The 9p21.3 coronary artery disease risk locus drives vascular smooth muscle cells to osteo-chondrogenic 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]Department of Public Health Science; University of Virginia; Charlottesville, VA 22904 USA

[†]Corresponding Author: losardo@wisc.edu

Abstract

Genome-wide association studies have identified common genetic variants at approximately 300 human genomic loci linked to CAD susceptibility. Among these genomic regions, the most impactful is the 9p21.3 CAD risk locus, which spans a 60 kb gene desert, encompasses about ~80 SNPs in high linkage disequilibrium, and has no defined function. We used induced pluripotent stem cell (iPSC) lines from risk and non-risk donors at 9p21.3, as well as isogenic lines with a full haplotype deletion. iPSC-derived vascular smooth muscle cells (iPSC-VSMCs) were used for single-cell transcriptomics. iPSC-VSMCs resemble the heterogeneity observed in human coronary arteries, establishing the robustness of this model. Our analysis revealed that the 9p21.3 risk haplotype prompts VSMCs to acquire a novel cellular state showing osteochondrogenic features, and we identified a set of signature genes crucial for defining this transcriptional program. Our study provides new insights into the 9p21.3 risk locus and its role in driving disease-prone states in VSMCs.

Introduction

Cardiovascular disease, particularly coronary artery disease (CAD), remains a significant medical concern and a major cause of death worldwide[1–3]. While lifestyle choices and environmental factors contribute to CAD risk, genetics also play a substantial role, accounting for 40-60% of cases[4–7]. Genome-wide association studies (GWAS) have identified ~300 genomic loci linked to susceptibility to CAD[8–10]. About 25% of these loci are predicted to function in the vascular wall, compared to only 10% linked to lipid metabolism, which has been the most pursued mechanism in CAD pathogenesis and therapy[10–13].

The 9p21.3 CAD risk locus was discovered in 2007 through three independent CAD GWAS studies[14–16]. It is unique to humans and encompasses a haplotype block of approximately 60 Kb, containing about 80 SNPs in strong linkage disequilibrium in most populations[14–16]. The SNPs are inherited together, making it challenging to pinpoint specific causal variants responsible for the CAD risk. The locus is located in a gene desert region and only partially overlaps with the extensively spliced long non-coding RNA (lncRNA) ANRIL[17]. 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[18]. A previous study reported the deletion of a putative syntenic region in the non-coding region downstream of the CDKN2A/B genes in the murine chromosome 4. These mice exhibited increased tumor incidence and increased aortic smooth muscle cell proliferation but not atherosclerosis[19]. The lack of conservation of this genomic region across most mammals, except primates, indicates the need for new approaches that maintain the complexity of the human genome to study the human 9p21.3 locus.

Vascular smooth muscle cells (VSMCs) are the most abundant cells in the arteries and exhibit remarkable plasticity, characterized by transition in diverse cell states, in physiological conditions and during CAD pathogenesis[20,21]. VSMCs have traditionally been 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 lower expression of contractile markers[20–22]. The reversible phenotypic switch of VSMCs between contractile and synthetic states is essential for vascular development and repair[20–24]. Synthetic VSMCs have also been implicated in the pathogenesis of atherosclerotic plaque formation, displaying a hyperproliferative phenotype and aberrant migration into the intima layer[21,25–27]. While historically VSMC transition has been described as a phenotypic switch between contractile and synthetic states, recent studies using single-cell technologies, such as single-cell RNA and ATAC sequencing, have revealed a higher degree of heterogeneity. Analyses of human coronary arteries, both healthy and affected by atherosclerosis, have suggested that VSMC plasticity is more complex than the classic dichotomy of contractile and synthetic phenotypes, and involves transcriptional and epigenetic landscapes that are more multifaceted[28–32]. To date, no report has described how single non-coding SNPs or large CAD haplotypes may drive or perturb VSMC plasticity and phenotypic switching. The impact of the 9p21.3 CAD locus on VSMC states and transitions has not been described. To address the role of 9p21.3 risk haplotype to influence cell state changes in VSMCs, we have leveraged a system we have previously generated and validated, based on induced pluripotent stem cells, vascular smooth muscle cells differentiation and haplotype editing. We previously derived induced pluripotent stem cells (iPSCs) from 9p21.3 CAD risk (RR) and non-risk (NN) homozygous donors. We performed a full haplotype editing in both backgrounds and generated isogenic knockout lines (RRKO and NNKO)[33]. Using this system, we conducted a comprehensive single-cell transcriptional profile to identify VSMC states and potential state trajectories mediated by the risk haplotype at the 9p21.3 locus.

Our single-cell profiling provides evidence that VSMCs carrying the 9p21.3 risk haplotype are characterized by less transcriptional heterogeneity than non-risk VSMCs or VSMCs with a complete deletion of the 9p21.3 haplotype, which show overlapping profiles. In an isogenic, environmentally controlled system, we demonstrate that the risk haplotype at 9p21.3 drives VSMCs to acquire an osteochondrogenic-like state. Risk iPSC-VSMCs show increased expression of SOX9 and upregulation of COL2A1, both markers of chondrocyte-like cells, despite maintaining VSMC identity. Our pipeline has revealed two signature genes, LIMCH1 and CRABP1, defining the 9p21.3-risk driven VSMC state. Moreover, our analysis shows that the risk haplotype at 9p21.3 alters the expression of genes in other CAD loci, suggesting an interconnected CAD gene network that may increase risk susceptibility in concert. Additionally, iPSC-VSMCs recapitulate the transcriptional heterogeneity and complexity of different VSMC states found in *ex vivo* coronary arteries, proposing iPSC-VSMCs as a tool to uncover the dynamics of cell state changes in disease-relevant cell types driven by human-specific CAD genomic regions.

RESULTS

iPSC-VSMC 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. Briefly, iPSC lines were derived from both risk and non-risk homozygous backgrounds at 9p21.3. TALE-mediated genome editing was used to delete the 60 Kb CAD region at 9p21.3[33]. These isogenic iPSC lines share identical genetic backgrounds and differ exclusively at the 9p21.3 CAD region, allowing strict and controlled evaluation of the 9p21.3 function in determining cellular phenotypes. We established a collection of lines composed of four genotypes, risk cells with intact haplotype (RR), isogenic risk knockouts (RRKO), non-risk cells with intact haplotype (NN), and non-risk knockouts (NNKO). To control for potential variability across iPSC lines, we used two clones per genotype for a parallel vascular smooth muscle cells (VSMCs) differentiation. A total of 8 lines (carrying the 4 genotypes) were differentiated to VSMC over the course of 17 days as we previously described[33]. At D17, cells (iPSC-VSMCs) were harvested and a total ~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 analyzed our data by Uniform Manifold Approximation and Projection (UMAP) to evaluate gene expression patterns across our samples.

To establish the robustness of this *in vitro* system in mimicking endogenous cellular phenotypes, we established a comparison between iPSC-VSMCs and *ex-vivo* VSMCs from human arteries. We evaluated the expression level of established SMC markers, namely ACTA2, CNN1, and TAGLN (Figure 1B) in parallel with publicly available scRNA sequencing data from healthy human coronary arteries[29]. After *de novo* clustering of the human artery data and cell identification through cell-type specific markers, we confirm that iPSC-VSMCs are an overall homogenous population expressing VSMC markers (Figure 1B and Supplemental Figure 1). To further support the validity of iPSC-VSMCs as a model, we leveraged the newly established Metaplaque atlas, a multi-dataset atlas of human coronary artery cells[32]. Transcriptomic projection of iPSC-VSMCs into the human coronary atlas showed strong overlap between iPSC-VSMCs and the *ex-vivo* SMC cluster. We detected a minimal overlap of few cells with the fibroblast population (Fibs) and small subset of iPSC-VSMCs projecting with the endothelial cell cluster (ECs), confirming the strong similarity of transcriptomic profiles between iPSC-VSMC and *ex-vivo* human SMC (Figure 1C).

We performed an unsupervised clustering analysis including all genotypes of iPSC-VSMCs and detected 10 transcriptionally distinct cellular clusters (Figure 2A). This initial analysis showed that, although iPSC-VSMCs exhibit overall homogeneous expression of cell-type-specific markers, our system can detect different subpopulations of cells, 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 NN, NNKO, and RRKO samples, while the RR cells are practically devoid of these clusters. Clusters 2 and 5 are equally represented across all genotypes. Cluster 8 was mostly represented in RRKO and NNKO, and cluster 7 was mainly present in

NNKO cells. While NN, NNKO, and RRKO shared similar cluster patterns, the RR VSMCs comprised clusters 3, 4, and 6 up to ~72% (Figure 2A). Grouping the cells by genotype highlighted 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 show intermixing in three regions of the UMAP. RR cells populate mostly three clusters 3, 4, 6 each comprised of 80% or more risk cells (Figure 2B). NN, NNKO, and RRKO are intermixed in multiple clusters, suggesting that the deletion of the 9p21.3 CAD risk region reverts the RR transcriptome to a NN status. Additionally, deletion of the non-risk haplotype seems to result in an additional cluster, cluster 7. 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. RR, RRKO, NN, and NNKO showed comparable differentiation performance 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 whether clusters uniquely populated by RR VSMCs had exclusive expression of these genes, explaining the divergent state. Interestingly, CDKN2A and CDKN2B expression did not localize to any specific clusters and were not enriched in RR VSMCs, excluding the possibility that the difference in RR VSMCs transcriptomic profile is due to a genotype-specific expression of these genes (Figure 3A). Previously, we described that ANRIL, the lncRNA partially overlapping the CAD region at 9p21.3, has genotype-specific isoforms expression in iPSC-VSMCs carrying different genotypes at 9p21.3. Specifically, the total amount of ANRIL is significantly upregulated in RR VSMCs compared to NN. While long ANRIL isoforms, terminating in exon 20, are equally expressed in RR and NN VSMCs, short ANRIL isoforms are expressed exclusively in RR VSMCs. Here, we confirmed increased ANRIL expression, attributable to short ANRIL isoforms, as the long isoforms did not show significant differences across genotypes (Figure 3B)[33,34].

The 9p21.3 risk haplotype reduces VSMC plasticity

To identify cluster-specific gene expression signatures, we conducted a Differential Expression Analysis (DEA) and identified the top 10% cluster-specific-genes across all clusters (Figure 4A). The DEA highlighted a set of genes differentially expressed specifically in the risk clusters 3-4-6 (Figure 4A). Gene Ontology (GO) enrichment analysis conducted with the R studio package clusterProfiler 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. Cluster 5 and 7 show upregulation of cell cycle related GO terms, mostly related to active mitosis, concomitant with downregulation of extracellular matrix organization. Conversely clusters 0, 1, 2, and 8 showed downregulation of cell cycle processes and upregulation of muscle development, would healing and adhesion (Supplemental Fig 2). Remarkably, this analysis uncovered a different trend in the RR cells. Several terms related to healthy SMC function such

as cell adhesion and wound healing were downregulated in two of the three risk clusters (4-6). This trend was accompanied by upregulation of cell cycle genes in cluster 4 and not-SMC lineages, including cartilage development and axonogenesis in cluster 3 and 6 (Figure 4B). Interestingly, Cluster 6 exhibited a strong enrichment for neural-like gene ontology. 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 tissue with high SMC content (Ovary, Bladder, Lung, etc.) (Supplementary Figure 3A). We found promiscuous expression of these genes across many tissues, leading to the hypothesis that the enrichment in neuron-like GO terms is a result of biased functional annotation of most of these genes, not necessarily an alternative differentiation of RR VSMCs into neuron-like derivatives (Supplementary Figure 3A). To corroborate the absence of an alternative neuronal differentiation, we chose a set of known neuronal markers and analyzed their expression in our RR VSMCs. This analysis revealed no significant enrichment of neuronal genes in the risk cells (Supplementary Figure 3B). The presence of genes related to axon guidance and neuronal projections suggest mostly a signature of morphological changes occurring in RR VSMCs.

VSMCs are known to exhibit high plasticity and have been shown to acquire numerous transitional states[21,35,36]. Guided by the remarkable gene expression changes in RR VSMCs, we wanted to test whether different genotypes at 9p21.3 could drive independent transcriptional trajectories. We evaluated the dynamic of cell cluster transition across each genotype independently (Figure 5A). We hypothesized that a divergent cell trajectory would occur in RR cells, while the other three genotypes would follow a similar trajectory. We perform a Monocle3 pseudotime analysis and chose an anchor, as starting point of the trajectory, in cluster 2, which is comprised of approximately equal proportions of all genotypes. We evaluated the trajectory of gene expression in all the genotypes, separately. NNKO, and RRKO lines converged on the same trajectory, showing that the absence of the 9p21.3 haplotype has equal effect on transcriptional states, independently of the genetic background. The KO trajectory is characterized by a transition from cluster 2 through 0 and terminating in 7. Gene ontology analysis of these clusters indicates a transition from a quiescent (clusters 2 and 0) to a more proliferative state (cluster 7). The non-risk trajectory mirrors the KO transition with the exception of the last stage, terminating in cluster 0 instead of 7, suggesting fewer NN cells transition to a proliferative (synthetic-like) state than NNKO or RRKO (Figure 5A-B). Analysis of the risk trajectory substantiated our hypothesis of a divergent state acquisition for the 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-VSMCs signatures in these clusters proceeding from cluster 3 to 6, with no cells following the non-risk and KO trajectories, suggesting that the Risk haplotype represses the non-risk trajectory and drive VSMC to acquire a divergent state (Figure 5A-B).

To date, VSMCs phenotypic modulation has been described mostly in murine models of atherosclerosis. Through rigorous lineage tracing studies in mice and, more recently, through single cell transcriptomic analysis of human arteries samples, numerous VSMC state nuances have been identified. To investigate

whether the cell state transition of our *in vitro* system recapitulates VSMCs states or phenotypic switch previously identified, we analyzed the expression of known master regulators of VSMCs phenotypic modulation. The expression of MYOCD and SRF traditional modulators of VSMC phenotypic switch[37,38], did not localize to any particular genotype. Moreover, the stem cell marker KLF4[25,39,40] was not detected in our dataset (Supplementary Figure 4A). Overall, the expression pattern of these transcriptional regulators previously identified to have a role in VSMC modulation do not show genotype specific or cluster specific pattern, 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[21,25,28,31,32,35]. We then sought to investigate markers of CAD-related VSMCs states in RR VSMCs from our dataset. We used the 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, & Monocyte intermediary state), or evidence of transdifferentiation to other cell types, including macrophages or endothelial cells. We used TNFRSF11B as marker of fibromyocytes, as previously described by Wirka and colleagues, and VCAM1, as described by Pan in 2020. When parallel expression was examined in iPSC-VSMCs we found no expression of both markers in iPSC-VSMCs overall, suggesting the absence of these two cell populations in our dataset and excluding that RR VSMCs acquire one of these known VSMC states (Supplementary Figure 4B). Similarly, analysis of transdifferentiation toward alternative cell types, namely, macrophages (CD68 and AIF1) and endothelial cells (PECAM1 and vWF), showed negative results, suggesting that the RR VSMCs are not a result of transdifferentiation or alternative differentiation to other cell types. This analysis overall showed no significant resemblance between the RR cells and any of the cell types and states we surveyed (Supplementary Figure 4).

In mouse model of atherosclerosis, LGALS3 was previously shown to be linked to a macrophage-like disease state of VSMCs during CAD[41]. More recently, LGALS3 has been described as a marker of VSMCs plasticity and indicator of transitional state, indicating subpopulation of VSMCs that have ability to undergo phenotypic modulation[40]. Interestingly, LGALS3 showed homogeneous expression in all clusters populated by RRKO, NN and NNKO cells while RR VSMCs exhibit notable decreased levels compared to the other genotypes (Figure 6A). This pattern suggests that NN, NNKO and RRKO are present in a dynamic transitional phenotype, while the RR cells likely lack state plasticity and do not undergo 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[8] and found 13 homogeneously expressed in the Risk population and significantly differentially expressed between the RR and other genotypes (Figure 6B). This suggests that the 9p21.3 locus controls multiple other genes

linked to CAD and the risk status of the cells may arise from a cumulative effect of gene expression alteration at other genomic loci.

Risk iPSC-VSMCs exhibit osteochondrogenic-like state and uncover 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 described to be upregulated in VSMCs transitional state toward osteogenic- and chondrogenic-like cells[35,42,43]. Recently, SOX9 has been linked to vascular stiffening and aging through its modulation of collagen expression[44]. Evaluation of key markers for osteo and chondrogenic lineages between RR cells and their isogenic RRKO revealed overall higher expression of both lineages markers in the presence of risk 9p21.3 (Figure 7A). Remarkably, gene expression distribution of SOX9 revealed uniformly higher expression in RR cells compared to RRKO (Figure 7B-C). Additionally, SOX9 upregulation is accompanied by significant upregulation of COL2A1 and concomitant downregulation of COL1A1 and COL1A2, suggesting that the cells are acquiring a chondrogenic fate and actively switching ECM production (Figure 7B-C).

Using the most comprehensive human coronary artery single cell atlas available, Metaplaque[32], we examined whether iPSC-VSMCs resemble VSMCs state recently described by Mosquera and colleagues in *ex vivo* coronary arteries samples. Leveraging the Metaplaque atlas we perform an integration of SMC and fibroblast from the atlas and the iPSC-VSMCs. This integration showed remarkable overlapping of iPSC-VSMCs and SMC from *ex vivo* arteries (Figure 7D and Supplementary Figure 5). We applied a secondary annotation to this data integration looking at different cellular states for SMCs and fibroblasts, according to the previous report from Mosquera and colleagues. This analysis revealed that iPSC-VSMCs overlap mostly with Transitional SMC, Fibromyocytes, Foam-like SMC ad Fibrochondrocytes (Figure 7E). Pseudotime trajectory analysis incorporating SMCs and fibroblast from human coronary arteries and iPSC-VSMCs from this study showed that the non-risk iPSC-VSMCs (NN, NNKO, RRKO) overlap mostly with transitional populations of SMC, corroborating the finding using LGALS3 as marker of state modulation. Interestingly, RR cells resemble fibrochondrogenic SMC and partially foam-like cells. Additionally, we found a group of cells, mostly populated by RR VSMCs that diverge from most state described (Figure 7F), indicating a potential dynamic transitional cell population not abundant in the 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 an agnostic analysis to search for features that may not have been described in previous literature. Given the highly intermixed transcriptional profiles of the RRKO, NN and NNKO cells, we grouped these genotype together as “not containing the risk haplotype”. To determine risk-specific genes, 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 differentially expressed genes, highlighted only 11 that were uniformly

and uniquely expressed in the entire RR cell population. We selected two that had the highest fold enrichment and pvalue and were strongly and uniquely expressed in the RR cells: LIMCH1 and CRABP1 (Figure 8A).

Immunocytochemistry confirmed that the expression of transcripts translates to high abundance of the LIMCH1 protein across the entire RR population and highlight the concordant expression of CRABP1 in ~30% of RR VSMCs. Their expression is absent in the NN cells as predicted by the scRNAseq analysis (Figure 8C). To confirm the unique expression of these signature genes across multiple samples, we generate iPSC-VSMCs from 2 additional RR donors and 2 NN donors. Gene expression analysis by qPCR confirmed that LIMCH1 and CRABP1 increased expression in the VSMCs carrying the risk haplotype compared to NN (Fig 8B). We performed a time course analysis of expression for both signature genes to inquire whether they are acquired early on and may drive the terminal phenotype of RR VSMCs. Interestingly, both start detectable expression around day 9 of the differentiation protocol and linearly increase expression over the next 7-9 days. Their expression remains negligible in other genotypes (Supplementary Figure 6). Interestingly, the timeline of expression for LIMCH1 and CRABP1 aligns with the timeline of ANRIL short expression, we previously showed[33], increasing in the RR cells around the 7 days of differentiation from iPSC to VSMC, and LIMCH1 and CRABP1 begin increasing immediately afterward, around day 9, suggesting a possible link between these genes. Finally, we also checked the expression of these genes in the multi-dataset atlas of endogenous human arteries and found that LIMCH1 and CRABP1 are enriched in few cell populations. LIMCH1 can be found throughout the SMC population, with a slight enrichment in the fibromyocyte population and a subpopulation of endothelial cells predicted previously to undergo EndoMT (Figure 8D). Interestingly, CRABP1 was mostly found in the fibroblasts, but in a subpopulation predicted to resemble myofibroblast. Our analysis shows that the signature genes we identified through the iPSC-VSMCs dataset could resemble novel cellular states in human tissue to be studied further in future studies.

DISCUSSION

In the past decade, GWAS have made significant advancements in identifying genetic variants associated with CAD[8,45,46]. However, the translation of this knowledge into functional validation and mechanistic insights remains a challenging task. One of the biggest limitations is the lack of a system to evaluate the causal link between human-specific genetic traits and the biology of the tissues and cell types mainly involved in the disease. As the majority of genetic variants associated with CAD are located in non-coding regions, it can be arduous to determine their precise functions[47]. Furthermore, the non-coding genome lacks conservation across species[48], making it difficult to use reliable animal models to understand the complex genetics of human-specific CAD risk region. How non-coding genetic risk factors for CAD influence the ability of vascular cells to acquire disease-prone states is not well defined. Recent efforts using single cell omics have identified CAD-linked cellular states in human coronary arteries[28,30–32].

However, how CAD-linked genetic variants or large CAD haplotypes interfere with cell state maintenance dynamic and cell plasticity (intended as the ability of the cells to rapidly adapt to physiological stimuli and pathological events) is largely unclear. Most studies, performed using human samples, capture a spatial and temporal snapshot of the vascular tissue, bearing the limitation 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 CAD locus was the first human locus to be identified as a CAD genetic risk factor almost two decades ago through GWAS[14,15,49], but its function is not yet fully defined. The risk haplotype frequency (~50% of non-African descent population) and its strong additive effect for each risk allele results in ~10-15% of CAD incidence in the US[50], approximately ~ 3 million people. Numerous studies have explored the function of the 9p21.3 locus using various approaches, including primary cultures[51–55], cell lines[54,56] and primary tissues[57,58]. 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[11,59]; ii) most studies using primary cultures have predominantly employed cell types not primarily associated with atherosclerosis (whole blood, T cells, PMBCs)[51–54]; and iii) functional studies on the 9p21.3 CAD locus employing genetic manipulation have been based on immortalized cell lines (HEK cells)[54] or mouse models that lack synteny with the human 9p21.3 region[19]. 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[33,34]. However, previous approaches have left many questions unanswered about how the 9p21.3 locus affects VSMC heterogeneity, state changes, and disease trajectories. In this study, we leverage the power of induced pluripotent stem cell (iPSC) and their vascular derivatives, to decode the role of the 9p21.3 risk haplotype in VSMC state transition. iPSCs accurately maintain the genetic background of the donor and erase the epigenetic marks of the somatic tissue they originate from. 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 an excellent tool for evaluating cell-autonomous mechanisms in specific cell types based on their genetic repertoire.

In this study, we show that iPSC-derived VSMCs reflect the overall transcriptional profile of *ex-vivo* SMCs and recapitulate the cell heterogeneity of native tissue. This highlights the power of this tool to explore cell state changes and cellular dynamics in the vascular wall. Using isogenic knockout lines in both risk and non-risk 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. For the first time, we provide causal evidence of the 9p21.3 locus driving VSMC cell state towards 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 SMCs states, characterized by loss of contractility and increased expression of SOX9, RUNX2 and BGLAP, have been previously reported in connection with CAD[42]. We previously showed that RR VSMCs have lower contraction force[33], and now provide data in support of upregulation of SOX9, COL2A1 and BGLAP. However, RR VSMCs are not enriched in RUNX2 cells, suggesting that the 9p21.3 risk haplotype may be prompting VSMCs toward a novel state that does not fully resemble previously described VSMCs 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 regulation of extracellular matrix composition, leading to arterial stiffness[44]. Moreover, LGALS3 has been described as a marker of SMC plasticity[40]. In our model RR VSMCs show increased expression of SOX9 and decreased levels of LGALS3. We speculate that the 9p21.3 CAD risk locus alters the vascular wall by reducing VSMC plasticity and increasing stiffness. This phenotype ultimately could affect the adaptability of the cells to external stimuli and potential damage, making the vessel more sensitive to recurrent insults and unable to transition to different phenotypes essential for maintaining healthy vessel function.

In this study, we also identify two new signature genes, LIMCH1 and CRABP1, that define the 9p21.3 risk status. These genes are uniquely expressed in risk VSMCs and are minimally detected in other genotypes, suggesting they are controlled by the 9p21.3 risk haplotype. Although LIMCH1 and CRABP1 have been previously detected during heart development[60,61], their role in VSMC differentiation and state transition has not been described. Further studies are needed to define the roles of these genes in VSMC biology.

Importantly, we provide evidence that these genes are expressed in *ex-vivo* human coronary samples, thus excluding a potential artifact of our system. Moreover, the increased expression of LIMCH1 and CRABP1 is consistently replicated in iPSC-VSMCs from additional donors, providing strong evidence of the linkage between the risk haplotype at 9p21.3 and the exclusive upregulation of these signature genes.

Our transcriptional profile integration, including 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. Interestingly, the RR VSMCs compose also a cluster that does not seem to recapitulate a specific *ex-vivo* population but resemble a continuous and advanced stage of *osteochondromyocyte* in the pseudotime trajectory. We propose two possible explanations for this observation. First, the atlas, which includes only a few datasets, might not contain representations of this extreme phenotype. Alternatively, it is possible that our iPSC-based culture enhances this phenotype to a degree that is not captured *in vivo*. Since our system of iPSC-VSMCs lacks other cell types and a “natural” vascular niche, we cannot exclude the possibility that this extreme phenotype *in vivo* is attenuated 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 VSMCs state transitions. Moreover, we highlight the power of the iPSC-based system to uncover new biology of VSMCs and potentially other disease-relevant cell types to understand cellular trajectories dynamics driven by the most impactful CAD genomic locus.

Methods

Ethical compliance and Cell lines

The authors have complied with all ethical regulations. Human subjects were enrolled and informed consent obtained under a study approved by the Scripps IRB (#115676) and cells were transferred and maintained under a study approved by 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 from a total of five donors. Lines included in the scRNA sequencing analysis are from two donors: C512 (RR clones 1-5 and 2-3 and isogenic RR KO clones 1-9 and WB46), 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 in Lo Sardo et al. Cell 2018. Additional iPSC-VSMCs were used to confirm main findings, including unedited clones from RR donors HE787 and C021 and C939 and NBD5433.

iPSCs derivation and differentiation

Induced pluripotent stem cells were previously derived and characterized (Lo Sardo et al. 2017). Peripheral blood mononuclear cells (PBMCs) were reprogrammed to induced pluripotent stem cells (iPSC) through episomal plasmid transfection of reprogramming factors (Lo Sardo et al. 2018). iPSC were cultured on Matrigel-coated (Corning) cell culture plates and mTeSR (Stem Cell Technologies) and cultured at 37℃ in normoxia 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 were cultured in mTeSR and fed daily with splits performed approximately every 4-5 days, or when the culture reached ~75% confluency. 0.5 mM EDTA in PBS was used for detachment of iPSCs for regular split.

iPSC were differentiated into VSMCs as previously described (Lo Sardo et al. 2018) adapted from Cheung et al. 2014. Briefly, iPSC were cultured in mTeSR until approximately 50% confluency. Media was changed in CDM-BSA (250 mL IMDM (GIBCO, Cat. 12440-053), 250 mL F12 (GIBCO, Cat. 11765-054), 5 mL pen/strep (GIBCO Cat. 15140122), 5 mL chemically defined lipid concentrate (Life Tech Cat. 11905-031), 20 µL monothioglycerol (Sigma-Aldritch, Cat. 96-27-5), 0.5 g BSA (Sigma-Aldritch, Cat. A9418), 250 µL transferrin (GIBCO, Cat. 51500-056), 350 µL insulin human recombinant (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 (same as CDM-BSA, replacing BSA with 0.5 g PVA (Sigma-Aldritch, 9002-89-50)) and differentiated to lateral mesoderm lineage for five

days. At day 5, 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. After 12 days since seeding, the cells were harvested for single cell sequencing and maintained in culture in DMEM (GIBCO Cat. 10566-016) supplemented with 10% FBS, Pen/strep (GIBCO Cat. 15140122), Glutamax (GIBCO Cat. 35050061), NEAA (GIBCO Cat. 11140050) and passaged every 5-6 days.

Single cell sequencing cell suspension and library preparation

On day 17, terminally differentiated VSMC were detached using Tryple E (Cat. 12604-013), and incubated 3-5 minutes at room temperature. Tryple E was diluted with CDM-PVA and cell suspension was centrifuged at 200 g for 5 minutes. Cells were gently resuspended in 0.04% BSA in PBS, passed on a 30 µm cell strainer then counted using Trypan Blue (Sigma-Aldritch Cat. 72-57-1) solution and an Invitrogen Countess II cell counter. The cell suspension was centrifuged at 200 rcf for 5 minutes, and then the supernatant was removed. 1 mL of PBS with 0.04% BSA was added and cells were counted again. After centrifugation, the samples were gently resuspended in 0.04% BSA in PBS at a concentration of 1000 cells/ul. Cells were maintained on ice and submitted for 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 as per manufacturer instructions. Target cell recovery (TCR) values were set according to quality and viability of samples and calculated for each sample to obtain ~3000 cells/sample. GEM generation and post-GEM clean-up procedures were completed as per manufacturer instructions.

Number of cycles for the thermocycler protocol during the cDNA amplification step was set to 12, based upon targeted cell capture. Quality control for cDNA amplification was then done using an Agilent HS DNA chip and 1 uL of sample diluted to 10 uL with nuclease-free water to calculate the total cDNA yield. Quality control results summarized in the table below.

Sample	**pg/uL**	**Size range (bp)**	**Yield (ng)**
C1-2	720.95	200-9000	288.38
C2-3	883.43	200-9000	353.37
CWB46	765.14	200-9000	306.06
C1-9	1058.48	200-9000	423.39
HE56	866.65	200-9000	346.66
HE14	938.9	200-9000	375.56

HE17	955.93	200-9000	382.37
HE50	1862.11	200-9000	744.84

Fragmentation end repair and A-tailing, adaptor ligation and cleanup, sample index PCR, post sample index PCR double sided size selection, and post-library construction quality control were all completed as per manufacturer instructions. SPRI reagent lot number was 18198500, and Dual Index TT Set A lot number was 162043. MiSeq was used for QC and balance of the samples was done using calculations based on UMI-tools (Smith et al, 2017). Samples libraries were balanced for the number of estimated reads per cell and run on an Illimina NovaSeq system on a NovaSeq S1 flow cell at the University of Wisconsin – Madison Biotechnology Center's Gene Expression Center Core Facility.

Single cell RNAseq data were firstly analyzed by the UW Bioinformatics Resource Center using Cell Ranger pipelines from 10x Genomics. Data was 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 cells sequencing of iPSC-VSMCs

Filtering. Data was 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 was 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 was conducted using Seurat V4.3.0. After filtering, counts were normalized using SC Transformation, analyzed through PCA, and clustered by the Louvain method. Variation from percent mitochondrial RNA was regressed out during the 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 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 made with ggplot2 (version 3.4.2) DotPlot and ggplot commands.

Gene Ontology. Gene ontology analysis was conducted using clusterProfiler (version 3.0.4) with statistically significant genes from the differential gene expression analyses. 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.). Data was made into a Monocle readable object using the as.cell_data_set command, cells were clustered using cluster_cells command, trajectories were calculated using learn_graph, starting nodes were chosen using order_cells, and 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 counts matrices and filtering/clustering using the metrics and filters below (detail in 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 in order 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. 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. Fixed samples were incubated with 0.5% Triton X-100 (T8787) in PBS for cytoplasmic and nuclear staining. Cells were then blocked with 5% FBS (GeminiBio 100-106) in PBS for 1 hour at room temperature. Blocking solution was aspirated and replaced with primary antibody for 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 4C. Primary antibody solution was then aspirated followed by three PBS washes. Secondary antibody 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 and incubated for 1 hour at room temperature. Three additional PBS washes

were performed after secondary antibody incubation, and the samples imaged with fluorescence microscopy.

RNA Extraction, cDNA synthesis, and qPCR

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

To make cDNA, 200 ng of RNA samples was diluted in Nuclease free water to a total volume of 15 ul. A master mix of 1ul iScript Reverse Transcription and 4ul iScript Reaction Mix (1708891, BioRad) per sample was made to have a total of 20ul reactions. Samples were placed in the thermocycler following the manufacturer protocol. This comprised 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. Samples were then moved directly to qPCR use or stored at -80°C.

For qPCR, each well of a 384 well plate was filled with, iTaq Universal SYBR Green Supermix (1725121), nuclease-free water, and cDNA samples of each line condition 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 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'

Metaplaque atlas Integration

Metaplaque data integration, projection and trajectory analysis was conducted as described in Mosquera et al. 2023.

Acknowledgements

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 lab 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. was supported by the UW-Madison Cardiovascular Research Center undergraduate fellowship and the American Heart Association undergraduate fellowship. 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, designed, performed and supervised all the experiments and analysis. C.D.M.V. performed qPCR analysis. R.Z. performed immunocytochemistry and fluorescence microscopy. E.S. and V.L.S. performed and analyzed scRNA sequencing and interpreted the data. J.V.M. and C.L.M. performed data integration with the Metaplaque dataset. C.L.M. provided input on 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. Tsao, C. W. *et al.* Heart Disease and Stroke Statistics—2022 Update: A Report From the American Heart Association. *Circulation* **145**, e153–e639 (2022).

2. Roth, G. A. *et al.* Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019 Update From the GBD 2019 Study. *J Am Coll Cardiol* **76**, 2982–3021 (2020).

3. Heron, M. & Anderson, R. N. Changes in the Leading Cause of Death: Recent Patterns in Heart Disease and Cancer Mortality. *Nchs Data Brief* 1–8 (2016).

4. Marenberg, M. E., Risch, N., Berkman, L. F., Floderus, B. & Faire, U. de. Genetic Susceptibility to Death from Coronary Heart Disease in a Study of Twins. *New Engl J Medicine* **330**, 1041–1046 (1994).

5. Won, H.-H. *et al.* Disproportionate Contributions of Select Genomic Compartments and Cell Types to Genetic Risk for Coronary Artery Disease. *Plos Genet* **11**, e1005622 (2015).

6. Lloyd-Jones, D. M. *et al.* Parental Cardiovascular Disease as a Risk Factor for Cardiovascular Disease in Middle-aged Adults: A Prospective Study of Parents and Offspring. *Jama* **291**, 2204–2211 (2004).

7. Myers, R. H., Kiely, D. K., Cupples, L. A. & Kannel, W. B. Parental history is an independent risk factor for coronary artery disease: The Framingham Study. *Am Heart J* **120**, 963–969 (1990).

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

9. 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).

10. 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).

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

12. Musunuru, K. *et al.* In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* **593**, 429–434 (2021).

13. Musunuru, K. & Kathiresan, S. Genetics of Common, Complex Coronary Artery Disease. *Cell* **177**, 132–145 (2019).

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

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

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

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

18. 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).

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

20. 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).

21. 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).

22. 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).

23. Shi, N., Mei, X. & Chen, S.-Y. Smooth Muscle Cells in Vascular Remodeling. *Arteriosclerosis Thrombosis Vasc Biology* **39**, e247–e252 (2019).

24. Louis, S. F. & Zahradka, P. Vascular smooth muscle cell motility: From migration to invasion. *Exp Clin Cardiol* **15**, e75-85 (2010).

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

26. Gomez, D. & Owens, G. K. Smooth muscle cell phenotypic switching in atherosclerosis. *Cardiovascular Research* **95**, 156–164 (2012).

27. Matic, L. P. *et al.* Phenotypic Modulation of Smooth Muscle Cells in Atherosclerosis Is Associated With Downregulation of LMOD1, SYNPO2, PDLIM7, PLN, and SYNM. *Arteriosclerosis Thrombosis Vasc Biology* **36**, 1947–1961 (2018).

28. 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).

29. 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.

30. 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.

31. 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).

32. 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).

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

34. 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).

35. 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).

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

37. 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).

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

39. 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).

40. 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).

41. 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).

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

43. 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).

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

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

46. 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).

47. 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).

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

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

50. 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).

51. 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).

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

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

54. 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).

55. 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).

56. 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).

57. 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).

58. 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).

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

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

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

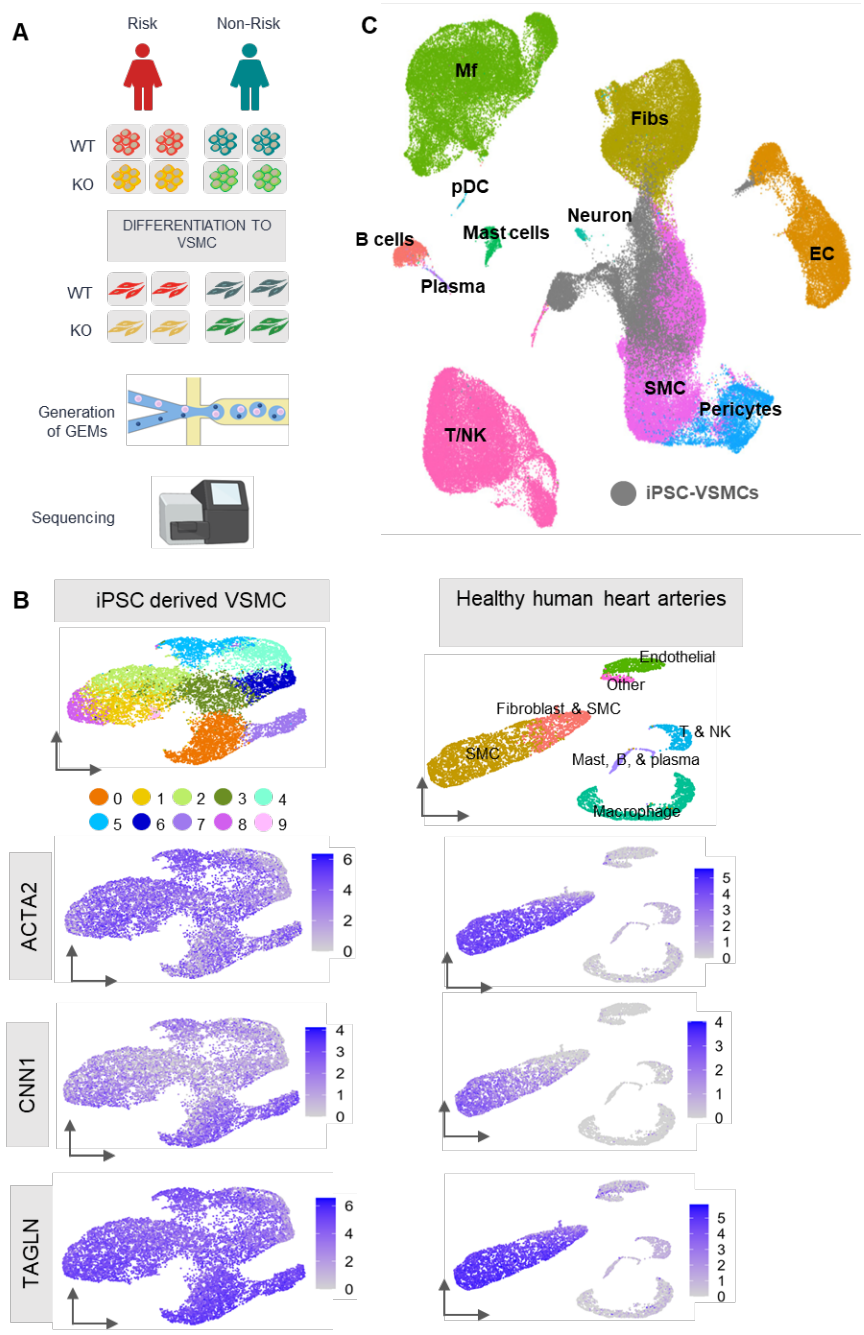

Figure 1. iPSC-VSMC are reflective of endogenous SMC. (A) Schematic of experimental workflow. iPSC are from one homozygous risk and one homozygous non-risk donors, both with intact 9p21.3 locus. Isogenic knockout lines for each genotype were derived by TALEN-mediated haplotype editing to delete the entire 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 Metaplaque atlas of *ex vivo* human artery cells from Mosquera et al., 2023.

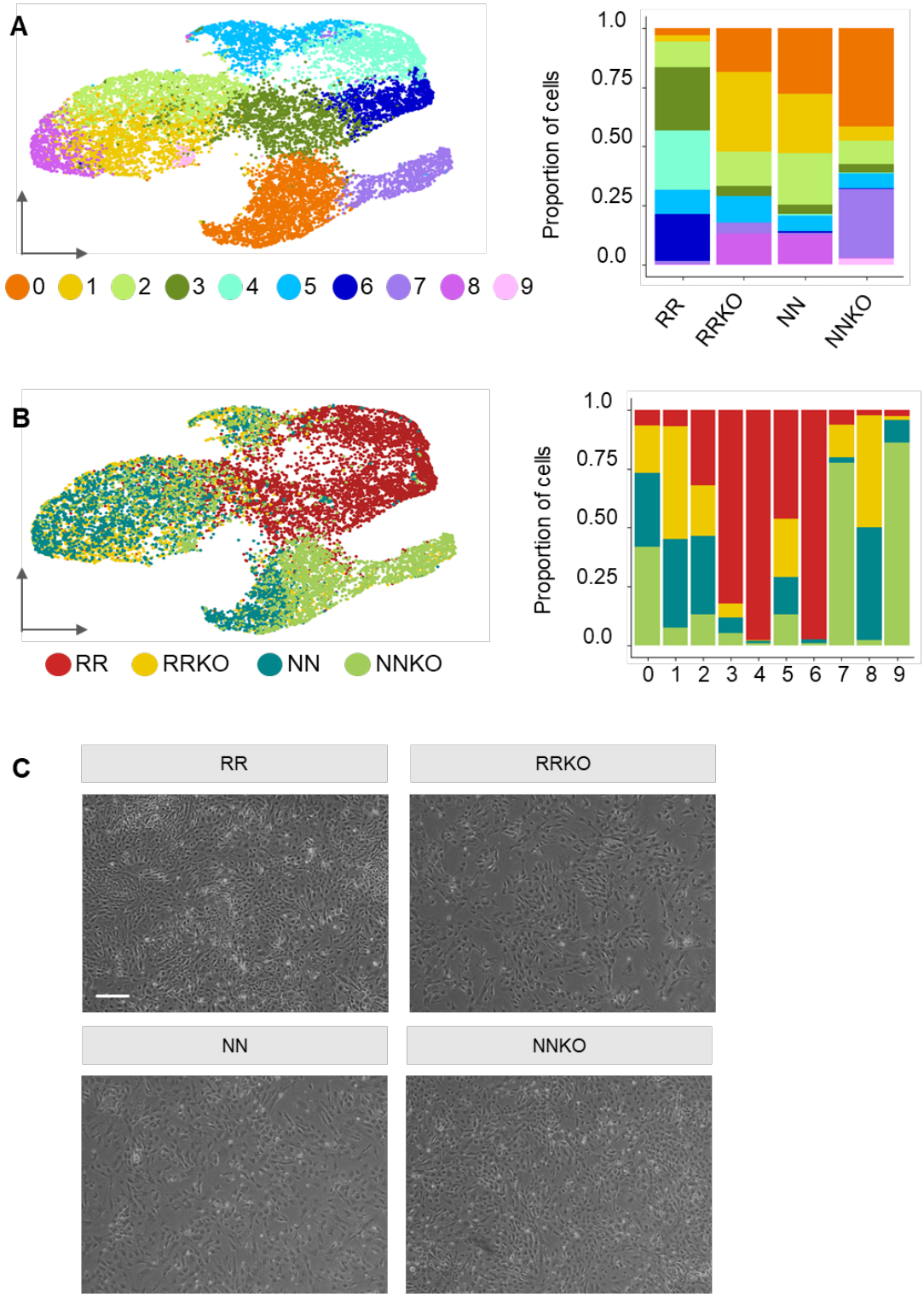

Figure 2. The transcriptional profile of Risk VSMCs diverges from the other genotypes. (A) UMAP of iPSC-derived VSMC colored by clustering (left) and barchart showing cluster composition of genotypes (right). (B) UMAP of iPSC-derived VSMC colored by genotype (left) and barchart showing genotype composition of clusters (right). (C) Images of cells used, captured at day 17 of differentiation. White bar shows 50µm.

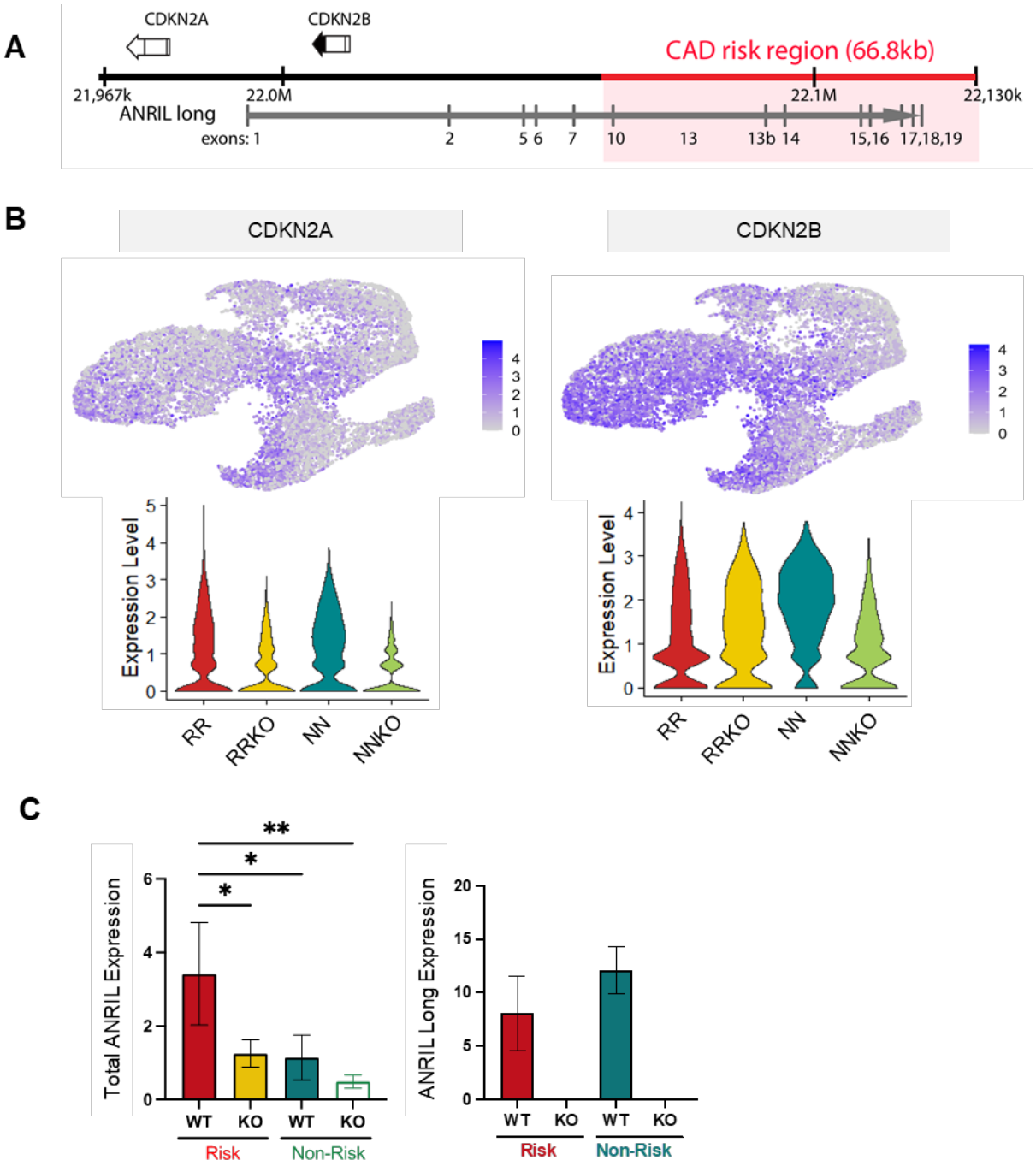

Figure 3. Expression of neighboring genes at the 9p21.3 locus. (A) Diagram of 9p21.3 CAD risk locus with nearest protein-coding genes CDKN2A/B and overlapping lncRNA ANRIL marked. (B) UMAPs of iPSC-derived VSMC single cell data showing the expression CDKN2A and CDKN2B, violin plots show quantification 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 < 0.05, **p value < 0.01, calculated with an unpaired t-test)

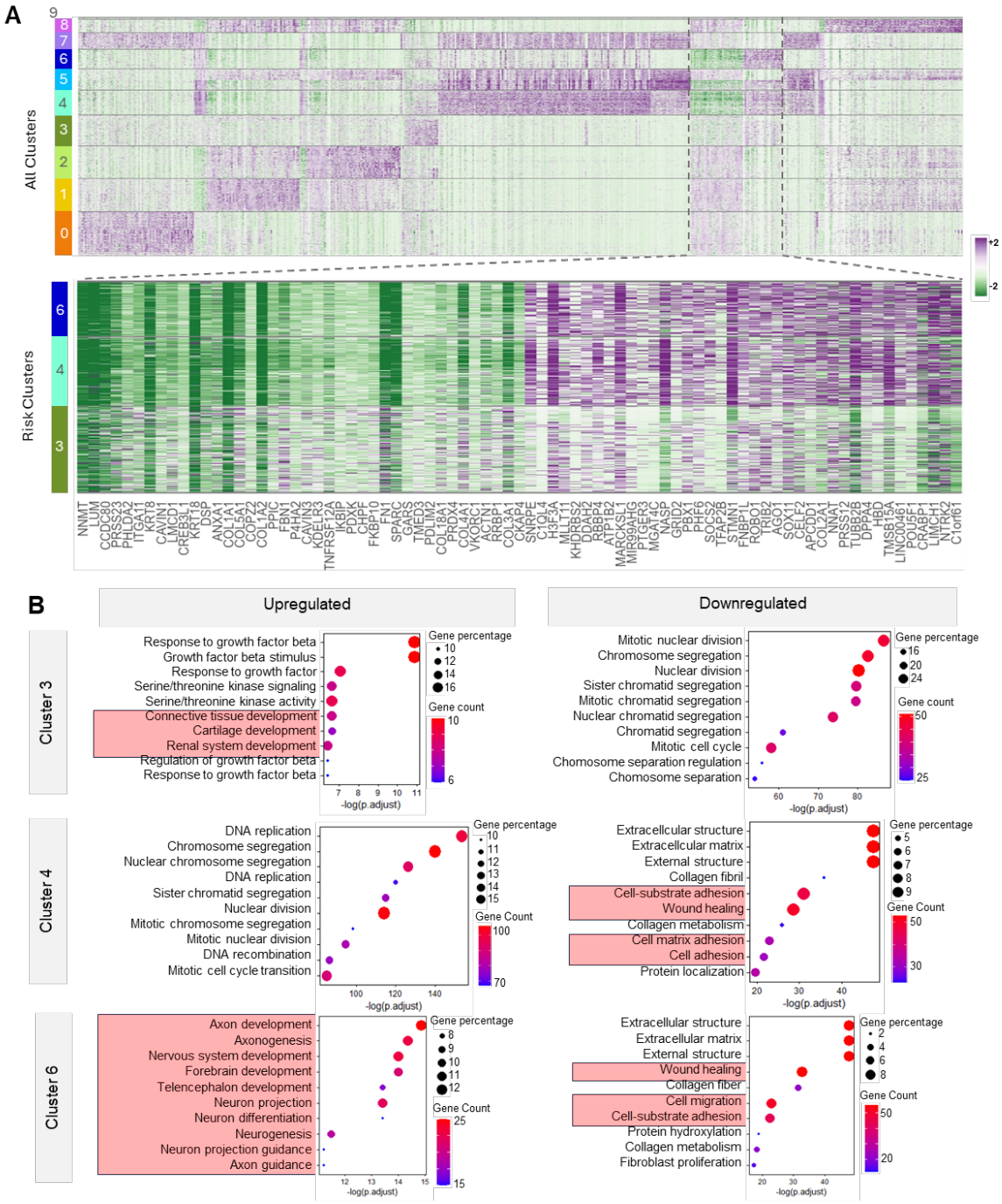

Figure 4. Differential gene expression analysis and Gene Ontology Enrichment reveal unique patterns of expression in RR VSMCs . (A) Top heatmap shows the top 10% genes by corrected p-value driving each cluster (differential expression analysis conducted using Seurat) as they are expressed across all 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 (generated with clusterProfiler). On the left terms associated with upregulated genes, on the right downregulated terms are shown.

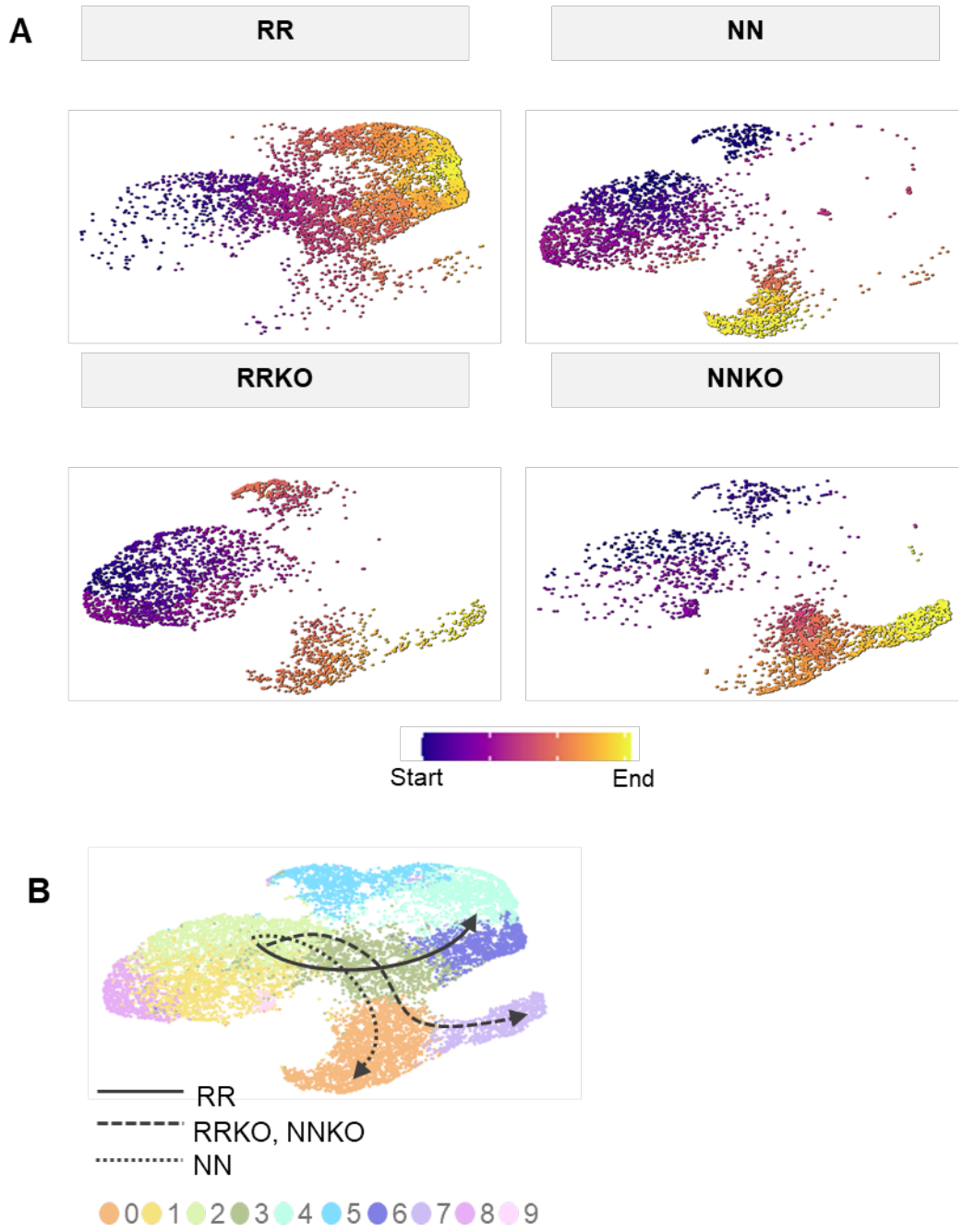

Figure 5. 9p21.3 RR VSMCs show a divergent state trajectory. (A) Pseudotime analysis (conducted with Monocle3) for each genotype. (B) Reconstruction of each genotype trajectory superimposed on a UMAP colored by cluster.

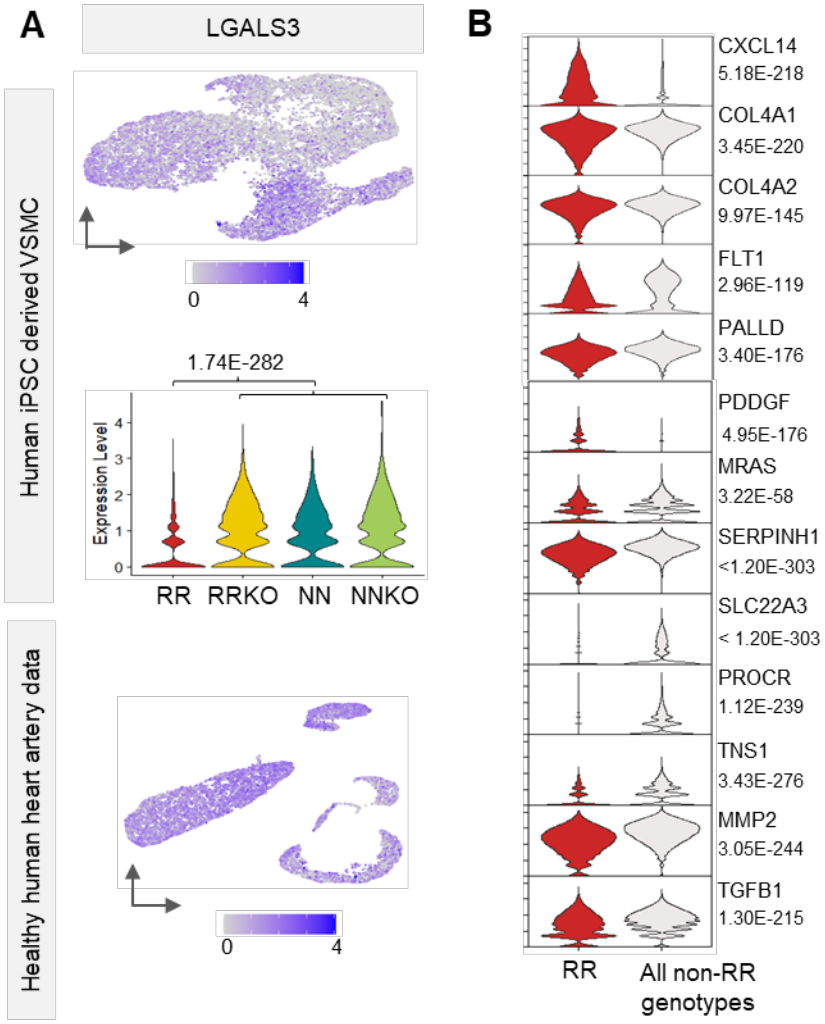

Figure 6. Effect of the 9p21.3 risk haplotype on VSMC transitional state and other CAD loci. (A) UMAP showing expression of LGALS3 in iPSC-VSMCs (top), violin plot showing the expression of LGALS3 across genotypes with p-value calculated via non-parametric Wilcoxon rank sum test (middle), and UMAP of 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.

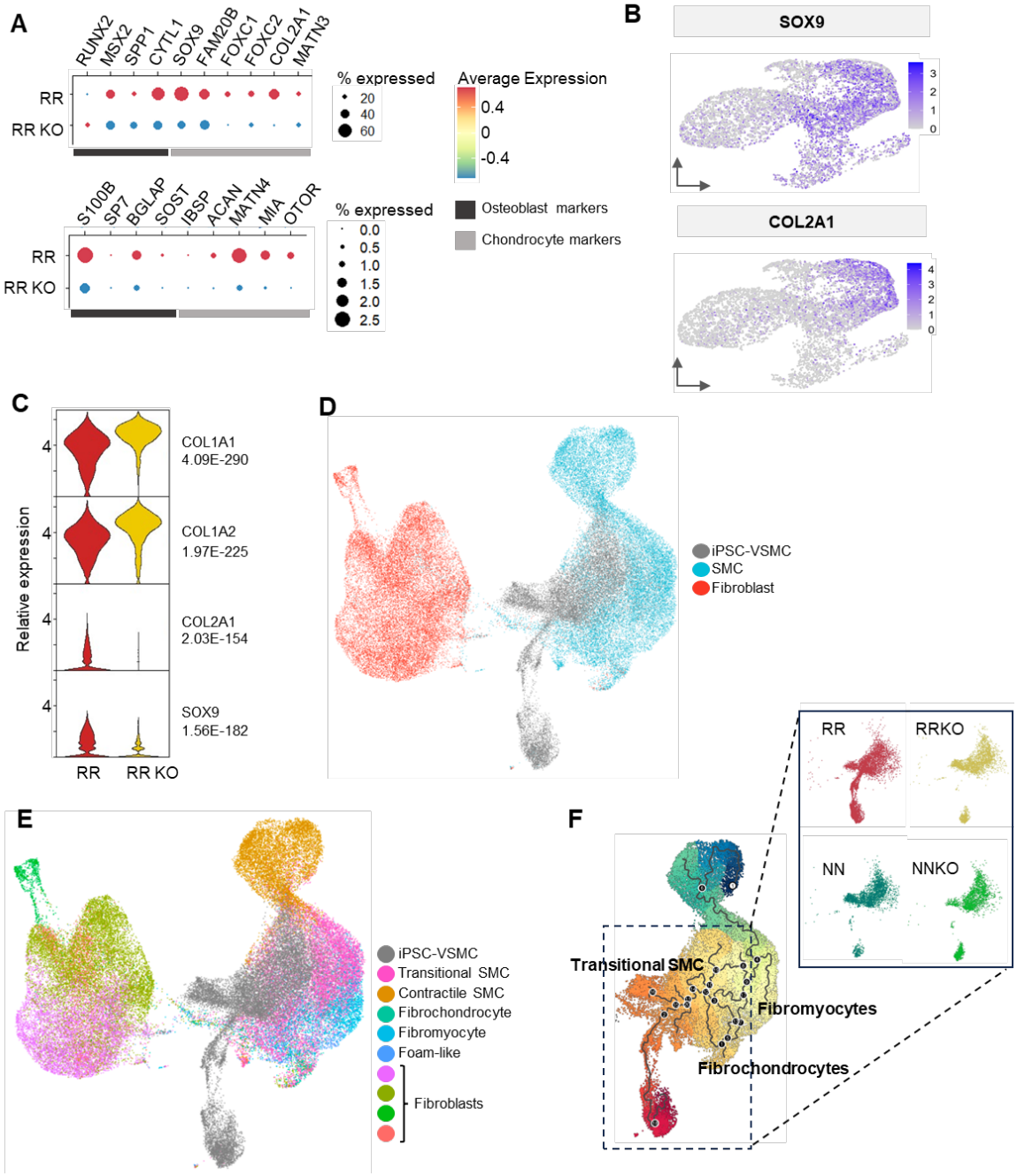

Figure 7. Risk VSMC acquire a osteochondrogenic-like state. (A) Bubble plots of osteoblast and chondrocyte markers in RR and RR KO. B) UMAPs showing expression of SOX9 and COL2A1 in RR and RRKO VSMCs. (B) Violin plots showing expression of SOX9, COL2A1, COL1A1, and COL1A2 in RR VSMC and isogenic RR KO,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

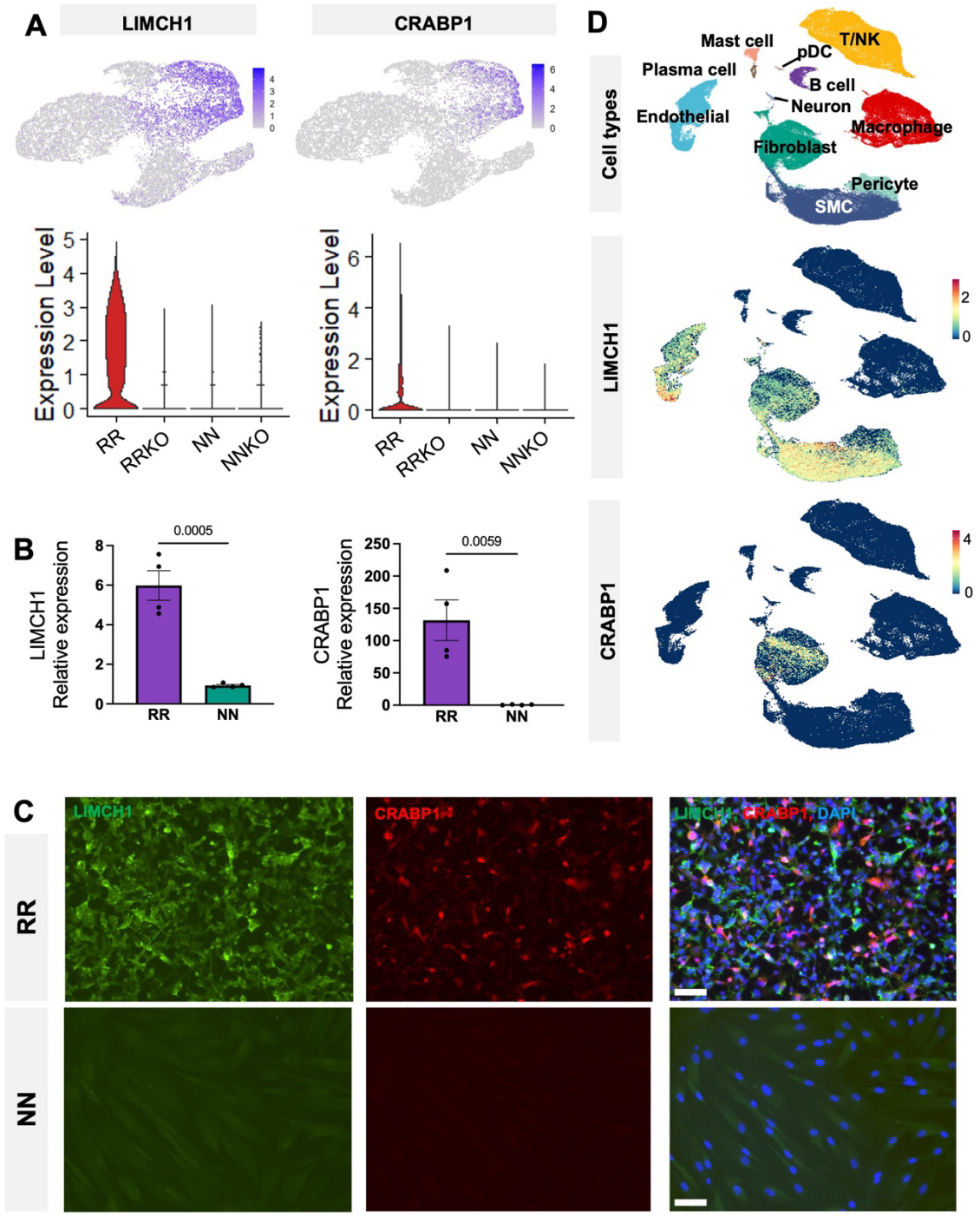

Figure 8. Risk VSMC exhibit signature genes that define the 9p21.3 risk phenotype. (A) UMAPs showing expression of LIMCH1 and CRABP1 in iPSC-VSMC (top), violin plots showing expression across genotypes (bottom). (B) qPCR of LIMCH1 and CRABP1 enrichment in 4 additional donors (2 RR and 2 NN) from 2 independent differentiation experiments. P values calculated with unpaired t-test. (C) Immunocytochemistry staining for LIMCH1 and CRABP1 in RR and NN. White bar indicates 50 μm. (D) Metaplaque atlas of human artery samples labeled with cell types (top), LIMCH1 and CRABP1 expression in Metaplaque atlas.

Supplementary Figures

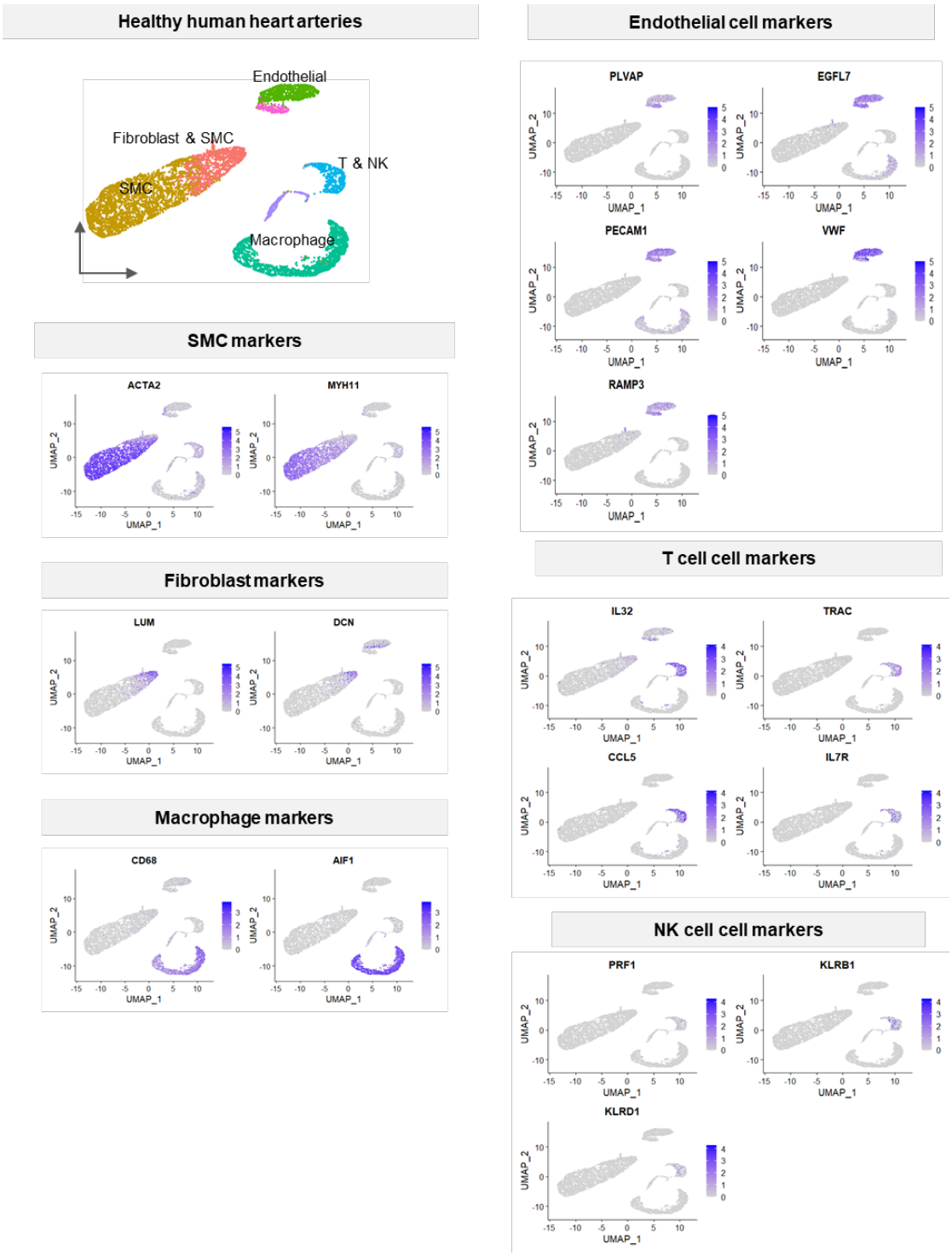

Figure S1. UMAPs showing expression of genetic markers used to identify cell types in published datasets. Top left UMAP shows cell type annotations for healthy human artery data (Hu et al, 2021), remaining UMAPs show expression of cell type-specific markers.

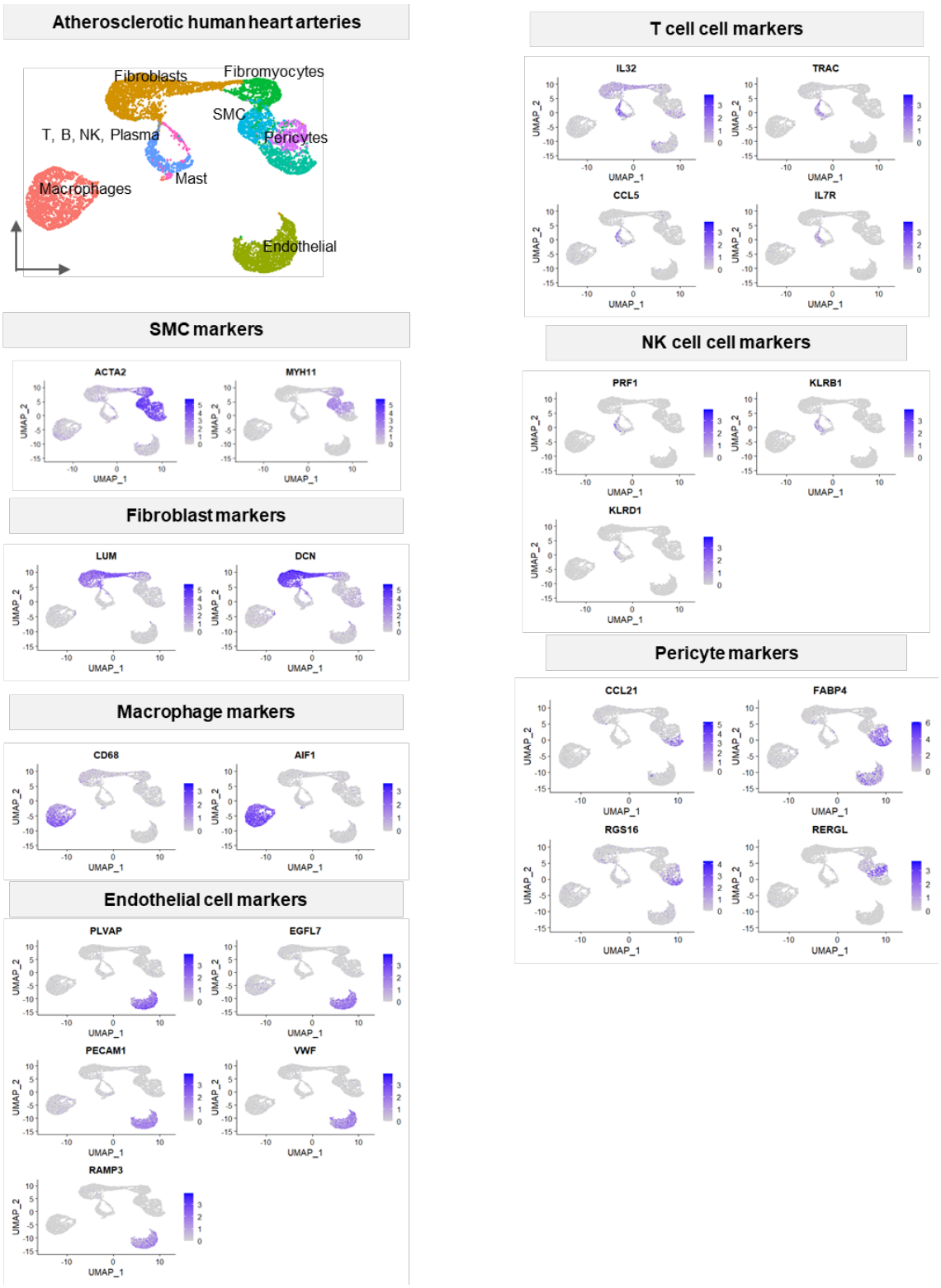

Figure S1 (cont.). UMAPs showing expression of genetic markers used to identify cell types in published datasets. Top left UMAP shows cell type annotations for atherosclerotic human artery data (Wirka et al. 2019), remaining UMAPs show expression of cell type-specific markers.

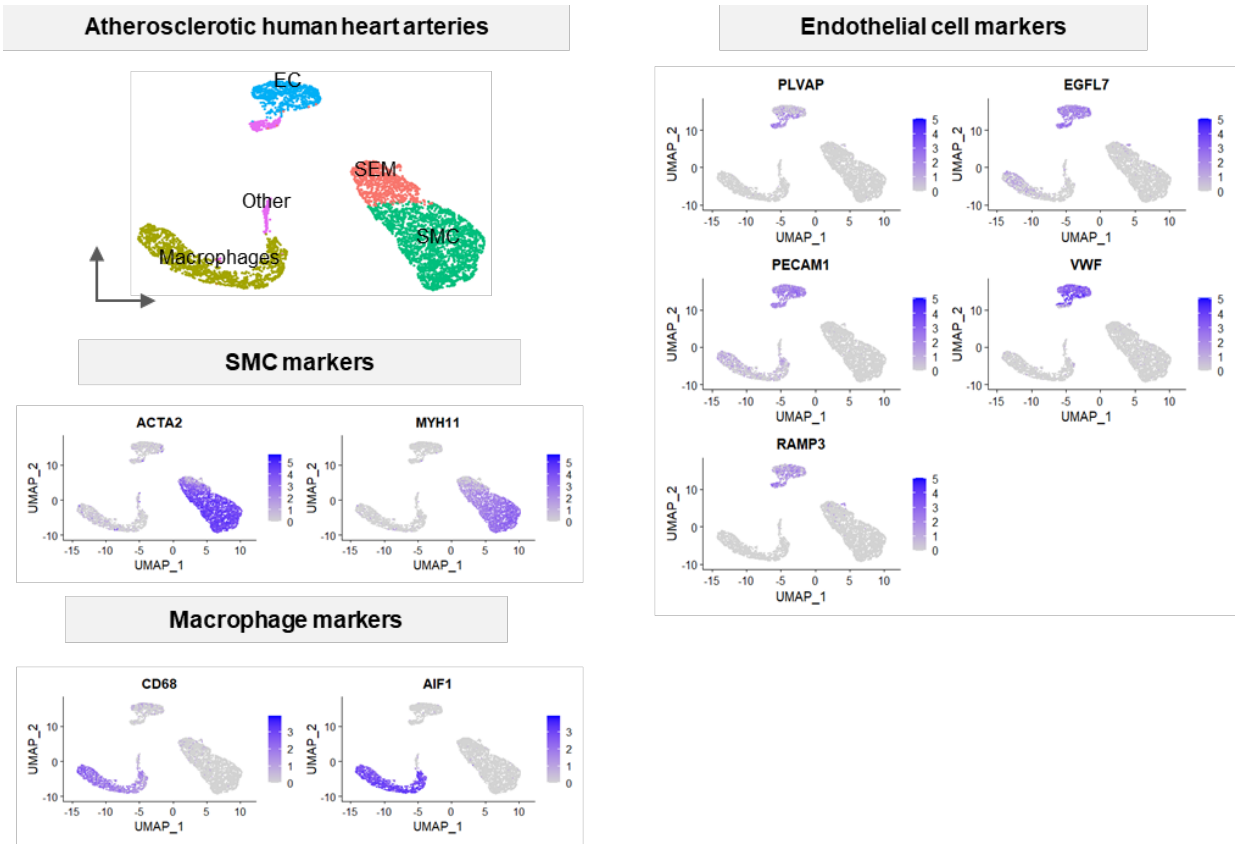

Figure S1 (cont.). UMAPs showing expression of genetic markers used to identify cell types in published datasets. Top left UMAP shows cell type annotations for additional atherosclerotic human artery data (Pan et al. 2020), remaining UMAPs show expression of cell type-specific markers.

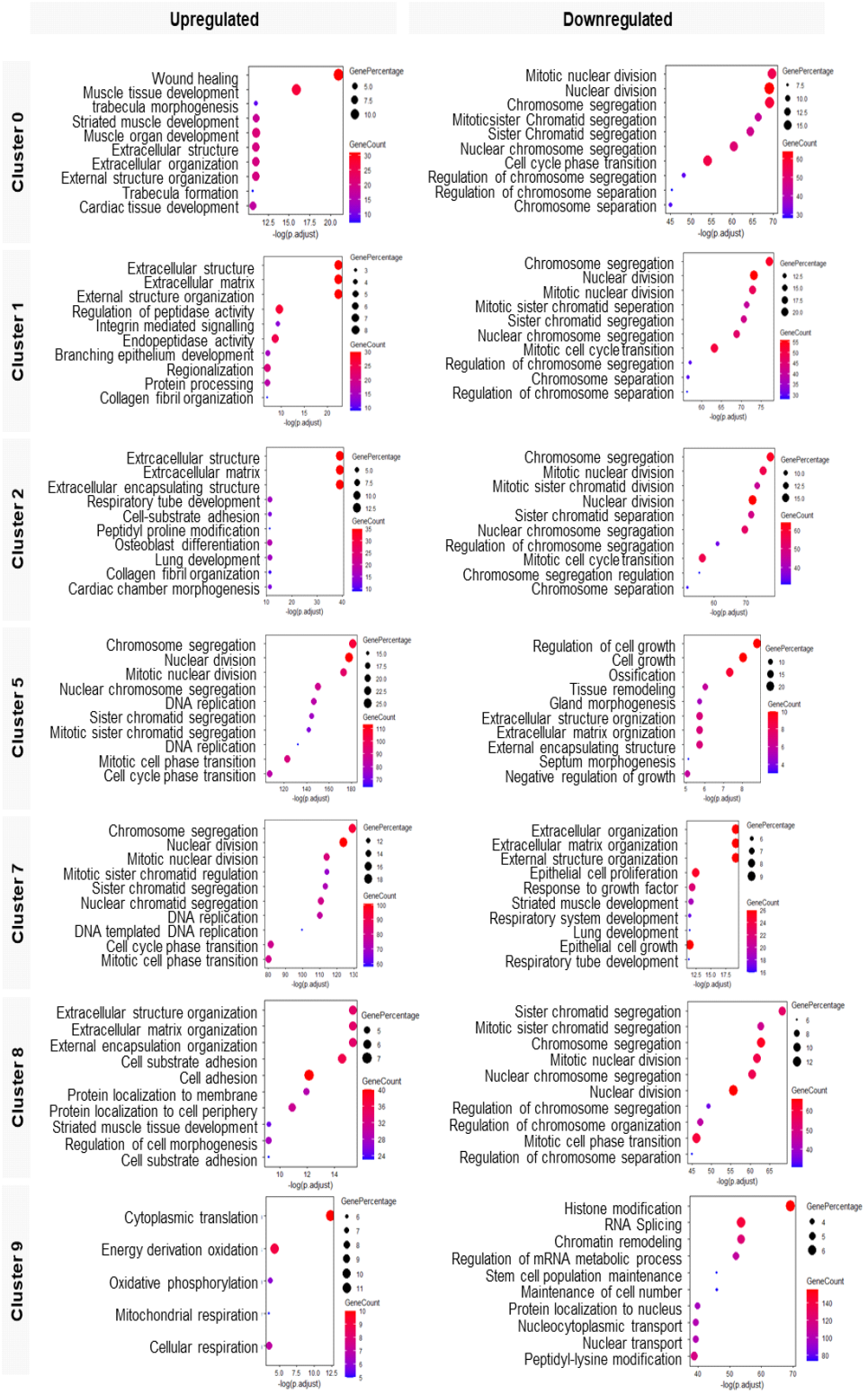

Figure S2. Gene Ontology enrichment analysis for non-RR clusters. Left hand column shows upregulated terms, right shows downregulated. Gene ontology analysis conducted with clusterProfiler

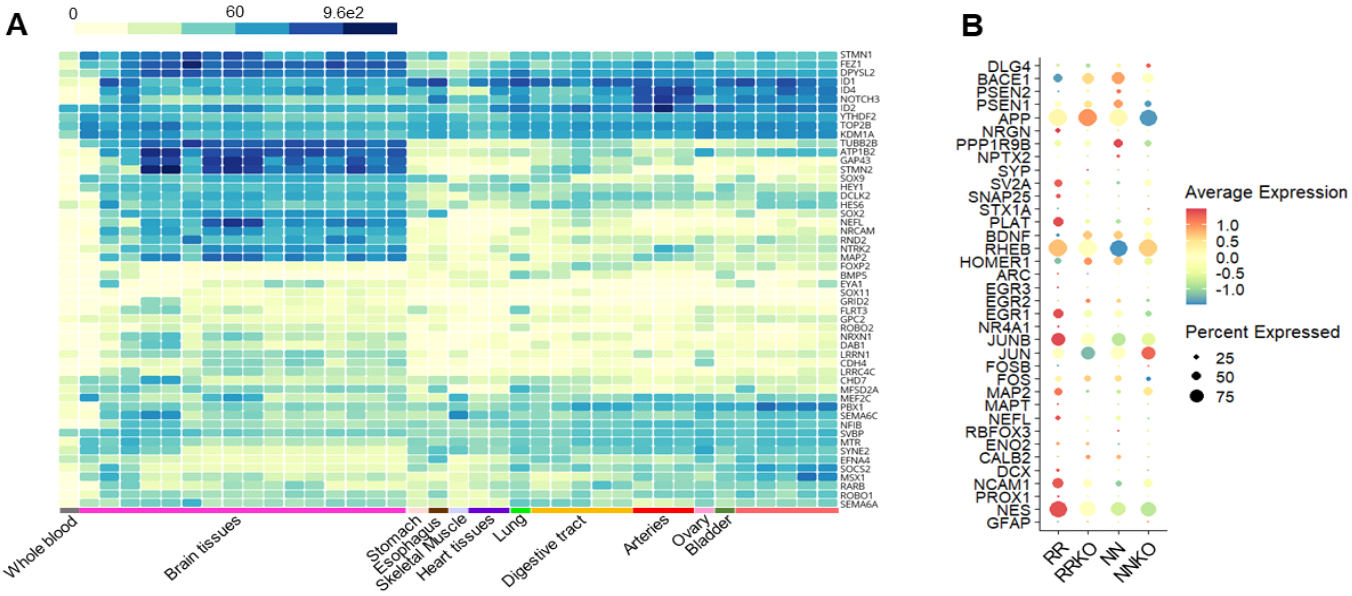

Figure S3. Analysis for neuronal differentiation of RR VSMC. (A) Heatmap of expression of genes associated with neuronal gene ontology terms across brain and VSMC-rich human tissues (Generated with the Genotype-Tissue Expression (GTEx) Portal) (Lonsdale et al. 2013). (B) Bubble plot showing expression of known neuronal markers across genotypes.

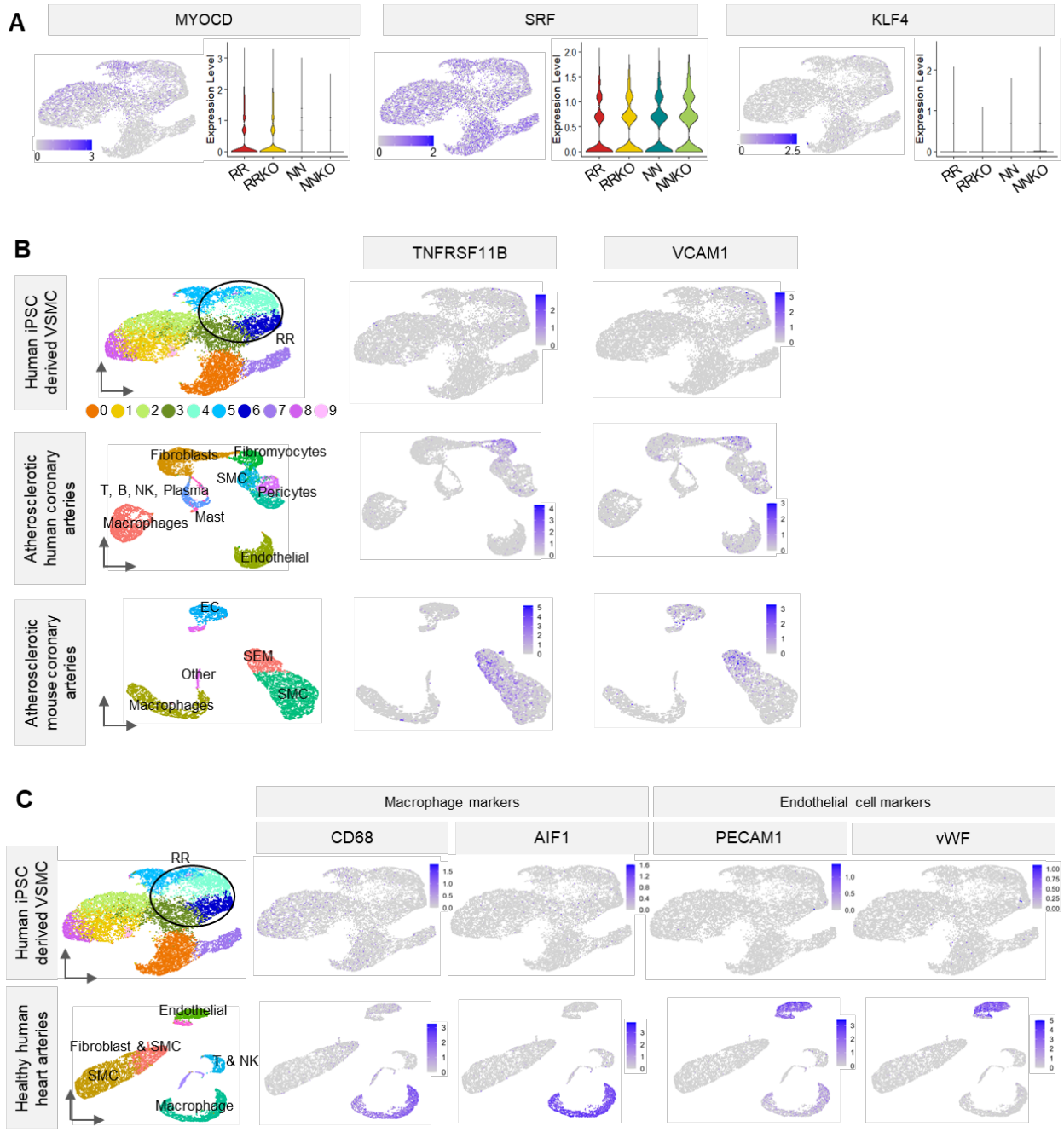

Figure S4. Analysis of known modulators of VSMCs phenotypic switch and known CAD-related VSMCs states. (A) UMAPS and violin plots showing expression for three classic VSMC phenotype modulators, MYOCD, SRF, and KLF4. (B) iPSC derived VSMC (top row) compared to atherosclerotic coronary arteries (Wirka 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.

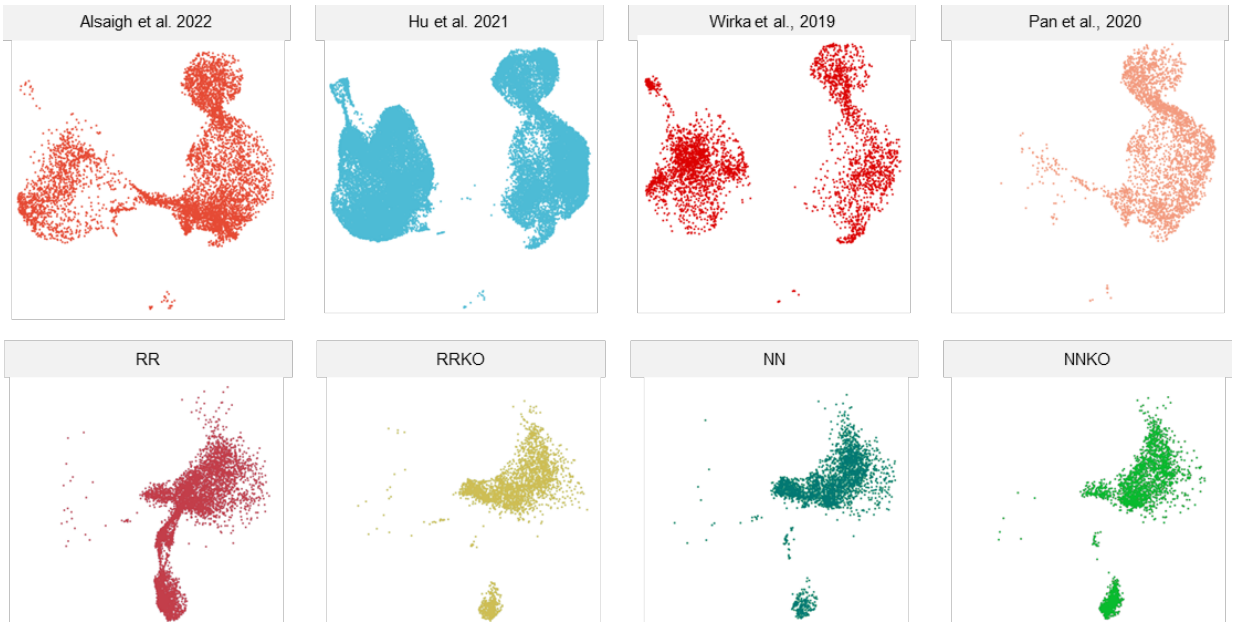

Figure S5. Breakdown of harmonized integration of iPSC-VSMC and Metaplaque atlas. SMC/Fibroblasts by contributing published dataset (top row) and by iPSC-derived VSMC genotype (bottom row).

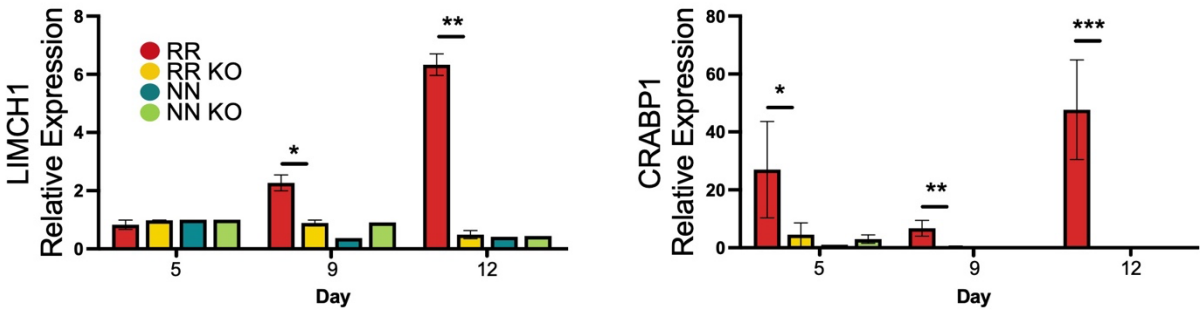

Figure S6. Signature RR genes expression overtime. Expression of signature genes over time, during iPSC differentiation to VSMCs in all genotypes. Statistical significance calculated with a one-way ANOVA with Bonferroni correction at each time point, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.