UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

Dipartimento di Medicina Molecolare e Traslazionale

DOTTORATO DI RICERCA IN PRECISION MEDICINE

Settore Scientifico

Bio/11

CICLO

XXXV

Single-cell RNA analysis to investigate the role of phenotypic modulation of smooth muscle cells in the onset of atherosclerosis

Tesi di dottorato di

Luca Lambroia

Tutor: **Prof. Leonardo Elia**

Anno accademico 2021-2022

ABSTRACT

Rationale: Atherosclerosis is a chronic inflammatory disease responsible for nearly 50% of all deaths in western countries. It is characterized by the formation of lipid-laden plaques in the arterial wall reducing the lumen, decreasing the elasticity of the walls and causing many cardiovascular diseases. In response to environmental cues, vascular smooth muscle cells (SMCs) can de-differentiate, proliferate and migrate in a process known as phenotypic modulation. Recently, it has been shown that they can transdifferentiate into other cell types which are important for the formation of atherosclerotic plaques and which determine their ability to rupture and induce thrombosis.

Objective: The purpose of this work is to use the most recent technologies in the field of transcriptomics, such as scRNAseq and spatial transcriptomics, to identify new SMCs-derived phenotypes induced by a pro-atherosclerotic environment and to characterize their role in the pathology, obtaining a specific gene signature that will help their identification in patients and consequentially to target them in the context of a precision medicine approach.

Methods and Results: The SMC-deriving cells dissociated from the aortic arch and thoracic artery of a cohort of atheroprone *ApoE* knockout mice and of a control one, were analyzed by scRNAseq with 10x Chromium platform. Then, the bioinformatic analysis of the data was performed identifying a new SMC-derived cell cluster, characterized by the expression of the *Adamtsl1* gene, not yet described in the literature. Furthermore, we found that this cell type is metabolically active, produces ECM, has a proinflammatory phenotype, and is a transitional state between contractile SMCs and other cellular phenotypes whose role within atherosclerotic plaques is already known. Subsequently, the presence of this cluster was also demonstrated in human and murine scRNAseq datasets from another work. Finally, two human atherosclerotic samples were analyzed using the Visium Spatial transcriptomics (10x genomics) highlighting the presence of this SMC phenotype within human plaques.

Conclusions: The identified SMC-derived *Adamtsl1*$^{+}$ cell type could be a good target to block the transdifferentiation of SMCs towards phenotypes involved in plaque formation and to improve the prognosis of patients with atherosclerosis.

TABLE OF CONTENTS

INTRODUCTION

1.1 Blood vessel morphology and composition

The arteries, together with the veins and the heart, form the cardiovascular system which carries out the task of transporting body fluids and nutrients to the various parts of the body. The cardiovascular system is the first functional organ system to appear within the embryo. its development begins during the third week of gestation, when the placenta is unable to meet the requirements of oxygen and nutrients to all tissues of the growing embryo[1].

Arteries, except for the pulmonary ones, carry oxygenated blood from the heart to the rest of the body. As the arteries move away from the heart they branch out into arterioles until they branch into capillaries where gaseous and nutrient exchanges take place between the blood and the peripheral district of the body reached. The wall of the arteries consists of three layers. the innermost layer is called the intima and consists of endothelial cells (ECs) that line the surface of the lumen of the arteries; immediately after we find the media which is made up of connective tissue and smooth muscle cells (SMCs) that allow the contraction and dilation of the arteries in order to keep the blood flow pressure consistent and finally we find the adventitia layer which is mainly composed of elastin, collagen, extracellular matrix (ECM), and fibroblasts[2, 3] (Figure 1).

The aorta is the largest vessel in the human body, it originates from the left ventricle of the heart from where it rises upwards, with a section that takes the name of ascending aorta, then curves backward, forming the aortic arch, and continues into the descending aorta composed by the thoracic aorta and the abdominal aorta.

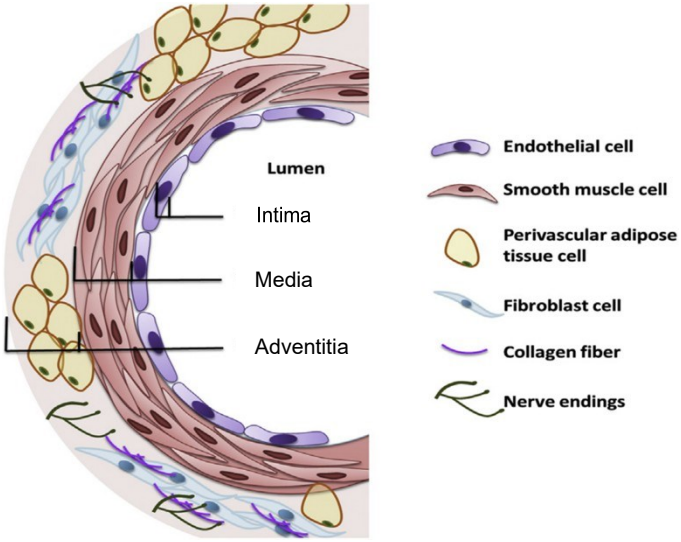


Figure 1. Schematic representation of the composition of the walls of arteries. The wall of the arteries is composed of three concentric tunics called the intima, composed mainly of EC, media, where we find the SMCs and adventitia, composed of collagen, fibroblasts, adipocytes and nerve endings. Adapted from Y. Zhao et al.;Journal of Pharmacological Sciences xxx (2015) 1e12[4].

1.2 Arteriosclerosis and Smooth Muscle Cells

Atherosclerosis is the leading cause of cardiovascular disease (CVD), which includes myocardial infarction (MI), angina and heart failure, and collectively constitute the leading cause of death globally[5, 6, 7]. Atherosclerosis is defined as a multifactorial degenerative disease affecting the medium and large arteries caused by the deposition of fat, white blood cells, calcium salts and fibrotic tissue in the intima layer that give rise to atherosclerotic plaques (or atheromas). Over time, atherosclerotic plaques increase in volume, reducing the elasticity of the arteries and hindering blood flow. Numerous epidemiological studies and animal experiments have established that atherosclerosis is a product of chronic inflammation of the intima where a lesion is created that causes a thickening of the endothelium with the infiltration of other elements at the origin of the plaques. The factors that contribute to the onset of this inflammatory state of the artery walls are diabetes, smoking, hypertension, hyperlipidemia, obesity, and first of all the high concentration of circulating cholesterol in the blood.

Cholesterol is transported in the blood by low-density lipoprotein (LDL). LDLs are composed of a hydrophobic core, made up of triglycerides and esterified cholesterol, coated with phospholipids, free cholesterol and a single apolipoprotein B100 (APOB)[7]. These factors cause an inflammatory state that can affect the walls of arteries, even in distant locations, through the release of soluble inflammatory mediators, such as cytokines.

The alteration of the endothelium occurs early during atherogenesis and manifests itself as follows: increased expression of adhesion molecules, chemokine secretion and leukocyte adhesion, increased LDL oxidation, platelet activation and increased cell permeability[8]. However, the main initiation process of atherogenesis is the subendothelial retention of LDL particles due to the binding between APOB and ECM molecules, in particular proteoglycans[9]. As a consequence of this retention, LDL particles are trapped in the intestine, where they are subject to oxidative modifications caused by enzymes such as myeloperoxidase and lipoxygenase, or by reactive oxygen species (ROS) generated during inflammation. Accumulation of oxidized LDL particles (oxLDL) together with abnormal haemodynamic forces cause monocytes and other immune cells to be recalled to the site of the atherogenic lesion[10,11]. The adhesion of monocytes to the lesion site depends on the very late antigen-4 (ITGA4), which can bind to certain fibronectin isoforms, and on the vascular cell adhesion molecule-1 (VCAM1), a molecule of the immunoglobulin family highly expressed in endothelial cells near atherosclerotic lesions[11]. Subsequently, the monocytes arriving in the intima are stimulated by the colony stimulating factor of macrophages (M-CSF), released by the activated endothelial cells, to differentiate into macrophages (Mϕs). Monocyte-derived Mϕ upregulates their scavenger receptors which can absorb oxLDLs. These cells engulf lipids and become foam cells, the hallmark of atherosclerotic lesions. Parallel to monocytes, T lymphocytes are recruited by similar mechanisms. T lymphocytes are not as abundant as Mϕs (with a Mϕs / T ratio between approximately 4:1 and 10:1 in human lesions), however, these have also been observed to contribute to the growth of atherogenic lesions and disease development.

Vascular smooth muscle cells (VSMCs) are highly specialized and differentiated cells in adult animals that have a very low rate of proliferation and low synthetic activity. The inflammatory cells present in the atherogenic lesion, together with platelets and EC, release mediators that cause the SMCs to become de-differentiated and lose their contractility. This phenomenon goes under the name of phenotypic switching and these SMCs become proliferating, metabolically active and migrate from the media to the intima where they contribute to the formation of the fibrous cap of the atherosclerotic plaques producing ECM[12] (Figure 2). Contractile SMCs have the function of regulating blood vessel diameter (vasodilation and vasoconstriction) and blood flow and are characterized by the spindle shape and the expression of contractile proteins such as alpha smooth muscle actin-2 (ACTA2), myosin-11 (MYH11), transgeline (SM22) and calponin (CNN1). Metabolically active SMCs are characterized by a rhomboid shape and the production of growth factors, such as platelet-derived growth factors (PDGF), endothelin-1 (EDN1), thrombin, fibroblast growth factor (FGF1) and interleukin-1 beta (IL1B)[3].

SMCs undergoing phenotypic switching therefore form a fibrous cap in the intima that wraps

around a central region, called the lipid core, composed of foam cells, cholesterol and other lipids, thus forming what is called atheroma. Over time, apoptotic cells and cellular debris also begin to accumulate in the plaque due to a reduced capacity for efferocytosis by phagocytes. Pro-inflammatory phenotype M1 may reduce the number of beneficial Mϕ, thus inducing a relative deficit of those Mϕ known to have a high phagocytic capacity[13]. Furthermore, ROS can impair the phagocytic abilities of Mϕs and dendritic cells. As a result, foam cells accumulate to promote lesion expansion and apoptotic tissue undergoes secondary necrosis to accelerate vascular inflammation and lesion instability. Many atherosclerotic plaques are prone to calcification, a process that consists of the accumulation of calcium minerals resulting from the dysregulation between deposition and clearance[14]. The progression of the lesion can occur silently for many decades. Eventually, the growing atherosclerotic plaque begins to invade the arterial lumen and can lead to the formation of flow-restricting lesions. These plaques can then face complications: for example, infiltrated immune cells help create a pro-inflammatory environment that can destabilize the plaque and make it prone to rupture[15, 16, 17]. When the material contained in the plaques comes into contact with the blood it is able to trigger coagulation and the formation of a thrombus which can lead to heart attack, ischemia, stroke, etc.

Recently it has been shown that SMCs in arteriosclerotic plaques can also undergo transdifferentiation phenomena in other cell types such as a macrophage-like type[18] that could contribute to the formation of foam cells (Figure 2) responsible for the formation of the lipid core into the plaques or to contribute creating a pro-inflammatory environment that destabilizes them, facilitating the rupture. On the other hand, there are other possible cell types in which SMCs can transdifferentiate such as fibromyocytes, which contribute to the formation of the fibrous cap making it more resistant to rupture[19]. Therefore SMCs and the cell types that derive from them seem to have a leading role both in the onset of arteriosclerosis and in its clinical impact on patients.

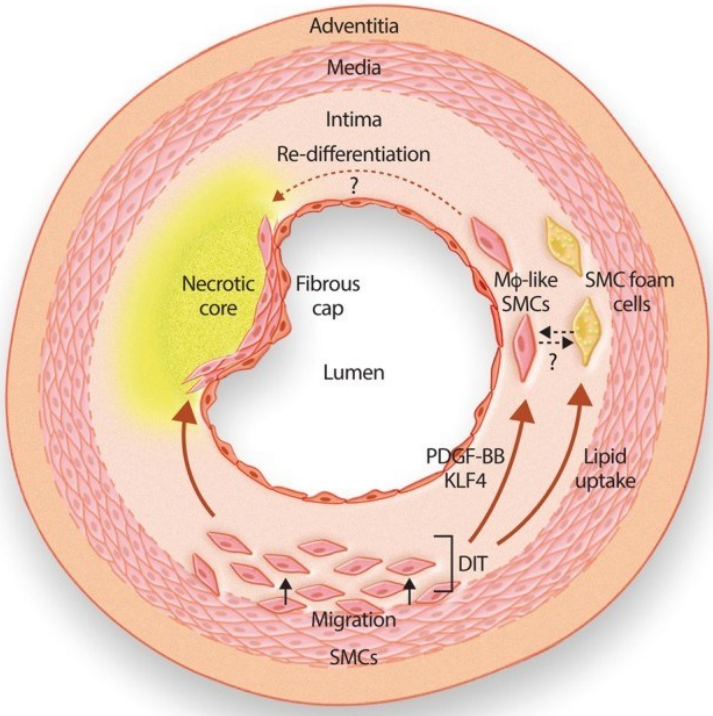


Figure 2. Phenotypic switching and transdifferentiation of SMCs in the arterial wall

VSMCs play a role of primary importance in the formation of plaques during atherosclerosis: they can undergo phenotypic switching, become proliferating and produce ECM constituting the external cap of the atheroma as well as undergoing transdifferentiation into macrophages, aggravating the inflammatory state and plaque development and increasing the foam cell population of the lipid core. Figure from Allahverdian S, et al. Cardiovasc Res. 2018 Mar 15;114(4):540-550[20].

Although genetic factors also contribute to the development of atherosclerosis, it has been observed that adopting a healthy lifestyle can prevent the development of the disease. From a therapeutic point of view, the introduction of statins has revolutionized the prevention and treatment of atherosclerosis. Statins are drugs that limit the synthesis of endogenous cholesterol by inhibiting the 3-hydroxy-3 methylglutaryl-CoA reductase (HMGCR) enzyme[7]. Although statins are a milestone in the treatment of atherosclerosis, identifying a SMC-derived cell type underlying plaque formation could offer a valuable therapeutic target to be used in precision medicine to further improve patient prognosis.

1.3 Transcriptomic

The 2000s saw the advent of Next Generation Sequencing which introduced the possibility of analyzing a huge amount of sequences in parallel quickly and at a low cost. This new technology has profoundly revolutionized the way of doing science in the clinical field and beyond, allowing

an omic approach to the study of genomes, mutations, proteins and transcripts. The study of all the sequences of the transcripts, therefore of all the mRNA produced by the transcription activity of the RNA polymerase from the DNA, is defined as transcriptomics.

The basic steps of a transcriptomics analysis involve the extraction of the mRNA from a sample, its conversion into cDNA by reverse transcription, the amplification of the cDNA and the preparation of libraries for sequencing, maximum sequencing and generation of a huge number of sequences called "reads", the use of bioinformatics techniques to align the reads obtained on the genome sequence of the reference organism in order to see from which genes the reads originate and to count the reads mapped to each gene. At the end of these steps, a matrix of raw counts is obtained which, with appropriate normalization, gives an inference of the level of expression of each gene in the sample.

Thanks to transcriptomics it is possible to identify how the gene expression of individual organs and organisms changes in response to certain stimuli, to reconstruct the interaction networks between genes, to study the isoforms of transcripts and much more. However, until the last few years, the study of transcriptomics involved starting from a tissue or cell culture and extracting the entire mRNA. This approach did not allow us to know which cell types were expressing the genes of interest but only returned what was the transcriptional profile of the average cell in the tissue and masked any cell subtypes present in a small number that eventually responded to the stimulus under examination during the experiment. So the average signal of a heterogeneous population of cells does not necessarily represent the state of each component of the tissue. Moreover, it is known that even the smallest population of cells can have a huge impact on personalized medicine. For these reasons, a single-cell approach has been introduced in modern analysis.

1.4 Single-cell RNA sequencing

To increase the resolution of NGS transcriptomic analyzes at the single-cell level, the scRNA-seq technique was introduced: this technique consists of the analysis of the transcriptome of every single cell obtained from a tissue after dissociation thus allowing to investigate tissue heterogeneity and the behavior and relationships of cell subtypes present.

This technology was first published in Nature Methods in 2009, but it became more popular after 2014 when sequencing costs decreased, and new more validated protocols were designed.

All platforms available on the market to perform a scRNA-seq analysis are mainly based on four types (Figure 3) that differ in the mode of cell capture[21]: Fluorescence-Activated Cell Sorting (FACS) based platforms which exploit antibodies and flow cytometry techniques; droplet-based

and microwell-based platforms where the cells are separated by microfluidic techniques and trapped, respectively, in oil droplets or in microwells and microfluidics platforms like Fluidigm C1 which separates the cells making them circulate along a fluidic path. Droplet-based platforms capture a large number of cells, they are the ones with the lowest cost but have a low number of genes detected due to the low depth of sequencing; platforms with microwells achieve a much greater sequencing depth than those droplets but capture fewer cells; microfluidic platforms, on the other hand, have a high number of cell capture and sequencing depth, however, their cost is a limiting factor in their use.

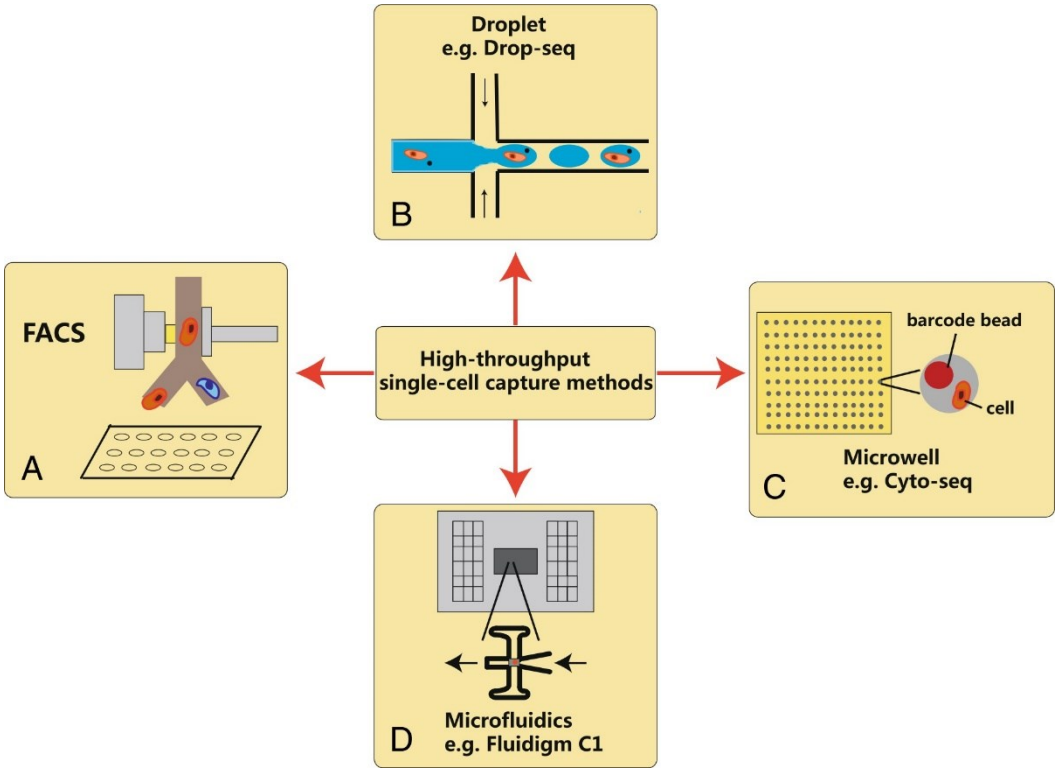


Figure 3. Main techniques of cell capture on which the platforms for scRNA-seq are based. The different platforms for the single-cell differ in the method of cell capture. The main isolation strategies are shown here: A) the FACS platform isolates the cells using flow cytometry; B) droplet based methods; C) methods based on microwells; D) methods based on microfluidics. Figure from Ye, F., et al. . J Hematol Oncol 10, 27 (2017)[22].

Nowadays, single-cell sequencing is used to study transcriptional dynamics by the identification of novel and rare cell types or unknown cellular states. In addition, these techniques allow the measurement at the single-cell level of chromatin accessibility, methylation, mutation and CNV (scDNA-seq). Currently, there are different protocols that can be used: SMART-seq2, CELL-seq, Drop-seq, etc. There are also commercial platforms available, including the Wafergen ICELL8, the Fluidigm C1 and the 10X Genomics Chromium.

Other works have already used the single-cell method to study the phenotypes assumed by SMCs during atherosclerosis. However, in these works, as in Wirka, R.C. et al.[19] and Huize Pan et al.[23], scRNAseq analysis was applied on the SMCs of *ApoE* KO mice fed with the WD diet alone, ignoring the possible effects of diet and genotype in influencing the transdifferentiation of SMCs towards various phenotypes they may assume. In this work, we apply the scRNAseq analysis only on tdT^{+} cells of the previously described experimental groups providing a complete and in-depth overview of all the phenotypes/cell-state assumed by SMCs and elucidating whether the $ApoE^{-/-}$ genotype predisposes to having a certain type of SMC-derived cells with respect to the others and the impact of diet on their distribution.

1.5 Bioinformatic analysis of scRNAseq data

1.5.1 Pre-processing

The scRNAseq technology has allowed transcriptomics to go down to single cell resolution, however, this method is characterized by low capture efficiency and high dropouts and the data obtained are much noisier and more variable than those obtained with a bulk RNAseq[24]. For these reasons the quality control (QC) of the obtained data is a key step that needs a lot of attention during the in silico analysis of this type of data. A schematic overview of the main steps of a scRNAseq data analysis is shown in Figure 4.

In general, once we proceeded with the alignment of the reads produced by the sequencing versus the reference genome and the assignment of the mRNA counts obtained to the corresponding cell, we generated a count matrix where in each row there is a gene and in each column a cell barcode. QC should be performed on this matrix to remove low quality cells in order to obtain clusters of cells with a transcriptional profile with as little noise as possible. To date, there is no standard pipeline to perform QC however, the main quality indices used are the number of genes found, the number of total counts and the fraction of reads that map to mitochondrial genes[24, 25]. These indices allow us to affirm, once a threshold based on the distribution of these values within the sample has been chosen, if a cell has reached a sufficient sequencing depth for the analysis or, in the case of droplet-based methods, if we are in front of an empty droplet; the fraction of mitochondrial counts also indicates us if the cell was damaged at the time of capture since in such case cytoplasmic mRNAs will escape from the cell while the mitochondrial mRNA will not and therefore the number of mitochondrial gene counts will be more high relatively to nuclear gene counts. Some protocols also use intrinsic spike-ins to estimate cell quality. Sometimes, it can happen that the cell isolation method does not work perfectly and therefore more than one cell ends up in the same micro-well or in the same droplet (doublet). This error could then lead to obtaining "dirty" gene signatures as

we may have assigned two different cell types to the same cell barcode and therefore the previous quality indices can also be used to identify a threshold beyond which to consider the cell barcode as a sequencing product of two cells.

Another typical QC practice is to remove the genes expressed by a few cells, these are poorly informative of cellular heterogeneity and could cause noise in the identification of cell clusters.

Following the QC, we proceed with the normalization of the count matrix, in order to have a transcriptional profile of each cell comparable with the others, and with the possible removal of the batch effect or the integration of more datasets. ScRNAseq experiments are often conducted on multiple samples analyzed at different times, prepared with different protocols and by different operators. These factors introduce a source of non-biological variation in the dataset that goes by the name of "batch effect". This variation can lead to mistakenly consider two cells completely identical from a biological point of view to be different and therefore it is necessary to eliminate this bias with suitable algorithms that correct the count matrix, making the data coming from several samples comparable. Furthermore, you may want to combine samples from different experiments, from different organisms or prepared from different platforms: in such case, a batch correction algorithm would risk eliminating the biological variance really existing between the samples and therefore it is appropriate to use integration algorithms such as Canonical Correlation Analysis (CCA)[26] or Mutual Nearest Neighbors (MNN)[27]. While batch effects are typically corrected using linear methods, non-linear approaches are used for data integration. A further correction of the counts, that is often done, is the retrieval of the expression in which algorithms are applied that are able to infer a dropout event, which are a common problem in single-cell RNA sequencing technologies, of a gene and impute a plausibly count for that gene.

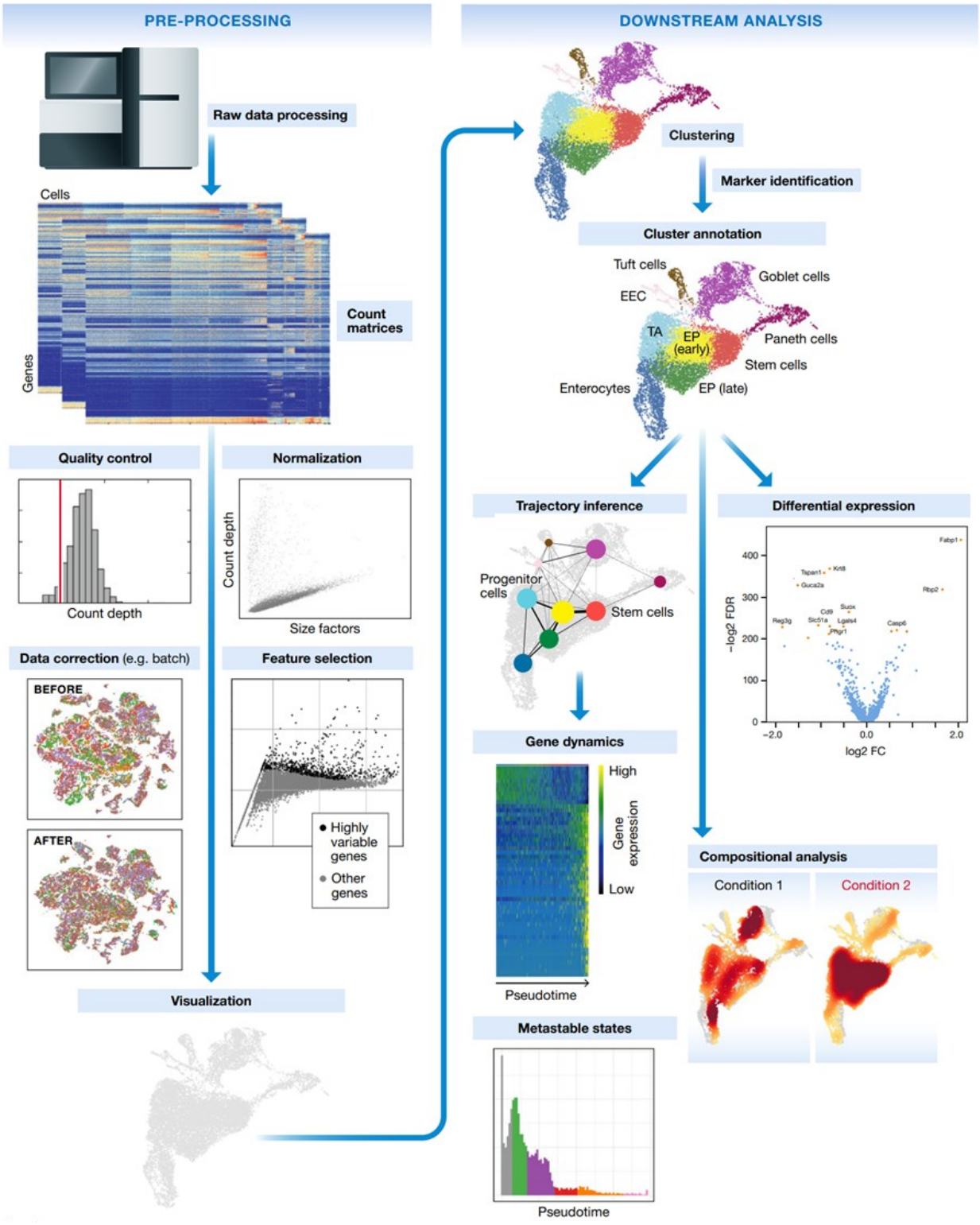


Figure 4. Schematic representation of a typical single-cell RNA-seq analysis workflow. The analysis of scRNAseq data consists mainly of two parts: a part of data pre-processing in which the alignment to the reference transcriptome is carried out, obtaining the matrices of the counts for each cell, the QC, the normalization and correction of the counts; and a second part, called downstream analysis, in which the biological meaning of the data is extrapolated by identifying the cell clusters present, the trajectories, the DEGs and the compositional analysis of the samples. Figure from Luecken MD, Theis FJ. Mol Syst Biol. 2019 Jun 19;15(6):e8746[25].

Some genomes, such as the human one, have tens of thousands of genes and not all genes are important to classify the cells into meaningful clusters based on their expression profiles. Thus, it is necessary to use, in terms of the data science field, a feature selection and to identify the most relevant features to reduce the number of dimensions used in downstream analysis and to reduce the complexity of the data. This allows subsequent analyzes to be computationally less expensive and easier to handle. Typically we select between 1000 and 5000 high variable genes (HVGs).

In order to view the characteristics of a dataset with many features such as those deriving from a scRNAseq analysis, it is necessary to execute dimensionality reduction methods that project the data into a lower dimensional space by optimally preserving some key properties of the original data. Once the size reduction has been performed, it is possible to project the dataset into a two-dimensional space where the cells are grouped into subpopulations. The methods that can be used to do this can be linear, such as the main component analyis (PCA), or non-linear, such as the t-distributed stochastic neighbor embedding (t-SNE) and the Uniform Approximation and Projection method (UMAP).

PCA is a linear approach that generates reduced dimensions by maximizing the captured residual variance in each further dimension. While PCA does not capture the structure of the data in few dimensions as well as non-linear methods, it is the basis of many currently available analysis tools for clustering or trajectory inference. Indeed, PCA is commonly used as a pre-processing step for non-linear dimensionality reduction methods[25]. The two main methods of non-linear dimensional reduction used are the t-SNE and the UMAP. The t-SNE uses Gaussian joint probability measures to estimate the pairwise distances between the data points in the original dimension and this allows to capture local similarity at the expense of global structure; it is a computationally expensive and slow method for large datasets and may also exaggerate differences between cell populations and overlook potential connections between these populations. UMAP uses combinatorial topological modeling to capture the data and applies Riemannian metrics to enforce the uniformity in the distribution. Fuzzy logic is also applied to the graph to adjust the probability distance if the radius grows. Once the graphs are built then optimization techniques are applied to make the embedded space graph very similar to the original space one. The UMAP method provides the fastest run times, the highest reproducibility, and the most meaningful organization of cell clusters than other dimensionality reduction approaches[24].

1.5.2. Downstream analysis

After pre-processing, the so-called downstream analysis is performed: i.e. the biological insights extraction and description of the underlying biological system.

The first step of the downstream analysis is the organization of the cells in clusters. Clustering is a classical unsupervised machine learning problem, based directly on a distance matrix. Cells are assigned to clusters by minimizing intracluster distances or finding dense regions in the reduced expression space. There are various algorithms for clustering scRNAseq data and they are all based on grouping cells based on a similarity score obtained on the basis of distance indices.

Then, once we have obtained the clusters, we must proceed with their annotation which can be done with an automatic annotator, which uses machine learning to compare the expression profile of the cells with the data on which the classifier was built, or manually identifying the expressed marker genes that all the cells of the cluster have in common and comparing themselves with the literature.

Once the clusters have been identified, the other downstream analyzes usually performed are the identification of differentially expressed genes (DEGs) between two or more clusters (or the same under different conditions) by means of a statistical test, the comparison between the proportions of the cell types present between two samples and the analysis of the trajectories.

When you perform a scRNAseq analysis what you get is a picture of a cell population in which the cells are not synchronous with each other and therefore they will be in different stages of their differentiation. Using algorithms for the analysis of trajectories it is possible to order the cells along a scale of pseudotimes that reflects their differentiation path or in any case their transcriptional similarity.

2. MATERIALS AND METHODS

2.1 Mouse strain

To identify the SMC-derived phenotypes that formed following an atheroprone insult we used a mouse model knock-out (KO) for the *ApoE* gene (C56BL/6 *ApoE*$^{-/-}$), an apolipoprotein that if mutated leads to hypercholesterolemia, and a group of wild-type (Wt) mice (C56BL/6 *ApoE*$^{+/+}$) as a control. Both groups of mice were engineered with a well-characterized BAC transgene expressing a tamoxifen-inducible Cre recombinase driven by the Myh11 promoter specific for SMCs. This BAC transgene contains a floxed tandem dimer tomato fluorescent reporter so that specific SMC-lineage cells can be traced. As the Cre-expressing BAC was integrated into the Y chromosome, all lineage-tracing mice in the study were male.

Five doses of tamoxifen, at 0.2 mg g^{-1} bodyweight, were administered by intraperitoneal injection at 7 weeks of age, with each dose separated by 24 h.

At 8 weeks of age, the two groups (KO and Wt) of mice were then fed two types of diets: a high-cholesterol diet that mimics the Western diet (WD) and a chow diet (CD). The mice were then maintained in these alimentary regimes for 16 weeks and finally, we obtained 4 experimental groups: CD Wt, CD KO, WD Wt and WD KO mice as shown in Figure 5A.

2.2 Mouse aortic root/ascending aorta and thoracic artery cell dissociation

Immediately after sacrifice, mice were perfused with phosphate buffered saline (PBS). The aortic root and ascending aorta were excised, up to the level of the brachiocephalic artery. Tissue was washed three times in PBS, placed into an enzymatic dissociation cocktail (2 U ml^{-1}Liberase (5401127001; Sigma–Aldrich) and 2 U ml^{-1} elastase (LS002279; Worthington) in Hank's Balanced Salt Solution (HBSS)) and minced. After incubation at 37 °C for 1 h, the cell suspension was strained and then pelleted by centrifugation at 500*g* for 5 min. The enzyme solution was then discarded, and cells were resuspended in fresh HBSS. For the SMC^{lin} genotype, five mice were used at baseline, and five mice were used at 16 weeks of disease. For the SMC$^{lin\text{-}KO}$ genotype (apoE$^{-/-}$), five mice were used at baseline (chow diet) and five mice were used at 16 weeks.

2.3 FACS of mouse aortic root/ascending aorta cells

Cells were sorted on a BD Aria II instrument. An overview of the cell-sorting process is illustrated in Figure 3B. Cells were gated on forward/side scatter parameters to exclude small debris and then gated on forward scatter height versus forward scatter area to exclude obvious doublet events.

Events passing these criteria were then sorted into one of two 1.5 ml Eppendorf tubes based on tdT fluorescence levels. *tdT*$^{+}$ cells (considered to be of SMC lineage), while *tdT*$^{-}$ cells were not collected (Figure 5B).

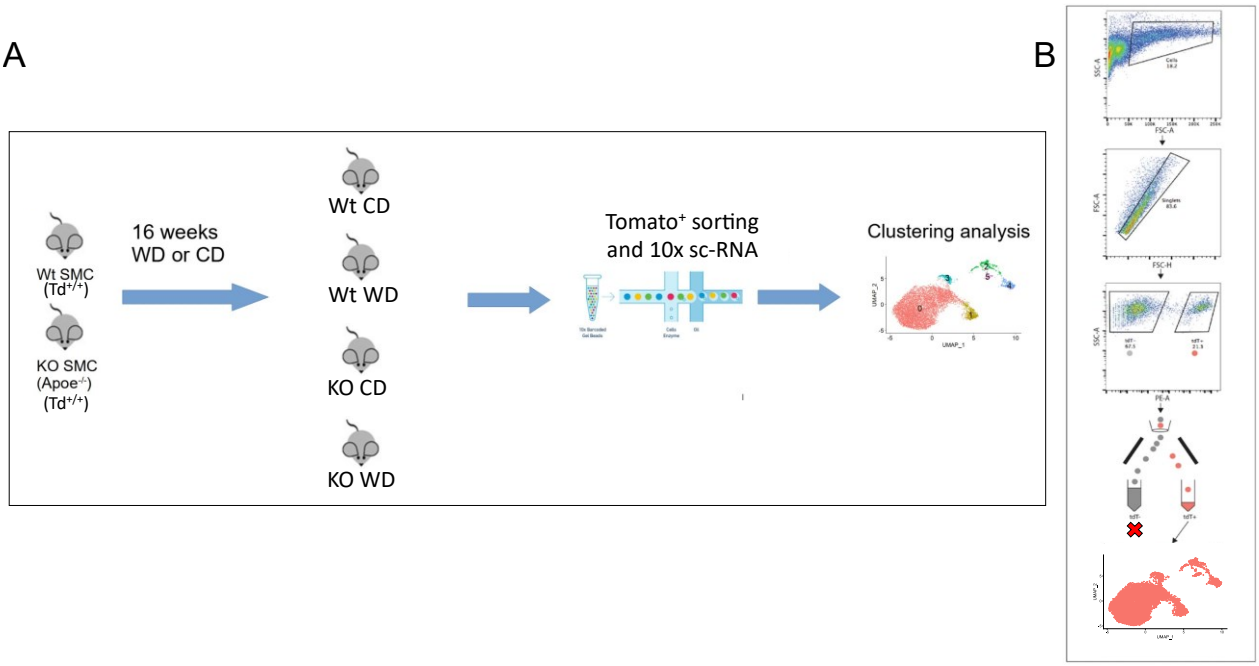


Figure 5: A, research workflow: the cells analyzed with scRNAseq were dissociated from the four experimental groups obtained with the two diets and finally the bioinformatic analysis of the data was carried out. B, sorting by FACS of tdT $^{+}$ cells. Figure B adapted from Wirka, et al. Nature medicine vol. 25,8 (2019): 1280-1289[19].

2.4 Single-cell 10× Genomics Chromium system

10x Genomics single-cell technology is a powerful tool that allows the identification of transcriptomes down to ~1% abundance. It's a droplet-based technique that enable processing of tens of thousands of cells in a single experiment. This technology breaks down reactions into nanoliter-scale droplets containing uniquely barcoded beads called GEMs (Gel Bead-In EMulsions). This core technology can be used to partition single-cells, nuclei, or high molecular weight gDNA to prepare next generation sequencing libraries in parallel. To achieve single-cell resolution, the cells are delivered at a limiting dilution: in this way, ~90- 99% of generated GEMs will contain no cell, while the remaining should contain a single-cell. Each functional GEM contains a single cell, a single Gel Bead, and RT reagents.

The surface of every microparticle bead has attached an oligo dT sequence, a unique molecular identifier (UMI), specific for each molecule of mRNA in order to allow the removal of artifacts due to cDNA amplification, a cell barcode, which identifies the cell from which the transcriptome

comes, and a PCR primer (Figure 6). Within each GEM reaction vesicle, a single cell is lysed, the Gel Bead is dissolved to free the identically barcoded RT oligonucleotides into solution, and reverse transcription of polyadenylated mRNA occurs.

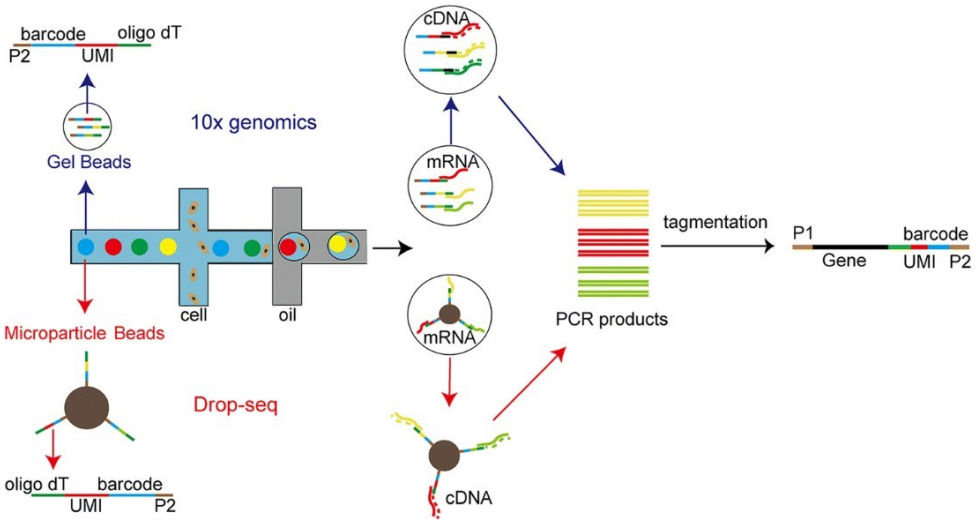


Figure 6. 10x Gel Bead-In EMulsions of Genomics' Chromium technology. Schematic representation of cell capture and library preparation with the 10x Genomics Chromium platform. in each GEM a cell is captured from which the mRNA is extracted and captured by probes linked to beads. the probes ensure that each cDNA that is produced by RT has a UMI and a cell barcode. Adapted from: Zeng Z et al., 2018[28].

2.5 scRNAseq cells' preparation, library preparation and sequencing

All single-cell capture and library preparation was performed and cells were loaded into a 10x Genomics microfluidics chip and encapsulated with barcoded oligo-dT-containing gel beads using the 10 Genomics Chromium controller according to the manufacturer's instructions (ChromiumTM Single Cell 3' Reagent Kits v2-rev). The libraries obtained were then evaluated with the TapeStation 4200 (Agilent, LifeTechnologies), they were then sequenced on the Illumina NextSeq550 platform; on average at least 40.000 reads per single cell were obtained.

2.6 10x Visium Spatial Transcriptomics technology

In recent years, the Visium Spatial Transcriptomics platform (10x Genomics) has been introduced which allows to map the transcriptome deriving from spots of 55µm in diameter and 100µm away from each other, in which the mRNAs of the cells of a section of tissue are captured, above the image of the section itself. This technology is based on a slide with up to 4 capture areas of 6.5mm x 6.5mm and contains ~ 5000 barcoded spots (Figure 7). In each spot there are probes that contain an oligo dT sequence, the UMI and the spatial barcode, which allows to trace the position of the spot in which the capture took place. Then the tissue section placed in the capture area is

permeabilized to release the mRNA from the cells, the mRNA is captured, retro-transcribed and the libraries prepared.

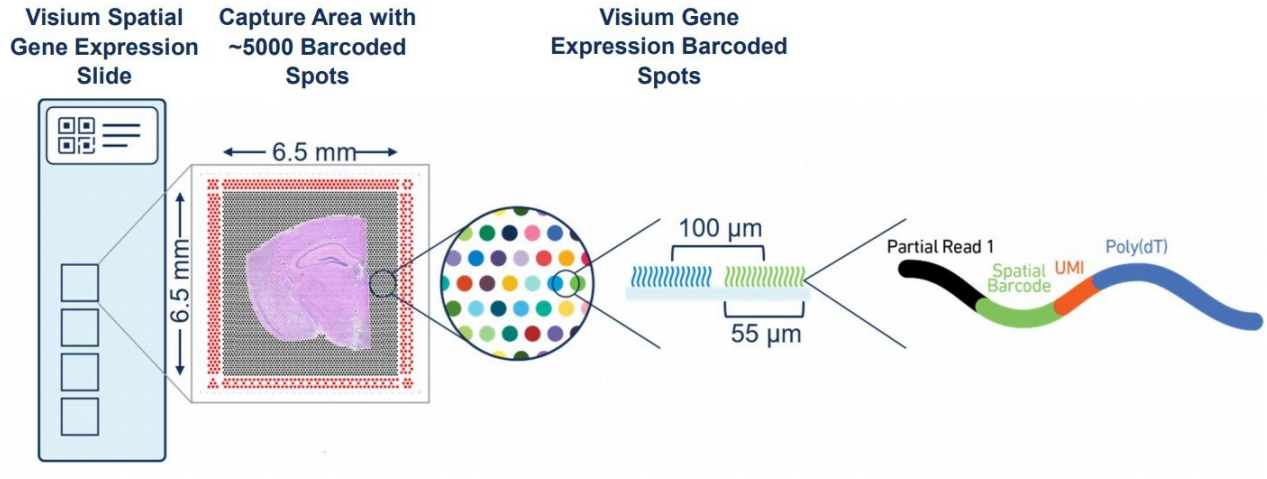


Figure 7. The Visium Spatial Gene Expression Slide (*https://www.10xgenomics.com/*)

2.7 Visium spatial transcriptomics tissue' preparation, library preparation and sequencing

We select 4 human plaque samples based on RIN >7. We mounted the selected OCT tissue blocks on the specimen stage aligned as previously based on cardinal signs and cut serial cryosections which have been placed onto Tissue Preparation Guide (CG000240). For the following steps we acted in accordance with Visium Spatial Gene Expression protocol (CG000239) with minor modifications. Visium Spatial Gene Expression slide, containing the four CRLM sections, was incubated following the Methanol Fixation + H&E Staining guide (CG000160). Finally, the slide was dried at 37°C for 5 minutes and mounted with Mounting Medium (22.5 µl RNase inhibitor, 7.5 µl Nuclease free water and 170 µl Glycerol). The slide was imaged with AxioScan, in brightfield with 20x magnification. After the acquisition, the Spatial Gene Expression slides was immersed in a 3x SSC solution in ultrapure water for 20 minutes, to remove the coverslip. 25 ng of the amplified cDNA were then used for each sample to construct Illumina sequencing libraries following the manufacturer's protocol (CG000239) using 17 cycles and Dual Index Plate TT set A for their generation. Final libraries were checked by Qubit dsDNA HS Assay Kit (Q32854) and the Agilent 4200 Tape Station system using the High Sensitivity D5000 ScreenTape (5067-5592) analysis kit (Agilent, Santa Clara, CA, USA). Sequencing was performed on the NextSeq550 Illumina sequencing platform.

2.8 Analysis of single-cell RNA sequencing data

2.8.1 QC, clustering and DEGs identification.

The reads obtained from the sequencing of the aortic cells were individually mapped for each murine model, to the reference genome mm10, with the addition of the sequence for the *Tomato* gene, using the CellRanger Software v3.1.0 (10x Genomics). The matrixes with the obtained raw counts were merged together and were analyzed and filtered using R v4.0.5 and the scater v1.24.0[29] and DropletUtils v1.14.2[30, 31] packages. We used the emptyDrops function to remove all the droplets with a false discovery rate greater than 0.05 and then we filtered out all cells with no count for *Tomato* gene. Then we removed all cells with a number of UMIs, detected genes or a percentage of mapped reads onto mitochondrial and ribosomal genes that were considered outliers compared to the median absolute deviation. All genes that did not have a minimum of one count in at least the 5% of the cells of the entire dataset were removed. Cells that were imputed arising from doublets through the computeDoubletDensity function were excluded. After the quality filter, cells were analyzed using the Seurat version 4.0.1 package[32]. The gene counts of each cell were normalized by dividing them to the library size of their cell; the resulting expression values were then multiplied by 10,000 and log transformed. Subsequent analyses to obtain the cell clusters were conducted using only the top 2000 most highly variable genes in the dataset excluding *Apoe* and *Tomato*. Fast-mutual nearest neighbors' correction, performed with batchelor package v1.12.3[33] was used for batch correction and dimensionality reduction, followed by clustering in MNN corrected principal component analysis space using a graph-based clustering approach. Subsequentially, Uniform Manifold Approximation and Projection (UMAP) was used for two-dimensional visualization of the resulting clusters. Genes differentially expressed among all cell clusters derived from SMCs were obtained with the Seurat function FindAllMarkers with logfc.threshold = 0.3 and min.pct = 0.2 parameters and using the MAST[34] statistical test. Only DEGs with p_val_adj <0.05 were kept.

2.8.2 Pathways enrichment analysis

Given a list of genes and a database of metabolic or functional pathways, it is possible to perform a pathway enrichment analysis through which a list of pathways is obtained whose genes are significantly enriched in the fixed list of genes of interest, as compared to all genes in the genome[35]. The P value of the enrichment of a pathway is computed using a Fisher's exact test and multiple-test correction is applied. We performed the pathway enrichment analysis by the online tool ENRICH[36, 37, 38] with the DEGs found by individually comparing the derived SMC clusters and the contractile SMCs with the previous parameters.

2.8.3 Inference of the activity of the regulons

A recently introduced method to confidently define cellular identity and how that identity is developed and maintained is Single-Cell Regulatory Network Inference and Clustering (SCENIC)[39]. This tool allows to link gene expression to the gene regulation networks on which the cell-state largely depends. Thus, SCENIC allows the identification of the activated regulons in a cell. In molecular genetics, a regulon is a group of genes that are regulated as a unit, generally controlled by the same regulatory gene that expresses a protein, called transcription factor (TF), acting as a repressor or activator. Briefly, the SCENIC pipeline consists of three steps: first, candidate regulatory modules are inferred from coexpression patterns between genes; second, coexpression modules are refined by the elimination of indirect targets using TF motif information.; third, the activity of these discovered regulons is measured in each individual cell[40]. We inferred the activity of the active regulons in the cells and calculated the Regulon specificity Score (RSS), a score of how much that regulon is activated only in the cells belonging to that cluster, with SCENIC v1.2.4.

2.8.4 Trajectory analysis

We performed the trajectory analysis by assigning a pseudotime to each cell using the slinghot package v2.4.0[41] on the umap reduction of the dataset. Slingshot consists of two main stages: 1) the inference of the key elements of the global lineage structure, the number of lineages and where they branch, using a cluster-based minimum spanning tree. This allows us to identify novel lineages while also accommodating the use of domain-specific knowledge to supervise parts of the tree (e.g., terminal cellular states); and 2) the inference of pseudotime variables for cells along each lineage. For the second stage, slingshot tool uses a method called simultaneous principal curves to fit smooth branching curves to these lineages, thereby translating the knowledge of global lineage structure into stable estimates of the underlying cell-level pseudotime variable for each lineage[40].

2.8.5 Integration with murine *tdT*$^{-}$ cells and with atherosclerotic patient scRNAseq data

In our analysis we have identified a cluster of cells characterized by the expression of the *Adamtsl1* gene and therefore called SMC *Adamtsl1*$^{+}$. To verify the presence of this cluster also in the mouse model of another work and in humans, we downloaded the human and mouse scRNAseq fastqs from Huize Pan et al.[23] and reanalyzed as we did for our dataset with the exception that the human

reads were mapped to the GRCh38 reference genome and not filtered for the *Tomato* gene. Once the datasets were analyzed independently, we proceeded to integrate them with our data to verify that our *Adamtsl1*$^{+}$ cluster really corresponded to theirs. The integration of the data of our KO WD model and the same model at 16 weeks and the human samples from other work was carried out by finding the correlation anchors and the Seurat IntegrateData function. For human data we transformed the matrix of human genes with murine orthologs by the biomaRt[42] package before integration.

2.8.6 Study of the communication between *Adamtsl1*$^{+}$ cells and macrophages

Numerous studies have shown that M1 Mϕs promote the inflammatory state of the plaques, allowing their growth and the onset of complications[11] and have been seen to be responsible for the decreased efferocytosis with consequent necrosis of the plaque[13]. For this reason, we wanted to test whether the new SMC *Adamtsl1*$^{+}$ phenotype was able to interact with Mϕ promoting their polarization in M1. For the study of the interaction between the *Adamtsl1*$^{+}$ cells and the Mϕ we used NicheNet algorithm implemented in the nichenetr package v1.0.0[43]. NicheNet is based on a prior model built on the basis of everything that is already known about that a ligand regulates the expression of a target gene. This previous model is generated by integrating data on ligand-receptor, signal transduction, and gene regulation interactions from multiple databases into weighted networks (Figure 8). These contain protein-protein interactions that cover the signaling pathways from ligands to downstream transcriptional regulators and gene regulatory interactions between transcriptional regulators and target genes. Finally, a regulatory potential score is assigned among all pairs of ligands and target genes based on whether the target gene regulators are downstream of the ligand signaling network using network propagation methods over integrated networks to propagate the signal from a ligand, on receptors, signaling proteins and transcriptional regulators, to terminate at target genes.

When applying NicheNet to study communication between cells in a scRNAseq experiment, this prior model of ligand-target regulatory potential is combined with cell gene expression data to prioritize ligands according to how well their prior target gene predictions correspond to the observed gene expression changes in receiver cells resulting from communication with sender cells. Then, to predict active ligand-target linkages, NicheNet looks for genes that are affected in receiver cells and that are possibly regulated by these priorityzed ligands, as indicated by a high regulatory potential score[43].

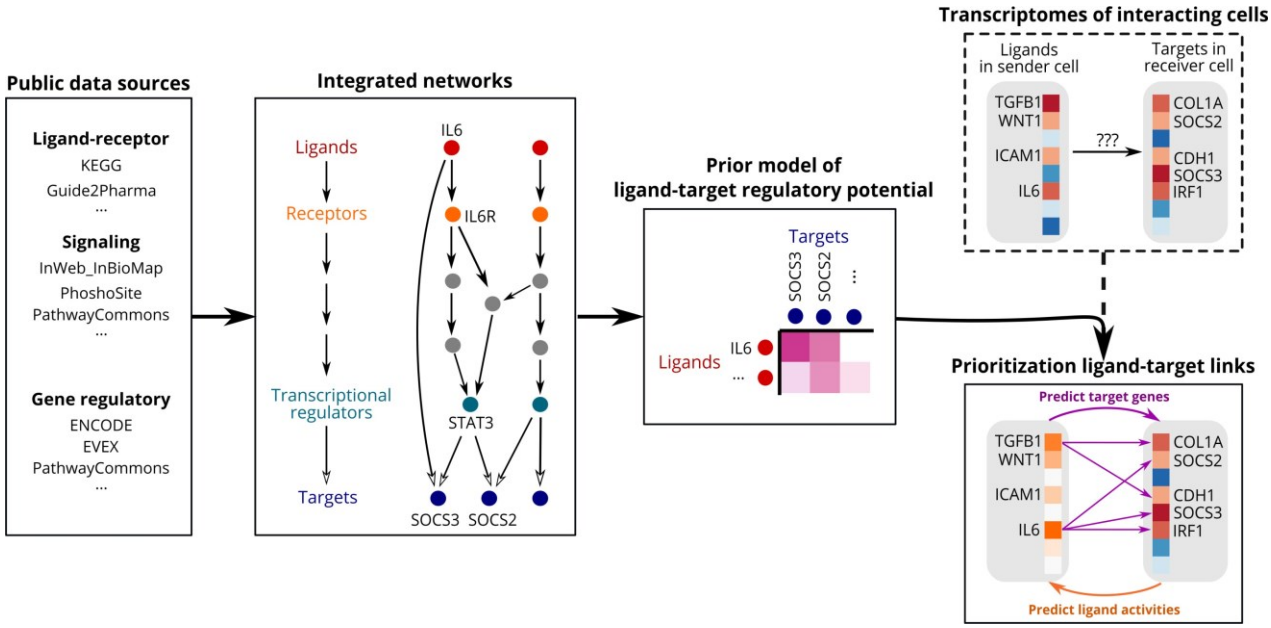


Figure 8. NicheNet workflow scheme. NicheNet uses a prior model obtained by integrating ligand-receptor data, signal transduction, and gene regulation interactions to build a weighted network where a score is assigned to each ligand-target pair based on whether the genes that regulate the target are downstream of the ligand signaling network. this prior model is then integrated with cell expression data to prioritize ligands produced by a sender cell population based on how they predict gene expression changes in receiver cells. the actual interactions between the targets and the ligands are subsequently verified on the basis of the regulatory potential score assigned to the pair in the prior model. Figure from Browaeys, R., et al. Nat Methods (2019)[43].

For the analysis we indicated as target genes in the Mϕ a signature of genes for M1 polarization common to in vitro and in vivo experiments coming from Orecchioni M., et al.[44] and the possible ligands produced by *Adamtsl1*$^{+}$ cells were selected from the DEGs of the signature obtained by comparing all the clusters of our dataset integrated with all four groups of mice.

2.9 Visium Spatial Gene Expression data analysis

2.9.1 QC of spatial transcriptomic data

Fastq files were generated with 10x Genomics software Space Ranger v1.2.2 with spaceranger mkfastq function. The reads were aligned to the human reference transcriptome GRCh38 with spaceranger count command. The obtained count matrices were processed using R v4.0.5 and Seurat package v4.0.1. After the removal of the low quality spots (number of counts < 100 and percentage of mitochondrial genes > 20) we proceeded to normalize the data with the SCTransform function by regressing for the n_Feature_spatial variable and for the percentage of mitochondrial genes.

2.9.2 Spatial localization of *Adamtsl1*$^{+}$ cells in human atherosclerotic plaque

The spots present in the visium spatial transcriptomic slide have a diameter of 55µm while some cell types, such as macrophages, are smaller. Therefore, more cells can be found in a spot and this will not allow us to have a resolution of the transcriptional profile at the single cell level. To solve this bias it is possible to apply a deconvolution of the transcriptional profile of each spot by integrating the expression data of scRNAseq through algorithms such as the deconvolution of Spatial Transcriptomics profiles using Variational Inference (DestVI) implanted in the scvi-tools package, this allows the tracing of the proportion of cell types present in each spot. The DestVI algorithm uses a Bayesian model for multi-resolution deconvolution of cell types in spatial transcriptomic data. DestVI learns both discrete cell-type-specific profiles and continuous sub-cell-type latent variations using a conditional deep generative model in this way, it recovers cell type proportions as well as a cell-type-specific snapshot of the transcriptional state of every spot. The deconvolution of the generated spatial transcriptomic data was performed with the scvi-tools[46] package and human scRNAseq data derived from the dataset of Huize Pan et al.[23]

3. RESULTS

3.1 Analysis of single-cell RNA sequencing data

3.1.1 QC, clustering and DEGs identification.

In order to identify cellular subpopulations deriving from the SMC lineage, the mice of the 4 experimental groups were sacrificed and the cells were dissociated from the aortic arch and selected for the presence of Tomato fluorophore by FACS and then we proceeded with scRNAseq (see Figure 3) using only the fluorophore positive cells. Then, the data obtained underwent a QC where empty droplets, doublets and cells with low expression were removed. We obtained the following transcriptomic data: 2379 cells for the group KO CD, 1699 for KO WD, 3485 for Wt CD and 970 for Wt WD. From these data we obtained 6 clusters, shown in Figure 4A, among which we proceeded to derive the differentially expressed genes and we searched among them by integrating with the bibliography the possible markers for the phenotype, the identified gene markers and their expression are shown in Figure 4C. It was observed that cluster 3 belonging to contractile SMCs shows the expression of some genes that characterize it compared to canonical SMCs such as *Adamtsl1*[47], a gene coding for a secreted protein belonging to the *Adamts* family, which is absent in contractile SMCs. In Figure 4B it is shown that the number of cells belonging to each cluster differs between the four groups of mice.

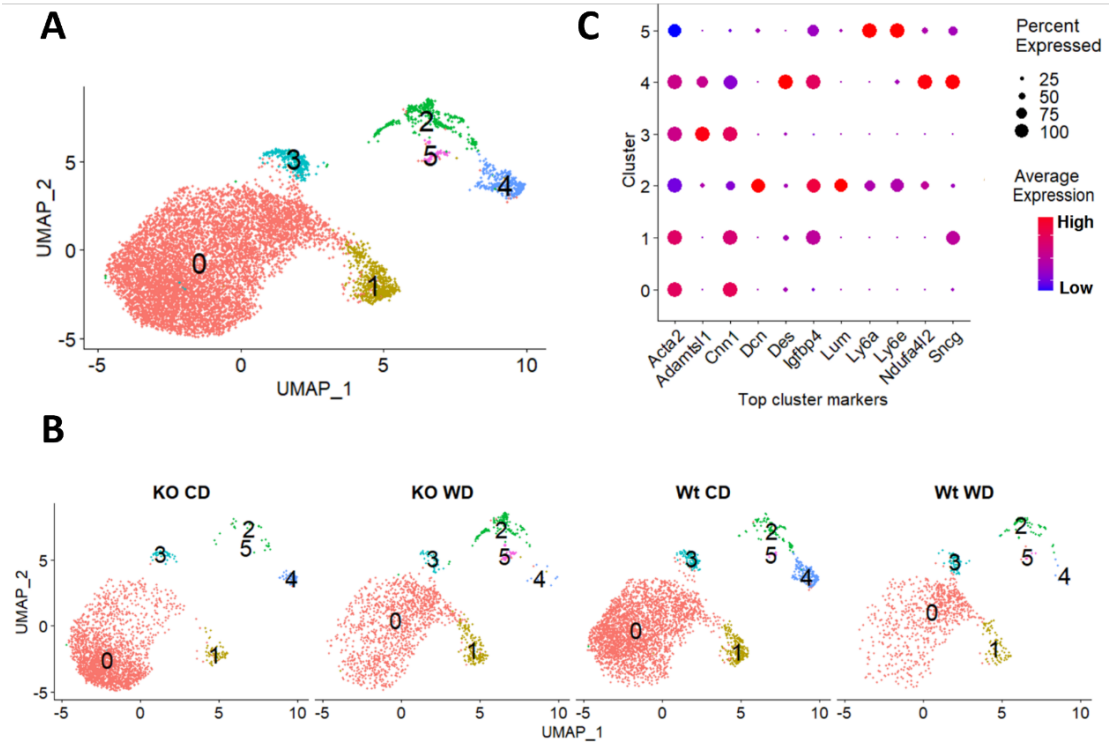


Figure 10. Clustering analysis and gene marker expression. A, UMAP visualization of the 6 clusters obtained; **B**, composition of the clusters by separating the cells by experimental group of origin; **C**, dotplot with the expression values of the top gene markers identified for each clusters, the color indicates the average

expression level of the gene and the size of the dot the percentage of cells in the cluster that express the considered gene.

Therefore, the clusters were annotated manually based on the expression of the markers obtained as follows (see Figure 5A): clusters 0 and 1 are the contractile SMC phenotype, cluster 2 is composed of fibromyocytes, cluster 3 has been noted as SMC *Adamtsl1*$^{+}$ as characterized in particular by the expression of this gene, cluster 4 corresponds to pericytes and cluster 5 to a phenotype called SEM[22].

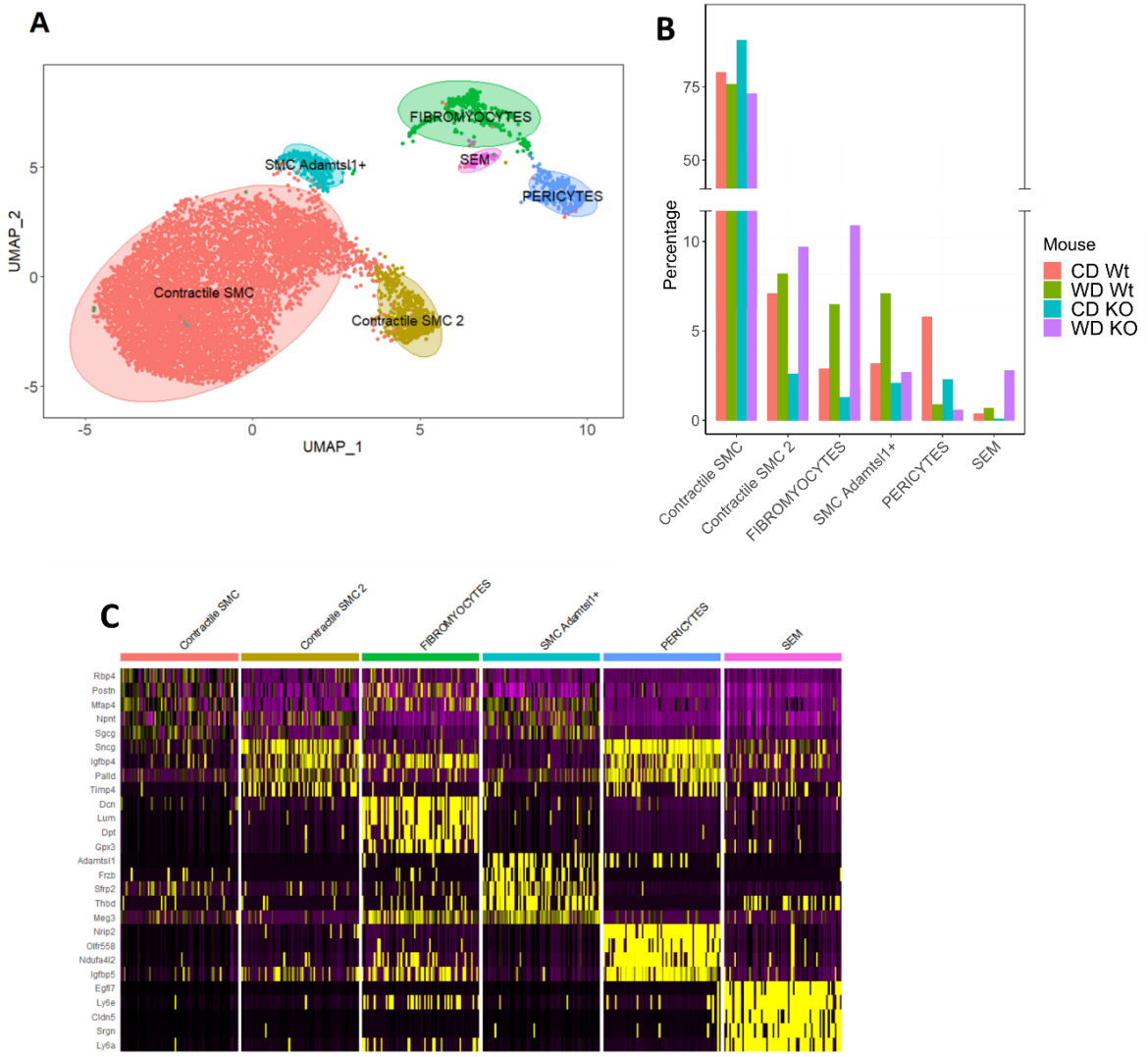


Figure 11. Annotation of clusters, compositional analysis and DEGs identification. A, UMAP with annotated cluster; B, percentage composition of the cell clusters for each experimental group; C, heatmap of the top DEGs among the clusters.

In order to verify how the diet influenced the abundance of the identified phenotypes, and therefore to verify if the SMC were stimulated by the diet to differentiate towards a certain condition rather than another, the relative abundances, in percentage, of the specific cell types was evaluated. We observed that both in KO and in Wt with WD there is an increase in contractile SMC2, fibromyocytes, SMC *Adamtsl1*$^{+}$ and SEM while pericytes and contractile SMC1 decrease.

Since the SMC *Adamtsl1*$^{+}$ phenotype is the only one not yet described in other works, we have concentrated the following analyzes only on this cluster.

3.1.2 Pathways enrichment analysis

The differentially expressed genes between the SMC *Adamtsl1*$^{+}$ and the contractile SMC1 and the contractile SMC2 have been calculated to highlight the genes that diversified this cluster from the canonical contractile SMCs and we proceeded with the analysis of the enriched pathways for the 334 up regulated and 261 down regulated genes. This analysis allows us to pass from a list of genes, which does not provide an immediate overview of the biological functions activated in the cells or of the role they could have in the biological context in which they are found, to a list of functional pathways. Therefore, in *Adamtsl1*$^{+}$ cells the pathways more enriched due to the up-regulated genes compared to the canonical contractile cells concern the organization of the extracellular matrix and the activation of neutrophils (Figure 12A) while the down-regulated ones are involved in muscle contraction (Figure 12B) suggesting that this phenotype is becoming metabolically active and pro- inflammatory.

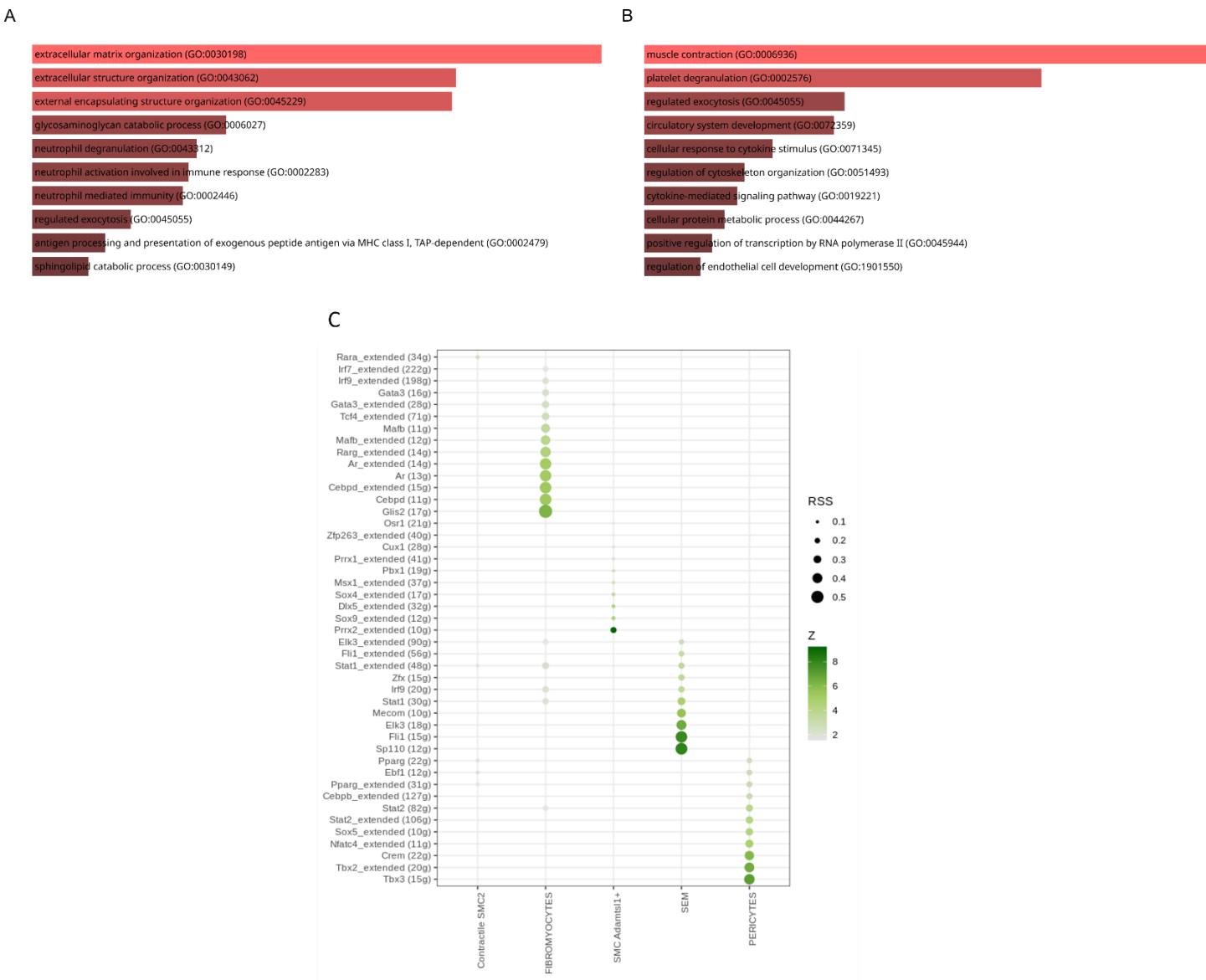


Figure 12. Pathways analysis and Regulon Specificity Score. A-B, respectively: the pathways enriched with up-regulated DEGs in *Adamtsl1*$^{+}$ and down in contractiles and vice versa; C, dotplot with the transcription factors of the regulons identified with SCENIC, the size of the dots reflect the RSS value and the color the z-score of their activity.

3.1.3 Inference of regulon activity

Due to the high number of dropouts and other technical biases of the scRNAseq technology, the pathways extrapolated from the expression data alone do not provide accurate information on the cell-state assumed by the cell. For this reason, it is necessary to link the expression data to the gene regulation networks, identifying the activated regulons, modules of genes regulated by the same transcription factor, and thus allowing to have a clearer picture of the functions performed by the cell. Therefore, we performed the workflow of the SCENIC tool, inferred regulon activity in each cell and computed the Regulon Specificity Score (RSS). The regulons with an RSS score greater than 0.1 in

the *Adamtsl1*$^{+}$ SMCs are: Prrx2, Sox9, Dlx5, Sox4, Msx1 and Pbx1 (Figure 12C). These transcription factors are involved in proliferation, extracellular matrix deposition and osteogenesis; this data confirms that this type of SMC is an intermediate state between contractile SMCs and proliferating SMCs

3.1.4 Trajectory analysis

To confirm that the *Adamtsl1*$^{+}$ cluster was an intermediate state of transdifferentiation from contractile SMCs to other phenotypes involved in the formation of the plaques, we performed a trajectory analysis with the R Slinghshot tool, ordering the cells along a scale of pseudotimes according to transcriptional similarity. As it is shown in Figure 13 we obtained three different trajectories: a trajectory that starts from the contractile SMC and through the *Adamtsl1*$^{+}$ cells reaches the fibromyocytes; a second lineage that starts from the contractile cells and through the *Adamtsl1*$^{+}$ reaches the pericytes and SEM cluster and finally a lineage that goes from the contractile SMC1 to the contractile SMC2. Therefore, the trajectories confirm the existence of a continuity pattern in the expression profile between the *Adamtsl1*$^{+}$ cells and the contractile SMCs and suggest that them could be a possible intermediate state between these and the more transdifferentiated phenotypes making us speculate that this phenotype could be a valid target where to block the transdifferentiation of SMCs and avoid the formation of atherogenic phenotypes.

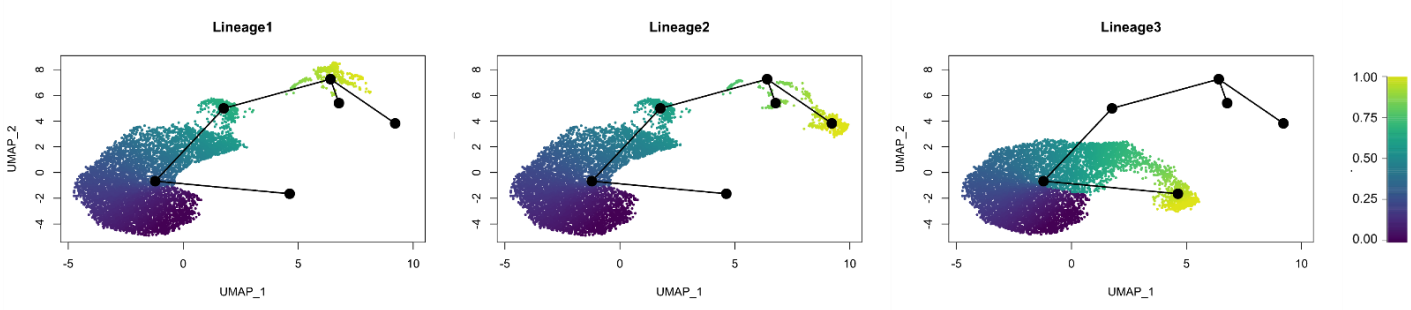


Figure 13. Trajectory analysis. Study of the trajectories in umap space of the clusters, the cells are coloured according to the pseudotime score.

3.1.5 Integration of murine *tdT*$^{-}$ cells and validation in human scRNAseq data

Once we identified the *Adamtsl1*$^{+}$ cluster, its characteristics and possible biological functions, we first wondered whether these cells were also present in other published scRNAseq datasets generated from the same model and if the same cluster was identified also in human atherosclerotic plaques. Therefore, we downloaded data from the work of Huize Pan et al.[23], where the authors performed a

scRNAseq analysis using the same mouse model as we did. They performed scRNAseq on the WD KO group, also fed for 16 weeks, also on using the *tdT*- cells, as well as on three plaques derived from patients with atherosclerosis. First of all, we performed the pre-processing of their raw data and downstream analysis in the same way as done for our dataset and we obtained that both in their mouse model and in the human dataset (Figure 14A-B) the *Adamtsl1*$^{+}$ cluster is present and from the signature transcriptional corresponds to the cluster that they noted as the minor SMC[23]. For an even more robust identification of our cluster in their dataset we performed the CCA integration implemented in Seurat between our WD KO group and their model with the same conditions and between our integrated dataset of all 4 groups with their human dataset. We found that our *Adamtsl1*$^{+}$ cells after integration with their cells, rebuilding the SMC *Adamtsl1*$^{+}$ cluster. These results confirm that, although this phenotype was not previously described, it is not an artifact of our experiment but is also present in other datasets. Furthermore, this shows also that *Adamtsl1*$^{+}$ cells are present in human atherosclerotic plaques.

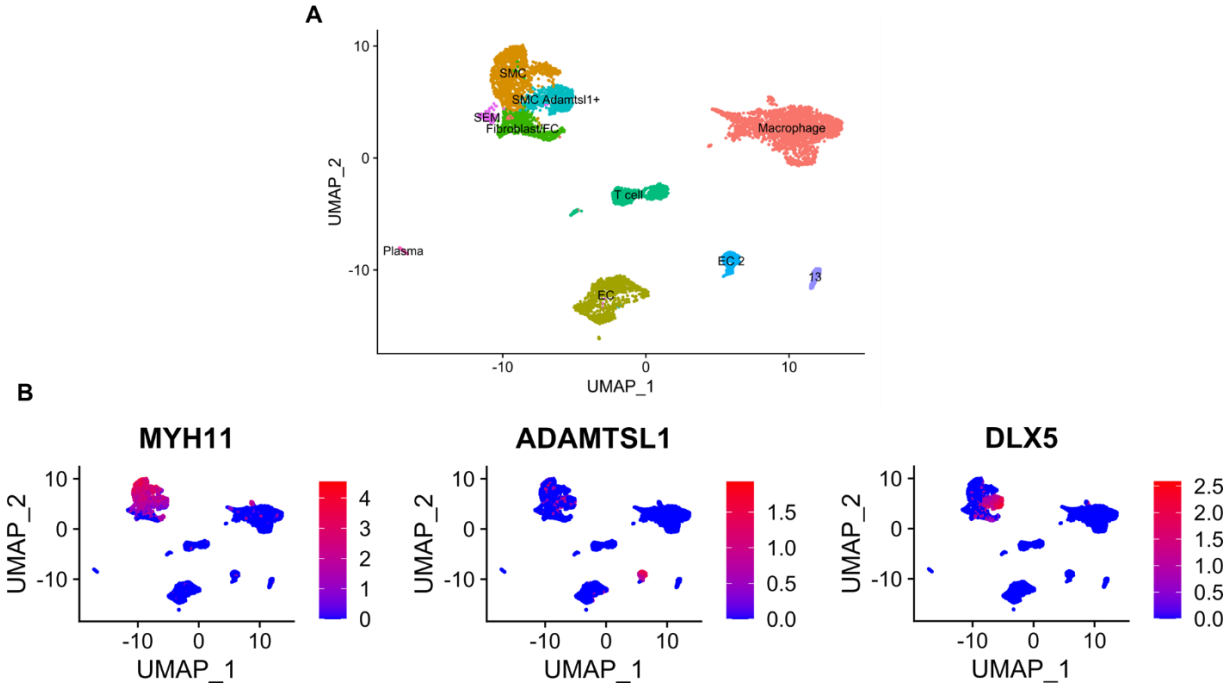


Figure 14. Clustering of human scRNAseq data. A) umap visualization of annotated clusters obtained with human scRNAseq data from Huize Pan et al.[23] ; B) Visualization of gene expression of MYH11, gene marker for SMC derived-cells, and of ADAMTSL1 and DLX5, main gene marker for ADAMTSL1^{+} phenotype.

3.1.6 Study of the communication between *Adamtsl1*$^{+}$ cells and macrophages

Finally, to verify the pro-inflammatory capacity of these *Adamtsl1*$^{+}$ cells, we applied the algorithm of the NicheNet R package to test their ability to induce an M1-type response in macrophages. Since in our experimental design we performed scRNAseq only on *tdT*$^{+}$ cells, we integrated our

dataset with scRNAseq data from the work of Huize Pan et al.[23] performed on the same model but also with *tdT*$^{-}$. The annotated clusters of integrated datasets are reported in Figure 15A. Then we applied NicheNet to prioritize the possible ligands which could predict changes in expression in target genes from a list of genes involved in the polarization of macrophages in M1[44], evaluating the interaction between the two cell-types among the genes differentially expressed by *Adamtsl1*$^{+}$ (Figure 15B). The interactions between the identified ligands and the possible receptors are shown in figure 15C. We repeated the procedure with a gene signature for anti-inflammatory M2 polarized Mϕ without finding a significant association with the expressed ligands, this confirmed that this cluster induces the activation of M1 Mϕ. These results suggest that Adamtsl1^{+} are able to activate the M1 phenotype and therefore they might be considered as proinflammatory SMCs. This phenotype could therefore stimulate the growth of plaques and to allow the development of complications such as their destabilization and rupture.

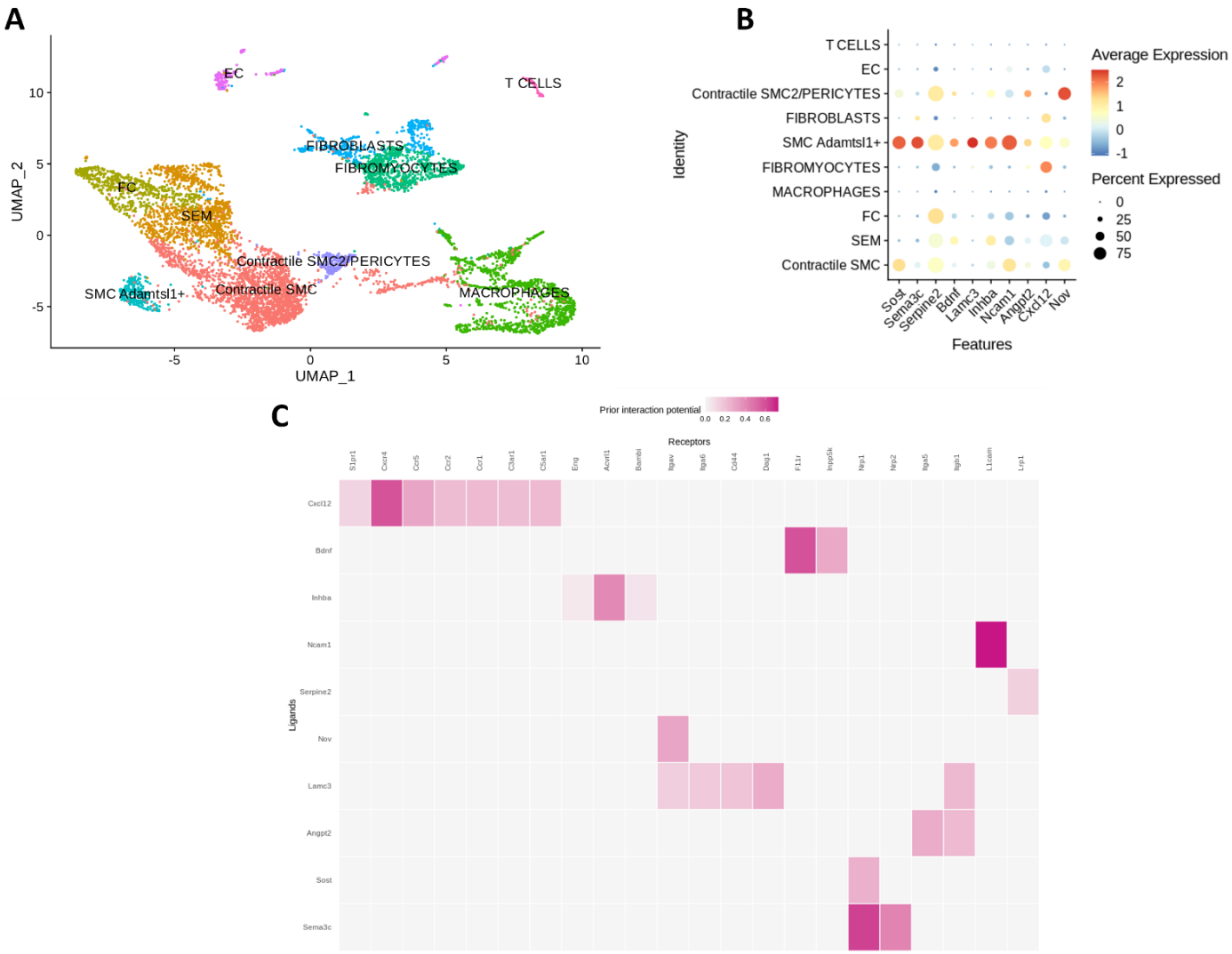


Figure 15. Integration with tdT^{-} data and interaction study. A) UMAP of the annotated clusters of the integrated dataset with data of mouse from Huize Pan et al.[23]; B) dotplot with the expression of priorized ligands produced by Adamtsl1^{+} cells, the size of the dots reflects the percentage of cells expressing the gene

and the color the mean expression value; C) heatmap with the priorized interactions between Adamtsl1$^+$ cell ligands and M1 receptors.

3.2 Visium Spatial Gene Expression data analysis

3.2.1 QC of spatial transcriptomic data

To verify whether SMC *Adamtsl1*$^+$ cells were also present in plaques of patients with arteriosclerosis, we performed a spatial Visium transcriptomics RNAseq on 4 sections of human plaques. This technique involves the permeabilization of a tissue to let the mRNAs spread out of the cells which will be captured by probes in the spots underlying the cell; in this way it is possible to see where the marker genes of the cell types of interest are expressed and to get an idea of their position and role.

Following the quality control of the obtained data, we decided to carry out the analyzes on two samples, shown in Figure 16A, for which an average number of UMIs per spot and sufficient detected genes was obtained. After aligning the transcriptomic data of the spots to the tissue using the spaceranger tool, the spots with low-quality data were removed. Then we checked the expression of some of the top *Adamtsl1*$^+$ cell marker genes such as *ADAMTSL1* and *DLX5*. As shown in Figure 16B, these genes are expressed within human plaques indicating that we have this phenotype in humans.

3.2.2 Spatial localization of *ADAMTSL1*$^+$ cells in human atherosclerotic plaque

Since the spots of the slide for visium transcriptomics are too large to have a single cell resolution, to understand where the *ADAMTSL1*$^+$ are located it was necessary to use a deconvolution algorithm such as DESTVI which, using the specific transcriptional profiles of the cell type obtained with a scRNAseq analysis, succeeds to infer the composition of the cell types present in the transcriptional profile of each spot. Therefore, we proceeded to apply DESTVI using scvi.tools package to deconvolve the transcriptional profile of the spots, using human scRNAseq data from Huize Pan et al.[23] as reference.

In this way it is possible to attribute the position of a cell type on the section in a more accurate way than observing only the expression of its marker genes since its entire transcriptional signature is used and not just a subset of genes which could also be expressed by similar cell types.

The results of this analysis are visible in the Figure 16C, where we observed an abundant presence of *ADAMTSL1*$^+$ cells in the plaque, that was more evident in at least one sample where there are spots where these cells constitute more than 75% of the mRNA content and therefore we can be confident

of the real presence of this phenotype in these spots. Hence *ADAMTSL1*$^{+}$ cells appear to be present in human plaques and appear to be distributed in the innermost layers of the plaque.

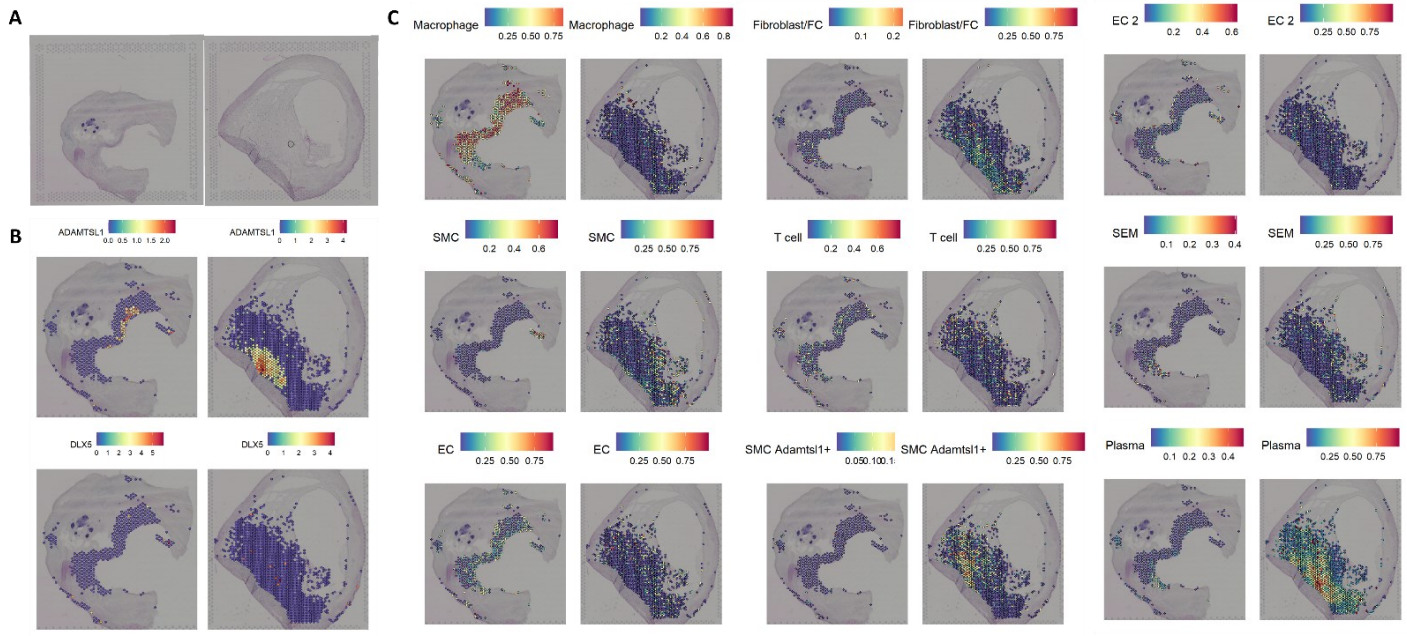


Figure 16. Localization of *ADAMTSL1*$^{+}$ cells into Visium Spatial Transcriptomic slide. A) H&E sections of the two human arteriosclerotic plaque samples on which Visium Spatial Transcriptomic analysis was performed; B) H&E sections with overlapped transcriptomic spots colored according to the expression of the two main markers of Adamtsl1^{+} cells; C) H&E sections with overlapped transcriptomic spots colored according to the proportions of cell types deconvolved with scvi-tools.

4. DISCUSSION AND FUTURE PROSPECTS

Arteriosclerosis is a problem of primary importance in Western countries and although current statin-based treatments have led to a decrease of morbidity and mortality rates[48], a considerable fraction of patients, does not benefit from this therapy. For this reason, we need to identify new molecular targets that allow us to block the onset of plaque formation.

It is known that the SMCs of the arteries play an important role in the formation of arteriosclerotic plaques thanks to their ability to undergo phenotypic switching and therefore to migrate into the intima forming the fibrous cap of the plaques[12]. Their ability to transdifferentiate also seems to play an important role in this pathology but this aspect has not been yet fully clarified. Some works, as the one of Wirka, R.C. et al.[19], have shown that SMCs can transdifferentiate into phenotypes that determine the stability and formation of plaques and that scRNAseq technology is a powerful tool for identifying new SMCs phenotypes. In our work we performed SMC lineage tracing in a mouse model fed with an atherogenic (WD) or a chow diet (CD) to identify phenotypically modulated cells of SMC origin in atherosclerotic lesions. scRNAseq analysis highlighted some phenotypes already described in the literature, such as fibromyocytes and SEM, and a new phenotype distinct from contractile SMCs characterized by the expression of some genes, in which the most selective was *Adamtsl1*.

Previous works from Wirka, R.C. et al.[19] and Huize Pan et al.[23] applied scRNA analysis only on *ApoE* KO mice fed with WD while we started from 4 experimental groups with Wt and *ApoE* KO mice that were also fed with CD in order to verify the impact of genotype and diet on the transdifferentiation of SMCs.

The values of the cluster percentages shown in Figure 11B show that: as far as the genotype is concerned, in CD the KO mice have a higher basal level of contractile SMCs while the Wt mice present a greater fraction of pericytes and of the cluster named contractile SMC 2; in mice fed with WD instead, it seems that the SMCs in KOs are guided to transdifferentiate towards fibromyocytes and SEMs while the Wt towards the *Adamtsl1*$^{+}$. Regarding the diet, WD seems to drive SMCs to transdifferentiate towards fibromyocytes, SEMs and *Adamtsl1*$^{+}$ rather than pericytes.

These results seem to tell us that the SMCs of the model used by the two previous works have a predisposition to become fibromyocytes and SEM which are the phenotypes they have principally analysed. Applying our scRNA data analysis pipeline to the dataset derived from mice fed for 16 weeks with WD obtained from Wirka, R.C. et al.[19], we found very few *Adamtsl1*$^{+}$ cells, and these cells did not cluster together. This may depend by the fact that their *tdT*$^{+}$ cells, although more numerous than ours - 11348 cells analyzed with our pipeline, versus 8533 of our dataset - have a lower sequencing depth which allowed us to identify fewer genes per cell: in our own dataset we have

a mean of detected genes of 1844 versus 1278 of theirs. Therefore, considering that the *Adamtsl1*$^{+}$ cluster is a rare population, and given that with KO mice SMCs mainly become fibromyocytes and SEM, we can conclude that a greater sequencing depth is required to observed the cluster analyzed in our work. On the other hand, analyzing the dataset derived from KO mice fed WD for 16 weeks published by Huize Pan et al.[23] both alone or integrated with our data, we observed that the cluster that they called "minor SMC" corresponds to our *Adamtsl1*$^{+}$ cluster. However, in that work they focused on SEMs and did not study this phenotype in detail.

Our analysis showed that the *Adamtsl1*$^{+}$ cellular phenotype, identified by scRNAseq in mice is also present in human atherosclerotic plaques and that, according to the results of the inference of regulons and enrichment of the pathways, these cells are losing their contractility capacities and are metabolically active, therefore, this population of SMCs could positively contribute to the formation of plaques.

From the literature we know that Mϕs are very important cells involved in the development of atherosclerosis[11, 15, 16]. They play an atherogenic role by infiltrating atherosclerotic lesions, promoting the inflammatory state of the tissue and they can overexpress scavenger receptors to absorb oxidized cholesterol and become foam cells, one of the main constituents of the lipid core of atheromas. Furthermore, the M1 phenotype can decrease the efferocytosis process, thus preventing the removal of apoptotic cells and causing the plaque necrosis with a consequent modulation of its stability. Our *Adamtsl1*$^{+}$ population seems to have activated pro-inflammatory pathways and through the study of cell-to-cell interactions, we have seen that they can produce ligands capable of interacting with receptors linked to the polarization of Mϕs into M1. Therefore, this phenotype could not only contribute to the development of the plaques but also might allow the onset of their destabilization due to the stimulation of Mϕs.

Finally, the analysis of the trajectories showed that *Adamtsl1*$^{+}$ cells could be an intermediate stage of SMCs that are transdifferentiating into other phenotypes already known in the literature for their atherogenicity. Being a cluster that expresses specific markers with respect to contractile SMCs, such as *Adamtsl1* itself, whose function is not known to date, and *Dlx5*, we can eventually think of applying the modulation of such genes as a possible therapy to block the transdifferentiation of SMC into the arteries and to improve the prognosis of atherosclerosis patients.

Therefore, in the next future we plan to test whether the *Adamtsl1* gene have a role in the transdifferentiation of SMCs towards the phenotype identified in this work. Its function will be tested through in vitro experiments on cultured SMCs in which its expression will be modulated through loss- or gain-of-function approaches and the cellular phenotype will be evaluated.

Then the presence of this protein will be also tested in plasma of mice and human patients to assess whether it can be used as early diagnostic tool on such conditions.

Finally, the role of *Adamtsl1* in the onset of atherosclerosis will be tested as therapeutic agent, and this will be done modulating *Adamtsl1* expression in *ApoE* KO mice in pro-aterosclerotic conditions by using a lentiviral system where its specific shRNA will be guided by the promoter of the SM22 gene.

If a role of *Adamtsl1*$^{+}$ cells in the onset of atherosclerosis will be confirmed, also in light of the discoveries made in this thesis, we might hypothesize that this gene could become an innovative and promising therapeutic target to prevent the aggravation of atherosclerosis development. Its modulation could block the pathology development, by counteracting the formation of plaques, but also reducing the proinflammatory environment. Furthermore, since ADAMTSL1 is a secreted protein, it could be used as a marker of plaque pathogenicity.

5. BIBLIOGRAPHY

1. Ribatti, D., Nico, B. & Crivellato, E. The Development of the Vascular System: A Historical Overview. Methods in Molecular Biology vol. 1214 (Springer New York, 2015).
2. Mercadante, Anthony A. and Avais Raja. "Anatomy, Arteries." StatPearls, StatPearls Publishing, 14 January 2022.
3. Rudijanto, A. The role of vascular smooth muscle cells on the pathogenesis of atherosclerosis. Acta Med. Indones. 39, 86–93 (2007).
4. Zhao, Yingzi & Vanhoutte, Paul & Leung, Susan. (2015). Vascular nitric oxide: Beyond eNOS. Journal of pharmacological sciences.
5. Frostegård, J. Immunity, atherosclerosis and cardiovascular disease. BMC Med. 1–8 (2013).
6. Wali JA, Jarzebska N, Raubenheimer D, Simpson SJ, Rodionov RN, O'Sullivan JF. Cardio-Metabolic Effects of High-Fat Diets and Their Underlying Mechanisms-A Narrative Review. Nutrients. 2020;12(5):1505. Published 2020 May 21.
7. Libby, P. et al. Atherosclerosis. Nat. Rev. Dis. Prim. 5, 1–18 (2019).
8. Grover-Páez, F. & Zavalza-Gómez, A. B. Endothelial dysfunction and cardiovascular risk factors. Diabetes Res. Clin. Pract. 84, 1–10 (2009).
9. Tabas, I., Williams, K. J. & Borén, J. Subendothelial lipoprotein retention as the initiating process in atherosclerosis: Update and therapeutic implications. Circulation 116, 1832–1844 (2007).
10. Libby, P. The changing landscape of atherosclerosis. *Nature* **592**, 524–533 (2021).
11. Ley, K., Miller, Y. I. & Hedrick, C. C. Monocyte and macrophage dynamics during atherogenesis. Arterioscler. Thromb. Vasc. Biol. 31, 1506–1516 (2011).
12. Gomez, Delphine, and Gary K Owens. "Smooth muscle cell phenotypic switching in atherosclerosis." Cardiovascular research vol. 95,2 (2012): 156-64.
13. Kojima, Y., Weissman, I. L. & Leeper, N. J. The Role of Efferocytosis in Atherosclerosis. Circulation 135, 476–489 (2017).
14. Ruiz, J. L., Hutcheson, J. D. & Aikawa, E. Cardiovascular calcification: Current controversies and novel concepts. Cardiovasc. Pathol. 24, 207–212 (2015).
15. Kobiyama K, Ley K. Atherosclerosis. Circ Res. 2018;123(10):1118-1120. doi:10.1161/CIRCRESAHA.118.313816.
16. Zhu Y, Xian X, Wang Z, et al. Research Progress on the Relationship between Atherosclerosis and Inflammation. Biomolecules. 2018;8(3):80. Published 2018 Aug 23.
17. Libby, P. Inflammation in atherosclerosis. Nature 420, 868–874 (2002).
18. Feil, Susanne et al. "Transdifferentiation of vascular smooth muscle cells to macrophage-like cells during atherogenesis." Circulation research vol. 115,7 (2014): 662-7.

19. Wirka, Robert C et al. “Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis.” Nature medicine vol. 25,8 (2019): 1280-1289.
20. Allahverdian, Sima et al. “Smooth muscle cell fate and plasticity in atherosclerosis.” Cardiovascular research vol. 114,4 (2018): 540-550.
21. Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. Single-cell RNA sequencing technologies and applications: A brief overview. Clin Transl Med. 2022 Mar;12(3):e694.
22. Ye, Fang & Huang, Wentao & Guo, Guoji. (2017). Studying hematopoiesis using single-cell technologies. Journal of Hematology & Oncology.
23. Huize Pan et al. Single-Cell Genomics Reveals a Novel Cell State During Smooth Muscle Cell Phenotypic Switching and Potential Therapeutic Targets for Atherosclerosis in Mouse and Human. Circulation. 2020;142:2060–2075.
24. Cheng G., Ning B., Shi T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front. Genet., 05 April 2019.
25. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019 Jun 19;15(6):e8746.
26. Butler, A., Hoffman, P., Smibert, P. et al. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36, 411–420 (2018).
27. Haghverdi, L., Lun, A., Morgan, M. et al. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 36, 421–427 (2018).
28. Zeng Z, Miao N, Sun T. Revealing cellular and molecular complexity of the central nervous system using single cell sequencing. Stem Cell Research and Therapy. 2018;9(1):234.
29. McCarthy DJ, Campbell KR, Lun ATL, Willis QF (2017). “Scater: pre-processing, quality control, normalisation and visualisation of single-cell RNA-seq data in R.” Bioinformatics, 33, 1179-1186.
30. Lun ATL, Riesenfeld S, Andrews T, Dao T, Gomes T, participants in the 1st Human Cell Atlas Jamboree, Marioni JC (2019). “EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data.” Genome Biol., 20, 63.
31. Griffiths JA, Richard AC, Bach K, Lun ATL, Marioni JC (2018). “Detection and removal of barcode swapping in single-cell RNA-seq data.” Nat. Commun., 9(1), 2667.
32. Hao*, Hao*, et al. “Integrated analysis of multimodal single-cell data”. Cell 2021.
33. Haghverdi L, Lun ATL, Morgan MD, Marioni JC (2018). “Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors.” Nat. Biotechnol., 36(5), 421–427.
34. Finak, G., McDavid, A., Yajima, M. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data.

Genome Biol 16, 278 (2015).

35. Reimand, J., Isserlin, R., Voisin, V. et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc 14, 482–517 (2019).Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013; 128(14).
36. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research. 2016; gkw377.
37. Xie Z, Bailey A, Kuleshov MV, Clarke DJB., Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, Jeon M, & Ma'ayan A. Gene set knowledge discovery with Enrichr. Current Protocols, 1, e90. 2021.
38. Aibar, Sara et al. "SCENIC: single-cell regulatory network inference and clustering." Nature methods vol. 14,11 (2017): 1083-1086.
39. Van de Sande, B., Flerin, C., Davie, K. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc 15, 2247–2276 (2020).
40. Street, Kelly et al. "Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics." BMC genomics vol. 19,1 477. 19 Jun. 2018.
41. Durinck S, Spellman P, Birney E, Huber W (2009). "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt." Nature Protocols, 4, 1184–1191.
42. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods (2019).
43. Orecchioni M., et al. "Macrophage Polarization: Different Gene Signatures in M1(LPS+) vs. Classically and M2(LPS–) vs. Alternatively Activated Macrophages". Front. Immunol., 24 May 2019 Sec. Molecular Innate Immunity.
44. Lopez, R., Li, B., Keren-Shaul, H. et al. DestVI identifies continuums of cell types in spatial transcriptomics data. Nat Biotechnol 40, 1360–1369 (2022).
45. Gayoso, A., Lopez, R., Xing, G. et al. A Python library for probabilistic analysis of single-cell omics data. Nat Biotechnol 40, 163–166 (2022).
46. Hirohata S., et al. "Punctin, a Novel ADAMTS-like Molecule, ADAMTSL-1, in Extracellular Matrix". JBC GLYCOBIOLOGY AND EXTRACELLULAR MATRICES| VOLUME 277, ISSUE 14, P12182-12189, APRIL 2002.
47. Kang, S., Wu, Y. & Li, X. Effects of statin therapy on the progression of carotid atherosclerosis: A systematic review and meta-analysis. Atherosclerosis 177, 433–442 (2004).

MODULO DI EMBARGO DELLA TESI
(da compilare solo se si richiede un periodo di segretazione della tesi)

Il sottoscritto Luca Lambroia Nato il 18/07/1994a
Pescia
provincia di Pistoia
Dottorato di Ricerca in Precision Medicine

DICHIARA

- che il contenuto della tesi **non può essere immediatamente consultabile per il seguente motivo**

Dati non ancora pubblicati

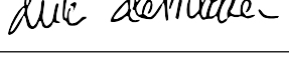
- che il testo completo della tesi potrà essere reso consultabile dopo:

□ 6 mesi dalla data di conseguimento titolo
x 12 mesi dalla data di conseguimento titolo
□ 24 mesi dalla data di conseguimento titolo
□ altro periodo _____

- che sarà comunque consultabile immediatamente l'abstract della tesi, che viene caricato in Esse3, profilo studente.

Luogo e Data
Milano 11/01/2023

Firma del Dichiarante



Controfirma del Tutor e/o Relatore del Dottoratoper la motivazione di embargo e il periodo.