Positional transcriptomics shed light on site-specific pathologies of the aorta

Milagros Romay
Department of Cell and Developmental Biology, Feinberg School of Medicine, Northwestern University, Feinberg School of Medicine

Feiyang Ma
University of California, Los Angeles https://orcid.org/0000-0002-6260-0787

Gloria Hernandez
University of California, Los Angeles / Northwestern University https://orcid.org/0000-0003-2882-811X

Marie Vandestienne
Medical Intensive Care Unit, Hopital Saint-Antoine, Assistance Publique-Hopitaux de Paris, Sorbonne Universite

Todd Kimball
Molecular, Cellular and Integrative Physiology Interdepartmental Graduate Program, University of California, Los Angeles

Michele Silvestro
Division of Endovascular surgery, Department of Surgery, New York University Langone Health

Tarik Hadi
Division of Endovascular surgery, Department of Surgery, New York University Langone Health

Andrew Reyes
Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles,

Margaret Ramirez
Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles,

Nathalie Bardin
Aix-Marseille University, INSERM

Marcel Blot-Chabaud
Aix-Marseille University, INSERM

Matteo Pellegrini
Department of Molecular, Cell and Developmental Biology, and Institute for Genomics and Proteomics, University of California, Los Angeles https://orcid.org/0000-0001-9355-9564

Aurelie Leroyer
Aix-Marseille University, INSERM

Hafid Ait-Oufella
Inserm

Bhama Ramkhelawon

Division of Endovascular surgery, Department of Surgery, New York University Langone Health

M. Luisa Iruela-Arispe (arispe@northwestern.edu)

Northwestern University Feinberg School of Medicine https://orcid.org/0000-0002-3050-4168

Article

Keywords: vascular smooth muscle cells, aneurysms, moyamoya, berry aneurysm, single-cell RNAseq

Posted Date: October 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-951818/v1

Abstract

Pathologies of large vessels, such as atherosclerosis and aneurysms tend to emerge at specific sites. The consistency in their distribution suggests that a combination of unique local stressors, including both physical forces and specific gene expression profiles are confounding factors in disease etiology. Here we used single-cell RNA sequencing to identify signatures in smooth muscle cells with site-restricted predominance to uncover potentially relevant gene products. We showed that a small cohort of transcripts (5.5%) display preferential expression at specific sites of the vascular tree and in accordance with their embryological origin. Importantly, in silico studies revealed that several of these genes mapped to linkage studies for which no specific disease-causing candidates had been previously found. One of these candidates was Mcam/CD146 that mapped to the familial aortic aneurysm 1 (FAA1) locus identified by linkage analysis two decades ago. We showed that Mcam was significantly reduced in the AngII / hypercholesterolemic model of aortic aneurysm and further demonstrated that absence of the gene in mice resulted in larger lesions and accelerated death due to dissection. Our study highlighted site-specific alterations in gene expression profiles of smooth muscle cells that yield important insight in understanding site-specific vascular pathologies.

Introduction

A growing body of evidence supports the notion that embryological origin can impact the gene expression profile of daughter cells through transcription factor selection and unique epigenetic modifications. In relation to the vascular tree, detailed cell-lineage tracing experiments uncovered complex and diverse embryological ancestry[1]. In particular, vascular smooth muscle cells (vSMC) that populate the aorta descend from neural crest, somites, second heart field, and lateral mesoderm[2–4]. Interestingly, vSMC derived from these precursors remain segregated in the adult tissue as revealed by lineage tracing studies[5]. In addition to ancestry, vSMCs like most cells, also possess positional identity[6] that is retained in the adult, as per expression of Hox family members[7,8]. Interestingly changing topographical expression of Hox leads to alterations in gene expression and vascular remodeling[9].

The link between embryological origin and location is thought to contribute to disease susceptibility particularly in the case of site-specific pathologies[10–13]. Furthermore, embryological origin can affect responses to the same stimuli even in adjacent cells. For example, in mouse models of Loeys-Dietz syndrome responses to TGF-b result in either activation or repression of Smad2/3 signaling depending on whether vSMCs are derived from secondary heart field or neural crest[14]. These findings underscore the notion that vSMCs derived from distinct progenitors appear to retain memory of their embryological ancestry with impact to disease susceptibility and responses to therapy. Thus, a more granular understanding of the transcriptional profile of vSMCs that examines location and ancestry is of high relevance. The advent of next-generation sequencing, and more specifically, single cell RNA-sequencing (scRNA-seq) offers the unprecedented opportunity to clarify coordinated transcriptional diversity of

vascular beds, vascular regions, vessel types and their relationship to unique physiological needs and/or emergence of pathology.

Amongst the pathologies that affect the aorta, aortic aneurysms show remarkable site-predilection, each type presenting with distinct etiology and history[15–17]. Of note, the prevalence of abdominal aortic aneurysms (AAA) is higher than that of thoracic aneurysms (TAA)[18] although there is an overlap in the risk-factors that predispose to both types of aneurysms[19]. This implies that intrinsic features associated with the topographical location of the vessel combined with risk factors modulate cellular response to vascular injury.

Here we use scRNA-seq to profile the transcriptome of adult vSMCs mice in three distinct locations of the arterial tree: carotids, aortic arch (ascending-thoracic) and descending (thoraco-abdominal) aorta. Using this approach, we identified a small, but important fraction of transcripts that displayed region-preferential patterns of expression. These regionally expressed genes defined site-specific signatures of adult vSMC subpopulations, in relation to their anatomic location. Within these small signatures, we found both known and novel disease-associated genes, linking regional skewed expression to susceptibility of disease emergence. Furthermore, delving into the network of mRNA signatures present in vSMCs during AAA, we identified CD146/*Mcam*, a gene with preferential expression on the ventral aspect of the thoraco-abdominal aorta. Notably, the *MCAM* gene resides in a locus previously associated with aneurysm susceptibility[20], but that had not been directly linked to development of the pathology. Importantly experimental aneurysm induction in the CD146/*Mcam* null mouse revealed that absence of this gene increases disease susceptibility and lethality. Overall, our data indicates that discrete expression patterns in regions of the aorta could guide identification of causative disease variants particularly from the multiple loci identified by genome-wide association studies.

Results

Transcriptional characterization of vascular smooth muscle cells

To identify regional-specific transcriptional identities in vSMCs populations, we performed single cell RNA sequencing (scRNA-seq) in 3 distinct sites: carotid arteries, aortic arch (ascending thoracic) and descending (thoraco-abdominal) aorta (Figure 1a and 1b). The rationale for selecting these sites was the combination of embryonic origin, hemodynamics, and disease emergence. The aortic root was excluded from the evaluation. The collected aortic arch region extended until the left subclavian branch, in concordance with the limits of neural crest origin[5]. In total, we sequenced between 3,732 to 4,632 cells per site with 72,685 to 91,597 average reads per cell (Supplementary Figure 1,2; Supplemental Table 2). Quality assessment of each of the three libraries was performed and those cells with low RNA counts and high levels of mitochondrial RNA were removed from the dataset prior to downstream analysis (Supplementary Figure 1,2).

To identify common and regional signatures, the samples were merged and analyzed using Seurat (Figure 1c). This analysis identified 11 distinct cell clusters of varying vessel composition (Figure 1d). To determine the cellular constituency present in the 11 cell clusters, expression of classical cell-type markers was applied identifying six distinct cell populations (Figure 1e, 1f, Supplemental Figure 1, Supplemental Table 3). Vascular smooth muscle cells (vSMCs) represented the predominant cell type in all samples with the carotid library showed the greatest degree of diversity as 32% of the carotid-identified cells were non-vSMC (Supplemental Table 3).

Populations of vSMCs were computationally extracted from the other cell types for in-depth analysis and comprised 1,740 to 3,336 cells per anatomic location (Supplementary Figure 2, Supplemental Table 4). Visualization of the data using t-Distributed Stochastic Neighbor Embedding (tSNE) highlighted the global uniqueness of each population of vSMC per anatomical region, as the data segregated into three distinct mostly contiguous but not overlapping groups that matched the referred anatomical location (Figure 2a, b, Supplementary Excel File 1).

From this cell population 18,940 transcripts were quantified and assessed for site-enriched expression. In this analysis, 1045 transcripts (5.5%) showed unique anatomic location-enriched expression (Figure 2c, Supplementary Excel File 1). Importantly, while vSMCs populations contained these regional-enriched expression patterns, their core identity as per expression of classical vSMC markers– *Myh11*, *Acta2*, and *Tagln*, were unchanged (Supplemental Figure 3).

Molecular drivers of the transcript regional specification were further evaluated by plotting their relative log fold change against the frequency of transcript-producing cells (Figure 2d and Supplementary Excel File 2). The predominant pattern was defined by relatively equal contributions of increased transcript production per cell as well as increased number of transcripts producing cells in a specific region (Figure 2d). For a smaller portion of transcripts, site-enrichment was driven by increased transcript counts per positive vSMC (fold change). This is exemplified by carotid enriched *Btg2,* which shows 1.5 - log fold increase in transcript quantity per cell in carotid vSMCs, relative to aortic arch and descending aorta (Figure 2d and Supplemental Figure 3). Some site-specific transcripts were characterized by differential increases in the frequency of cells expressing the transcript (cell enrichment). This expression pattern is exemplified by the HOX genes *Hoxb3os* (carotid) and *Hoxa7* (descending aorta) which show between 1.7 to 2.5-fold increases in transcript expressing cells, respectively (Figure 2d and Supplemental Figure 3). Importantly, 38 transcripts (3.6%); including carotid-enriched *Cebpb*, aortic arch-enriched *2210407C18Rik,* descending aorta-enriched *Rgs5* show substantial contributions of both features to their site-enrichment (Figure 2d and Supplemental Excel File 2).

Further analysis of the smooth muscle cellular population identified seven cell subclusters recognized by differences in their transcriptional profiles (Figure 2 e,f Supplemental Table 4 and Supplementary Excel File 3). Detailed analysis showed that while the vSMC subcluster derived transcriptional signatures were relatively unique, certain clusters showed high overlap with one of the three the anatomic location defined signatures (Figure 2g). Visualization of the contribution of vSMCs from each of the three anatomic

locations to the vSMC subcluster, revealed that subclusters with high gene signature overlap to a given anatomic signature contained substantial cell contributions from that anatomic location (Figure 2h). For example, subcluster 0, predominantly composed of cells isolated from the descending aorta, displayed a gene signature that is 80% identical with the descending aorta anatomic signature (Figure 2g and 2h). In contrast, Subcluster 6, predominantly composed of aortic arch cells (93.5% arch-derived vSMCs) shows minimal gene signature overlap (15%) with the aortic arch anatomic signature.

Expression of the HoxA cluster in combination with hierarchical clustering of the gene signature was utilized to further locate the topological distribution of cells (Figure 2i-k). The data uncovered striking links with developmental origin and topographical location despite maintaining strong core smooth muscle cell identity as highlighted by the high expression of common genes (Supplementary Figure 3). We next proceeded to evaluate unique transcriptional signatures associated with three specific anatomic sites: carotids, ascending aorta and descending aorta and explored the possible association of unique signature with emergence of site-specific vascular pathology.

Unique Carotid vSMCs Signature (caSMCs)

The carotids are a major branch point from the aortic arch responsible for the supply of blood to the brain (Figure 3a) and their tunica media is populated by vSMCs derived from the neural crest[1]. A total of 632 transcripts (3.3% from total transcripts) constituted the molecular signature of the carotid artery (caSMCs) distinguishing this population from their counterpart vSMCs in the aortic arch and descending aorta (Figure 2c).

caSMCs were the predominant population associated with two subclusters: subcluster 2, (87.81% carotid-derived) and subcluster 5 (100% carotid-derived) (Figure 3b,c and Supplementary Table 4). Differential gene expression analysis of these two subclusters identified 459 genes (subcluster 2), and 291 genes (subcluster 5) whose expression pattern separated these two clusters from other SMC (Supplemental Excel File 3). The overlap between carotid artery anatomic location signature and the gene signatures of the two caSMC dominated subclusters ranged from 22.3% (subcluster 5, 65 of the 291) to 94.1% (subcluster 2, 432 of the 459 genes) (Figure 2g, Supplementary Excel File 1 and 3).

Gene ontology analysis of the 632 anatomic location signature genes showed a significant enrichment for proteins involved with growth factor signaling, vascular development and AP-1 response (Figure 3d, Supplementary Excel File 4). From these pathways candidate gene validation using immunohistochemistry was performed for the genes *Klf4, Gucy1a3* and Gelsolin (*Gsn*).

Klf4, a member of the Kruppel-like factor transcription factor family, was represented in 6 of the 10 top ontological categories (response to growth factor, regulation of phosphate metabolic process, vascular development, AP-1 pathway, cellular response to organic cyclic compound and response to wounding) (Figure 3d and 3e). Immunofluorescence staining of KLF4 showed significantly increased expression in

all layers of carotid vascular smooth muscle relative to aortic arch and descending aorta (Figure 3f and 3g). Gelsolin (*Gsn*), a cytoplasmic, actin-regulating protein was identified as a significantly enriched carotid SMC gene through its membership in two top ontology categories - AP-1 response and response to wounding (Figure 3d and 3h). Immunofluorescence staining of GSN showed significantly increased cytosolic expression in the carotid artery with substantial increases in the layer of smooth muscle closest to the vessel lumen when compared to aortic arch and descending aorta (Figure 3i and 3j). *Gucy1a3,* the major nitric oxide receptor which was recently associated as causative for the carotid-restricted vasculopathy, Moyamoya[21], was also identified as a significantly enriched carotid SMC gene (Figure 3k). Immunofluorescence staining of GUCY1A3 showed high expression in carotid vascular smooth muscle with impressive levels in the smooth muscle layer in closest proximity to the endothelium (Figure 3l and 3m).

Unique Aortic Arch vSMCs Signature (aaSMCs)

The aortic arch is characterized by unique physical forces and flow patterns (Figure 4a) that comprise this region emerge from at least two distinct developmental origins: neural crest and secondary heart field[5]. Site-enriched expression of 226 genes (1.19 % of all transcripts) defined the anatomic location signature of aortic arch derived SMCs (aaSMCs) from their counterpart in the carotid artery and descending aorta (Figure 2c).

aaSMCs predominantly contribute to the formation of subclusters 1 and 6 (Figure 4b, 4c and Supplementary Table 4). Subcluster 1, the largest aaSMC-dominant subcluster, was composed of 91.8 % aaSMCs and was defined by the expression of 105 genes while subcluster 6 was composed of 93.5% aaSMCs and was defined by the expression of 473 genes (Figure 4c, Supplemental Excel File 3, Supplementary Table 4). The overlap between aortic arch location signature and the gene signatures of the two aaSMC dominated subclusters ranged from 15% (subcluster 6, 72 out 473 genes) to 70.4% (subcluster 1, 74 out of 105 genes) due to multiple factors including gene signature, subcluster size and cellular composition of subclusters (Figure 2g, Supplementary Excel File 1 and 3).

Gene ontology analysis of the aortic arch anatomic location signature showed significant enrichment for transcripts involved extracellular matrix organization, metabolism and autophagy (Figure 4d, Supplementary Excel File 4). From these identified signature genes, validation using immunohistochemistry was performed for the selected candidates: Aggrecan (*Acan*), extracellular superoxide dismutase (*Sod3*) and Thrombomodulin (*Thbd*). Aggrecan, a large proteoglycan, was identified through its membership in 3 ontological categories: extracellular matrix organization, collagen metabolic processes and pyruvate metabolic processes (Figure 4d and 4e). Immunofluorescence staining of ACAN displayed significantly increased expression in all layers of aortic arch vascular smooth muscle, characterized by intense staining surrounding each vSMC cell in the arch relative to the carotid arteries and descending aorta (Figure 4f and 4g). *Sod3*, was identified as part of the antigen presentation

ontology category, (Figure 4d and 4h). Immunofluorescence staining of SOD3 showed significantly increased expression of the protein in all layers of aortic arch vascular smooth muscle relative to the carotid arteries and descending aorta (Figure 4i and 4j). Thrombomodulin (*Thbd*) a critical component the coagulation cascade was identified as significantly enriched in the aortic arch, through membership in the platelet degranulation ontology category (Figure 4d and 4k). Immunofluorescence staining of THBD showed significantly expression in all layers of aortic arch vascular smooth muscle relative to carotid artery and descending aorta (Figure 4l and 4m).

Characterization of Descending Aorta vSMCs (daSMCs)

SMCs of the descending aorta derive from the somites and the lateral mesoderm [1](Figure 5a). The scRNA-seq data uncovered 187 (1% of total) genes that provided the molecular signature for daSMCs (Figure 2c). daSMCs represented the predominant SMC population of subclusters 0, 3 and 4 (Figure 5b, 5c). The majority of daSMCs resided in subcluster 0 (92.44% daSMC), the largest subcluster characterized by the expression profile of 140 genes (Figure 5c, Supplementary Table 4 and Supplementary Excel File 3). Subcluster 3 and 4 are mixed aortic vSMCs clusters, with subcluster 3 (59.5% descending-derived, 37.78% arch-derived) defined by the expression of 188 genes; while subcluster 4 (54.52% descending-derived, 42.46% arch-derived) was characterized by expression of 24 genes (Figure 5c, Supplementary Table 4 and Excel File 3). The overlap between the descending aorta signature and the gene signatures of the three daSMC dominated subclusters ranged from no overlap (subcluster 4, 0 out 24 genes) to 80% (subcluster 0, 113 out of 140 genes) due to multiple factors including gene signature, subcluster size and cellular composition of subclusters (Figure 2g, Supplementary Excel File 1 and 3)

Gene ontology on the descending aorta anatomic location signature identified significant enrichment for genes involved oxidative phosphorylation, extracellular matrix proteoglycans, and muscle contraction (Figure 5d, Supplementary Excel File 4). Three genes were selected to validate the signature using immunohistochemistry: Aquaporin 1 (*Aqp1)*, Ccdc3, and Perlecan (*Hsgp2*). *Aqp1*, a well characterized water channel, was identified in the gene ontology categories: oxidative phosphorylation, muscle contraction, blood vessel development and actin filament-based processes (Figure 5e). Immunofluorescence localization of AQP1 identified the protein preferentially in smooth muscle cells of descending aorta (Figure 5f and 5g). Surprisingly, AQP1 distribution in the endothelium was also site-enriched with limited expression in the descending aorta relative to the aortic arch (Supplemental Figure 4).

Ccdc3, a putative/predicted secretory factor, known for its anti-inflammatory role in the vasculature and effects on lipid accumulation was the top transcriptionally enriched descending aorta gene (Figure 5h). Immunofluorescent staining for CCDC3 showed predominant expression in the descending aorta location relative to aortic arch and carotid artery (Figure 5i and 5j). Perlecan (*Hsgp2*) a secreted heparan sulfate

proteoglycan was identified as descending aorta signature gene candidate through its membership in two ontology categories: ECM proteoglycans and blood vessel development (Figure 5d and 5k). Immunofluorescence identification of Perlecan (*Hsgp2*) revealed preferential expression in vSMC of the descencing aorta (Figure 5l and 5m).

Identification of CD146/MCAM as a novel site-restricted aneurysm-associated gene

The cohort of genes selectively expressed at specific vascular locations (5.5% of total) offered the opportunity to inquire about their potential contribution to regional-specific pathologies. Thus, we performed an *in-silico* forward genetic screen of archived loci previously associated with mendelian vascular disease through integration of entries from the Human Phenotype Ontology (HPO) database (https://hpo.jax.org/app/)[22] with data from Online Mendelian Inheritance of Man (omim.org)[23]. Following a detailed curation, 16 reported loci representing 6 vascular associated traits were identified (Figure 6a, Supplemental Figure 5 and Supplementary Table 5). From each locus, a list of regional candidate genes was obtained using the UCSC table browser and subsequently overlaid with the site-enriched single cell data to identity novel candidate genes for each genomic region. A total of 13 transcripts predominantly expressed in the specific vascular regions mapped to vascular disease loci (Supplementary Table 5). Given the discovery-driven nature of this approach, we sought to confirm the feasibility of validating putative disease candidate by further validating the regionally restricted expression of a disease-associated, site-enriched candidate, *Tfap2b* (Figure 6b, Supplementary Table 6)[24]. Immunohistochemistry at low magnification revealed a clear predominance of TFAP2B in the ascending portion of the arch leading towards the brachiocephalic artery, after which it becomes significantly reduced (Supplemental Figure 6). In the lower curvature however, expression persists into the region beyond the remnant of the ductus arteriosus (Supplemental Figure 6). At higher magnification, substantial increases in both level and frequency of TFAP2B positive nuclei were predominant to the aortic arch when compared to the descending aorta and carotid arteries (Figure 6c, 6d). These finding strongly supported the notion that site-specific expression might uncover disease associated gene candidates.

Amongst the identified candidate transcripts was the daSMC-enriched gene *Mcam*, which resided in the critical region of FAA1, a locus described for familial aneurysm identified two decades ago using linkage analysis (Figure 6e)[20]. Importantly, as reported in the original paper, this locus, unlike others, affect multiple aortic segments with dilation of the thoracic and abdominal aorta. Transcriptionally, *Mcam* (also known as CD146) was significantly enriched in daSMCs relative to other vascular sites (Figure 6f). Within the tunica media, CD146/MCAM protein was found to decorate the extracellular space of individual vSMCs, particularly concentrated in the ventral side of the descending aorta. The aortic arch and carotid artery showed more limited distribution of the protein with higher concentration in the innermost layers of the vessel (Figure 6g and 6h and Supplementary Figure 7). Upon further analysis of CD146/MCAM expression, we also noted uneven distribution across the circumference of the aortic segment.

Descending aortic segments in the ventral aspect displayed substantial increased in CD146/MCAM protein expression across all layers when compared when compared to vSMC layers on the opposite dorsal orientation of the vessel (Figure 6i).

Given that echocardiographic examination of the FAA1 families indicated the involvement of both the thoracic and abdominal aorta in aneurysms, we performed scRNA-seq profiling of vascular segments from dissecting aortic abdominal aneurysms (AAA) modeled by angiotensin II infusion (Figure 7a, b) to gain further insights into the expression profile of MCAM in aneurysmal SMCs. Using this approach 813 to 2440 cells per group were sequenced with 198,600 mean reads per cell (Supplemental Figure 8). We observed six groups of cells representing distinct cell type populations (Figure 7c, d, e). Expression analysis of CD146/*Mcam* across all identified cell types revealed that a substantial population of SMCs that expressed high levels of CD146/*Mcam* transcripts (Figure 7f).

SMCs positive for CD146/*Mcam* were identified in both control and AAA vessels where these cells expressed similar levels of classical SMC markers, *Acta2* and *Myh11* (Figure 7g). Direct comparison of the expression patterns of these CD146/*Mcam*+ SMCs from AAA and control showed a substantial decrease in the number of CD146/*Mcam* transcripts expressed per cell in the AAA derived SMCs (Figure 7h). Given the observed changes at the transcriptional level, expression of CD146/MCAM was assessed in control and AAA sections using immunofluorescence (Figure 7i). SMCs from the AAA samples showed substantially decreased CD146/MCAM at the protein level when compared to those from control aorta despite both SMC populations expressing substantial amounts of alpha smooth muscle actin (Figure 7i). Gene ontology enrichment for transcripts associated CD146/*Mcam*+ SMCs identified that significant fraction of genes that were upregulated in CD146/*Mcam*+ SMCs contribute to interactions with extracellular matrix, a crucial process that is altered in aneurysms (Figure 7j).

To confirm a causative role for CD146/MCAM in the pathogenesis of AAA, CD146/*Mcam* knockout mice were crossed to hypercholesterolemic ($ApoE^{-/-}$) animals and the resulting progeny were treated with angiotensin II infusion for 28 days (Figure 8a). Mice lacking CD146/MCAM showed substantially decreased survival during the treatment period, with only 30% of CD146/MCAM-null animals reaching the end of the 28-day period, relative to 60% of wild-type littermates (Figure 8b). Measurement of the external diameter of the thoracic aorta at sacrifice showed no significant differences between $ApoE^{-/-}Mcam^{+/+}$ and $ApoE^{-/-}Mcam^{-/-}$groups (Figure 8c and 8d). In contrast, abdominal aorta measurements showed significant increases in width between the two groups, with $ApoE^{-/-}Mcam^{-/-}$ animals showing substantially and statistically significant increased dilation of the abdominal aorta (Figure 8e and 8f). Taken together, these findings support the conclusion that expression of CD146/MCAM in the descending-abdominal aorta is protective against the development of AAA.

Discussion

Heterogeneity in smooth muscle cell populations has been established between distinct vessel types with important physiological outcomes[25] and even within the same vascular wall segment with pathological

consequences[26,27]. It is also acknowledged that SMCs are endowed with significant transcriptional plasticity particularly when exposed to stressors[28–32] Nonetheless, heterogeneity based on distinct regions of the same vessel has only recently started to be explored[33]. Single cell RNA sequencing has emerged as a powerful research tool that provides multi-layered transcriptomic information at single-cell resolution allowing for the deconvolution of heterogenous tissue samples. Here, we sought to characterize the molecular signatures of adult vSMCs associated with carotid arteries, aortic arch and thoraco-abdominal (descending) aorta using scRNA-seq taking into consideration embryological origin of those sites and positional identity. Using this approach, we profiled 18,940 transcripts across 12,305 cells. While the bulk of the transcriptome was significantly similar, 1045 transcripts (5.5%) showed regionalized expression patterns. Most of those transcripts exhibited increases in both expression levels and percentage of cells expressing the given transcript in the regional vSMC population, although a few were skewed in only one of the two patterns. The analysis provided the opportunity to explore whether these site-enriched transcripts bear potential relevance in disease emergence.

A question posed from the onset of this work was the intersection of embryological origin, topological identity and adaptation to physical forces. While distinct in structure, both the carotids and a large portion of the aortic arch share neural crest ancestry, thus the interest in including carotids in the analysis. Despite the common origin, scRNA-seq data, showed clear segregation between vSMCs from the aortic arch and carotids, however, hierarchical clustering demonstrated that carotids were closer to aortic arch than to the descending aorta. Also critical to the interpretation of these data are the unique hemodynamics to each vascular segment which notably impacts transcription. In fact, evaluation of the aortic arch revealed important adaptations to flow dynamics. Specifically, we identified two clearly distinct subclusters associated with the greater and lesser curvatures (subclusters 1 and 6), the later in locations of high turbulent flow. Interestingly, positional information, as per Hox gene expression, was retained and detectable by scRNA-seq. *Hoxa* genes, in particular, were highly informative to identify the topological distribution of the subclusters and also helped us uncovered three additional subclusters in the transition from the aortic arch (neural crest origin) and descending aorta (mesodermal origin), namely subclusters 3, 4 and 0.

Importantly many of the region-specific genes identified have been previously associated with vascular disease (Supplemental Table 6). For example: *Gucy1a3*, a transcript enriched in the carotids has been linked to moyamoya[21]; *Prdm6* and *Tfap2b* both specific to the arch were associated with patent ductus arteriosus[24,34], while *Mfap5* was associated with Familial aortic aneurysm classified as thoracic, but that affected the aortic root[35]. These associations further lend support the central hypothesis being tested that, localized, slightly skewed, transcriptional expression patterns can be informative in exploring disease causation. Along these lines and to further scrutinize the hypothesis, we overlaid the information captured by scRNA-seq to inquire (in silico) whether loci previously associated with vascular disease included any of the site-specific transcripts identified.

The Online Mendelian Inheritance in Man (OMIM) database is a manually curated, daily updated catalog of information for all known mendelian diseases including known causative genes. Despite significant advancement in the identification of disease-causing mutations using next-generation sequencing, OMIM still contains several hundred mendelian disorders for which the underlying genetic mutation remains unknown[36]. Thus, we had the opportunity to overlap the identified transcripts with regional-specific disease, such as moyamoya, patent ductus arteriosus, and site-specific aneurysms (Supplementary Table 5). Through this process, we captured 13 genes (*Mcam, Atf5, Cyth2, Ppp1r15a, Tbl1x, Bhlhe4, Wwp1, Stk3, Ywhaz, Klf10, Azin1, Pabpc1, and Gem*) that overlapped between site specific enrichment (this work) and previously identified loci for localized vascular disease. While exciting, definitive, causative proof will require rigorous mechanistic experiments. Along these lines, we decided to undertake one such exploration in relation to *Mcam* and aneurysm.

Rupture of aortic aneurysms is a well-established cause of preventable cardiovascular death in the developed world. However, given the disease's asymptomatic presentation, improved methods for identification of individuals at risk is required to decrease disease-associated mortality. Significant evidence has strongly suggested a substantial role for genetic susceptibility in the development of both thoracic and abdominal aortic aneurysms (TAA and AAA)[37]. In fact, mutations in several structural genes coding for extracellular matrix and contractile proteins have been identified and replicated in animal models[38]. Nonetheless, not all aneurysms showed mutations in the pool of causative genes identified so far, thus either additional genes are yet to be identified and/or non-genetic causes of the pathology, such as inflammation, might be at fault. In addition, gene-modifiers are also acknowledged to contribute to the pathology and part-take in severity. Our systematic in silico approach showed that *Mcam* mapped to the FAA1 locus[20] and was specifically reduced in smooth muscle cells at sites of aneurysms, further supporting a possible contribution of MCAM in the structural integrity of the tunica media.

Mcam (also known as CD146) is a transmembrane glycoprotein expressed by endothelial, pericytes, and smooth muscle cells and that also functions as a co-receptor for PDGFRß[39,40]. During embryonic development, CD146/MCAM is highly expressed in early progenitors of smooth muscle cells when it was shown to regulate the balance between vascular smooth muscle cell proliferation and differentiation[41]. Expression in adult smooth muscle appears to be restricted to sites subjected to flow-mediated stress such as branches[41] and the ventral aspect of the dorsal aorta (this work). Our findings are consistent with a role for CD146/MCAM in resisting tissue stress and promoting resilience, also a function in accord to its expression at vascular branches. Importantly, two recent publications have also attributed immunomodulatory functions to CD146/MCAM[42,43], a role of importance in for prevention of tissue breakdown by inflammatory cells also known to promote or exacerbate aneurysm development. Definitive experimental support for the contribution of CD146/MCAM in the prevention of aneurysms came from the evaluation of disease development in the null mouse, that clearly showed earlier lethality and increased pathology in comparison to control littermates.

Overall, our work adjoints a growing list of scRNA-seq profiling studies aiming at improving our collective understanding of transcriptional heterogeneities present in the vascular tree. Importantly, we further demonstrated that detailed scRNA-seq data can be leveraged to advance previous GWAS studies that linked loci to vascular disease.

Methods

Mice

Mice were housed at the University of California, Los Angeles (UCLA), New York University Langone Health and at the Cardiovascular Research Center, Universite de Paris, France. They received food and water ad libitum and experimental procedures were reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at the respective institutions. Protocols were conducted in accordance with federal regulations as outlined in the “Guide for the Care and Use of Laboratory Animals” for US labs and European Community Guidelines for European labs. For studies using CD146/Mcam knock out mouse[44], animals were backcrossed for more than 10 generations on the C57BL/6J background, both males and females were used in the studies between 8-14 weeks and comparison was done with littermates. For induction of aneurysms in double knock out mice, CD146/Mcam deficient mice were crossed with ApoE null mice (C57BL/6J background, Janvier Labs) and backcrossed for more than 10 generations on the C57BL/6J background. Primers for genotyping CD146/Mcam alleles were as follows:5’-TCACTTGACAGTGTGATGGT-3' (forward primer used to detect CD146/Mcam WT, floxed and KO alleles), 5’-CCTTAGAAAGCAGGGATTCA-3' (reverse primer used to detect CD146/Mcam WT and floxed alleles) and 5’-CCCAAATCCTCTGGAAGACA-3' (reverse primer used to detect CD146/Mcam KO allele).Genotyping primers for the ApoE alleles are as followed: 5'-GCCTAGCCGAGGGAGAGCCG-3', 5'-TGTGACTTGGGAGCTCTGCAGC -3' and 5'-GCCGCCCCGACTGCATCT-3'.

AAA induction

For AAA induction, 8 week-old male apolipoprotein deficient (*ApoE*$^{-/-}$) mice, as well as littermates with genotypes: *ApoE*$^{-/-}$*Mcam*$^{+/+}$ and *ApoE*$^{-/-}$*Mcam*$^{-/-}$ were implanted with Alzet osmotic pumps (model 2004; 0000298, Durect Corporation) subcutaneously to deliver a dosage of 1 µg/kg/min of PBS or angiotensin II (H-1705, Bachem) for 28 days[45]. Aneurysm progression was monitored by Doppler ultrasound and aortic measurements were quantified on images of abdominal aorta captured prior to pump implantation and at the end of the experiment.

Immunostaining and Confocal Microscopy

Formalin fixed, paraffin embedded specimens from aortic and carotid vessels were sectioned at 5µm and incubated with primary antibodies overnight followed by species - specific secondary antibodies for 1 hour prior to mounting in Prolong Gold (Thermo Scientific). Specific antibody information including antibody concentration, vendor information and RRID are included in Supplementary Table 1. A subset of antibodies (*Tfap2b*) were amplified using tyramide signal amplification kits (T20934) (Thermo Scientific)

following manufacturer's recommendations. Samples were evaluated and photographed using a confocal microscope (Nikon 1AR) equipped with 20X and 60X oil objectives. For a subset of candidates (*Tfap2b* and *Mcam*), samples were evaluated and photographed using an LSM880 confocal microscope (Carl Zeiss) equipped with Zeiss Plan-Apochromat 20x/0.8 M27 for acquisition.

Image quantification: Images were first processed with the Imaris file converter then transferred to the Imaris 9.7 software. Visualization of all the channels was normalized across images from all three anatomic sites. For quantification of the relative average intensity of staining for candidate gene. Imaris surfaces were generated using the aSMA+ channel (for cytoplasmic staining) using background subtraction for thresholding. The surfaces were approximately of 10 voxels in size per surface across the entire image for all three regions. This approach allowed us to segerate and restrict our analysis of the channel intensity to only those regions in the image that expressed aSMA+. Subsequently, following surface creation, all of the aSMA+ surfaces in image of the anatomic location were selected and the mean intensity of the site-specific candidate (Red Channel) per surface was extracted from the statistics table and exported as a new tab delimited file. This process of surface generation, selection and channel intensity statistic export was repeated for all 3 site, per candidate gene investigated if the staining pattern appeared predominantly cytosolic. If the stanning of the candidate gene was nuclear, surfaces were generated using the DAPI channel (nuclear staining – *Tfap2b*) using background subtraction for thresholding. The surfaces were approximately of 10 voxels in size per surface across the entire image for all three regions. Then manually, only those surfaces from nuclei that resided in the tunica media (nuclei in cells positive for aSMA) were selected for data export and analysis. Once exported out of imaris, the mean channel intensity files for the candidate genes (3 files per candidate gene) were merged into a table (1 table per gene, 11 tables total) and imported into R Studio for statistical testing and data visualization.

Single Cell RNA Sequencing: Sample Preparation for Site Enriched Transcriptomics: 30 minutes following SQ injection of 400 i.u of heparin 3 C57BL/6J male mice were sacrificed and perfused with 10mL of Versene. Following perfusion, the aorta and carotid vessels were harvested from the mouse and further dissected under a microscope to remove the adventitial layer. Subsequently the three vascular fragments: carotid arteries (right and left common not their branches), aortic arch (from the root to the subclavian branch), and the descending aorta (below the subclavian up to the gonadal branches, but excluding all branches) were obtained, minced into small fragments and incubated under agitation into 1mL digestion buffer containing freshly prepared liberase solution (2.5% Liberase TH from frozen stock [Sigma Aldrich], 5 Kunitz Unit/mL DNase1, 1M HEPES IN 1X HBSS). All preparations were done at temperature except for the digestion, which was done at 37^{o}C for approximately 20min. The suspension was neutralized with DMEM + 10%FBS and ran through a 40uM filter and then centrifuged for 5 min at 10,000 x g. Following centrifugation, the cell pellet was resuspended in in 0.4% BSA-PBS and assessed for viability prior to library preparation using trypan blue staining with the TC20 cell counter (BioRad). Only samples with more than 85% viability were used in the generation of libraries. Cells were partitioned with Gel Beads into emulsion in the Chromium instrument where cell lysis and barcoded reverse transcription of RNA

occurred following amplification. Single-cell RNAseq libraries were prepared by using the Chromium single cell 3' library and gel bead kit v3 (10x Genomics).

Single Cell RNA Sequencing: Sample Preparation for AAA mouse model: At sacrifice, three aneurysmal and three control aortas were pooled and digested for 1 h at 37°C in 10 mg/ml Collagenase type II (C6885, Sigma Aldrich) and 1 mg/ml Elastase (LS002292, Worthington Biochemistry) and processed for single-cell RNA-sequencing.

Data Availability and Bioinformatic Analysis for Single Cell RNA Sequencing Data

The scRNA-seq data sets in this article are deposited in the international public repository Gene Expression Omnibus database under accession code GSE156731, as well as Hadi 2018 (AAA)[45]. Sequencing of all three libraries (positional transcriptomic dataset) was performed on the same run on Illumina HiSeq 4000, and the digital expression matrix was generated by demultiplexing, barcode processing, and gene unique molecular index counting. Further processing of scRNAseq involved quality control, normalization, confounding factor identification, dimensionality reduction and cell-gene level analysis. Preprocessing of the raw data was conducted following the Cell Ranger pipeline (10x Genomics). To identify different cell types and find signature genes for each cell type, the R package Seurat (version 3.2.2) was used to analyze the digital expression matrix. Cells with less than 500 unique molecular identifiers (UMIs) and greater than 10% mitochondrial expression were removed from further analysis. Seurat function NormalizeData was used to normalize the raw counts. Variable genes were identified using the FindVariableGenes function, the top 2000 variable genes were selected for further analysis. The Seurat ScaleData function was used to scale and center expression values in the dataset for dimensional reduction. Principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation and projection (UMAP) were used to reduce the dimensions of the data, and the first 2 dimensions were used in plots. A graph-based clustering approach was later used to cluster the cells; then signature genes were found and used to define cell types for each cluster. The FindAllMarkers Seurat function was used to find positive markers of each vSMC population with the Wilcoxon rank-sum test. Pathway enrichment analysis was performed using Metascape. Top 10 categories for each of the clusters investigated as Supplemental Excel File 4. Data visualization (graphs) were generated using ggplot2 package for R.

For Mcam-based AAA Analysis: To identify different cell types and find signature genes for each cell type, the R package Seurat (version 3.1.2) was used. Neighboring and clustering was performed on the 10 top most significant PCAs using a resolution of 0.5. For Mcam analysis, a cutoff of ≥ 1 or ≤ 1 was used to define Mcam high and low populations, respectively. For batch correction we used Harmony because it simultaneously accounts for multiple experimental and biological factors and has shown outstanding performance to previously published algorithms particularly for large cell number files. The integrated dataset was then used to t-SNE clustering utilizing the formerly generated Harmony embeddings.

Site Specific Gene Signature determination: The FindAllMarkers Seurat function was used to find positive markers of each vSMC population with the Wilcoxon rank-sum test for each anatomic location (3; Supplemental Excel File 1) and subcluster population (6; Supplemental Excel File 3). Following gene list generation, the datasets were filtered for significance (padj $\leq$0.05) and the presence of shared genes between anatomic sites with duplicate genes if identified allocated to the anatomic location (1,045) or vSMC subcluster (1,681) with the most significant padj.

Subcluster and Anatomic Location Signature overlap: To identify shared signature genes the final gene lists for both anatomic location (1,045) and vSMC subcluster (1,681) were overlayed individually using the identify duplicate functions in excel. Genes were characterized as one of 3 major categories: unique to anatomic location, unique to subcluster signature or shared between a specific anatomic location and specific vSMC subcluster. From this data, a tabular matrix was constructed with all unique permutations of the 3 major categories (31 patterns combinations were found) and the recording number of genes in each category and data visualization was performed using the tidyverse and ggalluvial packages in R.

Gene ontology analysis: Pathway enrichment analysis was performed using Metascape express analysis using default options. Metascape utilizes the hypergeometric distribution in the calculation of significance for each gene enrichment category. Scores (log p-value) are reflective of the number of genes in the gene list defined by the category in question relative to the total number of genes in the category relative to background. From the output, the top 10 summary gene ontology categories (as ranked by p-value of enrichment) for each of the anatomic locations were used for visualization of gene ontology following curations and removal of duplicate/highly similar terms (Supplemental Excel File 4). From this data, into two tabular matrices were constructed. The first matrix (the vertex file): contained the fold enrichment for the individual gene ontology categories as well as top gene ontology category membership, fold enrichment for anatomic location and p-value for anatomic location. The second matrix (Edge file) contained all gene ontology memberships for visualized genes. Both matrices were then visualized into a custom ontology network using a combination of ggplot2, igraph and ggraph packages for R.

“In-silico” forward genetic screen: Genomic loci for mendelian vascular anomalies were identified through query, integration, and manual curation of data from multiple online biomedical databases as follows. Given the complexity of mendelian disease nomenclature, vascular diseases/traits were identified using the Human Phenotype Ontology (HPO) database (https://hpo.jax.org/app/) using the search term “vascular anomalies” (HP:0002597) which generated 1043 entries. HPO entries were filtered to sub-select for those entries represented in Online Mendelian Inheritance of Man (OMIM) (omim.org). Of the 540 OMIM entries, 193 entries were classified as phenotypic series entries, which is used to identify those disorders with known genetic heterogeneity. Given that the fact that the initial HPO query restricted OMIM entries to those phenotypic series entries with known-disease genes associations we reintegrated the missing entries for each disease/trait, representing and 924 additional entries. These 1464 entries were then manually curated, and entries were excluded based on the following criteria: 1) predominance of vascular feature, 2) known disease gene 3) case reports 4) chromosomal abnormalities 5) association or

ambiguous mapping to genome. This exclusion left 16 entries representing 16 unique genomic loci representing 6 distinct vascular traits. Using the published genomic markers that defined the critical region (either STS or SNPs depending on the publication/loci in question), the coordinates of these loci were translated to hg38 using the STS track on the UCSC genome browser. Using the UCSC table browser (https://genome.ucsc.edu/) we queried the UCSC SQL database to identify all genes residing in each of the given loci using the knownGene table based on the most current build of the human genome (hg38). The “name” column data from knownGene table for each query was then used to pull data from the KgxRef table to convert gene identifiers to HUGO approved gene symbols. The gene symbols from the regional loci data were then overlaid with the scRNASeq anatomic origin markers for each of the 3 locations. Genes were identified as positive hits if they resided withing the genomic coordinates of the loci in question and defined vSMCs in the arterial bed relevant to the trait.

Statistics: For image analysis and animal studies, statistical tests were performed using stats and rstatix packages. Gaussian distribution of the data was assessed with Shapiro–Wilk test. Normally distributed data were compared by unpaired t-test (two-group comparison). Data that departed from the Gaussian distribution were analyzed using Kruskal-Wallis test (multiple groups).

Declarations

Author Contributions

MCR and MLIA designed the experimental design. MCR, FM, GEH, MV, TK, MS, TH, AR, MR, NB and MBC performed experiments; MCR, MLIA, MP, ASL, HAO, and BR analyzed and interpreted the data. MCR and MLIA wrote the manuscript with comments and edits from all the authors. All authors approved the final version of the manuscript.

Acknowledgments

The authors would like to thank the UCLA sequencing core facility and members of the Arispe lab for valuable discussions.

Sources of Funding

This work was supported by funds from the Ruth L. Kirschstein National Research Service Award (T32 HL069766 and T32 HL1346334) to Milagros Romay and grants from the National Institutes of Health (NIH) to MLIA (R35HL140014) and BR (R01HL146627).

Disclosures

The authors declare no competing interests.

References

1. Majesky, M. W. Developmental basis of vascular smooth muscle diversity. *Arterioscler Thromb Vasc Biol* **27**, 1248–1258, doi:10.1161/ATVBAHA.107.141069 (2007).
2. Wasteson, P. *et al.* Developmental origin of smooth muscle cells in the descending aorta in mice. *Development* **135**, 1823–1832, doi:10.1242/dev.020958 (2008).
3. Harmon, A. W. & Nakano, A. Nkx2-5 lineage tracing visualizes the distribution of second heart field-derived aortic smooth muscle. *Genesis* **51**, 862–869, doi:10.1002/dvg.22721 (2013).
4. Etchevers, H. C., Vincent, C., Le Douarin, N. M. & Couly, G. F. The cephalic neural crest provides pericytes and smooth muscle cells to all blood vessels of the face and forebrain. *Development* **128**, 1059–1068 (2001).
5. Sawada, H., Rateri, D. L., Moorleghen, J. J., Majesky, M. W. & Daugherty, A. Smooth Muscle Cells Derived From Second Heart Field and Cardiac Neural Crest Reside in Spatially Distinct Domains in the Media of the Ascending Aorta-Brief Report. *Arterioscler Thromb Vasc Biol* **37**, 1722–1726, doi:10.1161/ATVBAHA.117.309599 (2017).
6. Chang, H. Y. Anatomic demarcation of cells: genes to patterns. *Science* **326**, 1206–1207, doi:10.1126/science.1175686 (2009).
7. Trigueros-Motos, L. *et al.* Embryological-origin-dependent differences in homeobox expression in adult aorta: role in regional phenotypic variability and regulation of NF-kappaB activity. *Arterioscler Thromb Vasc Biol* **33**, 1248–1256, doi:10.1161/ATVBAHA.112.300539 (2013).
8. Visconti, R. P. & Awgulewitsch, A. Topographic patterns of vascular disease: HOX proteins as determining factors? *World J Biol Chem* **6**, 65–70, doi:10.4331/wjbc.v6.i3.65 (2015).
9. Pruett, N. D. *et al.* Changing topographic Hox expression in blood vessels results in regionally distinct vessel wall remodeling. *Biol Open* **1**, 430–435, doi:10.1242/bio.2012039 (2012).
10. Cheung, C., Bernardo, A. S., Trotter, M. W., Pedersen, R. A. & Sinha, S. Generation of human vascular smooth muscle subtypes provides insight into embryological origin-dependent disease susceptibility. *Nat Biotechnol* **30**, 165–173, doi:10.1038/nbt.2107 (2012).
11. Owens, A. P., 3rd *et al.* Angiotensin II induces a region-specific hyperplasia of the ascending aorta through regulation of inhibitor of differentiation 3. *Circ Res* **106**, 611-619, doi:10.1161/CIRCRESAHA.109.212837 (2010).
12. Topouzis, S. & Majesky, M. W. Smooth muscle lineage diversity in the chick embryo. Two types of aortic smooth muscle cell differ in growth and receptor-mediated transcriptional responses to transforming growth factor-beta. *Dev Biol* **178**, 430–445, doi:10.1006/dbio.1996.0229 (1996).
13. Gadson, P. F., Jr. *et al.* Differential response of mesoderm- and neural crest-derived smooth muscle to TGF-beta1: regulation of c-myb and alpha1 (I) procollagen genes. *Exp Cell Res* **230**, 169–180, doi:10.1006/excr.1996.3398 (1997).
14. MacFarlane, E. G. *et al.* Lineage-specific events underlie aortic root aneurysm pathogenesis in Loeys-Dietz syndrome. *J Clin Invest* **129**, 659–675, doi:10.1172/JCI123547 (2019).

15. Davis, F. M., Daugherty, A. & Lu, H. S. Updates of Recent Aortic Aneurysm Research. *Arterioscler Thromb Vasc Biol* **39**, e83-e90, doi:10.1161/ATVBAHA.119.312000 (2019).
16. Quintana, R. A. & Taylor, W. R. Introduction to the Compendium on Aortic Aneurysms. *Circ Res* **124**, 470–471, doi:10.1161/CIRCRESAHA.119.314765 (2019).
17. Quintana, R. A. & Taylor, W. R. Cellular Mechanisms of Aortic Aneurysm Formation. *Circ Res* **124**, 607–618, doi:10.1161/CIRCRESAHA.118.313187 (2019).
18. Evans, G. H., Stansby, G. & Hamilton, G. Suggested standards for reporting on arterial aneurysms. *J Vasc Surg* **15**, 456, doi:10.1016/0741-5214(92)90269-e (1992).
19. Isselbacher, E. M. Thoracic and abdominal aortic aneurysms. *Circulation* **111**, 816–828, doi:10.1161/01.CIR.0000154569.08857.7A (2005).
20. Vaughan, C. J. *et al.* Identification of a chromosome 11q23.2-q24 locus for familial aortic aneurysm disease, a genetically heterogeneous disorder. *Circulation* **103**, 2469–2475, doi:10.1161/01.cir.103.20.2469 (2001).
21. Herve, D. *et al.* Loss of alpha1beta1 soluble guanylate cyclase, the major nitric oxide receptor, leads to moyamoya and achalasia. *Am J Hum Genet* **94**, 385–394, doi:10.1016/j.ajhg.2014.01.018 (2014).
22. Kohler, S. *et al.* Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res* **47**, D1018-D1027, doi:10.1093/nar/gky1105 (2019).
23. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* **43**, D789-798, doi:10.1093/nar/gku1205 (2015).
24. Khetyar, M., Syrris, P., Tinworth, L., Abushaban, L. & Carter, N. Novel TFAP2B mutation in nonsyndromic patent ductus arteriosus. *Genet Test* **12**, 457–459, doi:10.1089/gte.2008.0015 (2008).
25. Deng, D. X. *et al.* Molecular signatures determining coronary artery and saphenous vein smooth muscle cell phenotypes: distinct responses to stimuli. *Arterioscler Thromb Vasc Biol* **26**, 1058–1065, doi:10.1161/01.ATV.0000208185.16371.97 (2006).
26. Yoshida, T. & Owens, G. K. Molecular determinants of vascular smooth muscle cell diversity. *Circ Res* **96**, 280–291, doi:10.1161/01.RES.0000155951.62152.2e (2005).
27. Liu, M. & Gomez, D. Smooth Muscle Cell Phenotypic Diversity. *Arterioscler Thromb Vasc Biol* **39**, 1715–1723, doi:10.1161/ATVBAHA.119.312131 (2019).
28. Allahverdian, S., Chaabane, C., Boukais, K., Francis, G. A. & Bochaton-Piallat, M. L. Smooth muscle cell fate and plasticity in atherosclerosis. *Cardiovasc Res* **114**, 540–550, doi:10.1093/cvr/cvy022 (2018).
29. Bennett, M. R., Sinha, S. & Owens, G. K. Vascular Smooth Muscle Cells in Atherosclerosis. *Circ Res* **118**, 692–702, doi:10.1161/CIRCRESAHA.115.306361 (2016).
30. Albarran-Juarez, J., Kaur, H., Grimm, M., Offermanns, S. & Wettschureck, N. Lineage tracing of cells involved in atherosclerosis. *Atherosclerosis* **251**, 445–453, doi:10.1016/j.atherosclerosis.2016.06.012 (2016).

31. Feil, S. *et al.* Transdifferentiation of vascular smooth muscle cells to macrophage-like cells during atherogenesis. *Circ Res* **115**, 662–667, doi:10.1161/CIRCRESAHA.115.304634 (2014).
32. Speer, M. Y. *et al.* Smooth muscle cells give rise to osteochondrogenic precursors and chondrocytes in calcifying arteries. *Circ Res* **104**, 733–741, doi:10.1161/CIRCRESAHA.108.183053 (2009).
33. Dobnikar, L. *et al.* Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. *Nat Commun* **9**, 4567, doi:10.1038/s41467-018-06891-x (2018).
34. Li, N. *et al.* Mutations in the Histone Modifier PRDM6 Are Associated with Isolated Nonsyndromic Patent Ductus Arteriosus. *Am J Hum Genet* **98**, 1082–1091, doi:10.1016/j.ajhg.2016.03.022 (2016).
35. Barbier, M. *et al.* MFAP5 loss-of-function mutations underscore the involvement of matrix alteration in the pathogenesis of familial thoracic aortic aneurysms and dissections. *Am J Hum Genet* **95**, 736–743, doi:10.1016/j.ajhg.2014.10.018 (2014).
36. Bamshad, M. J., Nickerson, D. A. & Chong, J. X. Mendelian Gene Discovery: Fast and Furious with No End in Sight. *Am J Hum Genet* **105**, 448–455, doi:10.1016/j.ajhg.2019.07.011 (2019).
37. Pinard, A., Jones, G. T. & Milewicz, D. M. Genetics of Thoracic and Abdominal Aortic Diseases. *Circ Res* **124**, 588–606, doi:10.1161/CIRCRESAHA.118.312436 (2019).
38. Lindsay, M. E. & Dietz, H. C. Lessons on the pathogenesis of aneurysm from heritable conditions. *Nature* **473**, 308–316, doi:10.1038/nature10145 (2011).
39. Chen, J. *et al.* CD146 coordinates brain endothelial cell-pericyte communication for blood-brain barrier development. *Proc Natl Acad Sci U S A* **114**, E7622-E7631, doi:10.1073/pnas.1710848114 (2017).
40. Leroyer, A. S. *et al.* CD146 (Cluster of Differentiation 146). *Arterioscler Thromb Vasc Biol* **39**, 1026–1033, doi:10.1161/ATVBAHA.119.312653 (2019).
41. Roostalu, U. *et al.* Distinct Cellular Mechanisms Underlie Smooth Muscle Turnover in Vascular Development and Repair. *Circ Res* **122**, 267–281, doi:10.1161/CIRCRESAHA.117.312111 (2018).
42. Kranjc Brezar, S. *et al.* Intratumoral Gene Electrotransfer of Plasmid DNA Encoding shRNA against Melanoma Cell Adhesion Molecule Radiosensitizes Tumors by Antivascular Effects and Activation of an Immune Response. *Vaccines (Basel)* **8**, doi:10.3390/vaccines8010135 (2020).
43. Wang, Z. *et al.* CD146, from a melanoma cell adhesion molecule to a signaling receptor. *Signal Transduct Target Ther* **5**, 148, doi:10.1038/s41392-020-00259-8 (2020).
44. Jouve, N. *et al.* CD146 mediates VEGF-induced melanoma cell extravasation through FAK activation. *Int J Cancer* **137**, 50–60, doi:10.1002/ijc.29370 (2015).
45. Hadi, T. *et al.* Macrophage-derived netrin-1 promotes abdominal aortic aneurysm formation by activating MMP3 in vascular smooth muscle cells. *Nat Commun* **9**, 5022, doi:10.1038/s41467-018-07495-1 (2018).

Figures

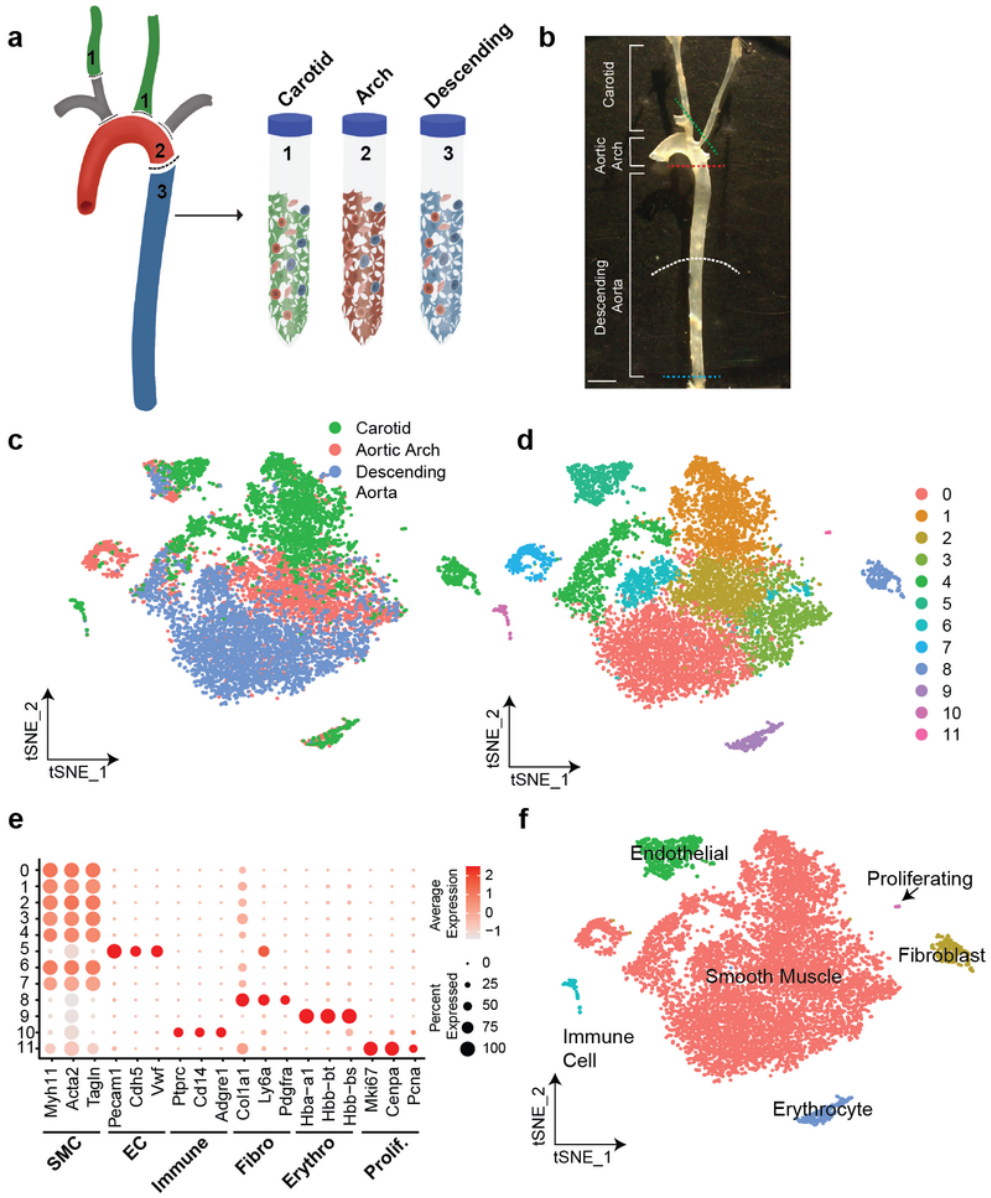

Figure 1

Cell type diversity across three vascular beds. a. Graphical description of experimental design. Carotid arteries (green), aortic arch (red) and descending aorta (blue) were dissected from three C57BL6/J male mice, combined, and enzymatically dissociated to obtain single cell suspensions for scRNA-seq using the 10X genomics platform. b. Image of an intact, isolated mouse aorta and carotid vessels following removal of the adventitial and prior to enzymatic digestion. Relative locations prior to dissociation are indicated as follow: carotid artery terminal point (green dotted line), aortic arch terminal point (red dotted line), and descending aorta terminal point (blue dotted line). Note that the descending aorta includes both thoracic and abdominal (up to gonadal artery). Diaphragm (curved white dotted line) is also indicated. Scale bar= 500 μM. c. tSNE plot of scRNA-seq data visualizing spread of data from all identified cells in the 3 anatomic locations characterized. d. tSNE plot identifying unique cell clusters from the data presented in C as identified using Seurat. e. Dot plot of visualizing the expression of classical cell type

markers in each of the identified clusters used to determine cellular identity of the individual clusters. f. tSNE plot of single cell clusters overlaying the cellular identity generated using expression of the markers showed in E.

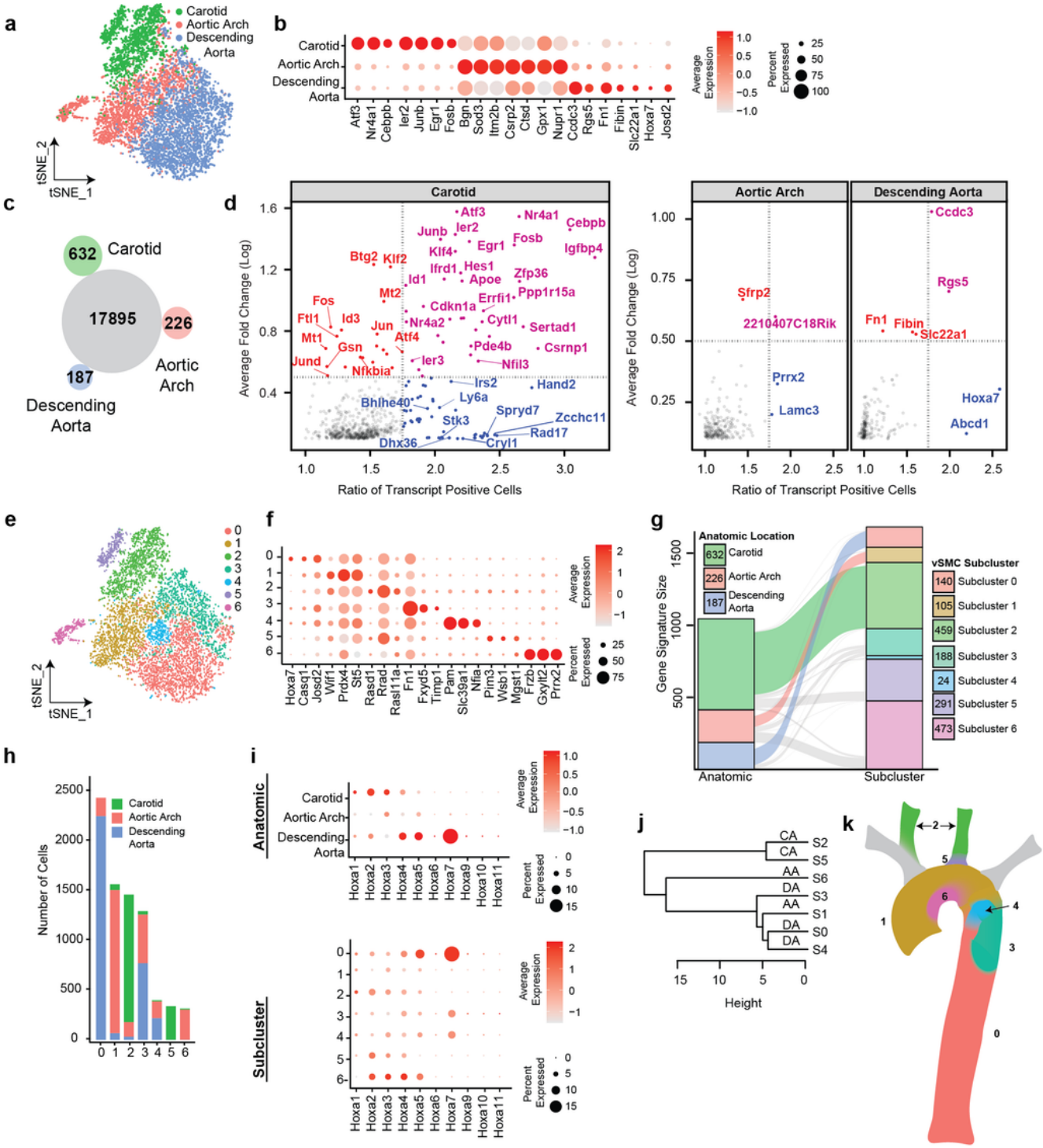

Figure 2

Identification of vSMC site-specific signatures. a. tSNE plot of all vSMCs identified from the three anatomic locations following sub-selection from the other cellular populations within the data (shown in

Fig. 1). b. Dot plot visualizing the expression of a subset of the most significant signature genes. c. Radial Venn diagram showing number of common transcripts (central gray circle) and site-enriched transcripts according to anatomic location. d. Scatter plot of site-enriched transcripts (Carotid: 632; Aortic Arch: 226; Descending Aorta: 187) displayed as average transcript fold change (y-axis) per cell relative to ratio of transcript expressing cells (x-axis). Class I genes are dark gray. Class II genes (transcript positive cells) are labeled in blue. Class III genes (fold change) are labeled in red. Class IV genes (fold change and transcript positive cells) are labeled in purple. For visualization purposes, only a subset of the genes named in the figure, for further detail please refer to supplementary excel file 2. e. tSNE plot of identifying unique SMC subclusters as determined by seurat. f. Dot plot visualizing the top differentially expressed genes that form the unique gene signatures in the seven vSMC subclusters. g. Alluvial plot visualizing the relationship between anatomic location and their respective contributions to the vSMC subcluster gene signatures. The strata (stacked box plots) indicate the size of the individual gene signatures as colored by the legend. The alluvia (ribbon-like structures connecting the two strata) indicate the number of gene shared by the interconnecting signatures. The three colored alluvia indicate the dominant subcluster signature linked to each anatomic location defined gene signature as follows: carotid-subcluster 2 (green), aortic arch-subcluster 1 (pink) and descending aorta – subcluster 0 (blue). h. Absolute number of vSMC present into each vSMC subclusters and their respective locations (as per color). i. Dot plot of Hoxa cluster gene expression in the vSMC population by both anatomic location and subcluster identity. j. Hierarchical Clustering of vSMC subclusters k. Diagram of subcluster location based on Hoxa expression (2I), hierarchical clustering (2J), and known anatomic location used to generate the initial libraries.

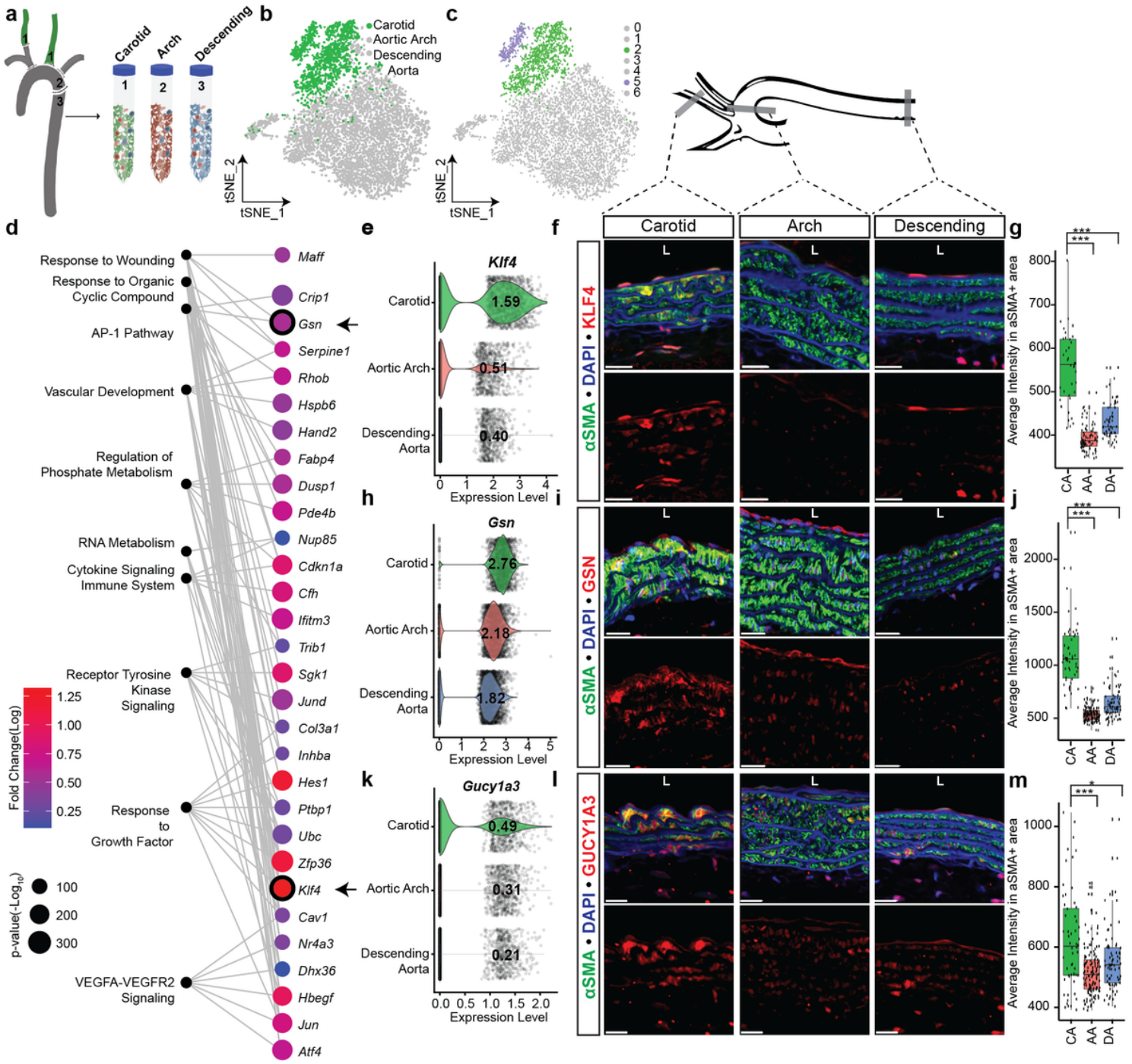

Figure 3

Carotid transcriptional signature. a. Experimental design overview highlighting the carotid arteries. b. tSNE plot visualizing the location-defined Carotid vSMC cluster. c. tSNE plot illustrating the vSMC subclusters with majority Carotid vSMC membership (subclusters 2 and 5). d. Gene ontology enrichment of Carotid vSMC signature including the top 10 unique ontology categories and selected member genes within each category. Dot color indicates average enrichment (log Fold Change), while size indicates significance of enrichment. For Klf4 and Gsn, the black circle surrounding the dots indicates that p-value of enrichment is smaller than the limit imposed by R ($p \leq 2.225074e\text{-}308$). Arrows highlight the genes

that were further validated using immunofluorescence (IF). e. Violin plot comparing vSMC expression of Klf4 in the three anatomic locations. Bold numbers within the violin indicate the average Klf4 expression per vSMC in the indicated location. f. Immunofluorescent staining of KLF4 (red) in histological sections representing the three anatomic locations. Note that KLF4 protein is also noted in the endothelium of carotid and descending aorta but not in the arch. Expression in smooth muscle, however, is exclusive to the carotid. g. Quantification of average KLF4 expression (as measured by fluorescent intensity) in relation to alpha smooth muscle actin. h. Violin plot comparing vSMC expression of Gsn in the three anatomic locations. Bold numbers within the violin indicate average expression. i. Immunofluorescent staining of GSN (red). j. Quantification of GSN expression (as measured by fluorescent intensity) in relation to alpha smooth muscle actin. k. Violin plot comparing vSMC expression of Gucy1a3 transcript in the three anatomic location Bold numbers within the violin indicate the average Gucy1a3 expression per vSMC in the indicated anatomic location. l. Immunofluorescent identification of GUCY1A3 (red) in sections. m. Quantification of GUCY1A3 (as per fluorescent intensity) in relation to alpha smooth muscle actin. Error bars represent ± SD. n = 40-62 (g), 53-114(j),56-130(m) 10 – voxel aSMA+ surface per anatomic location. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p < 0.001$ using the Kruskal-Wallis test. CA = carotid, AA = aortic arch, DA= descending aorta, L = lumen. Scale bar = 20μM.

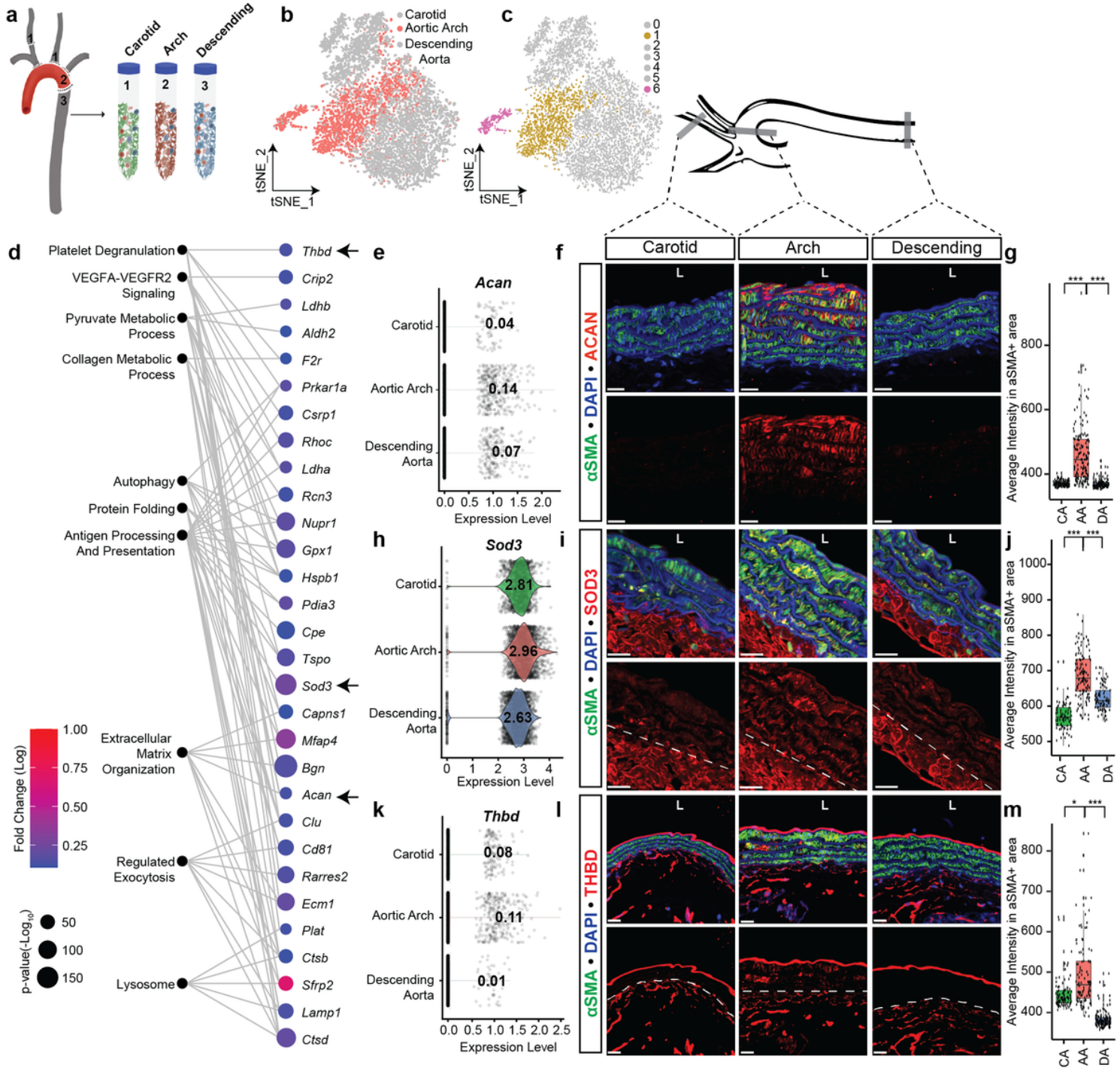

Figure 4

Aortic arch transcriptional signature. a. Overview of experimental design highlighting the aortic arch. b. tSNE plot visualizing the location-defined aortic arch vSMC cluster. c. tSNE plot visualizing the vSMC subclusters with majority aortic arch membership (subclusters 1 and 6). d. Gene ontology enrichment of Aortic Arch vSMC signature including the top 10 unique ontology categories as well as the expression of selected genes within each category. Dot Color indicates average enrichment (log Fold Change), while dot size indicates significance of enrichment. Arrows indicate genes selected for further validation via

immunofluorescence (IF). e. Violin plot comparing vSMC expression of Acan in the three anatomic locations. Bold numbers in the violin indicate the average Acan expression. f. Immunofluorescent staining for ACAN (red). g. Quantification of ACAN expression (as measured by fluorescent intensity) in relation to alpha smooth muscle actin. h. Violin comparing vSMC expression of Sod3 in the three anatomic locations. Bold numbers indicate the average Sod3 expression. i. Visualization of SOD3 by immunofluorescence (red) in sections from the three anatomic locations. The white dotted line indicates the boundary between the vessel's medial and adventitia layers. j. Quantification of SOD3 expression in vascular smooth muscle across three anatomic sites. k. Violin plot comparing vSMC expression of Thbd in the three anatomic locations. Bold numbers indicate average of Thbd expression per vSMC. l. Immunofluorescent identification of THBD (red) in three anatomic locations. The white dotted line indicates the boundary between the vessel's medial and adventitia layers. ,. Quantification of average THBD expression (as measured by fluorescent intensity) in vSMCs across three anatomic sites. Error bars represent ± SD. n = 124-155 (g), 76-113(j), 89-137(m) 10 – voxel aSMA+ surface per anatomic location. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p < 0.001$ using the Kruskal-Wallis test. CA = carotid, AA = aortic arch, DA= descending aorta , L = lumen. * $p \leq 0.05$, *** $p < 0.001$. Scale bar = 20μM.

Figure 5

Descending aorta transcriptional signature. a. Overview of experimental design highlighting the descending aorta. b. tSNE plot visualizing the location-defined descending aorta vSMC cluster. c. tSNE plot visualizing the subclusters with majority cellular membership from the descending aorta (subclusters 0, 3 and 4). d. Gene ontology enrichment of descending aorta vSMC signature including the top 10 unique ontology categories as well as the expression of selected genes within each category. Dot color indicates average enrichment (log Fold Change), while dot size indicates significance of enrichment. Arrows highlight the proteins selected for further validation on the right. e. Violin plot comparing vSMC expression of Aqp1 across the three anatomic locations. Bold numbers indicate the average Aqp1 expression. f. Immunofluorescent staining for AQP1 (red also white arrow) in the three anatomic locations. L indicates the location of the lumen of the vessel. g. Average AQP1 protein expression in vSMC across the three anatomic sites. h. Violin plot comparing vSMC expression of Ccdc3 in the three anatomic locations. Bold numbers indicate average Ccdc3 expression. i. Immunofluorescent staining for CCDC3 (red). j. Quantification of CCDC3 protein expression in relation to alpha smooth muscle actin. k. Violin plot comparing vSMC expression of Hsgp2 transcripts across the three anatomic locations. Bold numbers indicate the average Hsgp2 expression. l. Immunofluorescent staining for PERLECAN (Hsgp2) (red) in three anatomic locations. m. Quantification of PERLECAN(Hsgp2) in relation to alpha smooth muscle actin. Error bars represent ± SD. n = 86-153 (g), 97-138 (j), 66-137 (m) 10 – voxel aSMA+ surface per anatomic location. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p < 0.001$ using the Kruskal-Wallis test. CA = carotid, AA = aortic arch, DA= descending aorta, L = lumen. *** $p < 0.001$. Scale bar = 20μM.

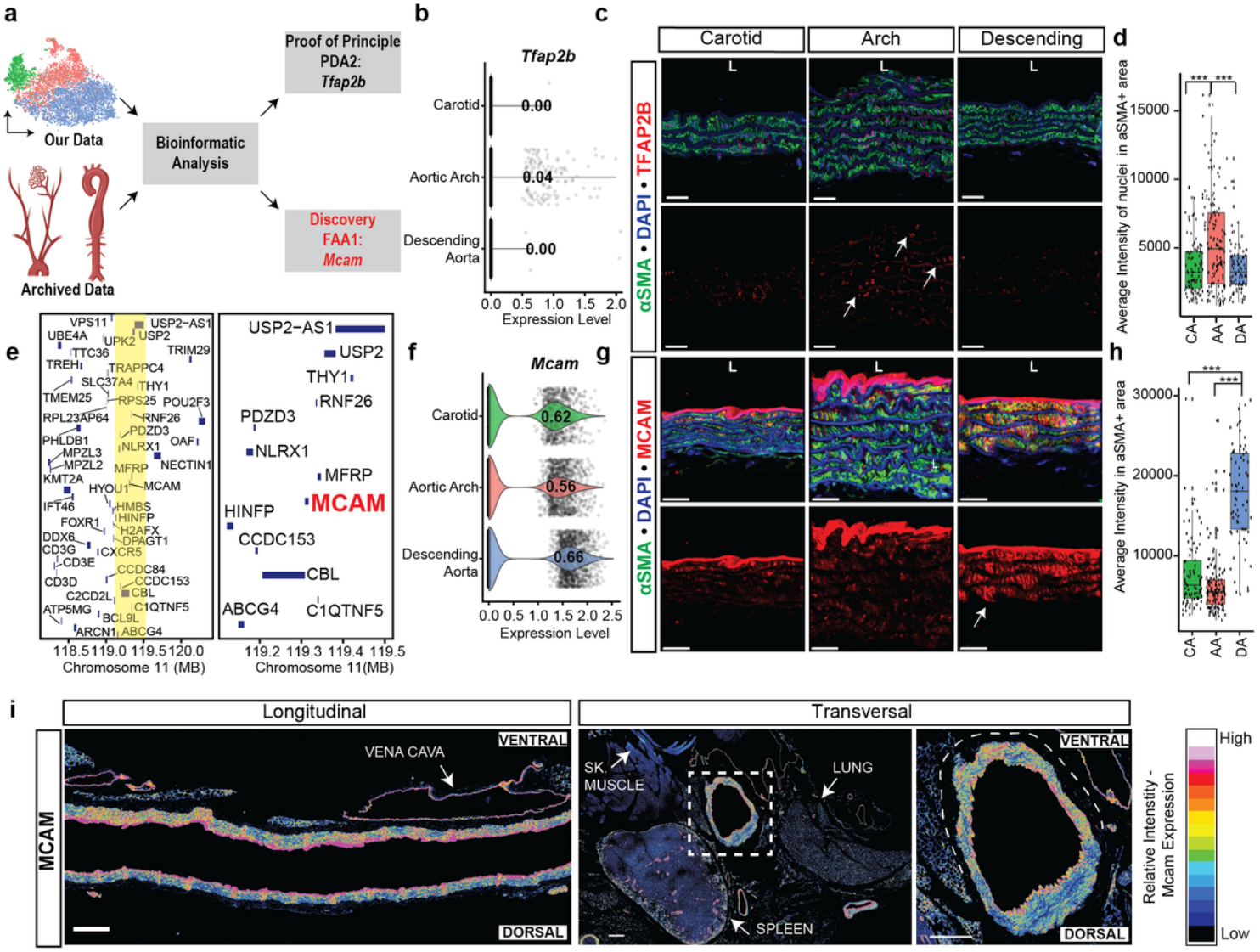

Figure 6

Identification of Mcam, a novel region-restricted aneurysmal candidate gene a. Overview of the approach utilized to determine if site-specific signatures might be associated with site-specific disease incidence (Proof of Principle: Tfap2b) and to uncover novel mendelian vascular disease candidates (FAA1-Mcam) through bioinformatic integration with archival data. b. Violin plot comparing Tfap2b transcript expression in vSMCs across the three anatomic locations. Bold numbers indicate average Tfap2b expression per SMC. c. Staining of TFAP2B protein (red) a known PDA-causing gene in the three anatomic locations showing significant enrichment in aortic arch vSMCs. d. Quantification of average TFAP2B protein expression in relation to aSMA. e. Visualization of FAA1 locus critical region including gene locations and current names as of genome build Hg38. f. Violin plot comparing expression of Mcam transcripts across the three anatomic locations. Bold numbers indicate average Mcam expression per vSMC. g. IF staining for MCAM protein (red) in the three anatomic locations. h. Quantification of average MCAM protein expression normalized to aSMA across three anatomic sites. i. IF staining of MCAM in the

thoracic aorta in longitudinal and transverse using intensity pseudocolor. Error bars represent ± SD. n = 90-145 (d), 86-124 (h), 10 – voxel aSMA+ surface per anatomic location. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p < 0.001$ using the Kruskal-Wallis test. *** $p < 0.001$. CA = carotid, AA = aortic arch, DA= descending aorta. L= lumen. Scale bar = 20µm (C and G), 200µm (I).

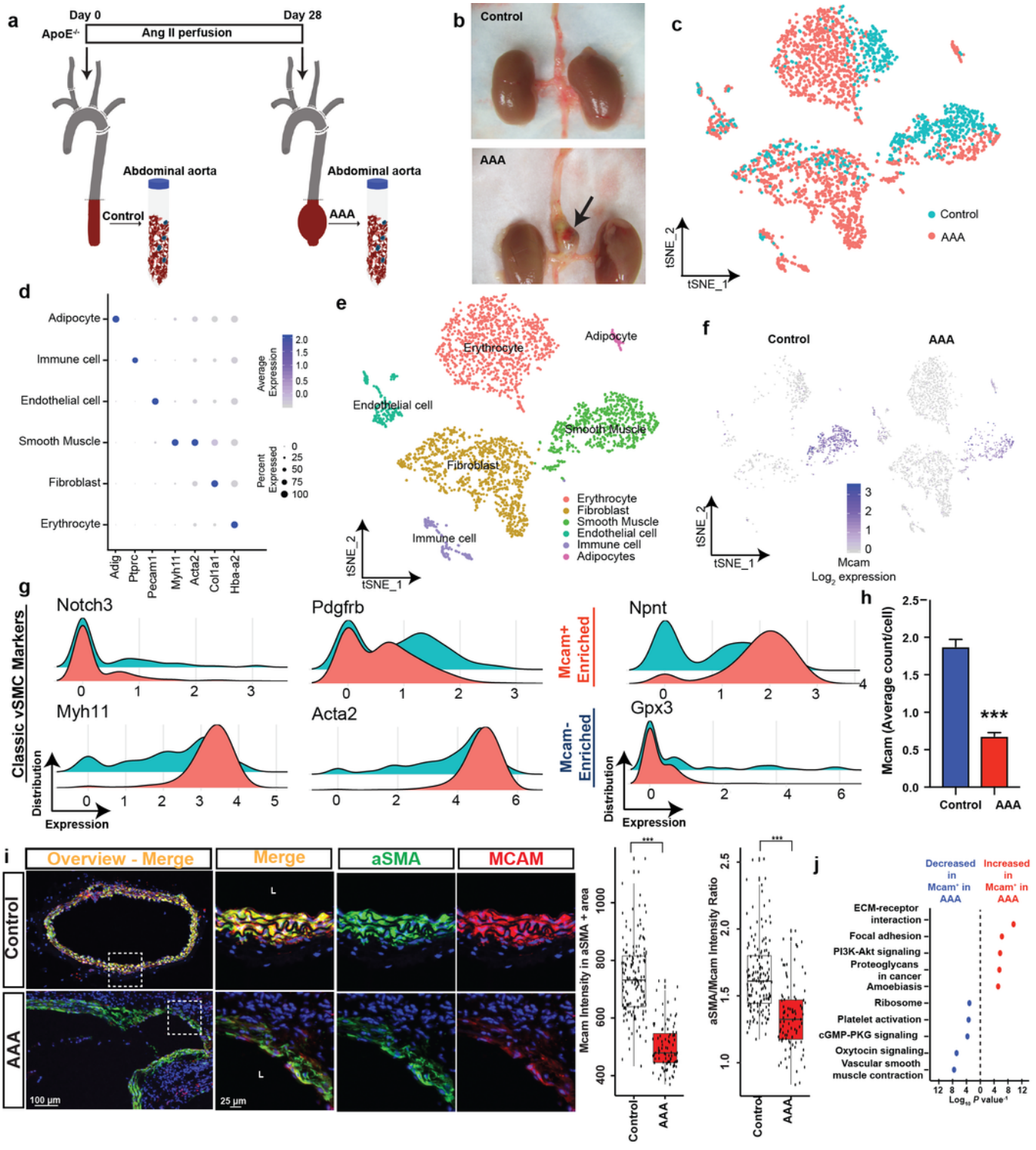

Figure 7

Characterization of Mcam/CD146 in AAA. a. Schematic representation of study design. ApoE -/- mice were implanted subcutaneously with osmotic pumps delivering constant rate of angiotensin II (n=3) or PBS (n=3) for 28 days. Aortas were collected, digested, and analyzed by single cell RNA sequencing. b. Image of control versus disease aorta prior to digestion. Arrow indicates region of aneurysm formation. c. t-distributed stochastic neighbor embedding (t-SNE) plot of single-cell RNA-sequencing of diseased aortas (n = 3, pooled) and healthy aortas (n = 3, pooled). d. Dot plot of classical cell type markers in identified cell populations in single-cell RNA-Seq of control versus diseased aortas. e. t-SNE plot of single-cell cell clusters showing cell type enrichment. f. Density plot representation of Mcam transcripts level in healthy and diseased aortas. g. Average global Mcam expression in healthy and diseased aortas. h. Distribution and expression of classical vSMC markers in Mcam+ and Mcam- vSMCs. i. IF staining of MCAM in control vs AAA aortas at low (20X) and high magnification (60X). Quantification of average MCAM expression (as measured by fluorescent intensity) and the relative ratio of aSMA to MCAM in vSMA across anatomic sites j. Pathway enrichment analysis showing most significantly decreased and increased pathways in Mcam positive cells from AAA aortas compared to Mcam positive cells from Healthy tissue. Error bars represent ± SD. n = 95-146 (i) 10 – voxel aSMA+ surface per anatomic location. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p < 0.001$ using the Wilcoxon rank sum test

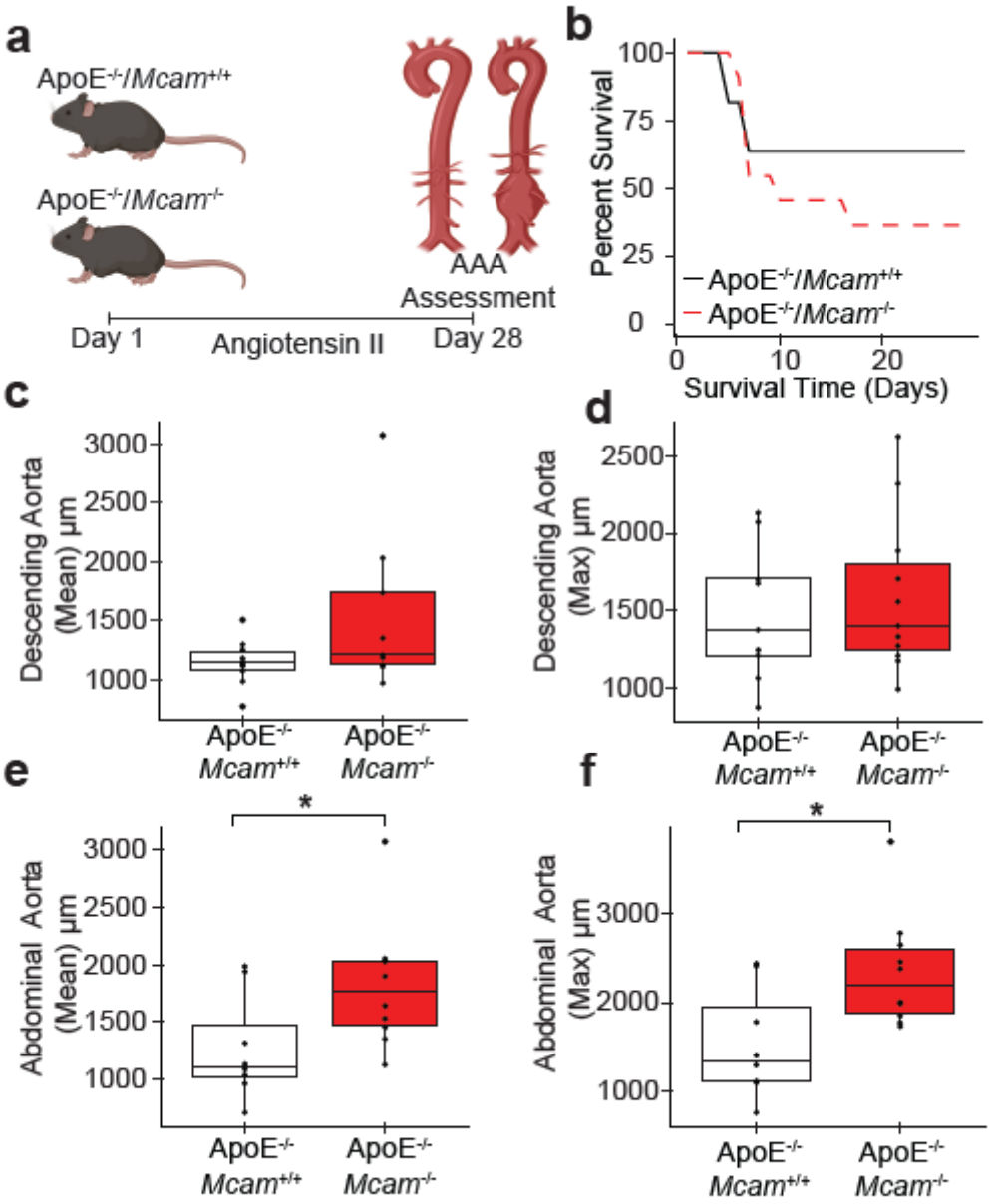

Figure 8

Loss of CD146/Mcam accelerates development of AAA. a. Diagram of experimental design. ApoE knockout mice lacking CD146/ Mcam (ApoE-/-/Mcam -/-) and CD146/Mcam wild-type litermates (ApoE-/-/Mcam +/+) were treated with angiotensin II for 28 days via osmotic pump to promote the development of abdominal aortic aneurysms (AAA). Following the treatment period, the animals were sacrificed, and measurements of pathologically non-compromised descending aorta were taken to assess vessel dilation. b. Kaplan Meier curves of control experimental groups over the course of angiotensinogen exposure. c. Box plot visualizing the mean descending aorta (non-aneurysmal) diameter following 28d of angiotensin II treatment. Dots indicate the size of individual animals in the cohort. d. Boxplot visualizing the max descending aorta (non-aneurysmal) diameter following 28d of angiotensin II

treatment. e. Mean abdominal aorta (aneurysm) diameter after 28d of angiotensin II treatment. f. Max abdominal aorta (aneurysmal) diameter after 28d of angiotensin II treatment. Error bars represent ± SD. n = 8-11 animals per group. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p < 0.001$ using unpaired t-test.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RomaySupplementaryExcelFile1AnatomicLocationSignature.xlsx
- RomaySupplementaryExcelFile2AnatomicLocationSignatureClassification.xlsx
- RomaySupplementaryExcelFile3vSMCSubclusterSignature.xlsx
- RomaySupplementaryExcelFile4Top10GeneOntology.xlsx
- 2021ROMAYSuppMaterialsFinalNatComms.pdf