensen, H. F., Clarke, M. C. H., Bennett, M. R. & Mallat, Z. Vascular smooth muscle cells in atherosclerosis. *Nature reviews. Cardiology* **16**, 727–744 (2019).

16. 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).

17. Hu, Z. *et al.* Single-Cell Transcriptomic Atlas of Different Human Cardiac Arteries Identifies Cell Types Associated With Vascular Physiology. *Arteriosclerosis, Thrombosis, and Vascular Biology* ATVBAHA120315373 (2021) doi:10.1161/atvbaha.120.315373.

18. Turner, A. W. *et al.* Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat Genet* 1–13 (2022) doi:10.1038/s41588-022-01069-0.

19. Wirka, R. C. *et al.* Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nature medicine* **25**, 1280–1289 (2019).

20. Mosquera, J. V. *et al.* Integrative single-cell meta-analysis reveals disease-relevant vascular cell states and markers in human atherosclerosis. *Cell Rep.* **42**, 113380 (2023).

21. Feil, S. *et al.* Transdifferentiation of vascular smooth muscle cells to macrophage-like cells during atherogenesis. *Circulation research* **115**, 662–667 (2014).

22. Speer, M. Y. *et al.* Smooth Muscle Cells Give Rise to Osteochondrogenic Precursors and Chondrocytes in Calcifying Arteries. *Circ. Res.* **104**, 733–741 (2009).

23. Bobryshev, Y. V. Transdifferentiation of smooth muscle cells into chondrocytes in atherosclerotic arteries in situ: implications for diffuse intimal calcification. *J. Pathol.* **205**, 641–650 (2005).

24. LoSardo, V. *et al.* Unveiling the Role of the Most Impactful Cardiovascular Risk Locus through Haplotype Editing. *Cell* **175**, 1796-1810.e20 (2018).

25. Mayner, J. M. *et al.* Heterogeneous expression of alternatively spliced lncRNA mediates vascular smooth cell plasticity. *Proc. Natl. Acad. Sci. United States Am.* **120**, e2217122120 (2023).

26. Chen, R., McVey, D. G., Shen, D., Huang, X. & Ye, S. Phenotypic Switching of Vascular Smooth Muscle Cells in Atherosclerosis. *J. Am. Hear. Assoc.* **12**, e031121 (2023).

27. Miano, J. M., Fisher, E. A. & Majesky, M. W. Fate and State of Vascular Smooth Muscle Cells in Atherosclerosis. *Circulation* **143**, 2110–2116 (2021).

28. Wang, Z., Wang, D.-Z., Pipes, G. C. T. & Olson, E. N. Myocardin is a master regulator of smooth muscle gene expression. *Proc. Natl. Acad. Sci.* **100**, 7129–7134 (2003).

29. Miano, J. M. Myocardin in biology and disease. *J. Biomed. Res.* **29**, 3–19 (2015).

30. Shankman, L. S. *et al.* KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nature medicine* **21**, 628–637 (2015).

31. Alencar, G. F. *et al.* Stem Cell Pluripotency Genes Klf4 and Oct4 Regulate Complex SMC Phenotypic Changes Critical in Late-Stage Atherosclerotic Lesion Pathogenesis. *Circulation* **142**, 2045–2059 (2020).

32. Bennett, M. R., Sinha, S. & Owens, G. K. Vascular Smooth Muscle Cells in Atherosclerosis. *Circulation research* **118**, 692–702 (2016).

33. 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.* **100**, 13531–13536 (2003).

34. Faleeva, M. *et al.* Sox9 Accelerates Vascular Aging by Regulating Extracellular Matrix Composition and Stiffness. *Circ. Res.* **134**, 307–324 (2024).

35. Harst, P. van der & Verweij, N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circulation research* **122**, 433–443 (2018).

36. Erdmann, J., Kessler, T., Venegas, L. M. & Schunkert, H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc Res* **114**, 1241–1257 (2018).

37. Samani, N. J. & Schunkert, H. Chromosome 9p21 and Cardiovascular Disease: The Story Unfolds. *Circulation: Cardiovascular Genetics* **1**, 81–84 (2008).

38. Gränsbo, K. *et al.* Chromosome 9p21 genetic variation explains 13% of cardiovascular disease incidence but does not improve risk prediction. *J Intern Med* **274**, 233–40 (2013).

39. Jarinova, O. *et al.* Functional Analysis of the Chromosome 9p21.3 Coronary Artery Disease Risk Locus. *Arteriosclerosis, Thrombosis, and Vascular Biology* **29**, 1671–1677 (2009).

40. Liu, Y. *et al.* INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS ONE* **4**, e5027 (2009).

41. Holdt, L. M. *et al.* ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arteriosclerosis, Thrombosis, and Vascular Biology* **30**, 620–627 (2010).

42. Holdt, L. M. *et al.* Alu Elements in ANRIL Non-Coding RNA at Chromosome 9p21 Modulate Atherogenic Cell Functions through Trans-Regulation of Gene Networks. *PLoS Genetics* **9**, e1003588-12 (2013).

43. Almontashiri, N. A. M. *et al.* 9p21.3 Coronary Artery Disease Risk Variants Disrupt TEAD Transcription Factor–Dependent Transforming Growth Factor &bgr; Regulation of p16 Expression in Human Aortic Smooth Muscle Cells. *Circulation* **132**, 1969–1978 (2015).

44. Folkersen, L. *et al.* Relationship between CAD Risk Genotype in the Chromosome 9p21 Locus and Gene Expression. Identification of Eight New ANRIL Splice Variants. *PLoS ONE* **4**, e7677-6 (2009).

45. Holdt, L. M. *et al.* Expression of Chr9p21 genes CDKN2B (p15INK4b), CDKN2A (p16INK4a, p14ARF) and MTAP in human atherosclerotic plaque. *Atherosclerosis* **214**, 264–270 (2011).

46. Motterle, A. *et al.* Functional analyses of coronary artery disease associated variation on chromosome 9p21 in vascular smooth muscle cells. *Human Molecular Genetics* **21**, 4021–4029 (2012).

47. Khera, A. V. & Kathiresan, S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet* **18**, 331–344 (2017).

48. McPherson, R. & Tybjaerg-Hansen, A. Genetics of Coronary Artery Disease. *Circ Res* **118**, 564–578 (2016).

49. Phansalkar, R. *et al.* Coronary blood vessels from distinct origins converge to equivalent states during mouse and human development. *eLife* **10**, e70246 (2021).

50. Soysa, T. Y. de *et al.* Single-cell analysis of cardiogenesis reveals basis for organ level developmental defects. *Nature* **572**, 120–124 (2019).

51. Andrews, T. S., Kiselev, V. Y., McCarthy, D. & Hemberg, M. Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data. *Nat. Protoc.* **16**, 1–9 (2021).

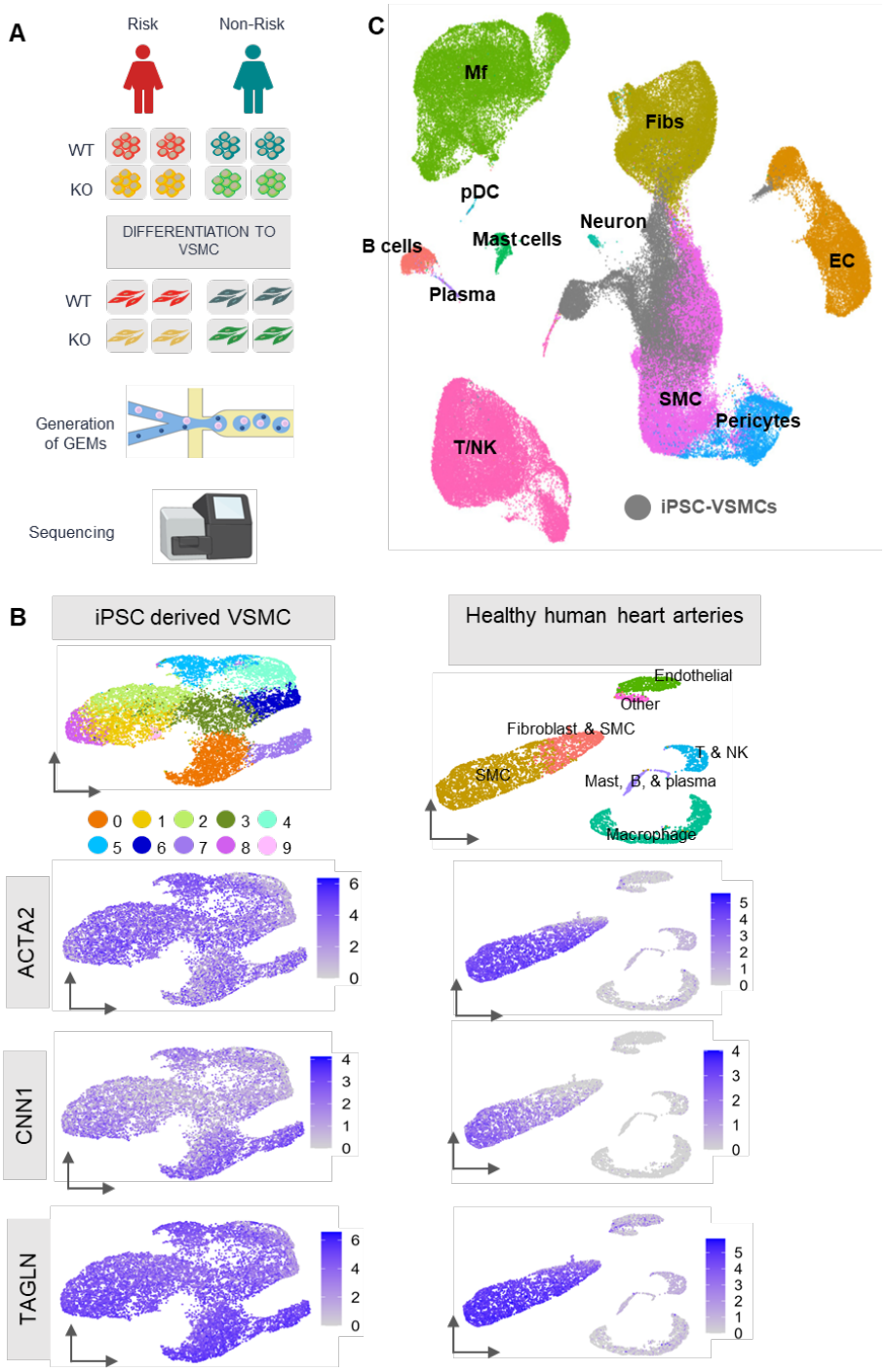


Figure 1. iPSC-VSMC are reflective of ex-vivo human SMC. (A) Schematic of experimental workflow. iPSC are from one homozygous risk and one homozygous non-risk donors, both with intact 9p21.3 locus. Isogenic knockout lines for each genotype were derived by TALEN-mediated haplotype editing to delete the entire CAD region. Two clones/genotype (N=8) were differentiated to VSMCs and used for single-cell sequencing. (B) UMAPs of scRNAseq data from iPSC-derived VSMCs collected at day 17 in 4 genotypes (RR-NN-RRKO-NNKO) (left) and UMAPs for *ex-vivo* healthy human artery data from Hu et al, 2021 (right). Topmost UMAPs are colored by unbiased clustering in the iPSC-VSMC and by cell type in the ex-vivo data. Expression of SMC markers shown in both datasets. (C) Projection of iPSC-derived VSMC in the Metaplaq atlas of *ex vivo* human artery cells from Mosquera et al., 2023.

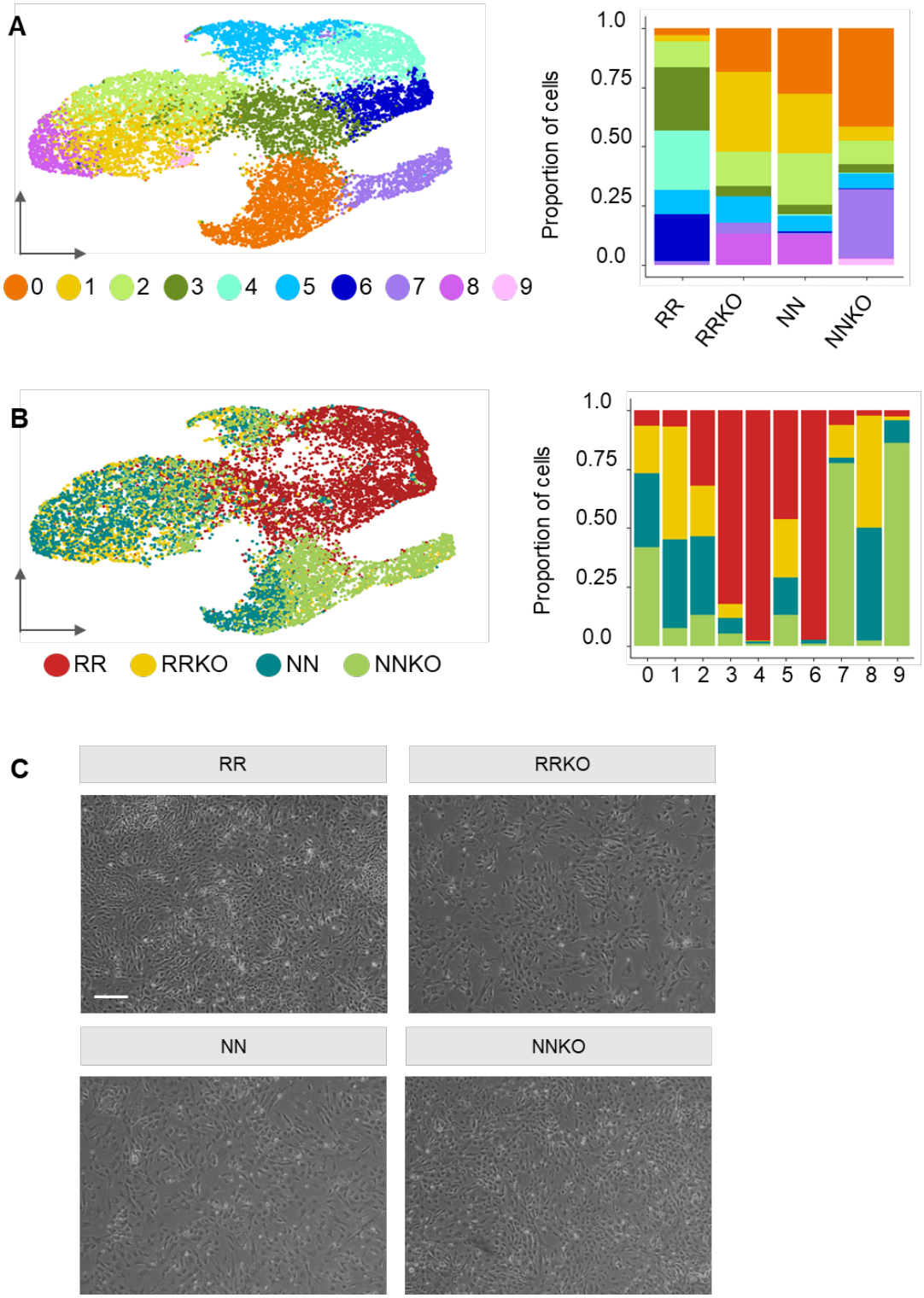


Figure 2. The 9p21.3 CAD risk haplotype alters the transcriptional profile of VSMCs. (A) UMAP of iPSC-derived VSMC colored by unbiased clustering (left) and barchart showing the quantification of cluster composition for each genotype (right). (B) UMAP of iPSC-derived VSMC colored by genotype (left) and barchart quantifying the genotype composition for each cluster (right). (C) Images of cells used, captured at day 17 of differentiation. White bar shows 50µm.

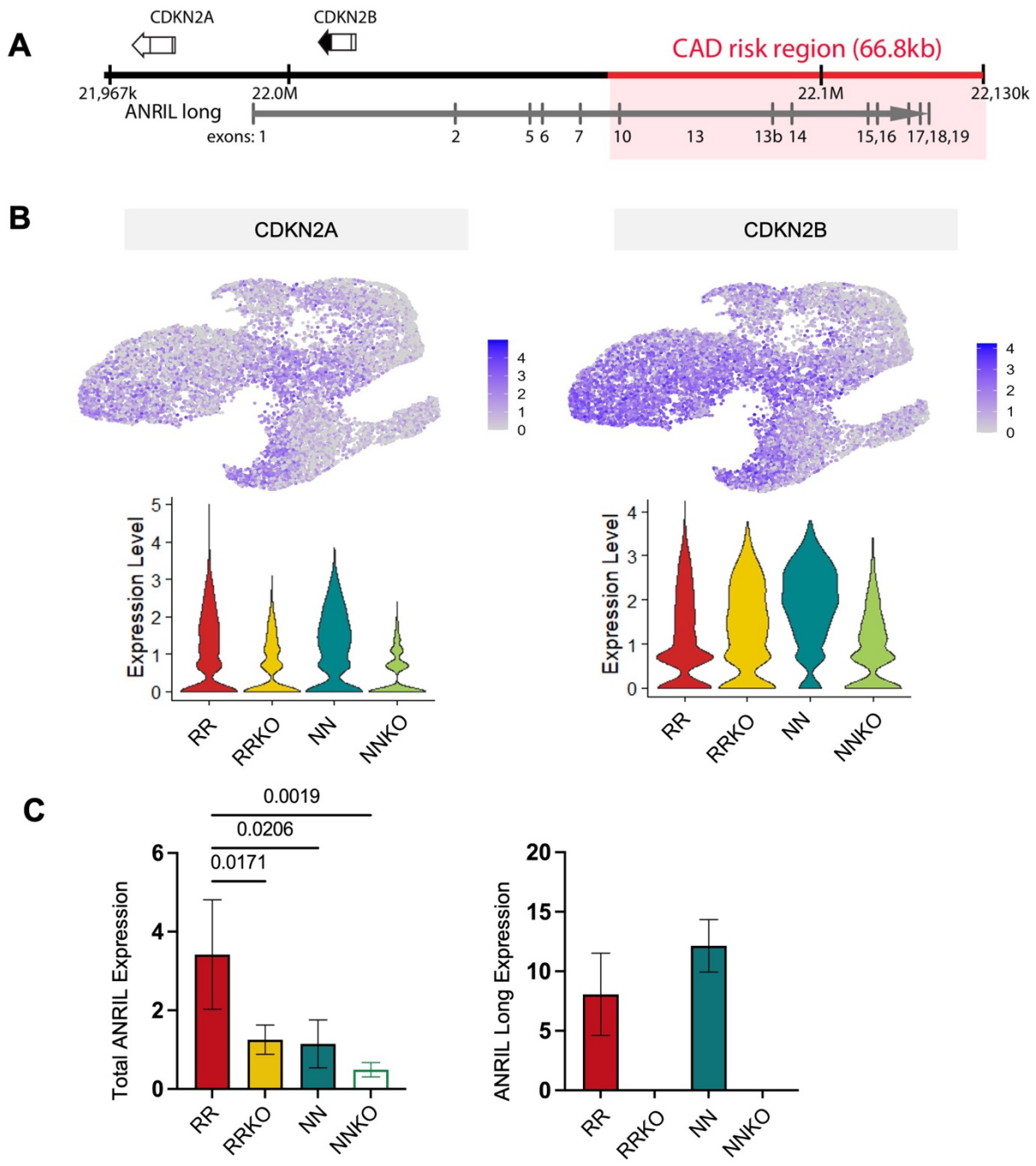


Figure 3. Expression of neighboring genes at the 9p21.3 locus. (A) Diagram of 9p21.3 CAD risk locus with nearest protein-coding genes CDKN2A/B and overlapping lncRNA ANRIL marked. (B) UMAPs of iPSC-derived VSMC single cell data showing the expression CDKN2A and CDKN2B, violin plots show quantification of the same genes across genotypes. (C) qPCR quantification of all ANRIL isoforms (primers select for exons 5-6) in iPSC-derived VSMCs used in single-cell dataset, and qPCR quantification of long isoforms (primers select for exons 19-20) in the same samples. P value calculated with unpaired t-test

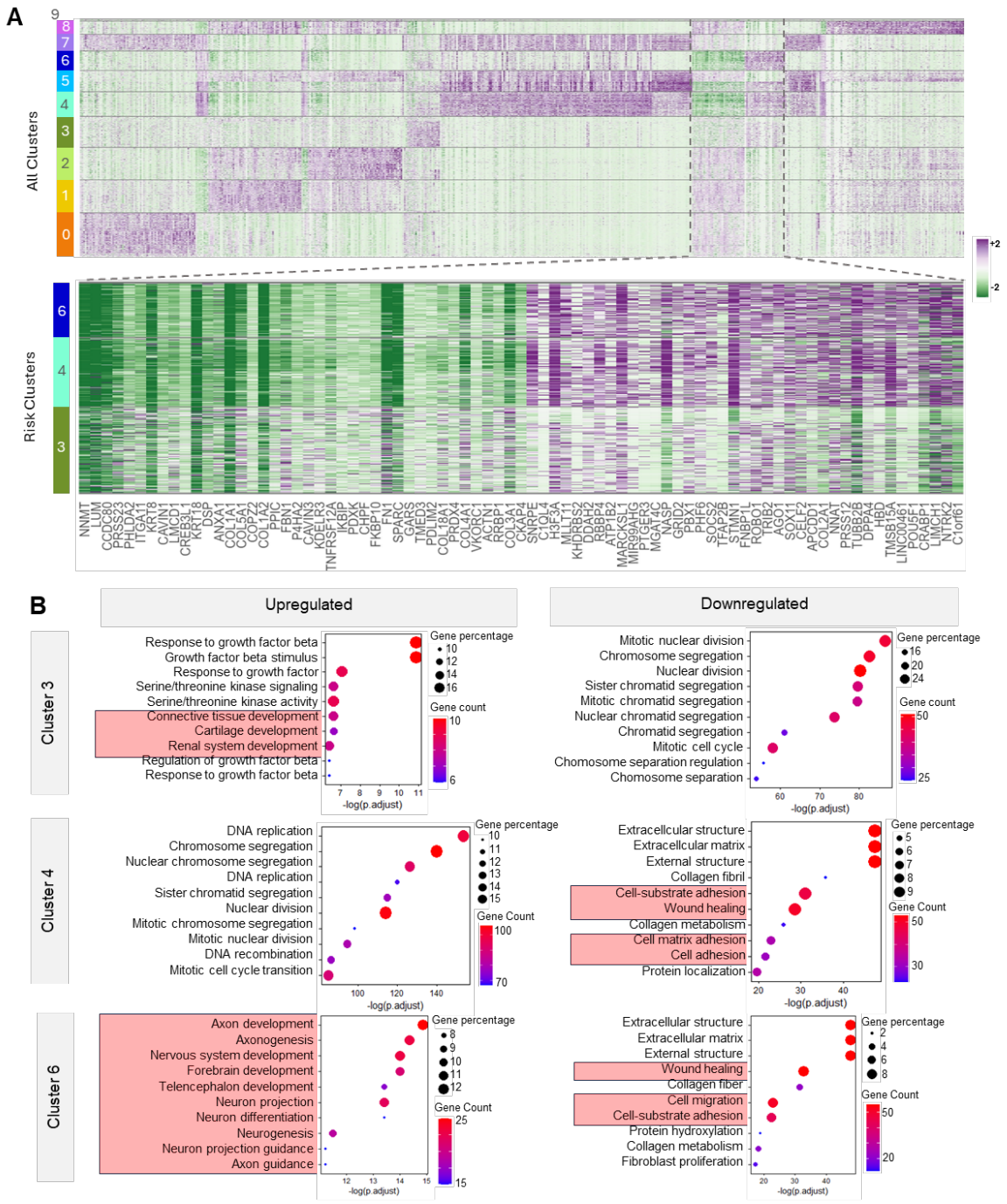


Figure 4. Differential gene expression analysis and Gene Ontology Enrichment reveal unique patterns of expression in RR VSMCs . (A) Top heatmap shows the top 10% genes by corrected p-value driving each cluster (differential expression analysis conducted using Seurat) as they are expressed across all clusters. Bottom heatmap shows a block of genes expressing uniformly and uniquely among risk clusters. (B) The top 10 gene ontology terms enriched in the differentially expressed genes driving each RR cluster (3-4-6) (generated with clusterProfiler). On the left terms associated with upregulated genes, on the right downregulated terms are shown.

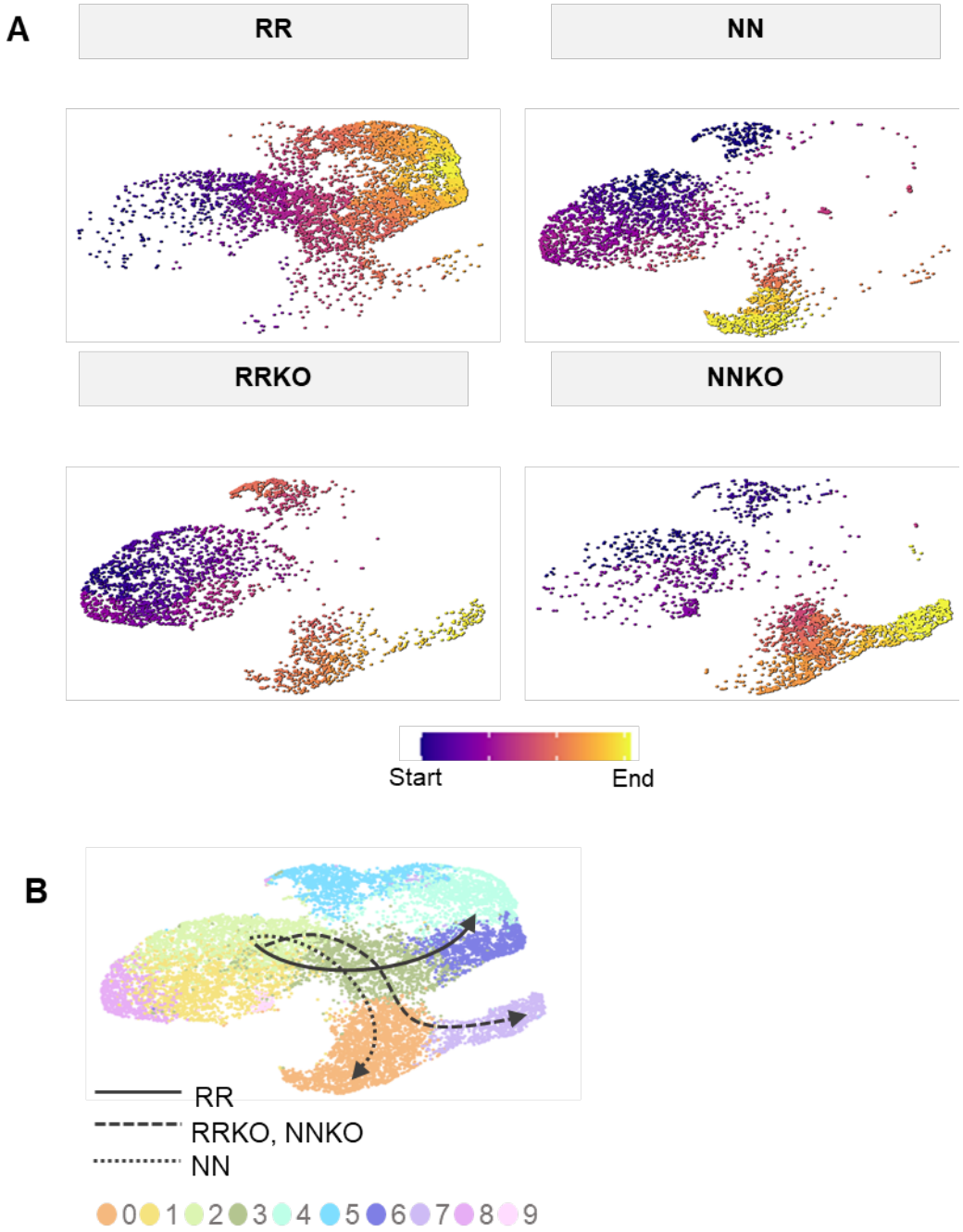


Figure 5. 9p21.3 RR VSMCs show a divergent state trajectory. (A) Pseudotime analysis (conducted with Monocle3) for each genotype. (B) Reconstruction of each genotype trajectory superimposed on a UMAP colored by cluster.

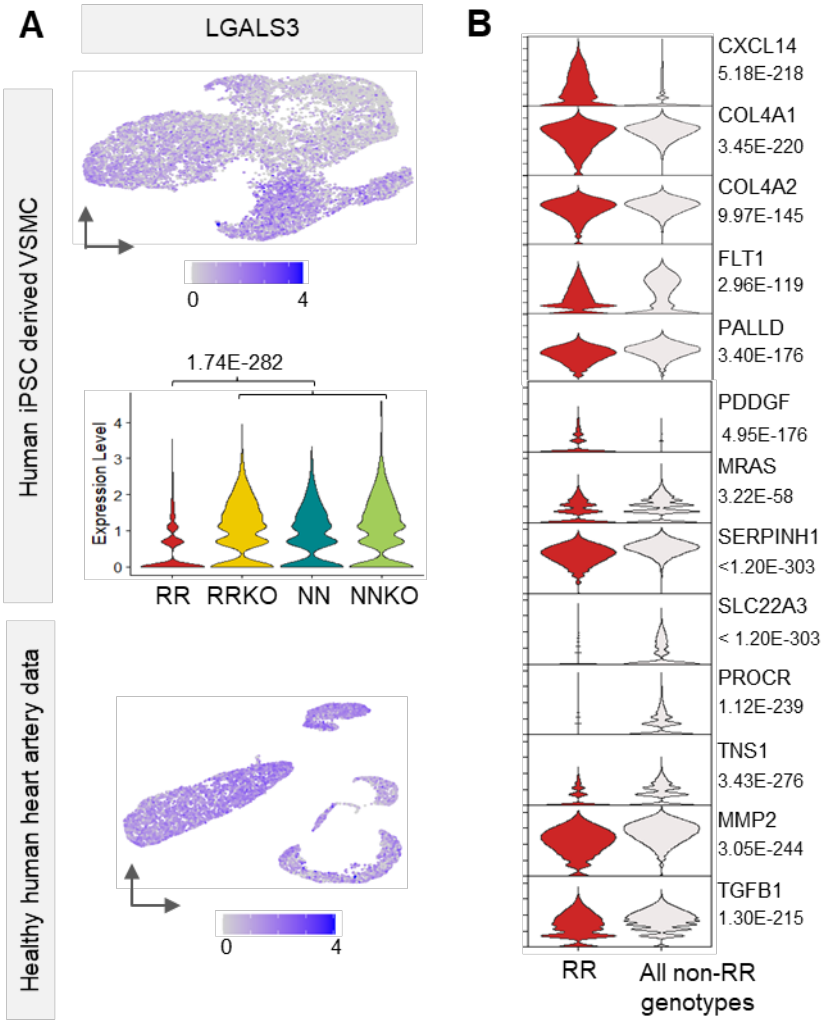


Figure 6. Effect of the 9p21.3 risk haplotype on VSMC transitional state and other CAD loci. (A) UMAP showing expression of LGALS3 in iPSC-VSMCs (top), violin plot showing the expression of LGALS3 across genotypes with p-value calculated via non-parametric Wilcoxon rank sum test from Seurat (middle), and UMAP of LGALS3 expression in healthy human artery data from Hu et al., 2019. (bottom). (B) Violin plots showing the expression of known CAD genes in RR vs all other genotypes with p-values calculated with non-parametric Wilcoxon rank sum test (Seurat).

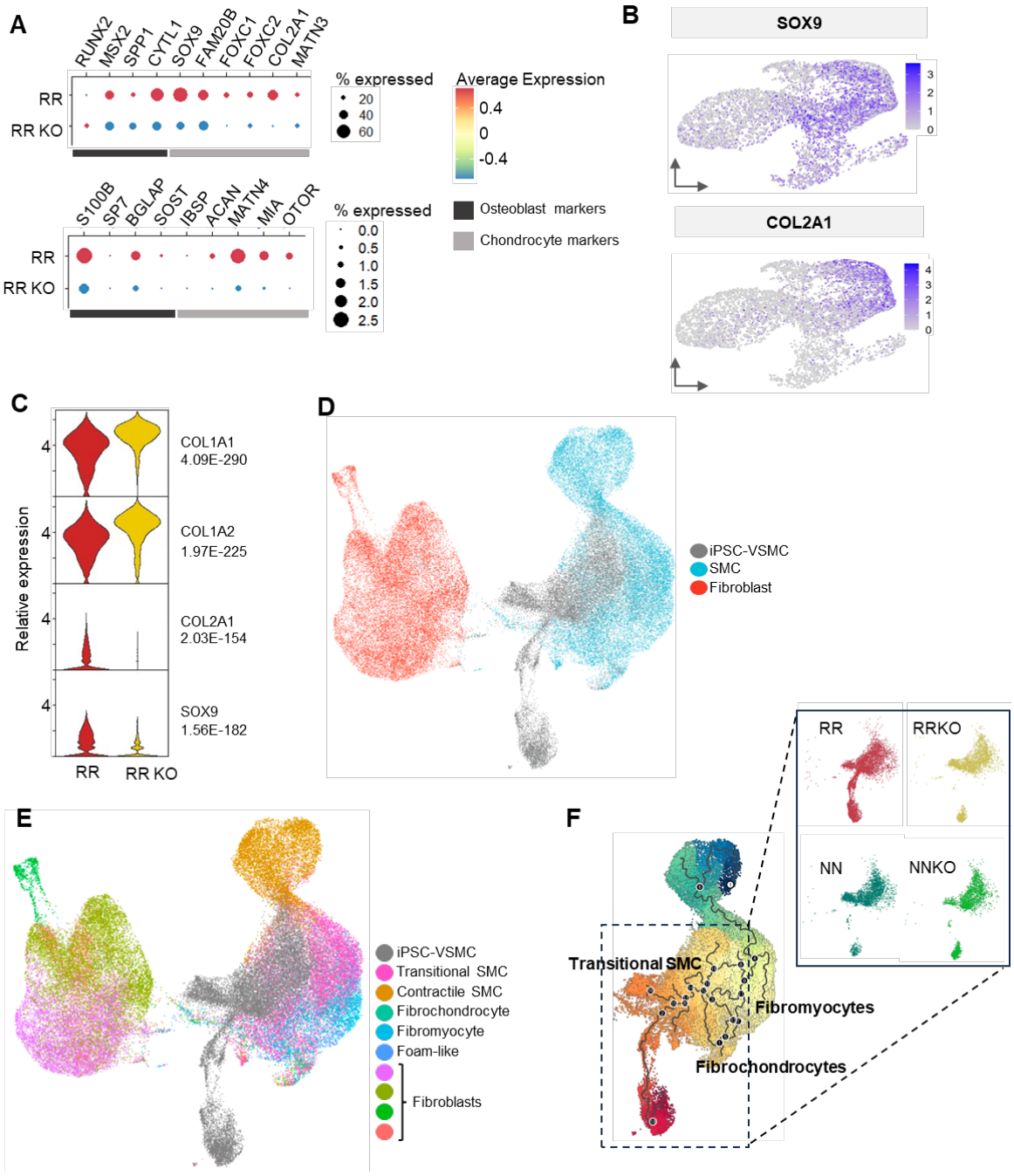


Figure 7. Risk VSMC acquire an osteochondrogenic-like state. A) Bubble plots of osteoblast and chondrocyte markers in RR and RR KO. (B) UMAPs showing upregulation of SOX9 and COL2A1 in risk VSMC. (C) Violin plots showing expression of SOX9, COL2A1, COL1A1, and COL1A2 in risk VSMC and isogenic knock-out, with p-values calculated by non-parametric Wilcoxon rank sum test. (D) Harmonization of iPSC-VSMC from this study with SMC and Fibroblasts from Metaplaq. (E) Secondary annotation of iPSC-VSMC and ex-vivo SMC/Fibroblasts harmonization showing modulated subpopulations. (F) Pseudotime analysis of harmonized iPSC-VSMC and Metaplaq SMC/Fibroblasts showing the trajectory of iPSC-VSMC and profiles of specific genotypes in the insert to the right.

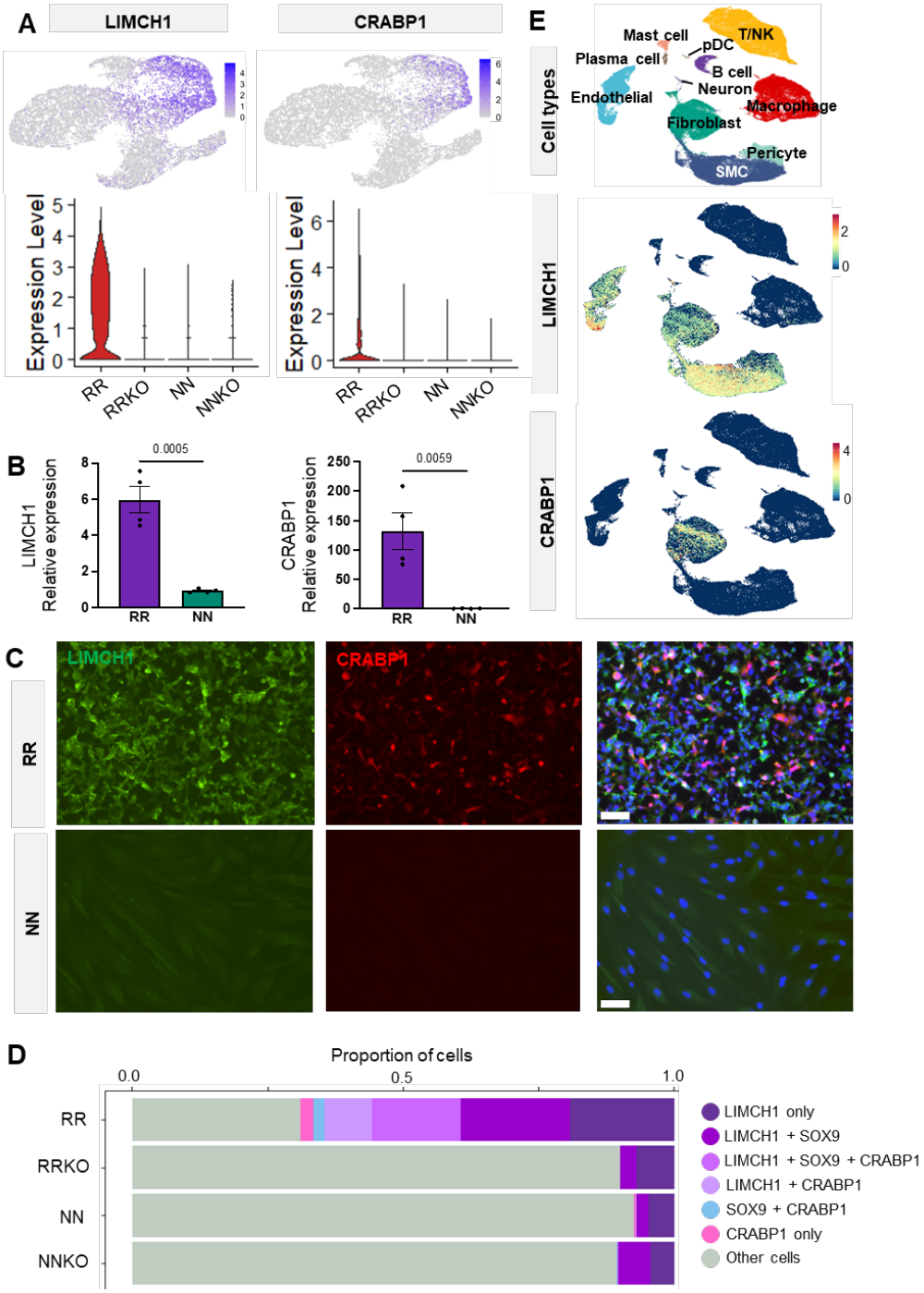
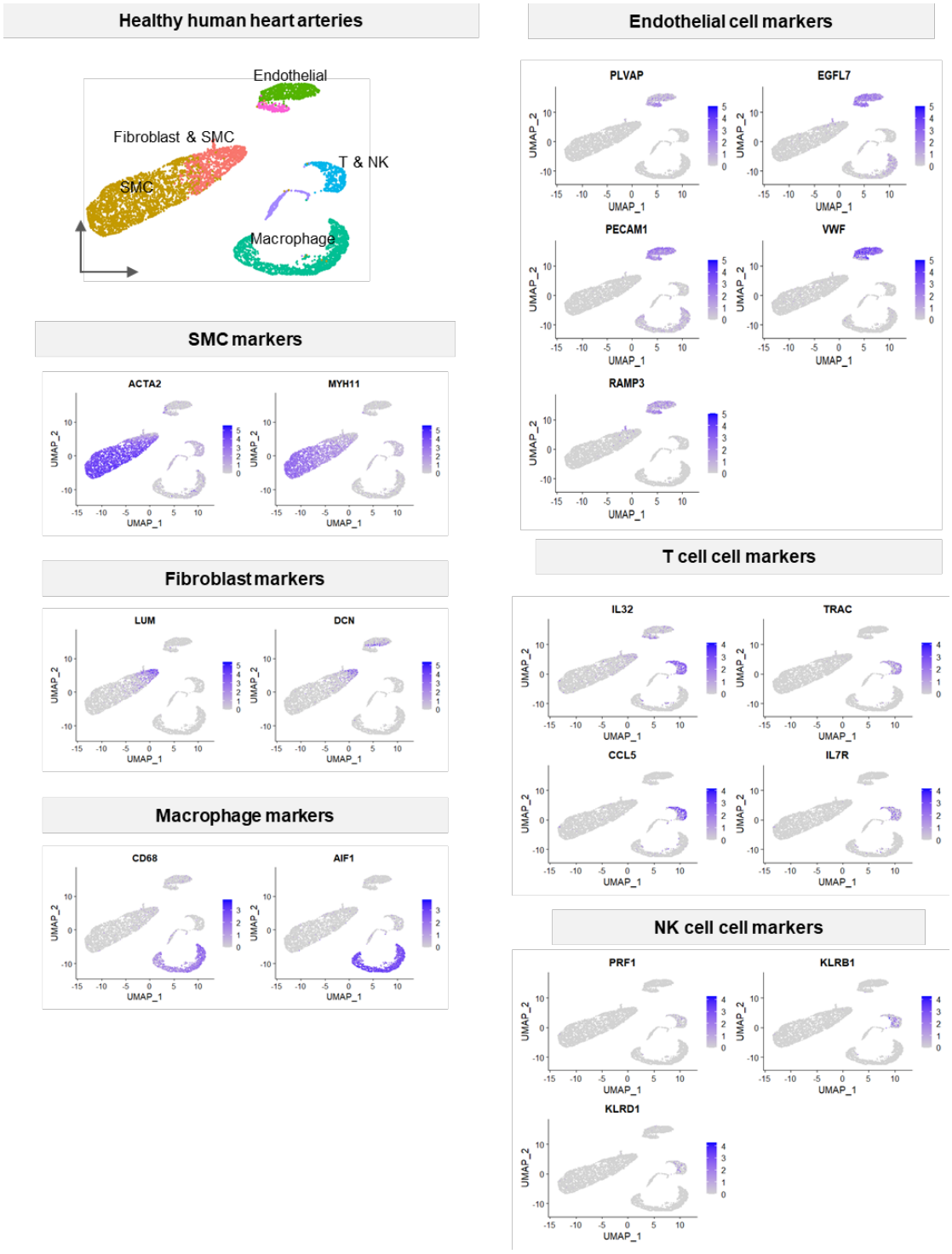
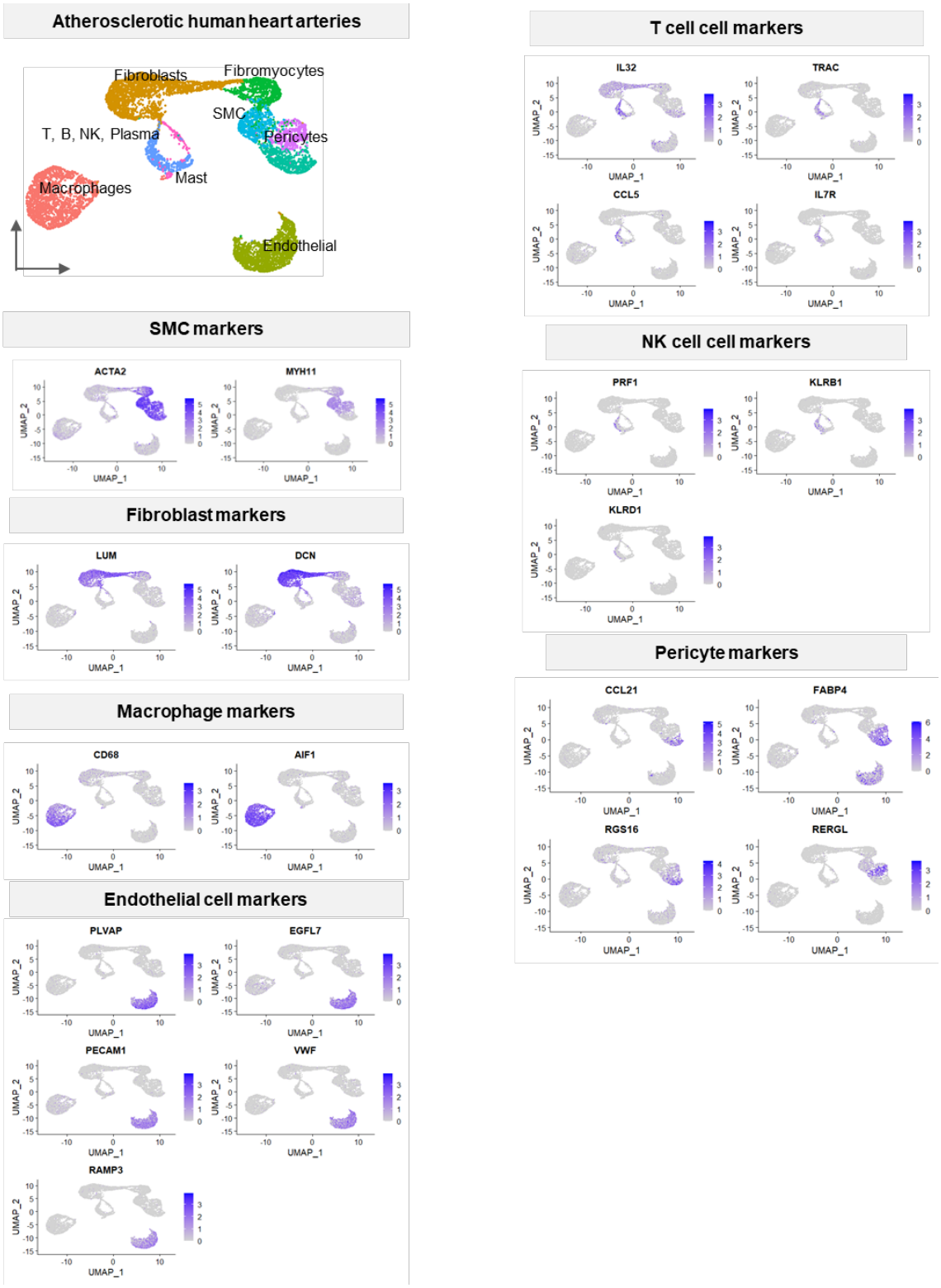


Figure 8. Identification of 9p21.3 risk VSMCs signature. (A) UMAPs showing upregulation of LIMCH1 and CRABP1 in iPSC-VSMCs (top), violin plots showing expression across genotypes (bottom), (B) qPCR of LIMCH1 and CRABP1 enrichment in 4 additional donors (2 RR and 2 NN) in 2 independent experiments. P values calculated with an unpaired t-test. (C) Immunocytochemistry staining for LIMCH1 and CRABP1 in RR and NN. White bar indicates 50 µm. (D) Metaplaq atlas of human artery samples labeled with cell types (top), LIMCH1 and CRABP1 expression in Metaplaq atlas. (E) Bar chart showing proportions of cells in different genotypes expressing LIMCH1, SOX9, and CRABP1.

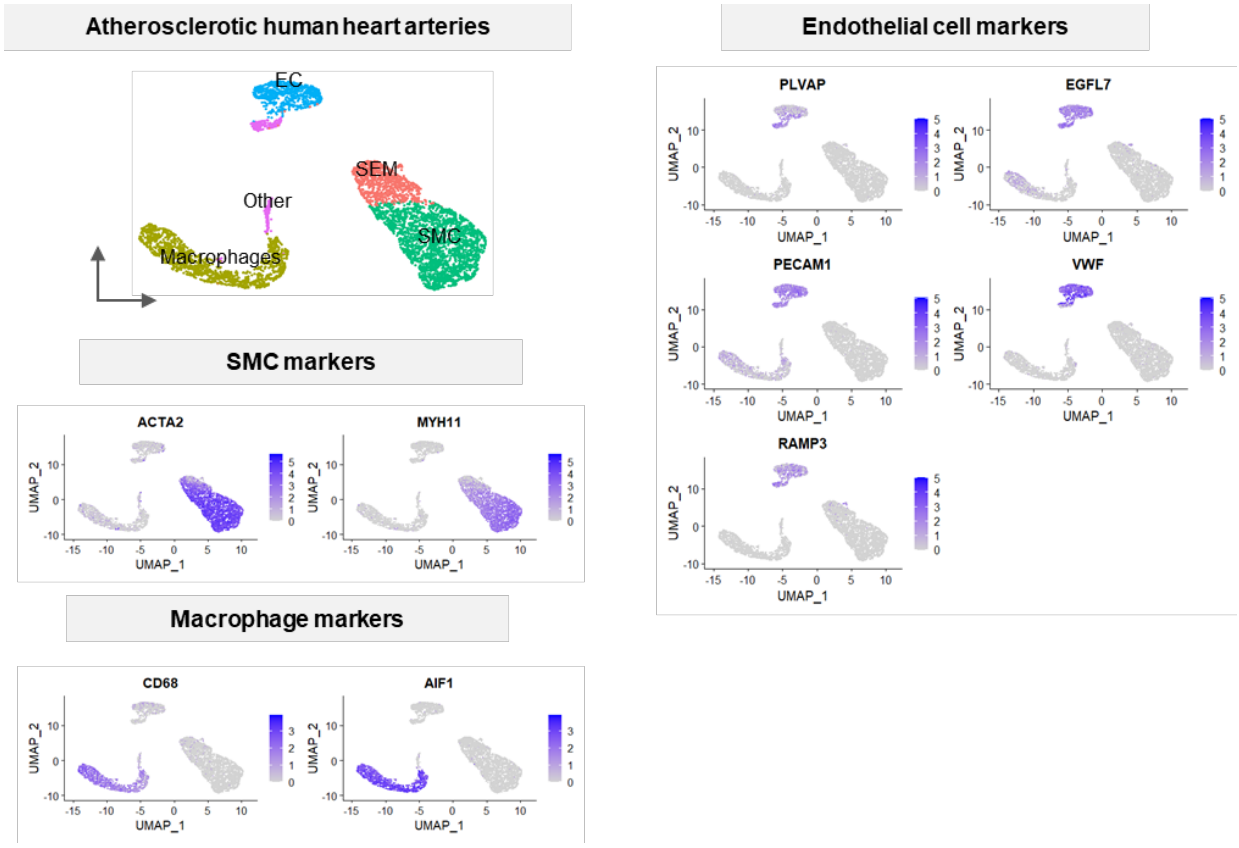
Supplementary Figures



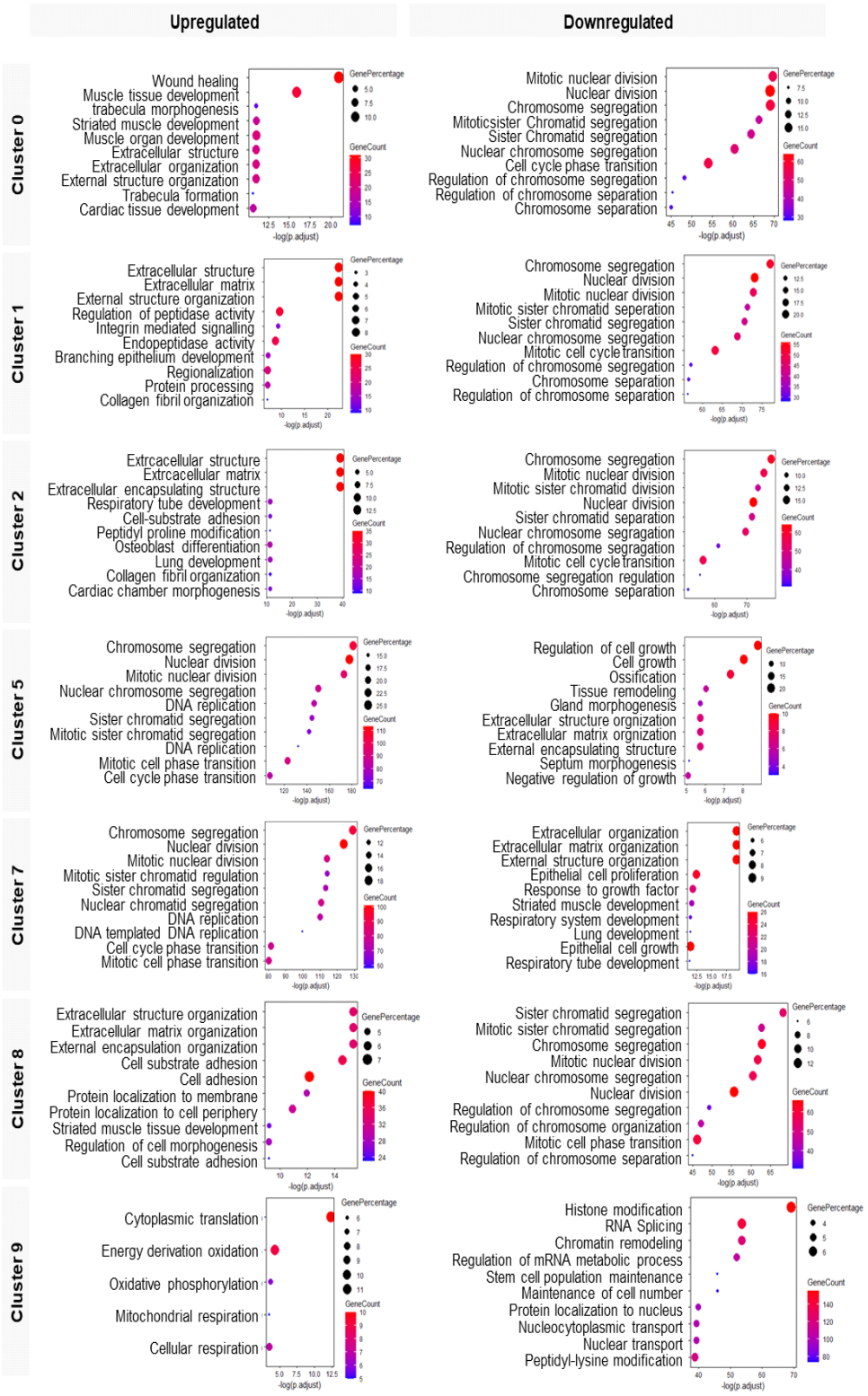
Extended data 1. UMAPs showing expression of genetic markers used to identify cell types in published datasets. Top left UMAP shows cell type annotations for healthy human artery data (Hu et al, 2021), remaining UMAPs show expression of cell type-specific markers.



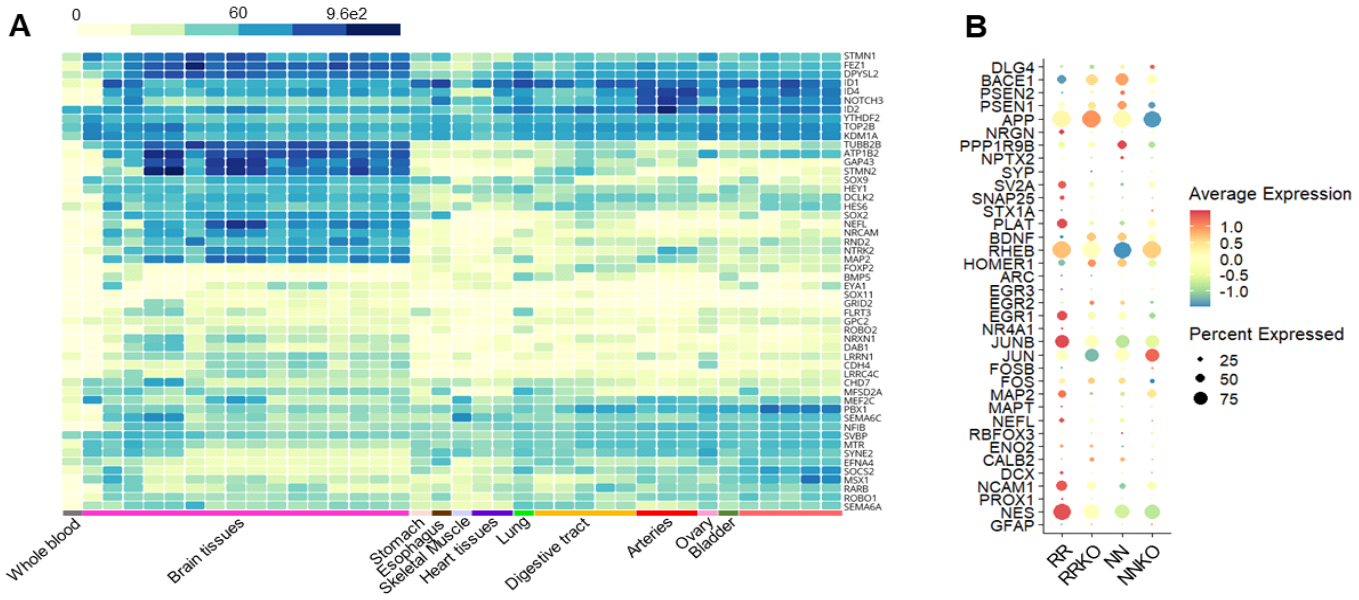
Extended data 1 (cont.). UMAPs showing expression of genetic markers used to identify cell types in published datasets. Top left UMAP shows cell type annotations for atherosclerotic human artery data (Wirka et al. 2019), remaining UMAPs show expression of cell type-specific markers.



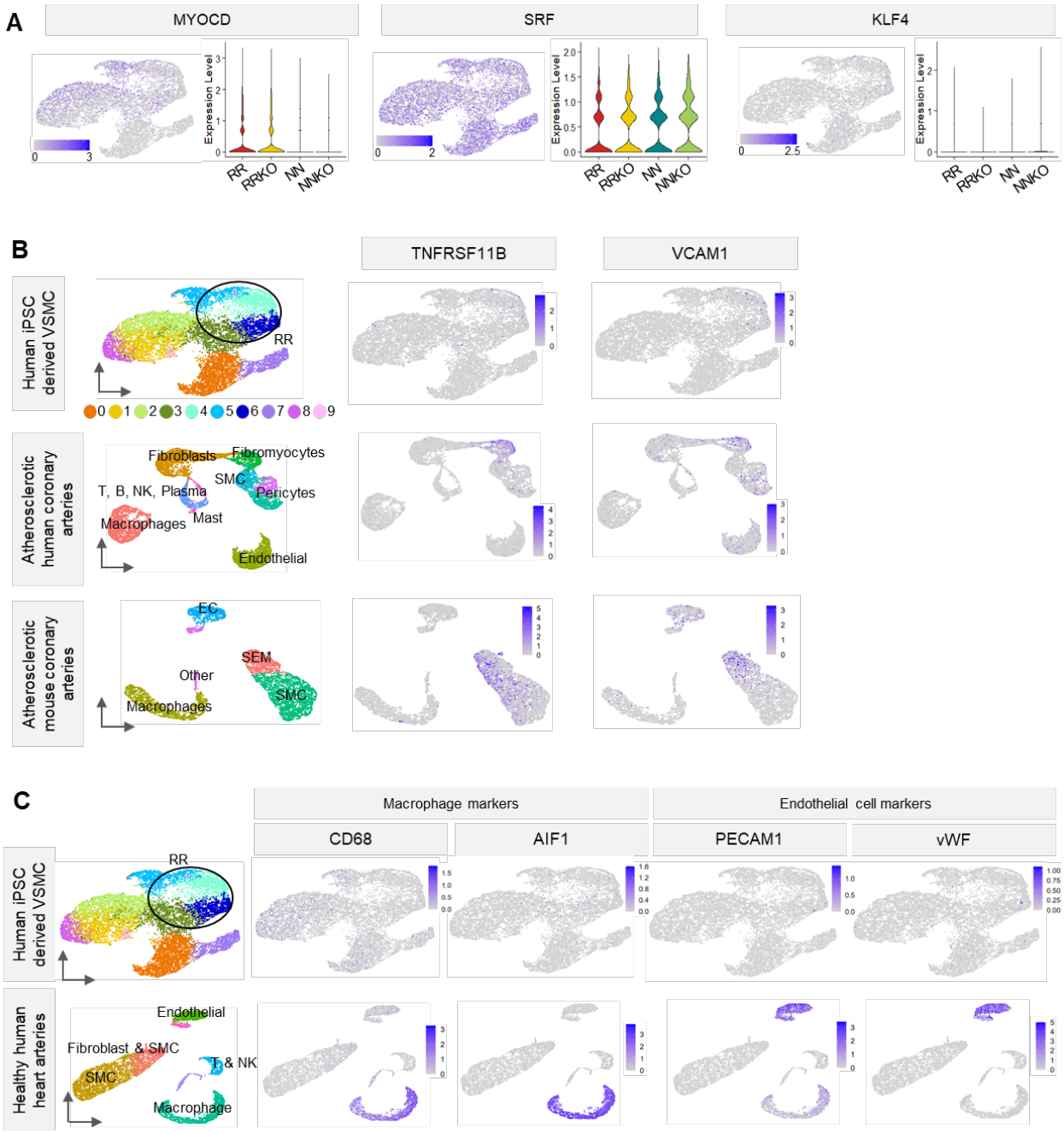
Extended data 1 (cont.). UMAPs showing expression of genetic markers used to identify cell types in published datasets. Top left UMAP shows cell type annotations for additional atherosclerotic human artery data (Pan et al. 2020), remaining UMAPs show expression of cell type-specific markers.



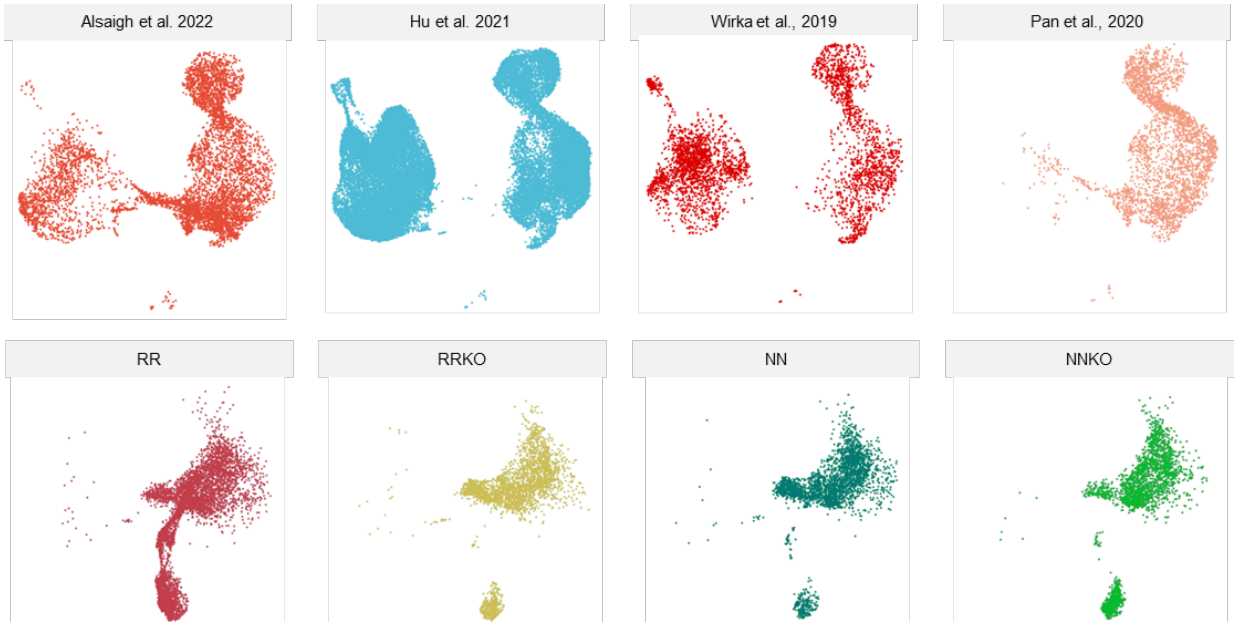
Extended data 2. Gene Ontology enrichment analysis for non-RR clusters. Left hand column shows upregulated terms, right shows downregulated. Gene ontology analysis conducted with clusterProfiler



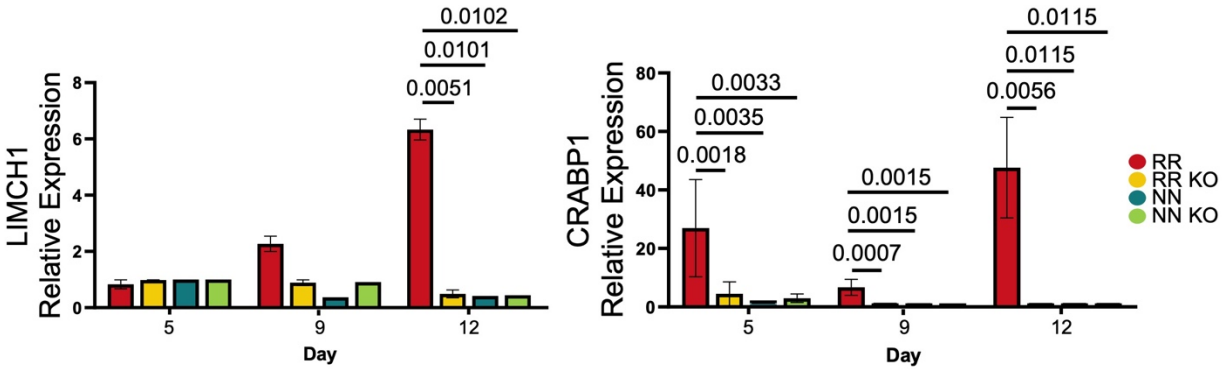
Extended data 3. Analysis for neuronal differentiation of RR VSMC. (A) Heatmap of expression of genes associated with neuronal gene ontology terms across brain and VSMC-rich human tissues (Generated with the Genotype-Tissue Expression (GTEx) Portal) (Lonsdale et al. 2013). (B) Bubble plot showing expression of known neuronal markers across genotypes.



Extended data 4. Analysis of known modulators of VSMCs phenotypic switch and known CAD-related VSMCs states. (A) UMAPS and violin plots showing expression for three classic VSMC phenotype modulators, MYOCD, SRF, and KLF4. (B) iPSC derived VSMC (top row) compared to atherosclerotic coronary arteries (Wirka at el. 2019, center row) and atherosclerotic carotid arteries (Pan et al. 2020, bottom row). Left hand UMAPs colored by clusters and labeled with cell types, other UMAPs colored according to expression of the gene named above the UMAP (C) Comparison of iPSC derived VSMC (top row) to endogenous healthy artery data (bottom row). Two examples of gene markers given for macrophages and endothelial cells.



Extended data 5. Breakdown of harmonized integration of iPSC-VSMC and Metaplaque atlas. SMC/Fibroblasts by contributing published dataset (top row) and by iPSC-derived VSMC genotype (bottom row).



Extended data 6. RR VSMCs signature genes expression overtime. Expression of signature genes over time, during iPSC differentiation to VSMCs in all genotypes. Statistical significance was calculated with a one-way ANOVA with Bonferroni correction at each time point. Comparison shown are RR vs other genotypes.