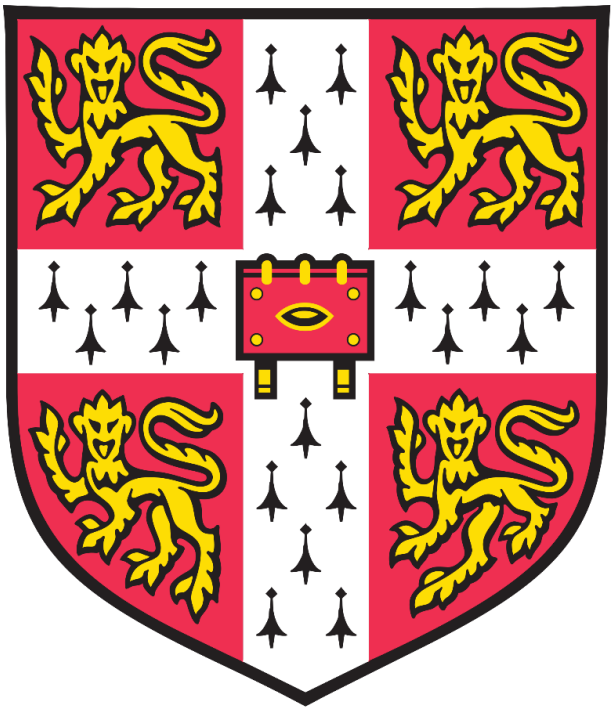
A genome-wide, single-cell analysis of vascular smooth muscle cell plasticity



Lina Dobnikar

Babraham Institute

Churchill College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

September 2019

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

This thesis is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

This thesis does not exceed 60,000 words.

Cambridge, United Kingdom, September 2019

Lina Dobnikar

Summary

A genome-wide, single-cell analysis of vascular smooth muscle cell plasticity *Lina Dobnikar*

Vascular smooth muscle cells (VSMCs) possess a remarkable capacity to change phenotype in response to injury or inflammation. In healthy arteries, VSMCs exist in a contractile state, but upon vascular inflammation or injury, they can switch into an activated state, in which they downregulate the contractile differentiation markers and show increased migration, proliferation and secretion of proinflammatory cytokines. This process is termed phenotypic switching and can lead to VSMC accumulation within atherosclerotic plaques. Previous observations of clonal expansion of a small number of VSMCs in atherosclerosis suggested that VSMCs were functionally heterogeneous. I hypothesised that functional heterogeneity of VSMCs in disease may originate from VSMC heterogeneity in healthy arteries.

In the first part of this thesis I explored the regional heterogeneity of VSMCs originating from different parts of the mouse aorta, as well as heterogeneity of VSMCs within a vascular bed using single-cell and bulk RNA sequencing. VSMCs originating from the atherosclerosis-prone aortic arch and atherosclerosis-resistant descending thoracic aorta were found to have distinct transcriptional signatures at the single-cell level. Additionally, several disease-relevant genes were observed to be heterogeneously expressed within both vascular beds.

In the second chapter I identified and characterised a rare subset of VSMCs expressing Stem cell antigen 1 (SCA1). Single-cell RNA-seq was combined with VSMC-specific lineage tracing to profile gene expression in individual VSMCs from healthy mouse arteries and to compare SCA1-expressing VSMCs to other cells. SCA1-positive VSMCs were heterogeneous, with many of them expressing low levels of contractile VSMC markers. Additionally, a subset of SCA1-positive VSMCs in healthy arteries expressed transcriptional signatures characteristic of activated VSMCs involved in phenotypic switching.

In the third chapter I investigated the involvement of SCA1-positive VSMCs in phenotypic switching. SCA1 upregulation was found to mark the process of VSMC phenotypic switching following *in vitro* culture and *in vivo* vascular injury. Single-cell RNA-seq profiling of VSMCs in atherosclerosis and following vascular injury showed that *Ly6a/Sca1*-expressing VSMCs were present and expressed transcriptional signatures similar to activated SCA1-positive cells observed in healthy arteries.

Overall the results presented in this thesis highlight the heterogeneous nature of VSMCs in healthy arteries, both regionally and within a vascular bed. I identified a rare subset of SCA1-positive VSMCs with activated transcriptional signatures in healthy arteries. I hypothesised that SCA1-positive VSMCs may be responsible for clonal expansion of VSMCs in atherosclerosis, which would have clinical implications for earlier detection and specific targeting of expanding VSMCs in atherosclerosis in the future. In support of this hypothesis I have shown that *Ly6a/Sca1* is upregulated in model systems of VSMC phenotypic switching and that transcriptional signatures of *Ly6a/Sca1*-expressing VSMCs in mouse atherosclerosis and vascular injury resemble those of healthy activated SCA1-positive VSMCs.

Acknowledgements

First I want to thank my supervisors, Mikhail and Helle, for giving me the opportunity to work with them and for providing support, mentorship and guidance throughout my PhD. They have both been wonderful supervisors and mentors throughout the past four years and always believed in me. Special thanks to Mikhail for his constant and infectious enthusiasm and motivation for science and for his commitment to supporting the wellbeing of everyone in the lab. Special thanks also go to Helle for her patience and kindness, dedication to science and for using her extensive knowledge to thoroughly assess and improve all scientific ideas.

I also want to thank all the current and past members of the Spivakov and Jørgensen labs for exciting scientific discussions and support. Specifically I want to thank Annabel, Joel and Jenny for their substantial contributions to this project. I would also like to thank Anne, who became my official supervisor during my PhD and has helped and supported me throughout the last year. Additionally, I thank the Babraham Bioinformatics Facility for their help, in particular Simon, Steven, Felix and Anne. I am also grateful to the BBSRC DTP for funding and for giving me the opportunity to learn so many new skills through rotation projects and the internship.

During my PhD I met many wonderful people, several of them I am now lucky to call my friends. In particular, I would like to thank Jo, Marco, Michiel, Stephen, Steven, Rachel, Jorg, Elena, Paula and Jonathan for the fun times.

Huge thanks go to my big family, for their constant love and support. My grandparents, parents, step-parents and uncles and aunts have supported my interest in science from a young age and dedicated their time to teaching me all sorts of academic and life skills, which provided a foundation for all my future learning.

Last but not least, big thanks to Ondrej, for the endless support and love and for making life great fun.

Acknowledgement of Assistance

1) Initial training in techniques and subsequent mentoring

Dr Mikhail Spivakov (Babraham Institute and MRC London Institute of Medical Sciences, Imperial College London) - supervision, mentoring, training and advice with computational analysis

Dr Helle F. Jørgensen (Department of Medicine, University of Cambridge) – supervision, mentoring, introduction to laboratory techniques used to generate scRNA-seq datasets

Dr Hashem Koohy (Babraham Institute) - advice on how to approach the random forest analysis

Dr Simon Andrews (Babraham Institute Bioinformatics Facility) - advice on computational analysis, assessment

Dr Gavin Kelsey (Babraham Institute) - assessment

Dr Anne Segonds-Pichon (Babraham Institute Bioinformatics Facility) - advice on statistical analysis

2) Data obtained from a technical service provider

All next generation sequencing data has been obtained from the Babraham Institute and the CRUK sequencing facilities. Babraham Bioinformatics facility processed the next generation sequencing data through the Babraham Institute pipeline or the cellranger pipeline, including aligning the reads to a reference genome.

3) Data produced jointly

Single-cell transcriptomes of VSMCs from the atherosclerotic plaques (10X Genomics Chromium platform) - generated jointly with Dr H.F. Jørgensen

4) Data/materials provided by someone else

Bulk RNA-seq of AA and DT regions – generated by Dr Jennifer L. Harman (Department of Medicine, University of Cambridge) and initial data processing by Phoebe Oldach (Babraham Institute)

RT-qPCR data validating the findings from bulk differential expression analysis between the AA and DT regions was produced by Annabel L. Taylor (Department of Medicine, University of Cambridge).

Single-cell *ex vivo* VSMC transcriptomes from the AA and DT regions (Fluidigm C1 platform) - generated by Dr H.F. Jørgensen and Dr M. Spivakov

Pooled transcriptomes of medial and adventitial layers - generated by Dr H.F. Jørgensen

Single-cell transcriptomes of S+L+, S+L- and S-L+ cells (Smart-seq2 platform) - generated by Dr Joel Chappell (Department of Medicine, University of Cambridge) and A.L. Taylor.

Dr M. Spivakov identified the final set of significantly highly variable genes in S+L+ cells and carried out the network analysis and gene correlation analysis with the cVSMC network

Single-cell transcriptomes of cells from healthy arteries (whole aorta and VSMC-only) and VSMCs after carotid ligation injury (10X Genomics Chromium platform) - generated by Dr H.F. Jørgensen and A.L. Taylor.

Flow cytometry analysis of SCA1-positive VSMCs in carotid arteries after carotid ligation injury – carried out by Dr J.L. Harman.

Flow cytometry analysis of SCA1-positive cells in healthy vessels using SCA1 antibody staining and GFP-Sca1 mice – carried out by Dr H.F. Jørgensen and A.L. Taylor.

Culture experiments of GFP-Sca1 medial and adventitial cells – carried out by A.L. Taylor

Analysis of the proportion of S+L+ VSMCs after varying time periods between the induction of lineage labelling and flow cytometry analysis – carried out by A.L. Taylor and Dr M. Spivakov

Table of Contents

1 Introduction 1

1.1 The cardiovascular system 1

1.2 Artery wall structure 2

1.3 Vascular smooth muscle cells (VSMCs) 3

1.4 Vascular development and the origins of VSMCs 5

1.5 VSMC heterogeneity 7

1.5.1 Regional heterogeneity of VSMCs 7

1.5.2 VSMC heterogeneity within a vascular bed 8

1.5.3 Functional heterogeneity of VSMCs 8

1.6 VSMC response to stimulus 9

1.6.1 Mechanisms of VSMC phenotypic switching 9

1.6.2 VSMCs in culture 11

1.6.3 VSMCs in atherosclerosis 12

1.6.4 VSMCs in carotid ligation injury 16

1.7 Origins of phenotypically switched VSMCs 16

1.8 Stem cell antigen 1 (SCA1): a marker of stem/progenitor populations expressed by a subset of vascular cells 18

1.9 Single-cell transcriptomics as a tool for studying cell heterogeneity and plasticity 19

1.9.1 Approaches for scRNA-seq library generation 20

1.9.2 scRNA-seq data analysis approaches 24

1.10 Aims and hypotheses 29

2 Methods 30

2.1 Animal models and tissue processing 30

2.2 Bulk RNA-seq analysis 31

2.2.1 Experimental protocol 31

2.2.2 Data processing and analysis 31
2.3 Flow cytometry analysis and scRNA-seq experimental methods 32
2.3.1 Fluidigm C1 platform 32
2.3.2 Smart-seq2 platform 33
2.3.3 10X Genomics Chromium platform 33
2.4 Processing, quality control and normalisation of scRNA-seq data 34
2.4.1 Fluidigm C1 and Smart-seq2 platforms 34
2.4.2 10X Genomics Chromium platform 35
2.5 Processing and analysis of publicly available cultured VSMC transcriptomes 36
2.6 Log transformation and PCA dimensionality reduction 36
2.7 Random forest analysis 37
2.8 Highly variable gene and co-expression analysis 37
2.9 Network analysis and visualisation 38
2.10 Summarising expression levels of gene signatures using PC1 scores 39
2.11 Identification of the cVSMCneg and cVSMCpos signatures 39
2.12 Clustering analysis and t-distributed Stochastic Neighbour Embedding (t-SNE) visualisation 39
2.13 Differential gene expression analysis 40
2.14 Gene ontology analysis 41
2.15 Fisher's exact test 41
2.16 Data availability 41
3 Regional differences between the aortic arch (AA) and descending thoracic aorta (DT) VSMCs 42
3.1 Introduction 42
3.2 Results 42
3.2.1 Single-cell analysis of VSMCs from the AA and DT aortic regions 42

3.2.2 Population-level differences in gene expression between the AA and DT regions ……49

3.2.3 Transcriptional signatures of AA and DT regions are detectable at the single-cell level ……53

3.2.4 Heterogeneity of VSMCs within the AA and DT vascular beds 57

3.2.5 Highly variable gene expression in independent VSMC profiles.....61

3.2.6 Processing of the 10X Genomics Chromium VSMC transcriptomes 61

3.2.7 Highly variable genes show co-expression in VSMCs profiled using the 10X Genomics Chromium platform.....64

3.2.8 Application of the random forest model to independent VSMC profiles 67

3.3 Discussion 71

3.3.1 Regional heterogeneity of VSMCs 71

3.3.2 Highly variable genes.....73

3.3.3 Conclusion 75

4 Rare subpopulation of SCA1-positive VSMCs detected in healthy arteries.....76

4.1 Introduction 76

4.2 Results.....77

4.2.1 Detection of SCA1-positive cells of VSMC lineage in healthy arteries 77

4.2.2 Targeted profiling of the medial SCA1-positive subpopulation 82

4.2.3 Processing and quality control of Smart-seq2 transcriptional profiles.....84

4.2.4 Heterogeneity of the medial SCA1-positive subpopulation of cells.....86

4.2.5 Transcriptional signatures of SCA1-positive VSMCs expressing low levels of contractile markers 93

4.3 Discussion 95

4.3.1 SCA1-positive cells within the adventitial and endothelial layers 95

4.3.2 SCA1-positive VSMCs within the medial layer 96

4.3.3 SCA1-positive cells of VSMC lineage within atherosclerotic plaques 97

4.3.4 Conclusions 98
5 SCA1-positive VSMCs in model systems of VSMC phenotypic switching 99
5.1 Introduction 99
5.2 Results 100
5.2.1 SCA1 is upregulated in VSMCs cultured *in vitro* 100
5.2.2 SCA1 is upregulated in VSMCs after carotid ligation injury 105
5.2.3 Transcriptional signatures of VSMCs within the atherosclerotic plaque 115
5.2.4 Alternative marker genes of *Ly6a/Sca1*-expressing VSMCs in models of phenotypic switching 123
5.3 Discussion 127
5.3.1 VSMC phenotypic switching in culture 127
5.3.2 Induction of VSMC phenotypic switching after carotid ligation injury 128
5.3.3 VSMCs are phenotypically modulated towards a variety of phenotypes in atherosclerosis 129
5.3.4 Alternative markers of *Ly6a/Sca1*-positive VSMCs 132
5.3.5 Conclusion 133
6 General Discussion 134
References 139

Table of Figures

Figure 1.1: Artery wall structure. 3
Figure 1.2: Phenotypic modulation of VSMCs. 4
Figure 1.3: Embryonic origins of VSMCs. 6
Figure 1.4: Development of the atherosclerotic plaque. 13
Figure 1.5: Phenotypic modulation of VSMCs in atherosclerosis. 15
Figure 1.6: Experimental protocols for scRNA-seq. 22
Figure 3.1: Schematic of the approach. 43
Figure 3.2: Quality control of scRNA-seq data of ex vivo VSMCs. 44
Figure 3.3: Profiled AA and DT single cells express VSMC contractile marker genes. 45
Figure 3.4: Profiled AA and DT single cells are distinct from pooled adventitial samples and share similarities with bulk RNA-seq VSMC profiles. 46
Figure 3.5: Transcriptional profiles of AA and DT *ex vivo* single cells are distinct from those of cultured VSMCs. 47
Figure 3.6: *Ex vivo* VSMCs retained higher levels of VSMC contractile marker gene expression than cultured VSMCs. 48
Figure 3.7: Differentially expressed genes between the AA and DT regions. 49
Figure 3.8: Enrichment for gene ontology terms (biological process) among genes upregulated in the AA and DT regions. 51
Figure 3.9: Comparison of the genes identified as differentially expressed using microarrays in a previous study (Trigueros-Motos et al. 2013) and bulk RNA-seq in our study. 52
Figure 3.10: Examples of expression patterns of genes identified as differentially expressed in bulk RNA-seq at the single-cell level. 53
Figure 3.11: Schematic of the random forest approach. 54
Figure 3.12: Random forest analysis distinguishes *ex vivo* VSMCs from the AA and DT regions at the single-cell level. 56
Figure 3.13: Examination of batch effects in *ex vivo* VSMCs profiled in separate AA and DT experiments. 57
Figure 3.14: Highly variable genes in AA and DT regions. 58
Figure 3.15: Gene ontology analysis of highly variable genes in the AA and DT regions. 60
Figure 3.16: Expression levels of highly variable genes in profiled VSMCs. 61

Figure 3.17: Quality control of the VSMCs only (red) and whole aorta (blue) 10X Genomics Chromium datasets. 62
Figure 3.18: Principal component analysis and selection of principal components for further analysis. 63
Figure 3.19: Clustering analysis and identification of cell types within the VSMC-only and whole aorta 10X Genomics Chromium datasets. 64
Figure 3.20: Highly variable gene expression in VSMC profiled as part of the 10X Genomics Chromium whole aorta and VSMCs-only datasets. 66
Figure 3.21: Random forest predictor genes are expressed sparsly in VSMCs-only and whole aorta 10X Genomics Chromium datasets. 68
Figure 3.22: Random forest model applied to 10X Genomics Chromium datasets. 69
Figure 4.1: *Ly6a/Sca1* transcript and SCA1 protein is expressed in a small number of medial cells..... 77
Figure 4.2: Schematic representation of the single-colour (top) and multi-colour (bottom) VSMC lineage labelling approaches. 78
Figure 4.3: A proportion of medial SCA1-positive cells express the VSMC lineage label. 79
Figure 4.4: A subset of medial cells from transgenic Sca1-GFP animals co-express GFP and VSMC marker ACTA2. 80
Figure 4.5: *Ly6a/Sca1* transcript is expressed in a subset of VSMCs. 82
Figure 4.6: Schematic representation of the approach used for profiling medial VSMC-lineage (L) and/or SCA1 (S) positive subpopulations of cells. 83
Figure 4.7: Quality control of scRNA-seq profiles generated using the Smart-seq2 protocol. 84
Figure 4.8: *Myh11* and *Ly6a/Sca1* expression levels in profiled cells..... 86
Figure 4.9: Principal component analysis of profiled subpopulations of cells. 87
Figure 4.10: Highly variable genes among the S+L+ cells. 89
Figure 4.11: Highly variable genes identified in S+L+ cells show high levels of co-expression and contain a contractile VSMC signature. 91
Figure 4.12: A subset of S+L+ cells shows characteristics of a synthetic VSMC phenotype. ... 94
Figure 5.1: *Ly6a/Sca1* expression levels are increased in cultured compared with *ex vivo* VSMCs. 101
Figure 5.2: SCA1 is upregulated in sorted GFP-negative medial cells during culture. 102
Figure 5.3: The proportion of S+L+ VSMCs increases with increased time period between lineage labelling and analysis. 103

Figure 5.4: Cultured VSMCs express transcriptional signatures of activated S+L+ cells from healthy arteries.....104
Figure 5.5: SCA1 is upregulated in VSMCs 8 days after carotid ligation injury. 105
Figure 5.6: Quality control of VSMCs profiled 7 days after carotid ligation injury. 106
Figure 5.7: Principal components 1 and 2 separate cells showing lower expression levels of the contractile marker *Myh11* and higher expression levels of *Ly6a/Sca1* and *Mki67*.....107
Figure 5.8: Clustering analysis of profiled VSMCs after carotid ligation injury. 108
Figure 5.9: Expression levels of VSMC marker genes *Myh11* and *Acta2* in VSMCs profiled following carotid ligation injury.....109
Figure 5.10: *Mki67* and *Ly6a/Sca1* expression levels among the cells profiled after carotid ligation injury. 110
Figure 5.11: Transcriptional signatures of identified clusters of VSMCs following carotid ligation injury. 112
Figure 5.12: Quality control metrics in identified clusters of cells following vascular injury.113
Figure 5.13: Summarised expression levels of the cVSMCneg and cVSMCpos signatures among cells profiled after carotid ligation injury. 114
Figure 5.14: *Col8a1* and *Spp1*, which are characteristic of the synthetic state of VSMCs, are expressed among the *Ly6a/Sca1*-positive cells. 115
Figure 5.15: Quality control of single-cell transcriptomes of VSMCs from atherosclerotic plaques. 117
Figure 5.16: Clustering analysis of VSMCs from atherosclerotic arteries. 118
Figure 5.17: Differential gene expression among identified clusters of VSMCs from atherosclerotic arteries. 120
Figure 5.18: Contractile VSMC marker and *Ly6a/Sca1* expression in profiled cells from atherosclerotic arteries. 121
Figure 5.19: Summarised expression levels of the cVSMCpos and cVSMCneg signatures in plaque VSMCs. 123
Figure 5.20: Identification of alternative markers of *Ly6a/Sca1*-positive VSMCs.....125
Figure 5.21: Expression levels of *Fbln2* and *Slco2a1* overlap with *Ly6a/Sca1* expression in model systems of VSMC phenotypic switching. 127

List of abbreviations

AA	Aortic arch
ACTA2	α-smooth muscle actin
Adv	Adventitial
APC	Allophycocyanin
ApoE	Apolipoprotein E
BAC	Bacterial artificial chromosome
bp	Base pair
CFP	Cyan fluorescent protein
COL8A1	Collagen α-1(VIII) chain
DT	Descending thoracic aorta
ECM	Extracellular matrix
ERCC	External RNA controls consortium
eYFP	Enhanced yellow fluorescent protein
FACS	Fluorescence activated cell sorting
FDR	False discovery rate
FPKM	Fragments per kilobase million
FSC	Forward scatter
GEO	Gene expression omnibus
GO	Gene ontology
GPF	Green fluorescent protein
GRCm38	Genome Reference Consortium Mouse Build 38
H3K4me2	Histone 3 lysine 4 dimethylation
H3K9me3	Histone 3 lysine 9 trimethylation
Hox	Homeobox
IgG	Immunoglobulin G
IVT	*In vitro* transcription
KLF4	Kruppel-like factor 4
L+/L-	VSMC lineage-label positive/negative
MGP	Matrix Gla protein
mRNA	Messenger RNA

MYH11	Smooth-muscle myosin heavy chain
NF-κB	Nuclear factor kappa B
NIR	Near infrared
NO	Nitric oxide
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDGF	Platelet-derived growth factor
Poly(A)	Polyadenylated
RFP	Red fluorescent protein
RNA-seq	RNA sequencing
ROC	Receiver operating characteristic
RPKM	Reads per kilobase million
RT	Reverse transcription
S+/S-	SCA1-positive/SCA1-negative
SCA1	Stem cell antigen 1
scRNA-seq	Single-cell RNA sequencing
SM22α	Transgelin
SM-MHC	Smooth-muscle myosin heavy chain
SPP1	Osteopontin
SRF	Serum response factor
TAGLN	Transgelin
TGF-β	Transforming growth factor beta
TPM	Transcripts per million
t-SNE	t-distributed stochastic neighbor embedding
UMAP	uniform manifold approximation and projection
UMI	Unique molecular identifier
VCAM1	Vascular adhesion molecule 1
VEGF	Vascular endothelial growth factor
VSMC	Vascular smooth muscle cell
αSMA	α smooth muscle actin

1 Introduction

1.1 The cardiovascular system

Single-cell organisms uptake oxygen and nutrients and excrete waste products through diffusion. As the organisms get larger and more complex, their volume to surface area ratio increases and diffusion across the outer surface of the organism becomes too inefficient to supply the entire organism with the required nutrients (Monahan-Earley et al. 2013). To overcome this problem, complex organisms evolved a circulatory system (Monahan-Earley et al. 2013).

In mammals, the cardiovascular system consists of the heart, the blood vessels and the blood. The latter acts as a carrier of oxygen, nutrients and waste products and is pumped through the blood vessels by the heart. The mammalian cardiovascular system is composed of two closed circuits; the pulmonary circuit and the systemic circuit. The role of the pulmonary circuit is to re-oxygenate the blood, which has returned from the systemic circuit. The right atrium of the heart receives the de-oxygenated blood, which is then pumped to the lungs through the pulmonary artery by the right ventricle. The large surface area of the lungs enables efficient diffusion of fresh oxygen and waste carbon dioxide between the air and the blood. Oxygenated blood returns to the heart through the left atrium and the left ventricle of the heart then pumps the freshly oxygenated blood into the systemic circuit. The blood is first pumped into the aorta, which branches into progressively narrower arteries and to the arterioles. These lead into the capillaries, where the exchange of nutrients, oxygen and waste takes place through diffusion. The deoxygenated blood then passes into the venules and returns to the right atrium of the heart through the veins (Aaronson et al. 2012).

Overall, the cardiovascular system vastly increases the effective surface area through which the diffusion can take place and allows cells throughout the body of complex organisms to have a supply of oxygen. Interestingly, the largest blood vessels in human themselves have a network of smaller blood vessels within their walls called *vasa vasorum*. These supply the cells within the blood vessel wall with blood as the wall is too thick for sufficient diffusion to take place through the lumen (Williams & Heistad 1996).

1.2 Artery wall structure

The large arteries need a strong and elastic wall in order to withstand and regulate the blood pressure. The artery wall consists of the intimal, medial and adventitial layers, which surround the lumen (Aaronson et al. 2012). The structure of the artery wall is shown in Figure 1.1. The intimal layer is the innermost layer of the artery wall. Its inner lining is formed of a single continuous layer of endothelial cells, which are supported by a subendothelial layer of connective tissue (Palotie et al. 1983). In the larger human arteries, the intimal layer also contains VSMCs (Aaronson et al. 2012). Endothelial cells act as a barrier between the blood and the rest of the artery wall and regulate the selective transport of molecules into the wall (Galley & Webster 2004). In addition, endothelial cells are involved in the regulation of vascular tone through production of vasoconstrictors and vasodilators, such as nitric oxide (Ignarro et al. 1987; Galley & Webster 2004). Overall, endothelial cells are involved in tissue homeostasis and respond to the local environment, such as shear stress induced by blood flow (Rajendran et al. 2013). Endothelial dysfunction contributes to many cardiovascular diseases, including atherosclerosis (Gimbrone & García-Cardeña 2016).

The medial layer of the aorta is composed of vascular smooth muscle cells (VSMCs) and extracellular matrix components and is separated from the intimal layer by the internal elastic lamina (Aaronson et al. 2012). Within the medial layer, VSMCs are arranged circumferentially in concentric layers around the lumen that are separated by the extracellular matrix (Wolinsky & Glagov 1967; Dingemans et al. 2000). The extracellular matrix is an ordered structure composed largely of elastin, collagens and proteoglycans and provides strength and flexibility to the medial layer (Stegemann et al. 2005). VSMCs are linked to the surrounding extracellular matrix via adhesion receptors, such as integrins, syndecans and dystroglycan (Moiseeva 2001). In addition to providing structural integrity, the interaction of VSMCs with the surrounding extracellular matrix plays a role in maintaining VSMC contractility through mechanical interactions (Moiseeva 2001; Stegemann et al. 2005).

The adventitial layer is the outermost layer of the artery wall and is separated from the medial layer by the external elastic lamina (Aaronson et al. 2012). Adventitia is the most heterogeneous layer of the vasculature and contains a diverse range of cell types and structures, such as fibroblasts, resident immune cells, progenitor cells, nerves as well as *vasa vasorum* in certain large vessels (Stenmark et al. 2013). Given the presence of nerves and the

vasa vasorum, the adventitial layer has an important function in transducing signals and nutrients to the inner layers of the artery wall (Stenmark et al. 2013). Adventitia has been observed to remodel extensively in response to injury or inflammation and is additionally thought to be involved in the process of vascular inflammation through resident immune cells, such as macrophages and dendritic cells (Maiellaro & Taylor 2007). Adventitia has also been identified as a niche for vascular stem cells and progenitor in recent years, which will be discussed in more detail in Section 1.7.

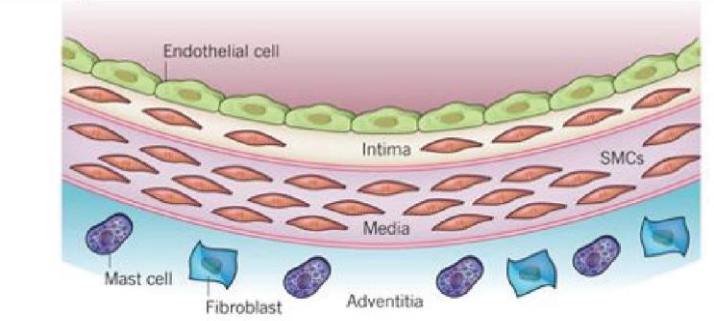


Figure 1.1: Artery wall structure.

Large arteries are composed of concentric intimal, medial and adventitial layers. A layer of endothelial cells in the intima surrounds the lumen of the artery and forms direct contact with the blood. The intimal layer of large arteries in some organisms also contain VSMCs. The medial layer is separated from the intima by the internal elastic lamina and is composed of layers of VSMCs and the extracellular matrix. This layer provides structural support to the blood vessel wall. The outermost layer is the adventitia, which contains a variety of cells and structures, such as fibroblasts and resident immune cells, as well as nerves and *vasa vasorum*. The figure is from Libby et al. (2011)

1.3 Vascular smooth muscle cells (VSMCs)

VSMCs are highly specialised cells, which play a key role in the regulation of blood flow and pressure through contraction (Owens et al. 2004). In healthy arteries, VSMCs generally exist in a so-called contractile state, in which they exhibit an elongated cell shape and contain an abundance of contractile filaments (Rensen et al. 2007). In this state, VSMCs express several characteristic marker proteins including smooth-muscle myosin heavy chain (MYH11/SM-MHC), α-smooth muscle actin (ACTA2/α-SMA) and transgelin (TAGLN/SM22α), with MYH11

being the most specific VSMC marker throughout development (Miano et al. 1994; Rensen et al. 2007). Contractile VSMCs show low levels of proliferation and synthesis of extracellular matrix components (Owens 1995).

The remarkable feature of VSMCs is their high degree of plasticity. Upon inflammation or injury, contractile VSMCs can transition to a so-called synthetic state during the process known as phenotypic switching or modulation (Figure 1.2). As VSMCs transition towards the synthetic state, they adopt a more rounded cell shape and show reduced contractile function and reduced expression of several VSMC markers proteins (Rensen et al. 2007). Synthetic VSMCs show higher levels of proliferation and migration than cells in the contractile state and increase the synthesis of extracellular matrix components, such as collagen, proteoglycans and elastin (Rensen et al. 2007). Phenotypically modulated VSMCs have been observed to express characteristic proteins, such as osteopontin (SPP1/OSTP), vascular cell adhesion molecule 1 (VCAM1), matrix Gla protein (MGP), collagen α-1(VIII) chain (COL8A1) and others (Rensen et al. 2007; Orr et al. 2010; Allahverdian et al. 2018). VSMC plasticity allows for dynamic blood vessel remodelling and injury repair. However it also poses a potential risk of misregulation, which may lead to cardiovascular disease (Bennett et al. 2016).

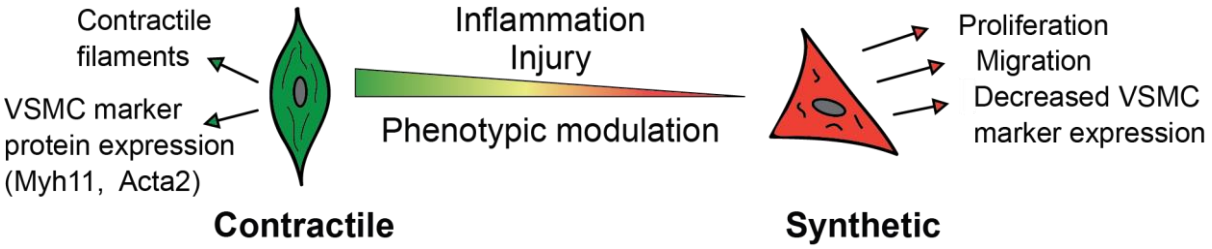


Figure 1.2: Phenotypic modulation of VSMCs.

In healthy vasculature VSMCs predominantly exist in the contractile state, where they express VSMC marker proteins and contain contractile filaments. VSMCs are capable of transitioning to a synthetic state after the stimulus of injury or inflammation. In the synthetic state they increase proliferation and migration, and decrease VSMC marker protein expression. The figure is based on Davis-Dusenbery et al. (2011).

1.4 Vascular development and the origins of VSMCs

The initial formation of the vascular system in the embryo occurs through the process of vasculogenesis. This process starts with the formation of blood islands in the extra-embryonic mesoderm of the yolk sac, which are lined by endothelial precursors (Ferguson et al. 2005). The endothelial precursors throughout the yolk sac subsequently merge to form the vascular plexus, which is the foundation for further development of the vascular system in the embryo (Coward & Wells 2013; Schmidt et al. 2007). This initial network is then extensively remodelled and extended through the process of angiogenesis during development (Ferguson et al. 2005). During vascular development, VSMC precursors are recruited to the forming vasculature and differentiated to provide structural support (Pfaltzgraff & Bader 2015). The recruitment process and the differentiation of VSMCs is mediated by many factors, including secretion of platelet-derived growth factor (PDGF)-BB by the endothelial cells (Hellstrom et al. 1999; Carmeliet 2000; Yao et al. 2014). PDGF-BB is also thought to be important for the patterning of the outer layers of the vascular wall (Greif et al. 2012). During development, VSMCs are involved in the maturation of the vascular system through the synthesis of extracellular matrix components, proliferation and migration (Owens et al. 2004).

VSMCs recruited to different parts of the vasculature have different embryonic origins (Majesky 2007). This observation was first made in experiments using quail and chick chimeras (Le Lièvre & Le Douarin 1975; Majesky 2007). In these experiments the neural tube and neural crest from one of the organisms were grafted into the embryo of the other and the different nuclear staining patterns enabled the tracking of neural cells in the recipient organism. These experiments revealed that VSMCs located in parts of the vasculature had neural origins (Le Lièvre & Le Douarin 1975). Later lineage tracing studies have confirmed that VSMCs in the carotid arteries and the aortic arch are derived from the neural crest (Jiang et al. 2000; Nakamura et al. 2006). Jiang et al. (2000) and Nakamura et al. (2006) used a Wnt1-Cre and P0-Cre dependent reporter systems respectively to trace cells originating from the neural crest in mammalian development and observed that VSMCs derived from the neural crest localise to the ascending aorta and the aortic arch, as well as the left and right carotid arteries in mature vasculature. Somites, which originate from the paraxial mesoderm, have also been shown to give rise to VSMCs during development (Pouget et al. 2006). VSMCs derived from the paraxial mesoderm have been observed in the aorta after transplantation of paraxial mesoderm to

create quail and chick chimeras (Pouget et al. 2006). Wasteson et al. (2008) observed that VSMCs can also be derived from somites in mouse embryos, and that somite-derived VSMCs localise to the descending thoracic aorta in the mature vasculature. Advances have also been made in the understanding of the embryonic origins of VSMCs in several other vascular beds. For example it was determined that aortic root VSMCs are derived from the secondary heart field and VSMCs located in the abdominal aorta are derived from splanchnic mesoderm (Majesky 2007). The embryonic origins of VSMCs located in different part of the vasculature are illustrated in Figure 1.3.

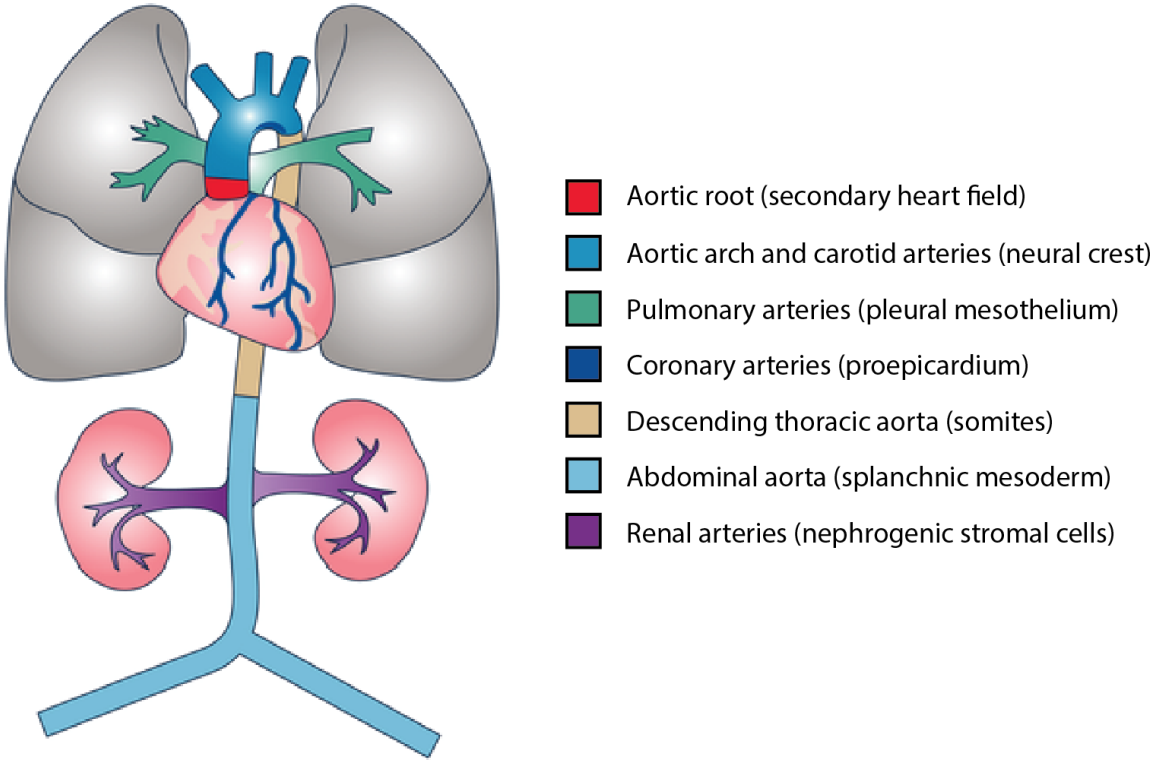


Figure 1.3: Embryonic origins of VSMCs.

VSMCs are recruited from a variety of embryonic sources. VSMCs in the aortic root (red) are derived from the secondary heart field, VSMCs in the aortic arch and the carotid arteries (medium blue) are derived from the neural crest, the pulmonary artery VSMCs (green) are from the pleural mesothelium, VSMCs in coronary arteries (dark blue) originate from the proepicardium, the descending thoracic aorta (brown) is composed of VSMCs originating from somites, VSMCs in the abdominal aorta (light blue) are derived from splanchnic mesoderm and the VSMCs in renal arteries (purple) originate from nephrogenic stromal cells. Figure adapted from Wang et al. (2015).

1.5 VSMC heterogeneity

1.5.1 Regional heterogeneity of VSMCs

There are regional differences in susceptibility to diseases, such as atherosclerosis, between different parts of the vasculature (Haimovici & Maier 1971; DeBakey & Glaeser 2000; Leroux-Berger et al. 2011). These differences are thought to originate from both the distinct embryonic origins of VSMCs located in different parts of the vasculature (section 1.4) and the unique environmental factors that various blood vessels are exposed to (Cunningham & Gotlieb 2005; Majesky 2007). In particular, lineage tracing studies have revealed that VSMCs in the aortic arch (AA) are derived from the neural crest, while VSMCs in the descending thoracic aorta (DT) originate from the paraxial mesoderm (Majesky 2007). The AA region is known to be more prone to atherosclerosis than the DT region (Van Assche et al. 2011; Trigueros-Motos et al. 2013; Sinha et al. 2014). This is likely to be at least partially due to environmental factors, with the curvature of the aortic arch and several branch sites creating areas of disturbed blood flow (Cunningham & Gotlieb 2005; Chiu & Chien 2011). However, intrinsic factors are also thought to influence different atherosclerotic susceptibilities of arteries from different parts of the body. VSMCs derived from the neural crest or the mesoderm, which were isolated from chick embryos, were observed to respond differently to the same culture conditions, particularly to exposure to the transforming growth factor (TGF)-β (Topouzis & Majesky 1996). Transcriptional differences between the neural crest and mesodermal VSMCs isolated from the AA and DT regions respectively were also observed, with a number of differentially expressed genes thought to be related to distinct developmental origins (Van Assche et al. 2011; Trigueros-Motos et al. 2013). For example, several *Hox* genes are expressed at higher levels in the DT region compared with the AA (Van Assche et al. 2011; Trigueros-Motos et al. 2013). Trigueros-Motos et al. (2013) observed an inhibitory relationship between HOXA9 and the proinflammatory transcription factor NF-κB (encoded by *Nfkb1*) in VSMCs, which has been previously also described in endothelial cells (Trivedi et al. 2007). Higher NF-κB activity in the AA region may provide a possible explanation for higher atherosclerosis susceptibility of the AA through increased inflammation (Trigueros-Motos et al. 2013).

To address the confounding enviromental factors affecting the AA and DT regions, Cheung et al. (2012) used *in vitro* differentiation to model VSMCs with neuroectodermal or mesodermal

origins. Several *Hox* genes, including *Hoxa9*, were found to be expressed at higher levels in mesodermal VSMCs compared with neuroectodermal VSMCs, both of which have been derived from human pluripotent stem cells *in vitro* (Cheung et al. 2012; Trigueros-Motos et al. 2013). In further support of the hypothesis that intrinsic differences between vascular beds party contribute to observed variation in disease susceptibility, a study in which grafts were transplanted from the disease-prone abdominal aorta to the disease-resistant jugular vein region, showed that grafts from the abdominal aorta retained a higher susceptibility for atherosclerosis in a different environment (Haimovici & Maier 1971). However, the mechanisms through which the observed intrinsic differences contribute to disease susceptibility are not yet fully understood.

1.5.2 VSMC heterogeneity within a vascular bed

There is also heterogeneity among VSMCs residing in the same vascular bed, despite them experiencing similar environmental conditions and sharing the same embryonic origin. VSMCs within a vascular bed show considerable cell-to-cell variability in the expression levels of contractile VSMC proteins, as well as adhesion molecules (Frid et al. 1994; Moiseeva 2001). Studies in a variety of organisms have further suggested that there were subpopulations of VSMCs resident in the media, which are characterised by different patterns of VSMC contractile marker expression or different responses to *in vitro* culture (Bochaton-Piallat et al. 1996; Frid et al. 1997; Li et al. 2001; Hao et al. 2002). These studies also observed that individual VSMCs had different propensities for phenotypic modulation, leading to speculation that VSMCs which adopt the extreme synthetic state in culture may be the VSMCs involved in neointima formation in disease (Bochaton-Piallat et al. 1996; Frid et al. 1997). These observations led to the hypothesis that VSMCs may exist on a spectrum between different phenotypic states (Rensen et al. 2007).

1.5.3 Functional heterogeneity of VSMCs

Benditt & Benditt (1973) observed that the entire plaque caps regularly showed inactivation of the same X chromosome, which was not observed to be the case in the underlying media. This observation led to the hypothesis that VSMCs in the plaque cap may have monoclonal origins (Benditt & Benditt 1973). However, a later study reported that larger patches of cells with the same inactivated X-chromosome can also exist in the media, which could explain the

previous observation without the implication of monoclonal origin of the plaque cap (Murry et al. 1997; Schwartz & Murry 1998). Recent advances in VSMC-specific lineage tracing approaches support the hypothesis that a small subset of VSMCs proliferate during plaque development, as well as during neointima formation following injury (Feil et al. 2014; Chappell et al. 2016; Jacobsen et al. 2017). Using a multi-colour lineage labelling strategy to study VSMC proliferation after injury in mice, Chappell et al. (2016) observed that the neointima induced by carotid ligation was composed of monochrome patches derived from a single VSMC. Furthermore, atherosclerotic plaques were observed to be composed of either single monochrome patches or a small number of monochrome patches, suggesting that only a fraction of VSMCs proliferated extensively during plaque formation (Chappell et al. 2016; Jacobsen et al. 2017). These observations suggest that VSMCs within a vascular bed may be functionally heterogeneous, with only a small subset of VSMCs responding to stimuli through extensive proliferation and accumulation in the lesion.

1.6 VSMC response to stimulus

In healthy arteries, VSMCs predominantly exist in the contractile state. As described in Section 1.3, certain stimuli such as inflammation, mechanical injury, or biochemical factors, can induce VSMC phenotypic modulation. This process occurs, for example, during *in vitro* culture as well as *in vivo* following vascular injury or development of atherosclerosis (Rensen et al. 2007).

1.6.1 Mechanisms of VSMC phenotypic switching

The process of VSMC phenotypic switching is regulated through a complex network of factors, many of which are not yet fully understood. Environmental factors, such as the surrounding extracellular matrix (ECM) components, biochemical factors and mechanical forces play an important role in influencing the phenotypic modulation of VSMCs (Rensen et al. 2007). Some of these factors promote the maintenance of the differentiated contractile VSMC phenotype, while others stimulate phenotypic modulation towards the synthetic state. The effect of the majority of studied ECM components on VSMC phenotype is to promote the contractile state (Rensen et al. 2007). For example, the deletion of heparan-sulfate side chains from the proteoglycan perlecan lead to increased VSMC proliferation *in vitro* as well as after carotid ligation injury *in vivo* (Tran et al. 2004). Collagen type IV and laminin are also thought to

promote the contractile state and lead to increased expression of contractile VSMC marker proteins when applied to cultured VSMCs (Thyberg & Hultgårdh-Nilsson 1994). There are also ECM components, however, which have been shown to promote the synthetic phenotype. For example, fibronectin and hyaluronan have both been shown to induce VSMC proliferation (Hedin et al. 1988; Evanko et al. 1999). Biochemical factors also influence the phenotypic state of VSMCs, with the most extensively studied examples being PDGF-BB and TGF-β. These two growth factors have diverging effects, with PDFG-BB promoting VSMC proliferation, while TGF-β promotes the contractile state of VSMCs (Rensen et al. 2007). Some effects of mechanical forces on the VSMC state are mediated through the endothelium, such as shear stress (Zhao et al. 2015). Mechanical stress can also act on VSMCs directly when the vessel wall is stretched periodically due to blood pressure. Mechanical stress has generally been observed to promote the contractile phenotype, with reduced mechanical stress levels, such as during culture, promoting phenotypic modulation towards the synthetic state (Rensen et al. 2007). Another environmental factor affecting VSMC phenotypic switching is vascular inflammation. NF-κB is one of the key inflammatory regulators in the vasculature (Brasier 2010), and a VSMC-specific inhibition of NF-κB has been shown to result in decreased neointima formation following injury (Yoshida et al. 2013).

Advances have also been made in the understanding of how transcriptional and epigenetic regulation impacts the phenotypic state of VSMCs. Transcription factors myocardin, SRF and KLF4 have so far been identified as central to transcriptional regulation of VSMC modulation (Allahverdian et al. 2018). One of the most extensively characterised models of transcriptional regulation in VSMCs is the SRF/myocardin regulation through CArG elements. Myocardin is expressed specifically in cardiac and smooth muscle cells during development and has been described as the master regulator of VSMC contractile phenotype (Yoshida et al. 2003; Wang et al. 2003). It acts by forming a complex with transcription factor SRF and promotes the expression of contractile VSMC genes through CArG regulatory elements (Yoshida et al. 2003; Wang et al. 2003). In contrast, KLF4 has been observed to be key in the process of de-differentiation of VSMCs during phenotypic switching (Yoshida et al. 2008; Shankman et al. 2015). KLF4 is thought to influence de-differentiation of VSMCs through multiple mechanisms, including disruption of the SRF/myocardin complex association with contractile gene promoters, repression of myocardin expression and mediation of chromatin modification (Liu et al. 2005; McDonald et al. 2006). The epigenetic regulation of VSMC phenotypic modulation

has also been investigated in recent years. For example, it has been observed that induction of VSMC phenotypic switching through PDGF-BB in culture resulted in increased repressive methylation (H3K9me3) and reduced activating histone acetylation at contractile gene promoters, with recruitment of histone deacetylases mediated by KLF4 (McDonald et al. 2006; Yoshida et al. 2007; Gomez et al. 2015). Interestingly, studies of epigenetic regulation of VSMC modulation have identified H3K4me2 at the contractile gene promoters as a possible lineage mark of VSMCs, which is retained following phenotypic switching (Gomez et al. 2015). Additionally, microRNAs have also been implicated in the regulation of VSMC phenotypic switching, with both pro-contractile and pro-synthetic effects. microRNA from cluster 143/145 promotes the contractile state (Cordes et al. 2009), whereas several other microRNAs have been observed to decrease VSMC differentiation and increase proliferation (Maegdefessel et al. 2015).

1.6.2 VSMCs in culture

In vitro culture is a commonly used model system for studying VSMC response to stimulus and subsequent phenotypic modulation. *In vitro* culture replicates some of the hallmarks of VSMC phenotypic switching, such as downregulation of contractile proteins, increased proliferation and a change in cell shape (Chamley-Campbell et al. 1979). It provides a practical way of mimicking and investigating the response of VSMC to stimuli, which are thought to influence their involvement in disease and vascular injury repair. Culture of VSMCs has provided numerous important insights into the biology of VSMC phenotypic modulation, several of which have also been verified *in vivo* (Rensen et al. 2007). However, in the absence of *in vivo* validation, findings from cultured VSMCs must be interpreted with caution, since culture conditions do not accurately represent the physiological conditions under which VSMCs are phenotypically modulated. For example, mechanical forces and ECM interactions occurring in intact arteries generally promote the contractile phenotype (discussed in the previous section, Rensen et al. 2007), therefore the process of phenotypic modulation *in vivo* is likely influenced by a complex set of opposing factors, some of which promote the contractile and some the synthetic state.

1.6.3 VSMCs in atherosclerosis

Atherosclerosis is an inflammatory disease, in which a number of different cell types are involved, including endothelial cells, VSMCs and immune cells (Aaronson et al. 2012). During atherosclerosis, plaques form in the blood vessel wall, which protrude into the lumen and may obstruct blood flow when advanced (Figure 1.4). The rupture of an atherosclerotic plaque can lead to a heart attack or a stroke (Aaronson et al. 2012). The endothelial layer plays an important role in mediating early development of atherosclerosis. It is thought that high lipid levels and disturbed blood flow in areas of increased atherosclerosis susceptibility, such as artery branch points, trigger atherosclerosis-promoting changes in endothelial cells (Chiu & Chien 2011; Libby et al. 2011). This includes increased lipid and monocyte infiltration into the artery wall and secretion of PDGFs, which induces increased proliferation in the endothelium and VSMCs (Chiu & Chien 2011; Libby et al. 2011).

Phenotypic switching and proliferation of VSMCs has long been thought to play an important part in the development of atherosclerotic plaques (Ross & Glomset 1973). VSMCs are known to be involved in different processes within the plaque, such as formation of the stabilising plaque cap, vascular calcification and accumulation within the necrotic core of the plaque (Bennett et al. 2016). However, the loss of VSMC contractile markers in many plaque VSMCs can make unambiguous identification and characterisation of VSMC-derived plaque cells difficult. To overcome this problem, VSMC-specific genetic lineage tracing approaches have been developed in animal models of atherosclerosis, which have shed light on the widespread presence and diversity of VSMCs within atherosclerotic lesions (Bennett et al. 2016).

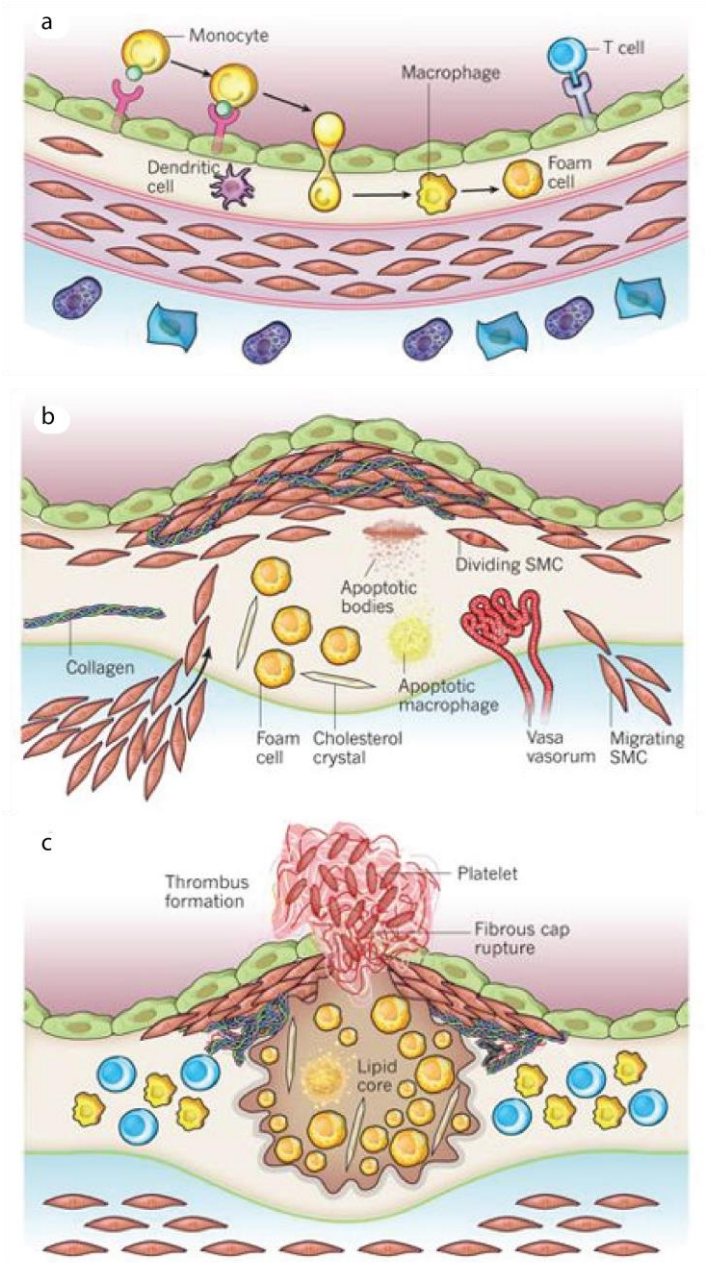


Figure 1.4: Development of the atherosclerotic plaque.

a) In the early stage of atherosclerosis circulating immune cells adhere to endothelial cells (green) and infiltrate into the blood vessel wall, alongside lipid deposits. b) VSMCs (red) migrate into the plaque where they form the fibrous cap. Plaque increases in size and protrudes into the lumen. c) Advanced atherosclerotic plaque have a lipid core and may rupture, which leads to thrombus formation. Figure from Libby et al. (2011)

The most commonly used model organism for studying atherosclerosis is the mouse, despite wild-type mice being relatively resistant to the development of atherosclerosis (Getz &

Reardon 2012). However, genetic modification led to the establishment of the ApoE$^{-/-}$ (Plump et al. 1992; Zhang et al. 1992) and Ldlr$^{-/-}$ models (Ishibashi et al. 1993) of atherosclerosis. The ApoE$^{-/-}$ model has been shown to develop progressive lesions, which are more similar in development to human atherosclerotic plaques and is the most widely used mouse model of atherosclerosis (Nakashima et al. 1994; Getz & Reardon 2012). Apart from enabling the use of genetic modifications to study progression of the disease, mouse models have the advantage of faster plaque development. In human, atherosclerotic plaques typically develop over many years, whereas in the ApoE$^{-/-}$ model plaques develop within 8 weeks of high fat diet, with advance lesions present after 15 weeks (Nakashima et al. 1994). There are, however, important limitations to the use of mouse models for studying atherosclerosis, which need to be considered when extrapolating the findings to human disease. For example, there are some differences in artery wall structure, such as thinner medial layer in mouse arteries and the absence of intimal VSMC and *vasa vasorum* in the mouse (Getz & Reardon 2012). Additionally, there are differences in atherosclerotic plaque distribution in different areas of the vasculature in the ApoE$^{-/-}$ model and in human (Nakashima et al. 1994; Getz & Reardon 2012). Despite these limitations, mouse models of atherosclerosis have greatly contributed to the understanding of this disease and have enabled investigations, such as VSMC-specific lineage tracing, which would not have been possible in human.

Studies using VSMC-specific lineage tracing in mice have suggested that 30-70% of cells within atherosclerotic plaques originate from VSMCs and that the majority of VSMC-derived cells lose VSMC marker proteins in the plaque (Gomez et al. 2013; Shankman et al. 2015; Chappell et al. 2016). In addition, recent studies have revealed that some macrophage-like cells within the atherosclerotic plaque could be derived from VSMCs as VSMC lineage-labelled cells were observed to express macrophage markers, such as LGALS3 and CD68 (Feil et al. 2014; Shankman et al. 2015; Chappell et al. 2016; Jacobsen et al. 2017).

On the whole, VSMCs within the plaque are thought to play both beneficial and harmful roles in the development of atherosclerotic plaques. For example, VSMCs in the fibrous cap synthesise extracellular matrix components and their proliferation in the event of initial plaque rupture can lead to repair and restoration of plaque stability (Bennett et al. 2016). Conversely, phenotypic switching of VSMCs to a macrophage-like state is thought to negatively contribute to the development of atherosclerosis (Shankman et al. 2015), as VSMC-

derived macrophages have been observed to clear lipids and apoptotic cells less efficiently than bone-marrow derived macrophages (Vengrenyuk et al. 2015).

The current view is that VSMCs can adopt multiple phenotypes within the atherosclerotic plaque, such as macrophage-like cells, fibroblast-like cells in the plaque cap, mesenchymal stem cell-like cells, as well as calcifying-like cells contributing to vascular calcification (Figure 1.5, Naik et al. 2012; Nguyen et al. 2013; Feil et al. 2014; Shankman et al. 2015; Chappell et al. 2016; Jacobsen et al. 2017). Observations of the clonal nature of VSMC-derived cells within the plaque and the wide range of phenotypes that VSMCs are thought to adopt in the plaque suggest that VSMCs involved in atherosclerosis have high levels of phenotypic plasticity and that a single VSMC can be phenotypically modulated and proliferate to form different subpopulations of VSMC-derived cells within the plaque (Chappell et al. 2016).

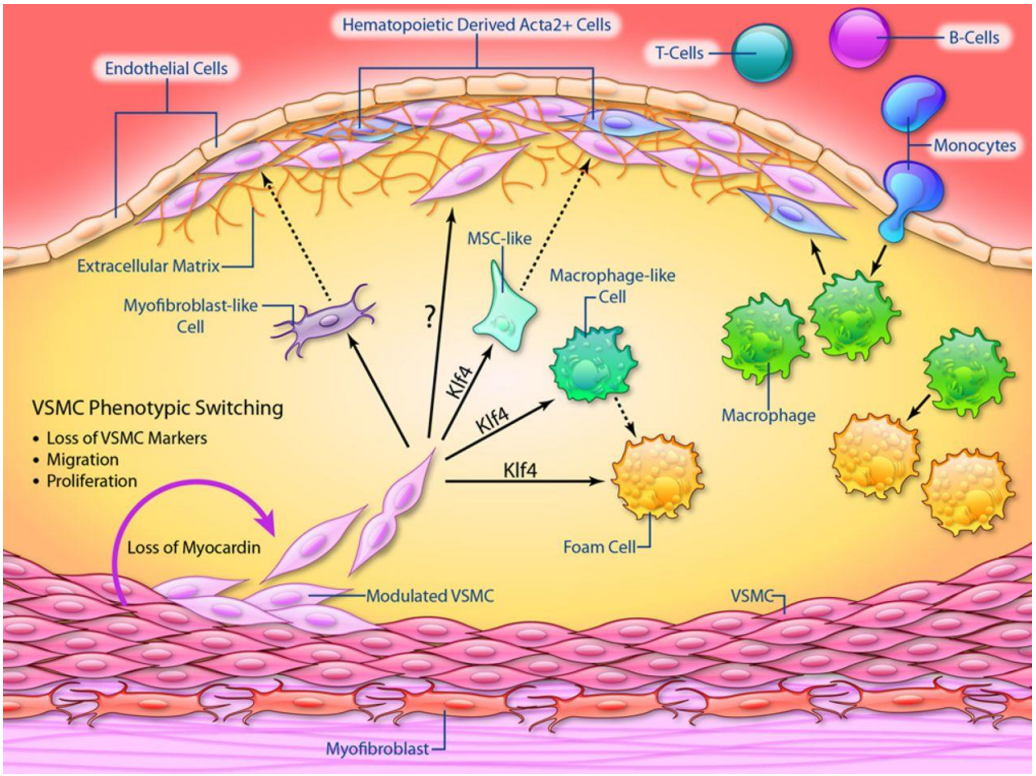


Figure 1.5: Phenotypic modulation of VSMCs in atherosclerosis.

Phenotypically modulated VSMCs migrate out of the media and proliferate during plaque development. VSMCs can undergo extensive phenotypic changes and transition to mesenchymal stem cell-like cells, macrophage-like cells, foam cells or osteochondrogenic-like cells involved in vascular calcification (not shown) among others. Figure from Bennett et al. (2016).

1.6.4 VSMCs in carotid ligation injury

Carotid ligation is a vascular injury model, which provides an acute and reproducible environment for studying VSMC phenotypic switching (Kumar & Lindner 1997). During carotid ligation surgery, the left carotid artery is mechanically tied to stop blood flow. Cessation of blood flow then triggers endothelial dysfunction, which induces vascular remodelling and VSMC response (Kumar & Lindner 1997). Alternative vascular injury models also exist, for example the wire injury model, where the endothelial layer is surgically removed (Lindner et al. 1993). However, the extent of VSMC proliferation and participation in the neointima formation appears to be greater in the carotid ligation model (Kumar & Lindner 1997).

The model of vascular injury provides an alternative setting for the investigation of VSMC plasticity to the atherosclerotic plaque. Studies have shown that VSMCs are extensively involved in the formation of neointima following carotid ligation injury and that neointimal VSMCs are derived from VSMC, which were in a differentiated state prior to injury (Herring et al. 2014; Chappell et al. 2016). Vascular injury models also provide a suitable setting for investigations of VSMC heterogeneity. Clonal or oligoclonal expansion of VSMCs (discussed in Section 1.5.3) is also observed following carotid ligation injury, which suggests that VSMCs are functionally heterogeneous in their response to injury (Chappell et al. 2016).

1.7 Origins of phenotypically switched VSMCs

The origin of VSMCs involved in the atherosclerotic plaque and neointima formation has been extensively debated in recent years. Several studies have asked whether VSMCs within lesions arise from previously differentiated VSMCs, from cells migrating from other vascular layers, or from alternative vascular progenitor populations. In support of the latter possibility, Sata et al. (2002) reported that the majority of VSMCs within the atherosclerotic plaque were derived from circulating hematopoietic stem cells derived from the bone marrow. However, a later report contradicted this observation and claimed that plaque VSMCs were derived exclusively from the local vasculature (Bentzon et al. 2006), although ACTA2 expression has since been observed in some bone-marrow-derived monocytes located in the neointima following wire injury (Iwata et al. 2010). Alternative origins of de-differentiated VSMCs from within the vasculature have also been proposed. VSMCs in the neointima of a pulmonary

hypertension model were proposed to originate from endothelial cells based on the observation that Tie2-Cre lineage traced cells upregulated ACTA2 and MYH11 (Qiao et al. 2014). Hu et al. (2004) suggested that adventitial progenitors could give rise to VSMCs within atherosclerotic plaques. These adventitial progenitors expressed Stem cell antigen 1 (SCA1) and were observed to migrate to the plaque and express VSMC markers after isolated and labelled adventitial SCA1+ cells were transplanted into the adventitia (Hu et al. 2004). Furthermore, Passman et al. (2008) reported that SCA1, CD34 and PDGFRB expressing adventitial cells were capable of differentiation into VSMCs *in vitro*. In addition, Klein et al. reported vascular-wall resident stem cells expressing CD44, CD90 and CD73, but not CD34 and CD45 markers in the adventitial layer (Klein et al. 2011). These vascular-wall resident stem cells upregulated VSMC markers following culture with TGF-β and were observed to contribute to neovascularisation after being grafted with endothelial cells *in vivo* (Klein et al. 2011). More recently, Kramann et al. (2016) used an inducible lineage tracing strategy to trace Gli1+ cells (adventitial location in healthy arteries) following vascular wire injury. Under these conditions, lineage labelled cells were observed to express VSMCs marker genes (Kramann et al. 2016). Given that Gli1+ cells also expressed *Sca1*, *Cd34* and *Pdgfrb*, these cells may overlap with previously reported adventitial progenitor populations (Kramann et al. 2016; Passman et al. 2008; Hu et al. 2004).

Vascular progenitor populations have also been proposed to exist in the medial layer. Sainz et al. (2006) reported a SCA1-positive subpopulation of medial cells, which were capable of differentiating into endothelial cells and VSMCs *in vitro*. Additionally, Tang et al. (2012) suggested that mature VSMCs do not contribute to neointima formation and that a resident progenitor population in the medial layer gives rise to phenotypically modulated VSMCs. However, the findings of Tang et al. have proven controversial and were questioned by experts in the field (Nguyen et al. 2013). Examples of raised concerns include the severity of the vascular injury model, which may have killed the medial VSMCs, and a lack of definitive lineage tracing approaches showing that the neointimal VSMCs originate from the proposed progenitor population (Nguyen et al. 2013).

Recent studies using inducible lineage tracing strategies have suggested that cells expressing differentiated VSMC markers prior to injury or atherosclerosis give rise to neointimal VSMCs. Nemenoff et al. (2011) used a tamoxifen-inducible Myh11-Cre lineage tracing strategy and

observed that VSMCs, which expressed contractile VSMC markers and thus the lineage label prior to injury, contributed to neointima formation. Differentiated VSMCs have also been reported to contribute to atherosclerotic plaques using a similar inducible lineage tracing strategy (Gomez et al. 2013). In addition, Gomez et al. (2013) reported that the contribution of VSMCs to the atherosclerotic plaque has likely been underestimated previously, as several VSMC-derived cells no longer expressed VSMC markers. Further studies using inducible and VSMC-specific lineage tracing approaches have confirmed these observations (Feil et al. 2014; Shankman et al. 2015; Chappell et al. 2016; Jacobsen et al. 2017).

A recent study, which investigated the origins of neointimal cells in vascular injury models of different severities, may provide a partial explanation for the contradicting observations about the origins of neointimal VSMCs reported in the literature (Roostalu et al. 2018). Roostalu et al. (2018) observed that VSMC-derived cells gave rise to the neointima in a limited wire injury model, but that adventitial cells contributed to repair following a more severe microanastomosis injury. Overall, there is strong evidence that cells of the VSMC lineage can expand to give rise to neointimal and plaque VSMCs. Further studies are required to determine whether the rare expanding VSMCs possess any special characteristics and to determine whether they are the sole contributor to lesional VSMCs, or whether multiple origins may contribute to the VSMC pool in disease simultaneously.

1.8 Stem cell antigen 1 (SCA1): a marker of stem/progenitor populations expressed by a subset of vascular cells

Stem cell antigen 1 (SCA1) is part of the Ly6 family and is a cell surface protein encoded by the gene *Ly6a* in mice (Holmes & Stanford 2007). SCA1 is used as a marker protein for the enrichment of hematopoietic stem cells (Spangrude et al. 1988) and is expressed in stem/progenitor populations in a wide range of tissues in the mouse (Holmes & Stanford 2007). For example, the precursor cells in adult skin tissue express SCA1 (Fernandes et al. 2004). Skin-derived precursor cells have been shown to be capable of *in vitro* differentiation into a variety of cell types, such as adipocytes and neural cells (Toma et al. 2001; Fernandes et al. 2004), vascular smooth muscle cells (Steinbach et al. 2011) and hepatocytes (Rodrigues et al. 2014). SCA1-positive progenitors have also been observed in the cardiovascular system, both in the vasculature (discussed in Section 1.7) and in the heart (Oh et al. 2003; Matsuura

et al. 2004). SCA1-positive cardiac progenitors were observed to have the capacity to differentiate into cardiomyocytes *in vitro* (Oh et al. 2003; Matsuura et al. 2004) and an increase in SCA1-positive cells was observed following myocardial infarction (Wang et al. 2006). However, recent studies suggested that SCA1-positive cells do not contribute to cardiomyocytes *in vivo* and are instead part of the endothelial cell population (Vagnozzi et al. 2018; Zhang et al. 2018).

SCA1 has no known human orthologue and its function in mice is not yet fully understood (Holmes & Stanford 2007). However, the relevance of SCA1 extends beyond being a marker of stem/progenitor populations, as demonstrated by the alterations observed in SCA1 knockout mice in several tissues (Holmes & Stanford 2007). For example, studies carried out in SCA1 knockout mice suggested an impact of this gene on haematopoiesis (Ito et al. 2003; Bradfute et al. 2005), T-cell proliferation (Stanford et al. 1997), osteoporosis (Bonyadi et al. 2003), as well as on cardiac function (Bailey et al. 2012). At the molecular level, SCA1 is thought to be involved in cell signalling, however its mechanism of action is poorly understood (Holmes & Stanford 2007). Better understanding of SCA1 function in mice may aid in the identification of its human orthologue, which may help in translating findings observed in mice to human in the future (Holmes & Stanford 2007).

1.9 Single-cell transcriptomics as a tool for studying cell heterogeneity and plasticity

Single-cell transcriptomics enables genome-wide measurements of mRNAs at the single-cell level, and as such can offer several advantages over the conventional population-level RNA-seq. For example, cell-to-cell heterogeneity at a global transcriptome-wide level can be addressed using single-cell RNA-seq (scRNA-seq) (Stegle et al. 2015). Transcriptome-wide analysis of cell-to-cell heterogeneity allows for the identification and characterisation of rare subpopulations of cells (Stegle et al. 2015), whereas such signals are typically lost in conventional RNA-seq analyses of heterogeneous cell populations. Single-cell analysis also allows for any contaminating cells to be identified and excluded from the analysis, while this is not possible in population-level approaches (Kaur et al. 2017).

1.9.1 Approaches for scRNA-seq library generation

scRNA-seq was first developed by Tang et al. (2009) to enable transcriptome-wide profiling in applications where only low amounts of biological material were available. Since then, several protocols for scRNA-seq have been developed and used extensively to profile tissue heterogeneity, characterise cell differentiation and identify new subpopulations of cells (Stegle et al. 2015). Several challenges needed to be addressed during the development of experimental protocols for wider application of scRNA-seq to become possible. These included the development of efficient methods for capturing single cells, and increasing the sensitivity of protocols to handle low amounts of mRNA present in individual cells (Kolodziejczyk et al. 2015).

The first step in scRNA-seq protocols is the capture of single cells (Figure 1.6). Several approaches for single-cell capture have been developed, each of which has certain advantages and disadvantages. Early studies used manual isolation of cells, which was low throughput but enabled the inspection of each captured cell to ensure viability and that only one cell was captured (Tang et al. 2009). To increase throughput, automated capture methods were also developed and used. For example, flow cytometry has been used to sort single cells into the wells of a plate, which additionally allows for the enrichment of desired populations of cells (Jaitin et al. 2014). However, subsequent library generation may be expensive due to relatively large volumes of reagents required in a 96-well plate format (Kolodziejczyk et al. 2015). Single cells can also be captured using microfluidic approaches. The commercial Fluidigm C1 platform captures single-cells on a microfluidics chip and enables subsequent processing to be carried out in small reaction volumes. The disadvantage of this method is that cells of different sizes cannot be captured on one chip (Kolodziejczyk et al. 2015), and that the number of captured cells per chip is limited. Droplet-based microfluidic approaches, which enable extremely high throughput of single-cell capture, have also been developed (Macosko et al. 2015; Klein et al. 2015). Droplet-based approaches tend to be cost effective for large numbers of profiled cells due to a small reaction volume and the high number of cells captured per experiment (Macosko et al. 2015; Klein et al. 2015). However, both the chip-based and droplet-based microfluidic approaches have relatively low capture efficiencies, and may not be suitable when the amount of starting material is the limiting factor (Kolodziejczyk et al. 2015).

Once single cells are captured, they are lysed and their mRNA is reverse-transcribed to generate cDNA (Figure 1.6). In most protocols this is achieved by exploiting the fact that most cellular mRNAs have poly(A) tails and efficient reverse transcription is an important step for the overall sensitivity of the protocol (Kolodziejczyk et al. 2015). Due to the low amounts of mRNA present in single-cells, resulting cDNA needs to be amplified prior to generation of sequencing libraries (Tang et al. 2009). Possible approaches for cDNA amplification are either *in vitro* transcription or PCR (Figure 1.6), with each method presenting certain advantages and disadvantages. PCR is an exponential amplification approach where amplification efficiencies may vary depending on sequence, while *in vitro* transcription is a linear amplification method but an additional round of reverse transcription could introduce additional 3' coverage biases (Kolodziejczyk et al. 2015). *In vitro* transcription is used in the CEL-seq (Hashimshony et al. 2012), CEL-seq2 (Hashimshony et al. 2016), MARS-seq (Jaitin et al. 2014) and inDrop (Klein et al. 2015) protocols. PCR is also used in several protocols, including the original Tang et al. (2009) protocol, Smart-seq (Ramsköld et al. 2012) , Smart-seq2 (Picelli et al. 2013; Picelli et al. 2014), STRT-seq (Islam et al. 2011), Drop-seq (Macosko et al. 2015) and the commercial 10X Genomics Chromium (Zheng et al. 2017) protocols. During PCR amplification, adapter sequences are incorporated into the cDNA to enable untargeted cDNA amplification (Tang et al. 2009). The first adapter is typically added during the reverse transcription to the poly(T) primer. The second adapter can be incorporated by different methods, for example through addition of a poly-A tail to the reverse-transcribed cDNA (Tang et al. 2009), or through template switching (Islam et al. 2011) (Figure 1.6). Template switching has the advantage of amplifying the full length of the transcript, therefore allowing the full transcript to be sequenced (Ramsköld et al. 2012; Picelli et al. 2013).

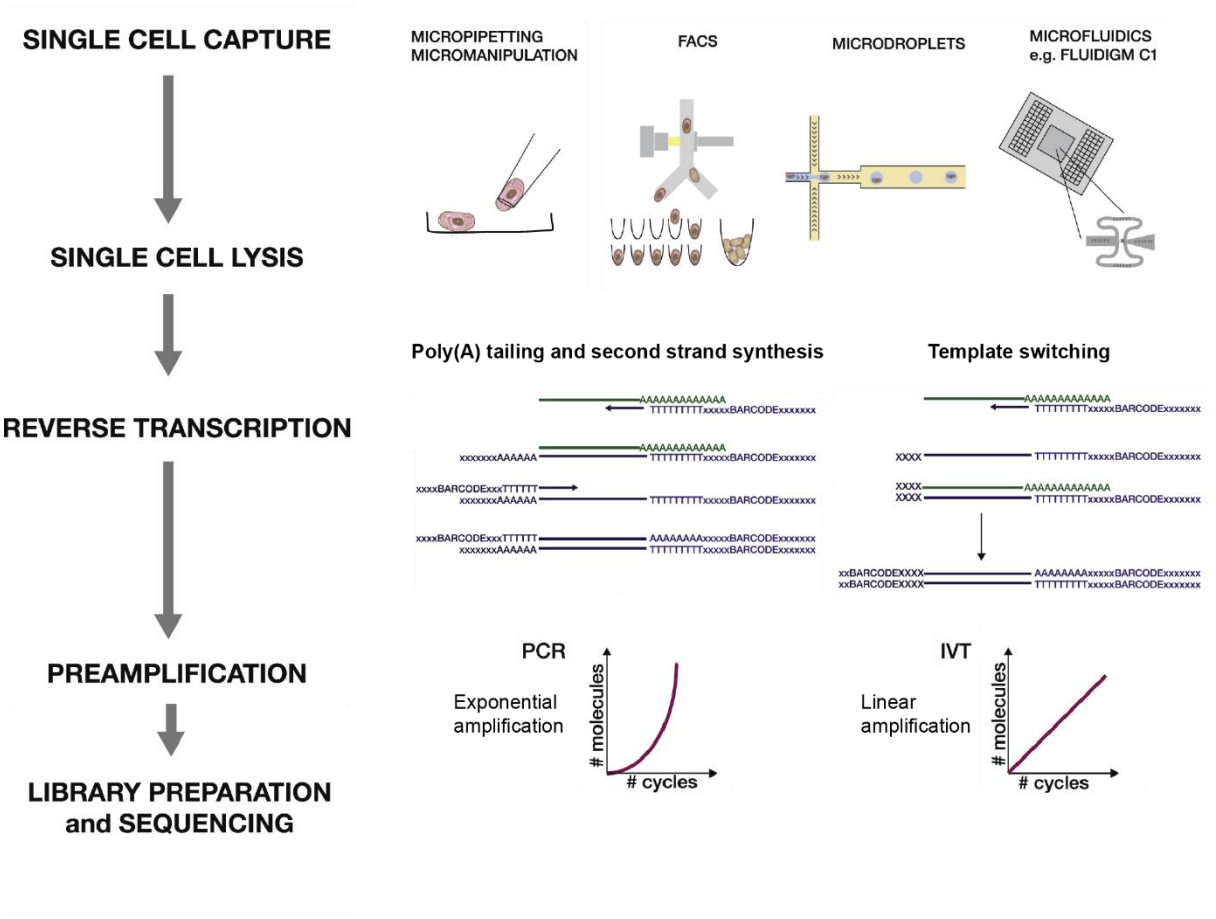


Figure 1.6: Experimental protocols for scRNA-seq.

The first step in scRNA-seq protocols is single-cell capture. This can be achieved using several different methods, including manual isolation by micropipetting, FACS, or microfluidics approaches where cells are captured in droplets or on a chip. After cell lysis, cellular mRNA is reverse-transcribed. Poly(A) tailing or template switching approaches can be used to insert barcodes into the cDNA. Resulting cDNA is amplified using either PCR or *in vitro* transcription (IVT), after which libraries are prepared for sequencing. Figure adapted from Kolodziejczyk et al. (2015).

An important part of scRNA-seq protocols, which has been crucial for increasing throughput and reducing the cost of scRNA-seq, is multiplexing (Svensson et al. 2018). Multiplexing enables libraries of many single cells to be sequenced together, and can reduce labour if introduced early in the protocol (Svensson et al. 2018). However, early introduction of multiplexing prevents the full length of the transcript from being sequenced. For the full transcript to be analysed, multiplexing needs to take place after cDNA amplification, as each fragment is required to contain the cellular barcode (Ramsköld et al. 2012). Examples of

protocols, which sequence the full-length transcript, are the Smart-seq (Ramsköld et al. 2012) and Smart-seq2 (Picelli et al. 2013; Picelli et al. 2014). The downside of multiplexing following cDNA amplification is that library preparation may be more labour intensive (Svensson et al. 2018). However, sequencing the full transcript can be advantageous in certain applications, such as when investigating alternative splicing (Ramsköld et al. 2012). Multiplexing can also be carried out prior to cDNA amplification, with cellular barcodes added during reverse transcription (Islam et al. 2011). cDNA can then be pooled for amplification. Early barcoding approaches also enable the addition of unique molecular identifiers (UMIs) to transcripts alongside the cellular barcode (Islam et al. 2014; Fu et al. 2011; Hug & Schuler 2003). Biases can be introduced during cDNA amplification and UMIs can be used to correct for uneven cDNA amplification (Islam et al. 2014). The disadvantage of this approach is that early barcoding is not compatible with the sequencing of full-length transcripts, as mentioned above. Every sequencing read needs to contain the cellular barcode to enable demultiplexing, therefore only the 3' or 5' tags of the transcript can be sequenced with early barcoding approaches (Svensson et al. 2018). Despite this limitation, early barcoding approaches are widely used and have enabled a remarkable increase in the numbers of profiled cells when combined with high-throughput capture methods (Svensson et al. 2018). Examples of protocols using early barcoding are STRT-seq (Islam et al. 2011), MARS-seq (Jaitin et al. 2014), CEL-seq2 (Hashimshony et al. 2016), inDrop (Klein et al. 2015), Drop-seq (Macosko et al. 2015) and the 10X Genomics Chromium platform (Zheng et al. 2017).

With so many different scRNA-seq protocols available, the key question is which one is best suited to answer a particular research question. It is challenging to directly compare the performance of protocols, due to different cell types and varying sequencing depths used in published studies. A benchmarking study by Svensson et al. (2017) attempted to overcome this challenge by using ERCC controls, which are a panel of artificial RNA molecules frequently added to single cells during library preparation for quality control purposes. By comparing reported numbers of added ERCCs with their read counts following sequencing, Svensson et al. (2017) estimated the accuracies and sensitivities of several scRNA-seq protocols. When varying sequencing depth was accounted for, the top performing protocols were CEL-seq2, STRT-seq and SMARTer protocols carried out on the Fluidigm C1 platform (Svensson et al. 2017). Droplet-based approaches, such as inDrop, Drop-seq and 10X Genomics Chromium, performed relatively well given their high throughput nature, particularly when varying

sequencing depth was accounted for (Svensson et al. 2017). However, there are differences between artificial ERCC controls and endogenous mRNAs, which may have impacted the performance of the protocols to different degrees (Svensson et al. 2017). Additionally, ERCCs can degrade during handling, which may have introduced technical noise to benchmarking (Svensson et al. 2017). Ziegenhain et al. (2017) used a different approach to benchmark the performance of a selection of scRNA-seq protocols. scRNA-seq libraries of mouse embryonic stem cells were generated for all six benchmarked protocols, which included Smart-seq2, Drop-seq and MARS-seq protocols (Ziegenhain et al. 2017). Smart-seq2 protocol was determined to have the highest sensitivity and captured the largest number of genes per cell (Ziegenhain et al. 2017).

Overall, different protocols for scRNA-seq library generation have various advantages and disadvantages. The most suitable protocol depends on the research question and any practical limitations, such as limited biological material. There is generally also a trade-off between the number of cells included in the study and the sequencing depth per cell due to the cost of sequencing. The droplet-based 10X Genomics Chromium platform has become widely used for single-cell transcriptome profiling in recent years, likely due to the practical convenience of an optimised commercial product and relatively high data quality (Zheng et al. 2017).

1.9.2 scRNA-seq data analysis approaches

The analysis of scRNA-seq data presents several challenges, which largely originate from the technical noise resulting from low amounts of cellular mRNA and the resulting sparsity of the data (Brennecke et al. 2013). A particular problem in scRNA-seq data are dropout events, which occur when a transcript was present in the cell at the time of reverse transcription, but was not included in the resulting library due to incomplete mRNA capture, or was not sequenced due to a low sequencing depth (Kharchenko et al. 2014). Technical challenges of noise and dropouts, alongside the large number of single cells included in most studies, mean that analysis approaches devised for bulk RNA-seq are not always applicable to scRNA-seq data. Since the development of high throughput scRNA-seq protocols, numerous analysis methods tailored specifically to scRNA-seq data have been developed instead (Stegle et al. 2015; Wagner et al. 2016; Kiselev et al. 2019).

An important early step in data analysis is quality control, as inclusion of low quality cells could skew the biological interpretation of the data (McCarthy et al. 2017). Quality control is typically carried out by filtering out cells with particularly high or low total read count, number of detected genes and proportion of mitochondrial reads. ERCC spike-in controls can also be used to eliminate poor quality cells (McCarthy et al. 2017). Good quality cells then need to be normalised to account for the differences in sequencing depths of individual cells and several different normalisation methods have been developed. Read counts of detected genes can be adjusted for sequencing depth based on the total library size or by additionally correcting for gene length using transcripts per million (TPM, Li et al. 2010), reads per kilobase million (RPKM, Mortazavi et al. 2008) or fragments per kilobase million (FPKM, Trapnell et al. 2010) approaches. Bulk RNA-seq normalisation approaches may not be well suited to scRNA-seq data, as the proportion of genes with zero counts is drastically higher in scRNA-seq (Vallejos et al. 2017). For example, the DESeq2 normalisation method (Love et al. 2014) developed for bulk RNA-seq estimates size factors for each sample based on the geometric mean of expressed genes. However, as a large proportion of genes are not detected consistently in scRNA-seq assays, this approach does not accurately reflect the sequencing depth in such sparse settings (Vallejos et al. 2017). Alternative approaches tailored specifically to sparser scRNA-seq datasets have been developed, for example the pooling normalisation approach, where size factors are estimated based on pools of single cell and subsequent deconvolution (Lun et al. 2016).

To address the problem of dropouts in scRNA-seq data, several imputation methods have been developed, such as MAGIC (van Dijk et al. 2018) and scImpute (Li & Li 2018). The goal of imputation methods is to determine whether a zero is a technical dropout or a true biological readout and to correct for dropouts arising due to technical effects. However, it may not be straightforward to distinguish between true and false negatives, which can lead to the introduction of false-positive values (Tallulah S. Andrews & Hemberg 2019). Any technical effects resulting from different batches of scRNA-seq libraries may also need to be accounted for prior to biological interpretation of the data (Stegle et al. 2015). Several approaches for batch correction have been developed for bulk RNA-seq, such as remove unwanted variation (RUV) (Risso et al. 2014) or limma (Ritchie et al. 2015) workflows. However, these approaches may remove interesting biological variation, which could result from varying cell types present in different batches or varying proportion of cell types sampled in each batch (Haghverdi et

al. 2018). scRNA-seq specific methods have been developed to address this problem, such as the mnnCorrect (Haghverdi et al. 2018), which is applicable if batches contain at least one overlapping subpopulation of cells. An alternative approach to comparing different scRNA-seq datasets is to use cell projection methods, which aim to map the transcriptional profiles onto a reference dataset and identify which cells originate from the same biological populations. For example, scmap (Kiselev et al. 2018) maps the cells in a dataset to the most similar cell or cluster of cells within a reference dataset. Another such tool for assessing similarities between cell populations profiled in different datasets is MetaNeighbour (Crow et al. 2018), which is tailored to assessing how reliable cell type annotations are across different datasets. Cell projection approaches are becoming particularly powerful with the development of large-scale reference datasets, such as the Human Cell Atlas (Regev et al. 2017) and the Tabula Muris (The Tabula Muris Consortium 2018). When reliable reference datasets are available, projection methods have the ability to assess whether cell types have been consistently annotated across datasets generated and analysed in different laboratories, as well as to assist with the assignment of cell types in newly generated datasets. Further analysis of scRNA-seq datasets is dependent on the biological question, but frequently includes dimensionality reduction to visualise the transcriptomes of profiled cells in a two or three dimensional space, clustering to identify subpopulations of cells, and differential expression analysis between identified subpopulations.

To visualise the highly dimensional scRNA-seq profiles, dimensionality reduction techniques can be employed. Various methods have been applied to scRNA-seq data analysis, such as the linear principal component analysis (PCA) and independent component analysis (ICA, Hyvärinen & Oja 2000; Trapnell et al. 2014), as well as non-linear t-distributed stochastic neighbour embedding (t-SNE, Maaten & Hinton 2008), diffusion maps (Coifman et al. 2005; Haghverdi et al. 2015) and uniform manifold approximation and projection (UMAP, McInnes et al. 2018; Becht et al. 2019). Recent benchmarking of different dimensionality reduction approaches for scRNA-seq data suggested that PCA may be best suited for applications where a low number of cells has been profiled, and t-SNE may give better results in datasets with larger numbers of cells (Li & Li 2019). However, diffusion maps and UMAP methods have not been included in this comparison (Li & Li 2019). Dimensionality reduction is frequently performed on a set of highly variable genes, which reduces the initial high dimensionality of scRNA-seq data. Highly variable genes can be estimated using several approaches, for example

based on their variance between cells, their squared coefficient of variation (Brennecke et al. 2013) or based on the relationship between their log mean expression and variance (Lun et al. 2016).

An important question arising during the analysis of heterogeneous datasets is which populations of cells have been profiled. Unsupervised clustering is a useful tool for answering such questions, and can be used for characterising the heterogeneity of the dataset and for identifying new subpopulations of cells. Numerous approaches for clustering have been developed and applied to scRNA-seq, with the majority of these approaches based on hierarchical, k-means or graph-based clustering (Kiselev et al. 2019). Several clustering approaches have been specifically adapted for scRNA-seq data to take the high dimensionality and sparsity of the data into account. These include CIDR (Lin et al. 2017) based on hierarchical clustering, SC3 (Kiselev et al. 2017) and RaceID (Grün et al. 2015) based on k-means clustering and PhenoGraph (Levine et al. 2015) and a graph-based clustering approach implemented in the Seurat package (Macosko et al. 2015; Butler et al. 2018) for graph-based clustering. Evaluating the results of clustering algorithms often requires a degree of biological knowledge about profiled cells. For example, known cell type markers are frequently used for interpreting identified clusters (Kolodziejczyk et al. 2015). Direct comparison of different clustering algorithms is therefore challenging, as there is a lack of gold standard ground truth datasets, and different clustering algorithms may specialise in certain applications, such as rare cell populations or large datasets (Kiselev et al. 2019). Due to these challenges, unsupervised clustering of scRNA-seq data is still an active area of research. Benchmarking of developed approaches on a variety of real and simulated datasets suggested that SC3 clustering and Seurat's graph-based clustering overall performed best, with the latter approach being faster and thus more applicable to larger datasets (Duò et al. 2018).

Differential gene expression analysis between clusters of cells is a common technique for investigating the characteristics and potential function of identified subpopulations of cells. Due to the sparsity of scRNA-seq data and the large numbers of cells profiled, approaches developed for bulk RNA-seq are not widely applicable. Several tailored approaches for differential gene expression analysis in scRNA-seq data have been developed, such as scde (Kharchenko et al. 2014) and MAST (Finak et al. 2015). Several standard statistical tests, such as the Wilcoxon rank sum test and the student's t-test, have also been shown to perform well

in identifying differentially expressed genes in scRNA-seq data (Soneson & Robinson 2018). Since scRNA-seq data profiles gene expression levels in individual cells, it is also possible to assess the relative variability in gene expression between groups of cells (Vallejos et al. 2015; Eling et al. 2018).

With increased throughput of scRNA-seq and advances in computational analyses, unbiased profiling of cellular heterogeneity in many different tissues became possible. scRNA-seq studies revealed that the transcriptional heterogeneity within many tissues is greater than previously thought. For example, Jaitin et al. (2014) profiled over a thousand spleen cells and characterised different subpopulations of dendritic cells, and the transcriptional heterogeneity of cells within a tumour was highlighted by Patel et al. (2014). The extent of heterogeneity of cell types in the mouse cortex and hippocampus was revealed using scRNA-seq by Zeisel et al. (2015), where several subtypes of oligodendrocytes were observed among what was previously thought to be a homogeneous cell population (Zeisel et al. 2015). Since then, many more scRNA-seq studies have been published and made their data publicly available, which created unprecedented resources for investigating the heterogeneity of different tissues and organs.

scRNA-seq is also a valuable tool for investigating continuous transitions between cell states or cell types, such as differentiation trajectories. Several analytical approaches for modelling differentiation trajectories tailored to scRNA-seq data have been developed, including Monocle (Trapnell et al. 2014), TSCAN (Ji & Ji 2016), diffusion pseudotime (Haghverdi et al. 2016) and Slingshot (Street et al. 2018). These analytical approaches have been applied to various biological systems. For example, Trapnell et al. (2014) gained insight into the transcriptional regulation of myoblast differentiation by constructing a differentiation trajectory of profiled cells and investigating the transcriptional changes along the resulting trajectory. Important insights using similar approaches have also been made in other systems, such as haematopoiesis (Paul et al. 2015). Paul et al. observed that there was a higher level of heterogeneity among myeloid progenitors than was previously known and that different progenitor populations were primed towards certain lineages, while retaining a degree of plasticity when the system was perturbed. Advancements in the understanding of differentiation and cell state transitions have also been made in many other biological

processes, including dendritic cell differentiation, T cell commitment and lymphoid differentiation (Papalexi & Satija 2018).

Computational analysis of scRNA-seq data is an area of active research and significant advances in performance and usability of the tools have been made in recent years. Comprehensive and user-friendly workflows, such as Seurat (Butler et al. 2018), scanpy (Wolf et al. 2018) and ASAP (Gardeux et al. 2017) have been developed, with Seurat in particular becoming widely used in scRNA-seq data analysis. Overall, single-cell transcriptomics is an exciting and rapidly developing field and the advances in the experimental and analytical approaches have enabled unprecedented resolution, coverage and scale in transcriptional profiling studies.

1.10 Aims and hypotheses

In this thesis I focused on the heterogeneity and plasticity of VSMC, both in healthy arteries, as well as during their response to injury and disease. I have taken advantage of scRNA-seq and lineage labelling approaches to investigate these questions at the single-cell level. Specifically, the aims of this thesis are:

1) Investigate the regional heterogeneity of VSMCs between the atherosclerosis-prone AA and atherosclerosis-resistant DT regions of the aorta, and to establish whether the observed population-level differences are detectable at the single-cell level.

2) Characterise the transcriptional heterogeneity of VSMCs within a vascular bed, with the aim of identifying potential sources of functional heterogeneity, which has been observed in VSMC response to injury and atherosclerosis. I hypothesised that the observed functional heterogeneity may stem from VSMC heterogeneity in healthy arteries.

3) Investigate the heterogeneity of VSMCs during phenotypic switching in the atherosclerotic plaque and following vascular injury.

2 Methods

2.1 Animal models and tissue processing

Animal work and tissue processing was carried out by H.F. Jørgensen, A.L. Taylor, J.L. Harman and J. Chappell. For VSMC-specific lineage tracing, both single-colour ROSA26-eYFP and multi-colour ROSA26-Confetti systems were used in combination with the tamoxifen-inducible Myh11-CreERt2 in C57Bl/6 mice (Wirth et al. 2008; Gomez et al. 2013; Chappell et al. 2016). To create a Myh11-CreERt2 construct, CreERt2 was inserted into the ATG codon of a bacterial artificial chromosome (BAC) carrying the *Myh11* gene (Wirth et al. 2008). The construct was injected into mouse oocytes, where it integrated into the Y chromosome (Wirth et al. 2008). Studies have suggested that CreERt2 expression is confined to mature VSMCs in healthy animals (Wirth et al. 2008; Nemenoff et al. 2011), however since CreERt2 is not under the control of the endogenous *Myh11* promoter, it should be noted that there may be differences in the activities of the endogenous and transgenic *Myh11* promoters (Chakraborty et al. 2019). CreERt2 has been mutated to bind tamoxifen, after which it is translocated to the nucleus, where recombination at the reporter allele takes place (Feil et al. 1996; Feil et al. 1997). Since CreERt2 is expressed selectively in VSMCs, reporter expression following induction of lineage labelling by tamoxifen is VSMC-specific.

For the analysis of VSMCs within atherosclerotic plaques, the Myh11-CreERt2/Confetti lineage tracing system was crossed with ApoE$^{-/-}$ mice, which has been described (Chappell et al. 2016). Carotid ligation experiments were carried out using Myh11-CreERt2/eYFP lineage traced mice containing the Ki67-RFP reporter (Basak et al. 2014). Sca1-GFP animals used have been described previously (Ma et al. 2002). As the Myh11-CreERt2 transgene is Y-linked, only male mice were used in this study. Tamoxifen (10 mg/animal) was administered intraperitoneally at 6-8 weeks of age over a 10 day period to induce recombination for lineage labelling. For analysis of VSMCs in atherosclerotic arteries, Myh11-CreERt2/Confetti/ApoE$^{-/-}$ mice were tamoxifen labelled as described and subsequently fed a high-fat diet (21% fat, 0.2% cholesterol) for 14-18 weeks (Chappell et al. 2016). For analysis of VSMCs following carotid ligation (Kumar & Lindner 1997), tamoxifen-treated Myh11-CreERt2/eYFP/Ki67-RFP mice underwent carotid ligation surgery, during which the left carotid artery was tied with a 6-0 silk suture to cease blood flow (Chappell et al. 2016). Prior to surgery the mice received analgesic

(Temgesic) and were anaesthetised using 2.5-3% isofluorane inhalation (Chappell et al. 2016). Prior to analysis, aortas or carotid arteries were isolated by dissection and the surrounding fatty tissue was removed.

2.2 Bulk RNA-seq analysis

2.2.1 Experimental protocol

Generation of bulk RNA-seq libraries was carried out by J.L. Harman. Dissected aortas were incubated in RNAlater, separated into the AA and DT sections and the adventitial and endothelial layers were removed. Medial layers from 3-5 animals were lysed in Trizol and RNeasy column (Qiagen) was used to clean the isolated RNA. 550 ng of total RNA was used to prepare sequencing libraries (TruSeq Stranded mRNA Library Prep Kit, Illumina), which were sequenced on the MiSeq sequencer (Illumina). Three independent libraries were generated from the medial layers of each of the AA and DT regions.

2.2.2 Data processing and analysis

Raw reads were processed through the Babraham Institute Bioinformatics pipeline, during which they were aligned to the GRCm38 mouse genome using *TopHat* aligner v2.1 (Trapnell et al. 2009). *SeqMonk* software was used to count the number of aligned reads per gene (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk, *SeqMonk* analysis was performed by P. Oldach). I used the R statistical environment for subsequent analysis. Differential gene expression analysis was carried out using the *DESeq2* Bioconductor R package v1.12 (Love et al. 2014). Specifically, functions *DESeqDataSetFromMatrix* (with design corresponding to the AA or DT origin of the samples) and *DESeq* (with default parameters) were used for the analysis. Genes were considered differentially expressed if their log2 fold change between regions was > 1 and their FDR-adjusted p-value was < 0.01 (Benjamini-Hochberg multiple testing correction).

2.3 Flow cytometry analysis and scRNA-seq experimental methods

Experimental work was carried out by H.F. Jørgensen, A.L. Taylor, J.L. Harman and J. Chappell, unless otherwise stated. In case of the carotid arteries and the whole aorta 10X Genomics Chromium experiments, the entire artery was dissociated to a single-cell suspension by incubation with collagenase type IV (2.5 mg/ml) and porcine pancreatic elastase (2.5 U/ml) for 1-2h, followed by filtering through a 40 µm cell strainer to remove any remaining clumps of cells. In other experiments, the adventitial and endothelial layers were removed prior to dissociation of medial cells to a single-cell suspension as described above.

For SCA1 antibody staining, single-cells were incubated with TruStain FcX (1:100, BioLegend) prior to staining with the APC-conjugated isotype control antibody (1:10 Miltenyi 130-102-655) or the anti-SCA1 antibody (1:10 Miltenyi 130-120-343). SCA1 antibody staining was analysed using an Accuri C6 (wild type and the single-colour reporter) or a Fortessa (multi-colour reporter) flow cytometer.

2.3.1 Fluidigm C1 platform

Medial layers of the AA and DT aortic sections from 5-7 male C57Bl/6 mice were isolated and dissociated to a single-cell suspension, as described above. Samples (100 cells/µl) were processed using the Fluidigm C1 Auto Prep Arrays system with medium-sized chips (17-25 µm) in line with manufacturer's instructions. Visual inspection of the chips was performed to select successfully captured single-cells. SMARTer Ultra Low RNA Kit (Clontech) was used to generate amplified cDNA. Sequencing libraries were then prepared using the Nextera Library Prep Kit (Illumina) and sequenced on a HiSeq 2500 sequencer (Illumina) using the 50bp paired-end sequencing protocol. Single cells from both the AA and DT regions were analysed in two independent experiments.

For the pooled VSMC and adventitial samples, single-cell suspension was prepared from the medial and adventitial layers of the aorta, as described above. Total RNA from 2000-4000 cells was then extracted with the RNeasy Plus Micro Kit (Qiagen). Amplified cDNA was generated using the SMARTer Ultra Low RNA Kit (Clontech) according to the tube control protocol. Nextera Library Prep Kit (Illumina) was used to prepare sequencing libraries from the amplified

cDNA. Libraries were then sequenced on a HiSeq 2500 sequencer (Illumina) using the 50bp paired-end sequencing protocol.

2.3.2 Smart-seq2 platform

Medial layers of 5-7 Myh11-CreERt2/Confetti male mice were dissociated to a single-cell suspension and stained for SCA1 as described above. Cells were additionally stained with Zombie-NIR (1:100, BioLegend) to identify dead cells. Aria-Fusion flow cytometer (BD Bioscience) was used to sort live single cells, which were lineage-labelled and/or SCA1-positive, into individual wells of a 96-well plate. Lineage-labelled cells were required to express only one of the four confetti colours to reduce the chances of profiling doublets of cells. Amplified cDNA was generated using the Smart-seq2 protocol (Picelli et al. 2014), with these modifications: for the reverse transcription Primescript (Clontech) was used, cDNA was amplified in 24 PCR cycles and ERCC controls (Invitrogen) were added at 1:40,000,000 or 1:80,000,000 dilution into the reverse transcription (RT) mix. Sequencing libraries were prepared with the Nextera Library Prep Kit (Illumina) and sequenced on a HiSeq 2500 sequencer (Illumina) through the 50bp paired-end sequencing. Analysed cells were from 3 independent experiments, with SCA1-positive, lineage-labelled cells included in all three experiments and cells, which were only positive for either SCA1 or the lineage label, included in two of the experiments.

2.3.3 10X Genomics Chromium platform

The 10X Genomics Chromium platform with the Gene Expression v2 kit was used to generate the datasets described in this section.

2.3.3.1 VSMCs in healthy aorta and whole aorta analysis

For both the whole aorta and VSMC-only samples, aortas of three tamoxifen-labelled Myh11-CreERt2/Confetti males were processed to a single-cell suspension as described above and stained with Zombie-NIR (1:100, BioLegend). Aria-Fusion flow cytometer (BD Bioscience) was used to isolate 20,000 live single cells for both samples. In case of the whole aorta sample, selection based on lineage label expression was not performed, while expression of one of the four fluorescent proteins was required for the VSMC-only sample to reduce the chances of profiling doublets. Samples were then processed through the 10X Genomics Chromium

platform. Sequencing libraries were sequenced on a HiSeq4000 instrument (Illumina) using paired-end sequencing.

2.3.3.2 Carotid ligation injury

The carotid ligation surgery has been performed on the left carotid artery as described previously and above (Kumar & Lindner 1997; Chappell et al. 2016). Left carotid arteries from 5 male tamoxifen-labelled Myh11-CreERt2/eYFP/Ki67-RFP mice were dissected 7 days following carotid ligation and dissociated to a single-cell suspension as described above. An Aria-Fusion flow cytometer (BD Bioscience) was used to isolate 20,000 live single eYFP+ cells. Due to the rare nature of RFP+ cells, the sample was enriched for about 100 eYFP+RFP+ cells. Isolated cells were processed through the 10X Genomics Chromium system and their libraries were sequenced on a HiSeq4000 sequencer (Illumina) using paired-end sequencing.

2.3.3.3 VSMCs in atherosclerotic arteries

Myh11-CreERt2/Confetti/ApoE$^{-/-}$ tamoxifen-labelled male mice were fed a high fat diet for either 14 or 18 weeks as described above. Aortas of 2-3 mice per time point were used, with atherosclerotic plaques manually isolated, dissociated essentially as described previously (Butcher et al. 2011), with the addition of porcine pancreatic elastase (2.5 U/ml) to the digestion cocktail, and passed through a 40 µm cell strainer to remove clumps. 20,000 single lineage-labelled cells, which expressed only one of the four fluorescent proteins, were isolated from each time point using an Aria-Fusion flow cytometer (BD Bioscience). Samples were processed through the 10X Genomics Chromium platform and sequenced using pair-end sequencing on a HiSeq4000 machine (Illumina). Experiment performed jointly with H.F. Jørgensen.

2.4 Processing, quality control and normalisation of scRNA-seq data

2.4.1 Fluidigm C1 and Smart-seq2 platforms

Raw reads were processed through the Babraham Institute Bioinformatics pipeline, during which they were aligned to the GRCm38 mouse genome using *TopHat* aligner v2.1 (Trapnell et al. 2009). *Htseq-count* v0.8 (Anders et al. 2015) was used to count the number of read alignments per gene in the resulting bam files. The *idattr* parameter was set to “gene_id” and

the *stranded* parameter to "no". Single cells processed through the Fluidigm C1 platform were considered of good quality if their total read count was over 1 million and less than 3.5 million, the number of genes detected in the cell was between 5000 and 9500 and the proportion of mitochondrial reads was less than 20%. Additionally, over 80% of reads were required to map to genes and over 50% to exons. The percentage of reads mapping to genes and exons was calculated using the *SeqMonk* software v1.42 from the aligned bam files (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk). Single cells processed using the Smart-seq2 protocol (Picelli et al. 2014) were considered of good quality if the total read count per cell exceeded 100,000, the number of genes detected exceeded 1500 and the percentage of ERCC control reads did not exceed 30%.

Prior to normalisation, the genes with mean expression levels below 1 count per cell were filtered out to reduce noise. The function *computeSumFactors* from the Bioconductor R package *scran* (Lun et al. 2016) was then used to compute the normalisation factors for individual cells (*scran* v1.2 for Fluidigm C1 data and v1.8 for Smart-seq2 data). For the PCA comparison of single VSMC transcriptomes with the pooled adventitial and VSMC samples, the normalisation method *estimateSizeFactorsForMatrix* from the R Bioconductor package *DESeq2* v1.12 was used, with the method *locfunc* = "shorth" from the *genefilter* R Bioconductor package v1.6.

Batch correction methods were not applied to Fluidigm C1 profiles of AA and DT VSMCs due to unbalanced experimental design, despite batch effects being noticeable for DT VSMCs. Batch correction could have introduced artificial differences between the AA and DT VSMC profiles, since individual batches contained transcriptional profiles from a single region. Instead, these technical issues were considered during downstream analyses to ensure that any conclusions drawn from the analyses were supported by all batches.

2.4.2 10X Genomics Chromium platform

10X Genomics *cellranger* pipeline (v2.0 for atherosclerosis samples and v2.1 for healthy aorta and carotid ligation samples) was used to demultiplex the raw BCL files, align the reads to the GRCm38 mouse genome using the *STAR* aligner (Dobin et al. 2013) and count the UMIs detected per gene (*cellranger* pipeline run by F. Krueger). I imported the gene-count matrices into R using the *Seurat* v2.3 package function *Read10X* (Butler et al. 2018).

Cells isolated from healthy arteries were considered of good quality if their total UMI count was between 1000 and 8000, the number of genes detected was between 500 and 2500 and the cell contained less than 8% of mitochondrial reads. VSMCs isolated from carotid arteries following ligation surgery were required to have at least 5000 UMIs and 2000 genes detected and to not exceed 6% of mitochondrial reads. VSMCs isolated from atherosclerotic plaques were required to have 5000-20,000 UMIs, 1000-5000 genes and less than 9% of mitochondrial reads detected. The remaining good quality cells were then normalised using the *NormalizeData* (method = "LogNormalize", scale.factor = 10,000) and *ScaleData* functions from the *Seurat* R package v2.3 (Butler et al. 2018).

2.5 Processing and analysis of publicly available cultured VSMC transcriptomes

Fastq files were downloaded from the Gene Expression Omnibus (GSE79436, Adhikari et al. 2015). Downloaded data was processed using the Babraham Institute Bioinformatics pipeline, which included alignment using *TopHat* v2.1 (Trapnell et al. 2009). *Htseq-count* v0.8 (same parameters as decribed above for the Fluidigm C1 and Smart-seq2 profiles) was then used to count the number of reads per gene. Transcriptomes from cultured and *ex vivo* VSMCs were normalised together using the *computeSumFactors* function from the Bioconductor R package *scran* v1.2 (Lun et al. 2016).

2.6 Log transformation and PCA dimensionality reduction

Normalised read counts were log2-transformed and a pseudocount of 1 was added to all counts to avoid log-transforming 0 counts. Principal component analysis of the Fluidigm C1 and Smart-seq2 profiles was carried out using the *plotPCA* function from the *scater* R Bioconductor package v1.8 (McCarthy et al. 2017) with the underlying *prcomp* R method. The features were scaled and the top 500 most variable genes were used for the analysis, apart from PCA based on the random forest predictors, where the 30 random forest predictor genes were used. For the 10X Genomics Chromium data PCA was carried out using the *Seurat* R package v2.3 (Butler et al. 2018). Specifically the function *RunPCA* was used and the *JackStraw* function (num.replicate = 100) was used to estimate whether PCs explained more variance than expected at random.

2.7 Random forest analysis

AA and DT cells were first split into the training (75% or 108 cells) and test (25% or 35 cells) sets randomly using the *createDataPartition* function from the CRAN R package caret v6.0. Recursive feature elimination was carried out on the training dataset using the *rfe* function from the *caret* R package with 10-fold cross validation. Performance of the random forest model based on the top 25, 30, 35, 40, 45, 50 features identified during recursive feature elimination was evaluated, with 30 predictor genes selected for further model training. The random forest model based on the identified 30 top predictor genes (1000 trees) was trained using the train function from the caret R package with the "rf" method from the *randomForest* R package v4.6. Following grid search optimisation the *mtry* parameter was set at 2. Repeated 10-fold cross-validation (10 repeats) was used during training. The performance of the final model on the test data was evaluated using the *predict* function from the *stats* R package. ROC plot was generated using the *ROCR* R package v1.0.

2.8 Highly variable gene and co-expression analysis

Highly variable gene analysis was performed using the highly variable gene identification workflow in the R Bioconductor package *scran* (Lun et al. 2016) v1.2 for Fluidigm C1 data and v1.8 for Smart-seq2 data. Analysis was carried out separately for the AA, DT and SCA1-positive, lineage-labelled (S+L+) cells. The design parameter in the *trendVar* and *decomposeVar* functions was used to account for differences between the batches. For the AA and DT VSMCs the technical variance for a given mean log expression level was estimated using the *trendVar* (method = "loess") function based on all genes under the assumption that the majority of the genes were not highly variable. For highly variable gene identification in S+L+ cells, ERCC controls were available and were used to estimate the technical variance using the *trendVar* (method = "loess") function. Based on the estimated relationship between technical variance and the mean log expression levels of a gene, the variance of each gene was then decomposed into the technical and biological components using the *decomposeVar* function. A gene was identified as highly variable if the biological component of variance was significantly greater than zero (FDR-adjusted p-value < 0.05, Benjamini-Hochberg correction for multiple testing) and greater than 0.5.

For increased stringency in highly variable gene identification, I repeated the highly variable gene analysis in S+L+ cells 1000 times using 90% of randomly selected cells during each iteration. To combine the p-values from the iterations, a method based on the work of Licht and Rubin (Licht 2010) was implemented by M. Spivakov. In this method, p-values from each sample were transformed into z-scores using the quantile function of the normal distribution. The combined z-score for each gene was computed as the mean z-score for this gene across iterations divided by the total variance of the gene's z-scores across the iterations. The probability density function of the normal distribution was then used to compute the p-values, which were adjusted for multiple testing using the Benjamini-Hochberg method. A gene was considered highly variable if the adjusted p-value was below 0.05.

Gene correlations within the cVSMC network of S+L+ cells from healthy arteries, as well as *Ly6a/Sca1* co-expressed genes in VSMCs post carotid ligation surgery, were identified using an approach based on the Spearman's rho implemented in the R Bioconductor package *scran* v1.8 (Lun et al. 2016). Function *correlatePairs* was used to assess the pairwise correlations either among the genes within the cVSMC network (S+L+ cells) or between *Ly6a/Sca1* and the remaining profiled genes (VSMC in injury). Gene pairs with Benjamini-Hochberg FDR-adjusted p-value < 0.05 (cVSMC network) or < 0.01 (VSMC in injury) and a positive rho value were considered to be positively correlated.

2.9 Network analysis and visualisation

Network analysis and visualisation of the resulting modules was performed by M. Spivakov. Network analysis was used to identify co-expressed modules of highly variable genes in S+L+ cells and was carried out using the *blockwiseModules* (power = 3, TOMType = "unsigned", minModuleSize = 5, reassignThreshold = 0, mergeCutHeight = 0.25) function of the R package WGCNA v1.63 (Langfelder & Horvath 2008). The soft-thresholding power was chosen using the *pickSoftThreshold* function. Resulting ME1 (cVSMC module) and ME2 modules were visualised using *Cytoscape* v2.6.1 (Shannon et al. 2003). The WGCNA co-expression weight of gene pairs is encoded by the edge thickness in visualised modules.

2.10 Summarising expression levels of gene signatures using PC1 scores

To summarise the overall expression levels of gene signatures, PC1 scores were used. Firstly, PCA was performed based on the genes contained in the gene signature using the *prcomp* (center = TRUE, scale. = TRUE) function from the R package *stats*. Since the sign of the PC values is arbitrary, PC1 values of individual cells were multiplied by -1 if the PC1 values correlated negatively with the total expression levels the gene signature.

2.11 Identification of the cVSMCneg and cVSMCpos signatures

Identification of genes that were positively or negatively correlated with the cVSMC score (summarised expression of module ME1) of S+L+ cells was carried out by M. Spivakov. A modified version of the approach for identification of trajectory associated genes from the *monocle* R Bioconductor package (Trapnell et al. 2014) was used. The *loess* (family = "symmetric") R function was used to estimate the mean-variance relationship of mean log expression levels and the log squared coefficient of variation. Negative binomial regression implemented in the *VGAM* R package v1.0 was then used to model the dependence of the expression level of a gene and the cell's cVSMC score. The size parameter of the negative binomial distribution was assumed to be the inverse fitted squared coefficient of variation described above. Likelihood-ratio test was used to assess the significance of the binomial regression fit compared with the intercept-only model. P-values were adjusted for multiple testing using the Benjamini-Hochberg procedure. Genes were considered significantly positively or negatively correlated if their adjusted p-values were below 0.05.

2.12 Clustering analysis and t-distributed Stochastic Neighbour Embedding (t-SNE) visualisation

Clustering analysis and t-SNE visualisation were performed using the R package *Seurat* v2.3 (Butler et al. 2018; Maaten & Hinton 2008). The dimensionality of the dataset was first reduced using PCA as described above. The principal components included in further analysis were chosen based on the elbow plot showing the standard deviations of calculated principal components, whether they explained biologically meaningful variation among the profiled cells, or whether they explained significantly more variance than expected at random (as

described above). Graph-based clustering implemented in the *FindClusters* function was used for the analysis. The resolution parameter was set based on the biologically meaningful outcomes of identified clusters as follows: at 0.6 for the analysis of the whole aorta and VSMC-only datasets and the initial analysis of plaque VSMCs; at 0.8 for re-analysis of plaque VSMCs following the removal of low quality cells (initial cluster 4) and for analysis of VSMCs following carotid ligation injury. t-SNE analysis implemented in the *RunTSNE* function was used to visualise profiled cells and the results of the clustering analysis. The same number of principal components was used to calculate t-SNE scores as for the clustering analysis in each dataset.

2.13 Differential gene expression analysis

Differential gene expression analysis between the identified clusters of cells in the 10X Genomics Chromium datasets was carried out using functions *FindMarkers* and *FindAllMarkers* of the R package *Seurat* (Butler et al. 2018), using the method "*wilcox*". For the analysis of differentially expressed genes between the VSMC clusters of the whole aorta and VSMC-only datasets from healthy arteries, each VSMC cluster was compared with the remaining VSMC clusters using the *FindMarkers* function. Genes expressed in at least 10% of the cells in either population and showing log2-fold change > 0.5 between populations were included in the analysis. Function *FindAllMarkers* was used for identifying cluster markers in the vascular injury and atherosclerosis datasets with log2 fold change required to be at least 0.5 between the compared populations of cells. Genes were required to be expressed in either of the compared populations in at least 10% of the cells. In all cases tested genes were considered differentially expressed if the Bonferroni adjusted p-value was < 0.05 according to the Wilcoxon rank sum test.

Differential gene expression analysis between cultured and *ex vivo* VSMCs was carried out using the R package *scde* (Kharchenko et al. 2014). Genes with greater than 10 read counts in at least 5 cells were included in the analysis. *scde.error.models* and *scde.expression.prior* functions were used to compute the error models and expression priors and the function *scde.expression.difference* (n.randomisations = 100) was used to analyse differential expression of genes between conditions. Genes were considered differentially expressed if their log fold change between conditions was greater than 1 and their absolute corrected Z score was greater than 1.96.

2.14 Gene ontology analysis

Gene ontology analysis of the differentially expressed genes between the AA and DT regions according to bulk RNA-seq, as well as of the identified highly variable genes within the AA and DT regions, was performed using the PANTHER overrepresentation test, with the gene ontology database released on 1/1/2019 (Mi et al. 2019). Gene ontology analysis of genes differentially expressed in the VSMC cluster 6 of the whole aorta and VSMC-only datasets and between cultured and *ex vivo* VSMCs was also performed using the PANTHER overrepresentation test, but with gene ontology databases released on 3/7/2019 and 8/10/2019 respectively. Gene ontology "biological process complete" was used for all analyses and the gene lists were analysed against the *Mus musculus* genome-wide background, unless otherwise stated. Fisher's exact test and the Bonferroni correction for multiple testing were used for determining whether a gene ontology term was significantly overrepresented (adjusted p-value < 0.05). Gene ontology analysis of cVSMCpos and cVSMCneg signatures was carried out by M. Spivakov based on the "biological process" gene ontology using the *clusterProfiler* R package v3.8 (Yu et al. 2012).

2.15 Fisher's exact test

Fisher's exact test was used to determine whether prediction of only AA regional identities would be expected at random, given the bias towards AA predictions when the random forest model was applied to the 10X Genomics Chromium profiles of VSMCs. The *fisher.test* function in the R package *stats* was used, with the contingency table constructed based on the number of predicted AA or DT cells in cluster 6 (136 and 0 respectively) and the number of predicted DT or AA cells in other VSMC clusters (690 and 4343 respectively).

2.16 Data availability

Next generation sequencing datasets used in this thesis are publicly available (GSE117963, GSE79436), with the exception of the VSMCs in carotid ligation injury dataset, which is unpublished.

3 Regional differences between the aortic arch (AA) and descending thoracic aorta (DT) VSMCs

3.1 Introduction

In this chapter I focused on exploring the regional heterogeneity of VSMCs. Regional heterogeneity among VSMCs is thought to stem partly from the environmental differences experienced by the different areas of the vasculature, as well as from diverse embryonic origins of VSMCs located in different parts of the organism (Cunningham & Gotlieb 2005; Majesky 2007). For example, VSMCs located in the AA are derived from the neural crest, while VSMCs in the neighbouring DT region of the aorta have mesodermal origins (Majesky 2007). In addition, the environmental factors related to the curvature of the AA section and branch sites of the aorta are thought to impact VSMC heterogeneity (Cunningham & Gotlieb 2005; Chiu & Chien 2011). Together, these factors likely underpin the observations that the AA and DT sections of the aorta have different susceptibilities for atherosclerosis, with AA being more susceptible to atherosclerotic plaque development (Van Assche et al. 2011; Trigueros-Motos et al. 2013; Sinha et al. 2014).

This chapter compares VSMCs from the AA and DT regions at the single-cell as well as at the population-wide level. The transcriptional differences between VSMCs derived from the AA and DT regions have been previously investigated at the population level (Van Assche et al. 2011; Trigueros-Motos et al. 2013). In this chapter I extended the understanding of regional VSMC heterogeneity in the aorta by showing that the transcriptional differences are consistently expressed at the level of single cells. Additionally, I explored the heterogeneity of VSMCs within both vascular regions and observed heterogeneous expression of several disease-relevant genes.

3.2 Results

3.2.1 Single-cell analysis of VSMCs from the AA and DT aortic regions

In collaboration with H.F. Jørgensen and M. Spivakov, we profiled individual *ex vivo* VSMCs from the AA and DT aortic regions of healthy mice using scRNA-seq. This approach allowed us

to address the differences between VSMCs from two regions of the aorta, as well as VSMC heterogeneity within a vascular bed. To obtain single *ex vivo* VSMCs, the aortas were first dissected from healthy mice and the endothelial cells were removed. The aorta was then separated into the AA and DT regions and the adventitial layer was removed from both regions. VSMCs resident in the medial layer were then dissociated to a single-cell suspension using enzymatic digestion separately for each region. We then used the Fluidigm C1 platform to capture single cells and to prepare the amplified cDNA for subsequent sequencing library preparation and sequencing (Figure 3.1).

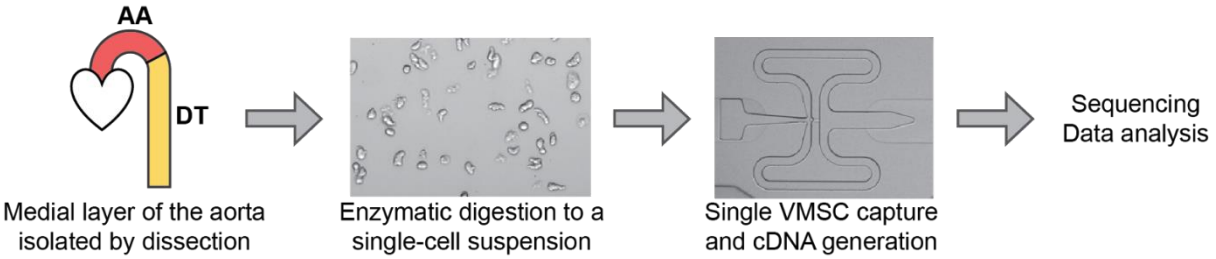


Figure 3.1: Schematic of the approach.

To obtain a single-cell suspension of VSMCs, the medial layer of the aorta from the AA and DT regions was dissected and enzymatically digested. Single cells were then captured and cDNA libraries were prepared followed by sequencing and data analysis. Microscopy images were prepared by H.F. Jørgensen. Experimental work was carried out by M. Spivakov and H.F. Jørgensen and the figure was adapted from Dobnikar and Taylor et al. 2018.

We obtained single-cell profiles of 173 cells. Raw sequencing reads were processed through the Babraham Institute pipeline and aligned to the GRCm38 mouse genome (details in Methods). I then generated the raw read count matrix and performed quality control to ensure that low quality cells do not affect the results of the analysis. Cells with very low or very high total read count (< 1,000,000 and > 3,500,000) and number of genes detected per cell (< 5000 and > 9500) were filtered out. This excluded the cells which may have burst and lost cytoplasmic mRNA, or were not sequenced to sufficient depth (low values) or could be doublets (high values). In addition, cells considered to be good quality had > 80% of reads mapping to genes, > 50% of reads mapping to exons and < 20% of mitochondrial reads (Figure 3.2). In total, 143 cells (79 AA and 64 DT) passed the quality control filters.

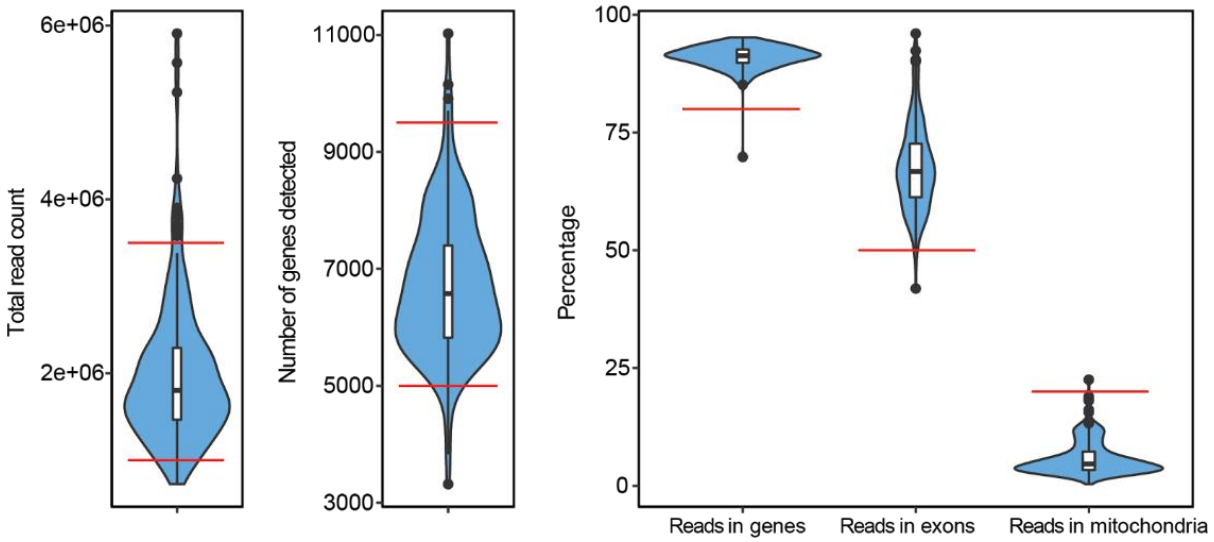


Figure 3.2: Quality control of scRNA-seq data of ex vivo VSMCs.

Violin plots showing distributions of total read count, number of genes detected, and the percentage of reads in genes, exons and mitochondria for the profiled AA and DT single cells. Red lines indicate quality control filtering thresholds. The figure is from Dobnikar and Taylor et al. 2018.

The transcriptional profiles of good-quality cells were normalised, with only genes showing mean expression > 1 read included in the normalised dataset to reduce noise. All profiled cells expressed key VSMC marker genes *Myh11*, *Acta2* and *Tagln* at high levels. *Myocd*, *Smtn*, *Vcl* and *Cnn1*, which were expressed at lower levels, were also detected in most cells (Figure 3.3, top panel). The fact that a few of the cells were negative for some of the VSMC marker genes may be due to dropout events, which are more frequent for lowly expressed genes. Dropout events can occur when a transcript is present in the cell, however it is not detected due to insufficient sequencing depth or imperfect mRNA capture and low amounts of starting material in single cells (Kharchenko et al. 2014). In agreement with this notion, detection rates of VSMC marker genes were comparable to housekeeping genes of similar expression levels (Figure 3.3, bottom panel).

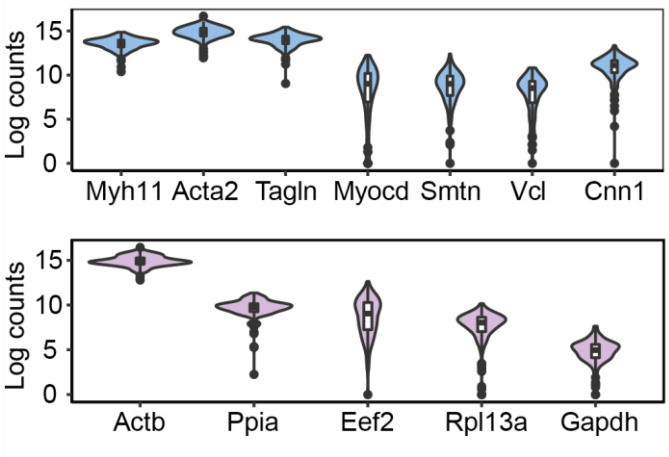


Figure 3.3: Profiled AA and DT single cells express VSMC contractile marker genes.

Violin plots showing the log2-transformed normalised read count distributions of VSMC marker genes *Myh11*, *Acta2*, *Tagln*, *Myocd*, *Smtn*, *Vcl* and *Cnn1* (blue) and housekeeping genes *Actb*, *Ppia*, *Eef2*, *Rpl13a*, *Gapdh* (purple) in profiled AA and DT single cells. The figure is from Dobnikar and Taylor et al. 2018.

To further verify that all of the profiled single cells were VSMCs, I compared the scRNA-seq transcriptomes with profiles generated from pooled medial and adventitial cells. Pooled samples were prepared by H.F. Jørgensen and involved pooling 2000-4000 cells prior to library preparation (details in Methods). Principal component analysis and unsupervised k-means clustering showed that all single VSMCs clustered with the pooled medial samples, suggesting a similarity of their transcriptomes, and away from pooled adventitial cell samples (Figure 3.4a and b). This suggested that none of the profiled single cells were likely contaminants from the neighbouring adventitial tissue. I additionally compared the averaged scRNA-seq transcriptomes with conventional bulk RNA-seq measurements of VSMCs from the AA and DT regions of the aorta. Although there were some discrepancies among lowly expressed genes, averaged scRNA-seq transcriptomes generally correlated with averaged bulk RNA-seq measurements (R^2 = 0.56, genes with mean expression greater than 3 log-transformed counts were included in the calculation, Figure 3.4c).

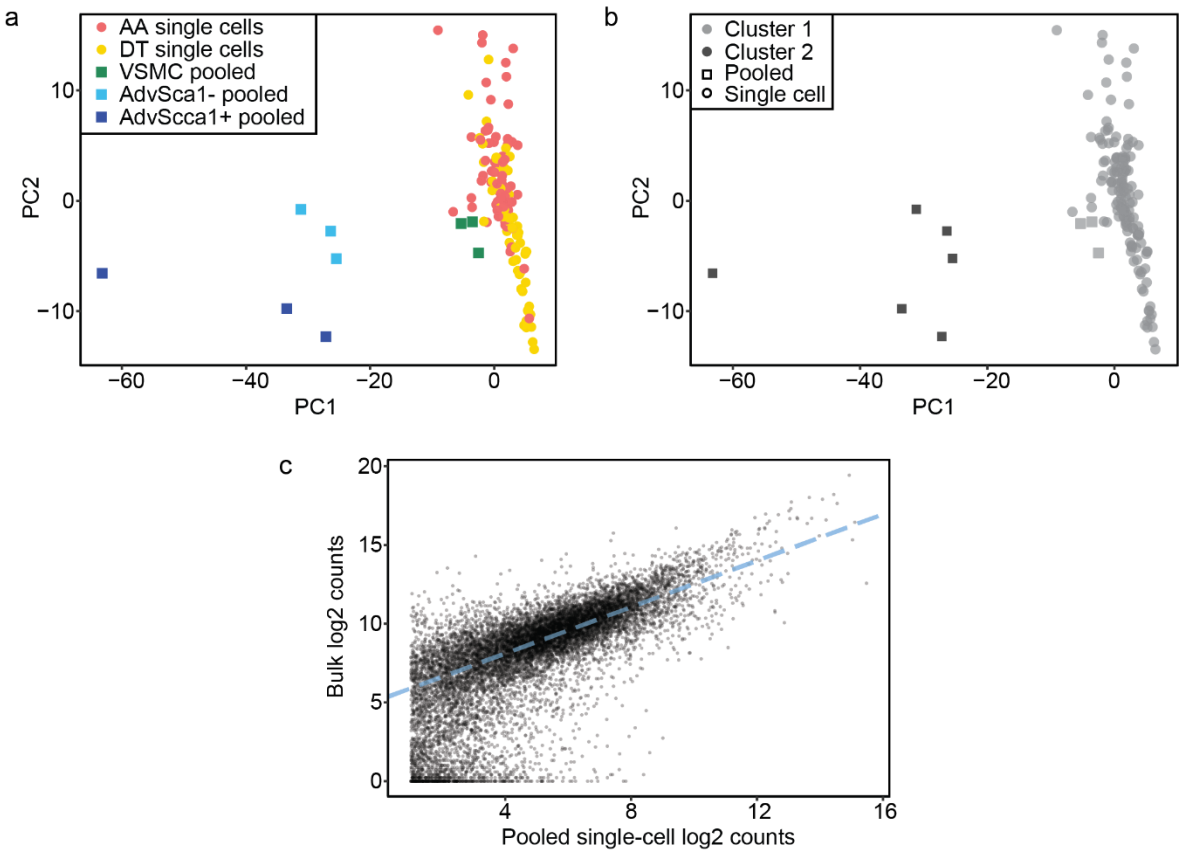


Figure 3.4: Profiled AA and DT single cells are distinct from pooled adventitial samples and share similarities with bulk RNA-seq VSMC profiles.

a) Principal component analysis comparing single-cell transcriptomes (yellow and red circles) to pooled medial (green squares), adventitial SCA1+ (dark blue squares) and adventitial SCA1^{-} (light blue squares) samples. PC1 explained 9% and PC2 explained 5% of the variance. b) Results of k-means clustering (k=2) of the pooled and single-cell samples presented in panel a. Pooled samples are represented by squares and single-cell samples by circles. c) A scatterplot showing the correlation between averaged scRNA-seq profiles and averaged bulk RNA-seq profiles from *ex vivo* AA and DT regions (R^2 = 0.56, where genes with mean expression greater than 3 log-transformed counts were included in the calculation).

Having established that the profiled single-cells were VSMCs, I next investigated whether the processing of the *ex vivo* samples had significant effect on the transcriptional profile of the cells. During culture, VSMCs acquire characteristics of the synthetic state and cultured VSMCs are frequently used as a model for phenotypic switching (Rensen et al. 2007). To assess whether VSMCs profiled in our study retained their contractile phenotype during sample preparation, I compared the transcriptional profiles of *ex vivo* single VSMCs with publicly

available transcriptional profiles of single cultured VSMCs (GEO accession GSE79436, Adhikari et al. 2015). Both the *ex vivo* and cultured datasets were generated using the Fluidigm C1 platform and the sequencing data was processed in an analogous way (details in Methods). The read counts were normalised jointly to account for differences in sequencing depth. Principal component analysis showed that transcriptional profiles of *ex vivo* and cultured VSMCs were clearly different, as the two populations clustered separately (Figure 3.5). Separate clustering of the cultured and *ex vivo* VSMCs may also be due to batch effects, as the two datasets were prepared in different laboratories. Although care has been taken to ensure that the data processing steps, such as alignment, counting the number of reads in genes and normalisation were carried out in the same way, batch effects cannot be ruled out.

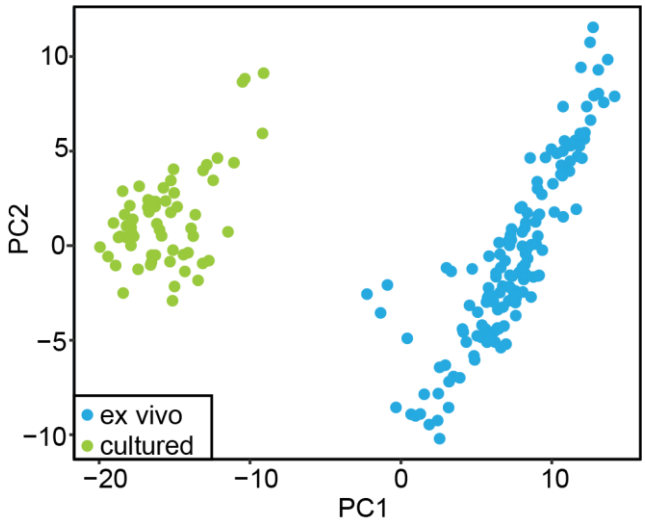


Figure 3.5: Transcriptional profiles of AA and DT *ex vivo* single cells are distinct from those of cultured VSMCs.

PCA comparing *ex vivo* VSMCs (blue) with cultured VSMCs (green). Colour-coded circles represent individual cells and PCA was carried out based on 500 most variable genes. PC1 explained 25% of the variance and PC2 explained 4%. Transcriptional profiles of cultured VSMCs are publicly available at the Gene Expression Omnibus (GSE79436, Adhikari et al. 2015). The figure is adapted from Dobnikar and Taylor et al. 2018.

To address this, I compared the expression of contractile marker genes as well as housekeeping genes between the *ex vivo* and cultured datasets. Cultured VSMCs showed reduced VSMC marker gene expression levels compared to *ex vivo* VSMCs (Figure 3.6a), while

the expression levels of all assessed housekeeping genes were comparable (Figure 3.6b). Differential gene expression analysis between the cultured and *ex vivo* cells identified 1278 genes as upregulated in cultured VSMCs and 2647 genes as upregulated in *ex vivo* VSMCs (details in Methods). Gene ontology analysis of the 200 most upregulated genes for each condition identified "muscle structure development", "muscle tissue development" and "potassium ion transport" as the three most overrepresented gene ontology terms for the *ex vivo* upregulated genes and "mitotic cell cycle", "mitotic cell cycle process" and "cell division" as the three most overrepresented gene ontology terms for the genes upregulated in cultured cells. This suggests that the differences between the samples are likely driven by biological rather than technical factors and that VSMCs profiled in our study retained their contractile phenotype and did not undergo significant phenotypic switching during experimental procedures. Overall, the results presented in this section show that profiled AA and DT single cells are of good quality and have the characteristics expected of contractile VSMCs.

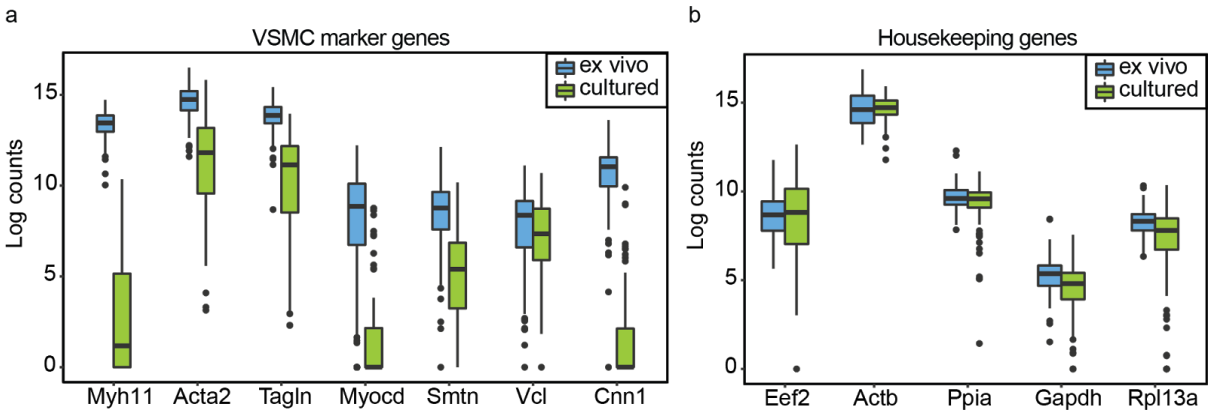


Figure 3.6: *Ex vivo* VSMCs retained higher levels of VSMC contractile marker gene expression than cultured VSMCs.

Boxplots show the log2-transformed normalised read count levels of VSMC marker genes (a) and housekeeping genes (b) in *ex vivo* (blue) and cultured (green) VSMCs. Transcriptional profiles of cultured VSMCs are publicly available at the Gene Expression Omnibus (GSE79436, Adhikari et al. 2015). The figure is from Dobnikar and Taylor et al. 2018.

3.2.2 Population-level differences in gene expression between the AA and DT regions

Given the low amounts of starting material and dropout events in single-cell data, bulk RNA-seq can give a clearer picture of the average gene expression differences between two regions and can be a useful resource to guide scRNA-seq data analysis. We therefore performed conventional bulk RNA-seq of the two regions to obtain a robust set of genes, which are differentially expressed between the AA and the DT regions (collaboration with J. L. Harman). I performed the differential gene expression analysis, which revealed 442 differentially expressed genes (adjusted p-value < 0.01, log2 fold change > 1), of which 386 were upregulated in the AA and 56 upregulated in the DT region (Figure 3.7a). Among the genes upregulated in the AA were *Pde1c* and *Spp1*, which have been associated with the synthetic state of VSMCs (Lesauskaite et al. 2001; Cai et al. 2015, Figure 3.7a). Gene expression differences between the two regions were validated using RT-qPCR in independent samples for *Dcn*, *Lum*, *Pde1c*, *Gpc3*, *Hoxa7* and *3632451O06Rik*, with RT-qPCR showing consistent changes in expression between regions with RNA-seq for all tested genes (Figure 3.7b, collaboration with A.L. Taylor).

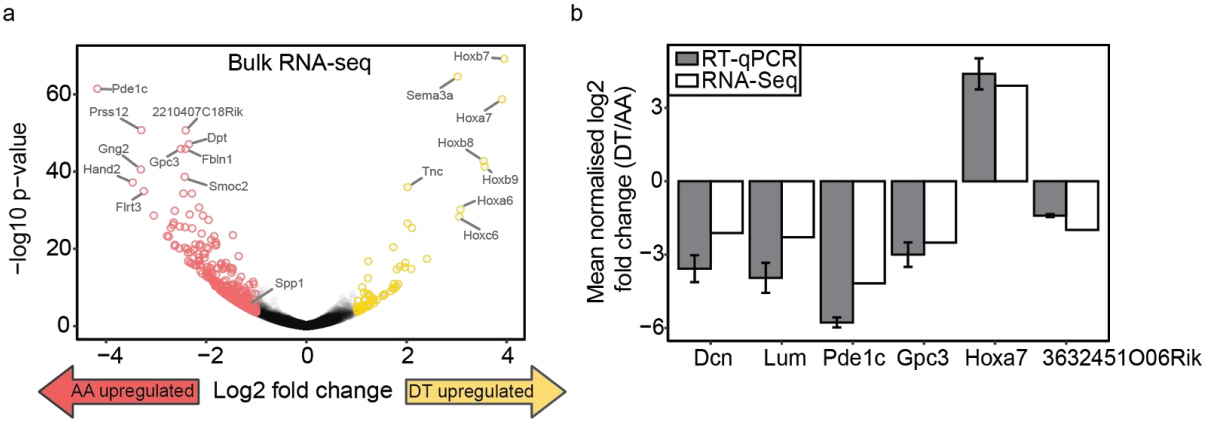


Figure 3.7: Differentially expressed genes between the AA and DT regions.

a) Volcano plot showing the –log10(p-value) versus log2 fold change for all genes detected using bulk RNA-seq. Genes identified as upregulated in the AA are shown in red and genes upregulated in DT are shown in yellow. Names of selected genes are indicated. b) Bar chart showing log2 fold change between the DT and AA as determined RT-qPCR (grey) compared with bulk RNA-seq (white). A.L. Taylor performed the RT-qPCR analysis and the figure was adapted from Dobnikar and Taylor et al. 2018.

Gene ontology (GO) analysis of genes upregulated in the DT region showed enrichment for GO terms related to regionalization and patterns specification, which was driven by higher expression levels of several *Hox* genes (Figure 3.8). Genes upregulated in the AA region showed enrichment for GO terms “regulation of cell population proliferation” and “inflammatory response” among others (Figure 3.8). This is consistent with previous reports of neural crest derived VSMCs (AA) being more responsive to treatment with inflammatory cytokines compared with mesoderm-derived (DT) VSMCs (Owens et al. 2010; Topouzis & Majesky 1996; Gadson et al. 1997; Cheung et al. 2012).

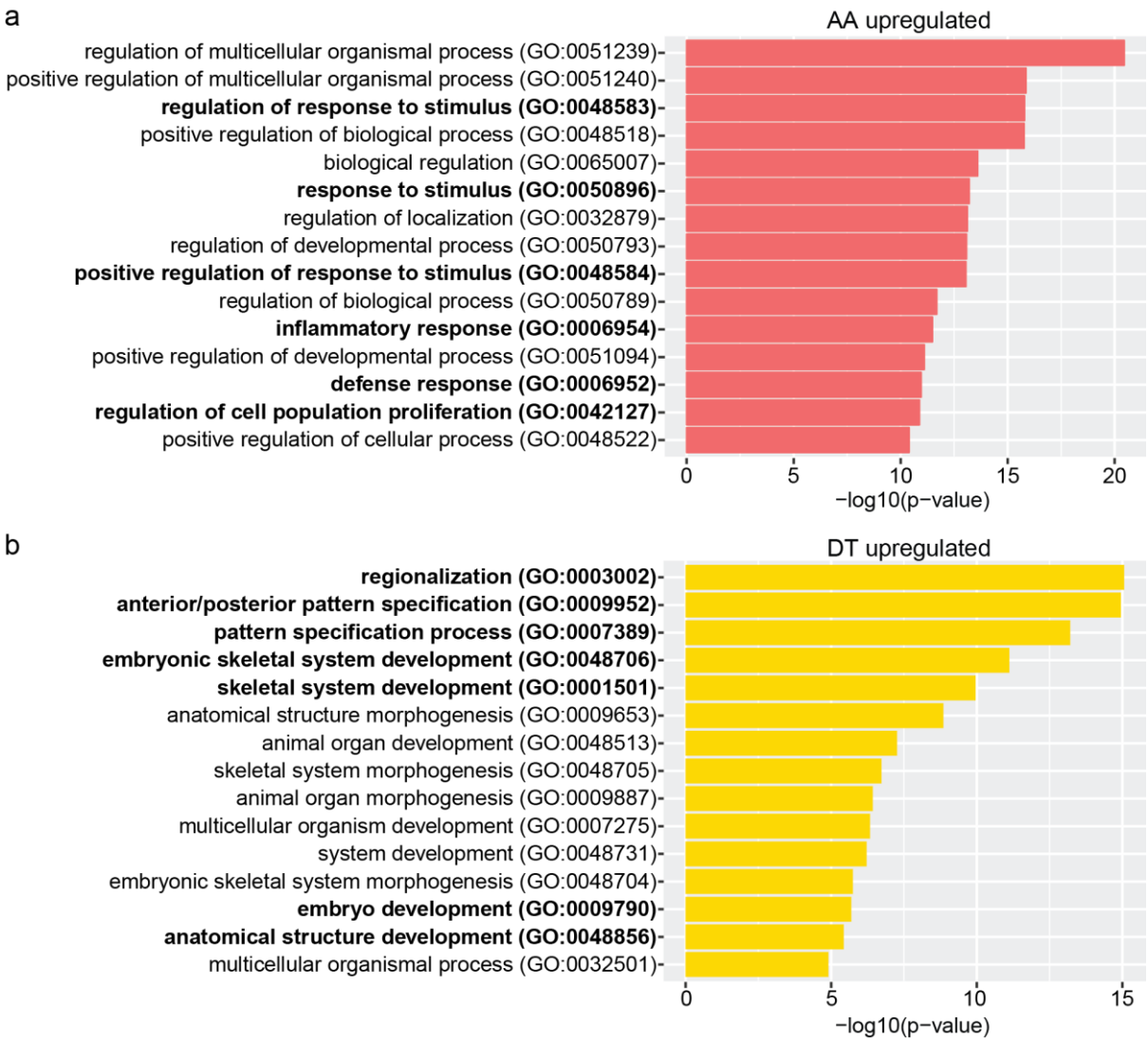


Figure 3.8: Enrichment for gene ontology terms (biological process) among genes upregulated in the AA and DT regions.

Bar plots show the top 15 gene ontology terms, for which there was the strongest enrichment among the genes upregulated in the AA (a) and DT (b) regions. Shown gene ontology terms are ranked according to their adjusted p-values (details in Methods).

Gene expression differences between the AA and DT regions were previously explored using microarrays (Trigueros-Motos et al. 2013). Consistent with previous findings (Trigueros-Motos et al. 2013), several *Hox* genes were found to be upregulated in the DT region (Figure 3.7a). Overall, the overlap between differentially expressed genes identified with microarrays and RNA-seq was relatively low, with 88 genes identified with both methods out of 442 and 246 genes identified as differentially expressed with RNA-seq and microarrays respectively (Figure

3.9a). This could be due to a large proportion of genes identified as differentially expressed only using microarrays showing low expression levels, with 70% of these genes showing log2-transformed read count < 4 in RNA-seq data (Figure 3.9b). Cross-hybridization and non-specific background hybridization of probes limit the accuracy of transcript detection with microarrays (Zhao et al. 2014). This is particularly problematic for lowly expressed genes, which could explain the discrepancy between the RNA-seq and microarray measurements for the genes expressed at low levels.

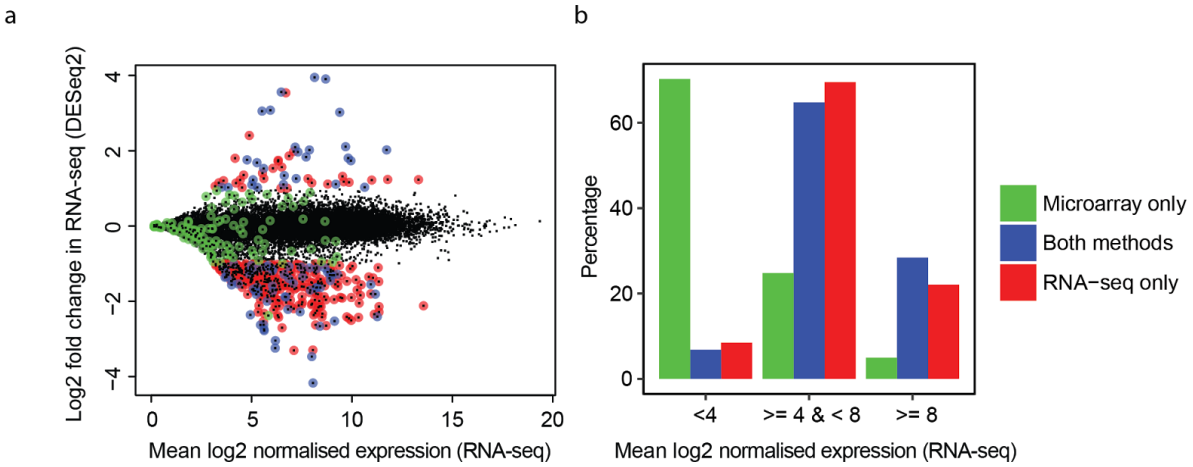


Figure 3.9: Comparison of the genes identified as differentially expressed using microarrays in a previous study (Trigueros-Motos et al. 2013) and bulk RNA-seq in our study.

a) Scatterplot showing log2 fold change in expression between the AA and DT versus log2-transformed mean expression levels for all genes detected using RNA-seq. Genes identified as differentially expressed using RNA-seq only (red), microarrays only (green) and with both methods (blue) are indicated. b) Bar plot showing the percentages of differentially expressed genes detected using microarrays only (green), RNA-seq only (red) or both methods (blue) for different expression level thresholds. The figure is from Dobnikar and Taylor et al. 2018.

In summary, I identified a refined list of differentially expressed genes between the AA and DT regions using bulk RNA-seq. I next explored how the region-specific gene expression is reflected at the single-cell level.

3.2.3 Transcriptional signatures of AA and DT regions are detectable at the single-cell level

Genes identified as differentially expressed using bulk RNA-seq had a variety of expression profiles at the single-cell level. For example, some genes showed almost exclusive expression in one region only (Figure 3.10, top panel), whereas others were expressed in different proportions of AA and DT cells (Figure 3.10, bottom panel). This led us to question whether each individual cell expressed a regional signature or whether population level differences might be reflected in different proportion of cells expressing the observed gene signatures associated with each region.

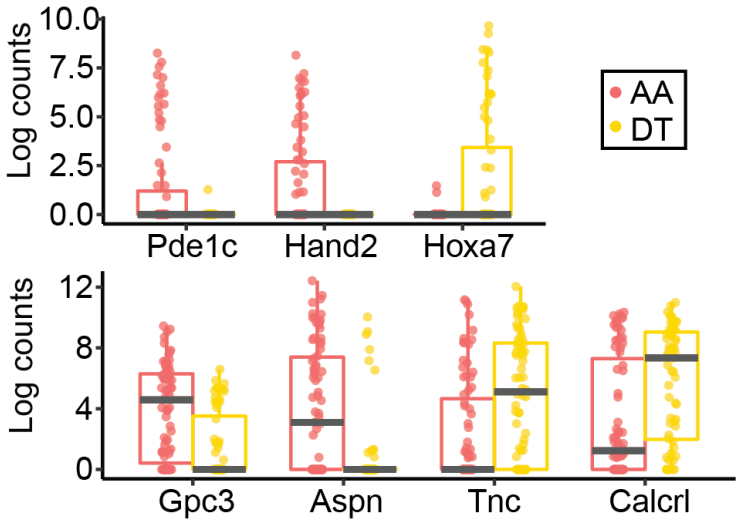


Figure 3.10: Examples of expression patterns of genes identified as differentially expressed in bulk RNA-seq at the single-cell level.

Boxplots showing log2-transformed normalised read counts for single VSMCs from the AA (red) and the DT (yellow). Top panel shows examples of genes, where only one of the regions shows appreciable expression. Examples of genes, which are expressed by different proportions of AA and DT cells, are shown in the bottom panel. The figure is from Dobnikar and Taylor et al. 2018.

To address this question, I assessed whether regional identity of VSMCs could be predicted with machine learning based on the transcriptomes of single cells. If a reliable predictive model could be built, this would suggest that each individual cell bears a signature of regional identity. I used a random forest model to classify each VSMC as either belonging to the AA or

the DT region. I chose to use a random forest approach as it provided insight into which genes were important for classification, in addition to the high performance of the random forest model in this setting. Firstly, single cells were randomly assigned to the training (75% of the cells) and test (25% of the cells) datasets (Figure 3.11). The test dataset was withheld during model construction and optimisation, and was only used to test the final model.

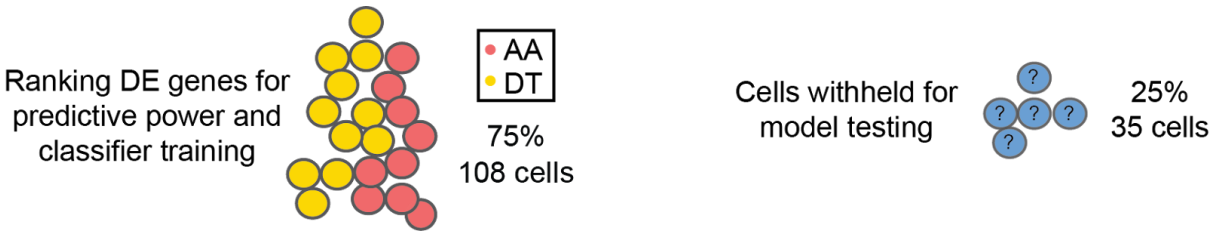


Figure 3.11: Schematic of the random forest approach.

A schematic showing the proportions of cells included in the training (75%) and test (25%) datasets during random forest model construction. The figure is adapted from Dobnikar and Taylor et al. 2018.

Due to a limited number of samples available for model training, it was necessary to select a reduced set of genes, on which the model would be based. This step removed redundant genes and ensured that the model was interpretable and not overly complex. I started with the list of genes identified as differentially expressed using bulk RNA-seq and then used the recursive feature elimination algorithm to reduce the number of genes further. Recursive feature elimination first constructs a random forest model and ranks the genes based on their importance for model accuracy. Further random forest models are then constructed based on a series of predetermined numbers of genes found to be most important for model accuracy. The performance of these models is then evaluated to assess which sets of genes lead to good performance. Using this approach, I identified a set of 30 genes with a strong predictive power and constructed the final random forest model based on the training data. Several of the 30 top predictive genes were among the most differentially expressed genes in bulk RNA-seq data (for example *Pde1c*, *Hand2*, *Hoxa7*, *Gpc3* and *22104C18Rik*), however the selection of other genes by recursive feature elimination was less obvious. This is likely due to gene expression differences between the bulk and single-cell datasets and the non-linear nature of the random forest algorithm, where performance is dependent on the whole set of genes used

in training rather than individual genes. The importance of the top 30 predictive genes for the accuracy of the final model is shown in Figure 3.12a.

Next, the cells initially withheld for model testing were used to evaluate the performance of the model on unseen data. The model showed high levels of accuracy, correctly predicting 17/18 AA cells and 14/15 DT cells from the test set. The accuracy of the random forest model is also demonstrated in Figure 3.12b, which shows that high levels of correct predictions are achieved while maintaining low levels of false positives. Consistent with the observation that the 30 genes used for random forest construction have the predictive power to classify cells as originating from the AA or DT regions, principal component analysis based on these genes showed a separation of the AA and DT cells (Figure 3.12c). Furthermore, Figure 3.12c shows that VSMCs from the test set (highlighted with black circles) were evenly distributed along the principal components 1 and 2 and thus form a representative test set. To examine whether the random forest model may be overfitting I performed an additional negative control. The regional labels of AA and DT single cells were randomly permuted and split into training and test datasets. An equivalent random forest model was than trained and tested on the transcriptomes with randomised regional labels. The performance of the random forest model on randomised data was markedly reduced compared with the original random forest model, and was only marginally better than random (Figure 3.12d), suggesting that significant overfitting during model training did not occur.

Ex vivo VSMCs from the both of the AA and DT regions were profiled in two separate experiments. VSMCs from both of the batches were present in both the training and test datasets and were accurately classified as originating from the AA or DT regions. However, for VSMCs originating from the DT region, batch effects were apparent on the PCA plot based on the 30 random forest predictor genes (Figure 3.13). I have chosen not to apply batch correction to single-cell transcriptomes to ensure that any technical factors, which have likely affected one of the DT batches, would not introduce artificial differences between the AA and DT profiles. Instead I have ensured that VSMCs from both experiments were accurately predicted in the test dataset before making conclusions about the effectiveness of the random forest model and kept the technical issues in mind during further analysis.

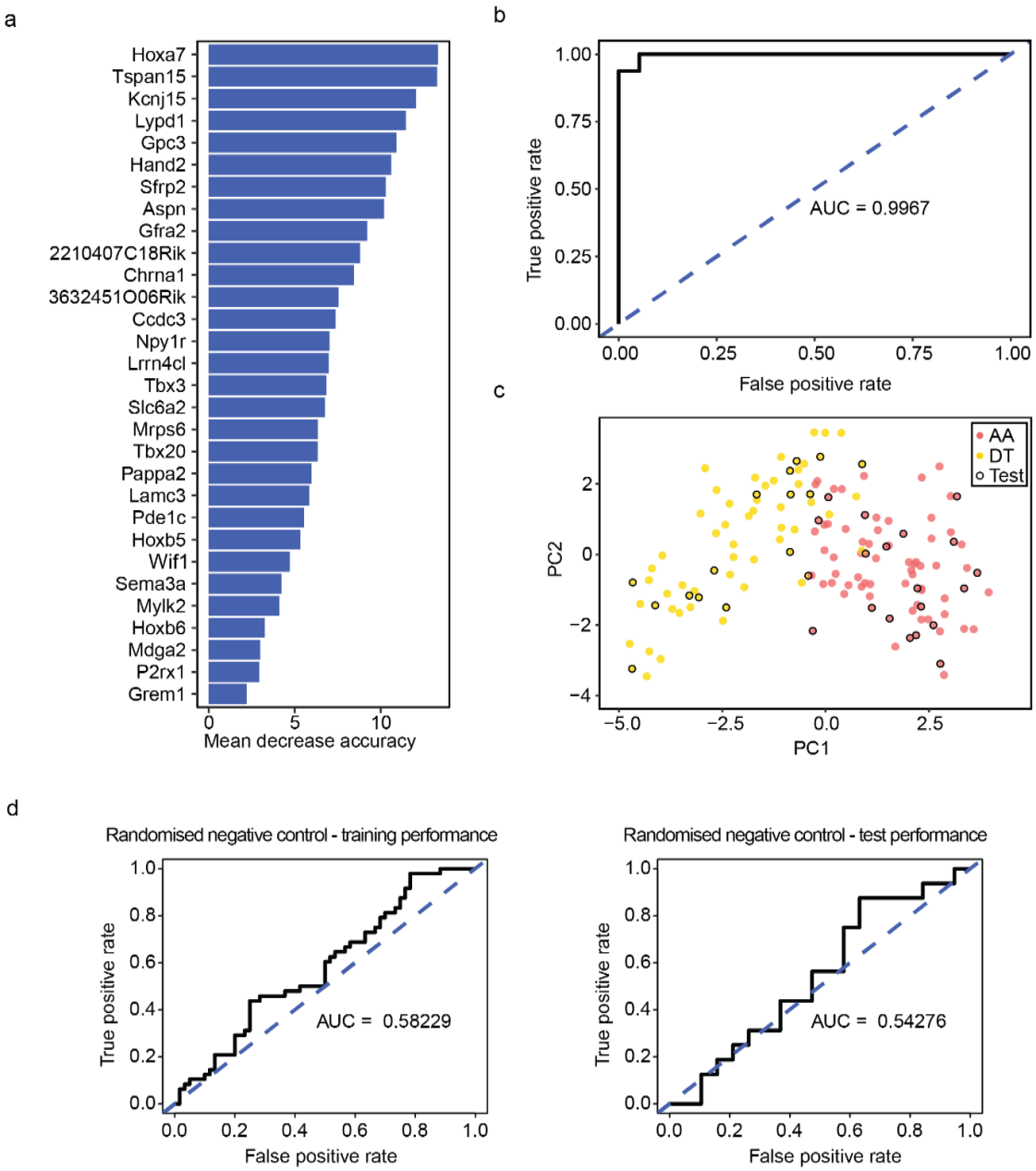


Figure 3.12: Random forest analysis distinguishes *ex vivo* VSMCs from the AA and DT regions at the single-cell level.

a) Bar plot showing the importance for model accuracy of genes used to construct the random forest model. b) Receiver operating characteristics curve showing the level of true positive rate of classification for a given level of false positive rate. The area under the curve (AUC) is indicated. c) PCA based on the 30 genes used for random forest construction. VSMCs from the AA are shown in red and VSMCs from the DT are shown in yellow. VSMCs assigned to the test set are circled in black. PC1 explained 17% and PC2 8% of the variance. d) Results of a randomised negative control, where the AA and DT regional labels of transcriptional profiles were randomly permuted for both the training (left) and test (right) datasets. Panels a-c are from Dobnikar and Taylor et al. 2018.

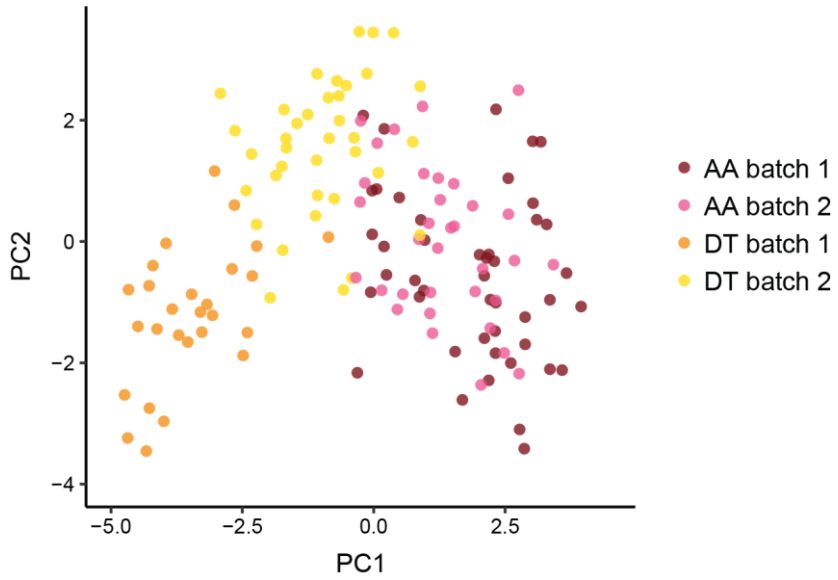


Figure 3.13: Examination of batch effects in *ex vivo* VSMCs profiled in separate AA and DT experiments.

PCA is based on the 30 predictor genes used in the random forest model. VSMCs from the AA are shown in dark red (batch 1) and pink (batch 2) and VSMCs from the DT are shown in orange (batch 1) or yellow (batch 2). PC1 explained 17% and PC2 8% of the variance.

Overall, high performance of the random forest classifier suggested that each individual cell within the AA and DT region could be assigned to the region of origin based on its transcriptional profile. If population level differences in gene expression profiles between the AA and DT regions were due to different proportions of cell populations present in each region, an accurate classification model would not be expected. These results therefore suggest that the regional identity signature is expressed at the level of individual cells.

3.2.4 Heterogeneity of VSMCs within the AA and DT vascular beds

My next objective was to explore whether VSMCs from the AA and DT regions showed heterogeneity within each vascular bed and if so, which genes were highly variable. Variability in observed gene expression levels between cells from the same vascular bed is due to both technical and biological effects. Low amount of starting material is one of the main technical factors contributing to the resulting cell-cell variability (Brennecke et al. 2013). Such technical variability is dependent on the mean expression level of a gene and decreases for highly

expressed genes as higher transcript levels are captured more consistently across cells and dropout effects become less dominant. I was interested in the genes, which showed variability in expression levels beyond the technical noise, given their mean expression level.

To assess which genes were expressed heterogeneously, I adopted an approach for identifying highly variable genes (Lun et al. 2016) based on the relationship between variance and mean log expression of profiled genes. To estimate the background technical variation I fitted a parametric trend to the variance of log expression levels versus mean log expression levels of all genes. The assumption of this method was that a large majority of genes were not variably expressed between the cells, therefore the minority of genes that were expressed heterogeneously should not skew the background estimation. Technical background variation was then subtracted from the total variance of each gene. A gene was identified as highly variable if the remaining biological component of variance was significantly greater than zero (adjusted p-value < 0.05, details in Methods, Lun et al. 2016). This resulted in 176 genes being identified as highly variable in the AA region and 120 in the DT region (Figure 3.14), with 65 of them common to both regions. Interestingly, only two of the genes identified as highly variable in at least one region (*Wif1* and *Lum*) were also identified as differentially expressed between AA and DT (Section 3.2.2), suggesting that heterogeneity within each vascular region is largely driven by factors independent of regional identity.

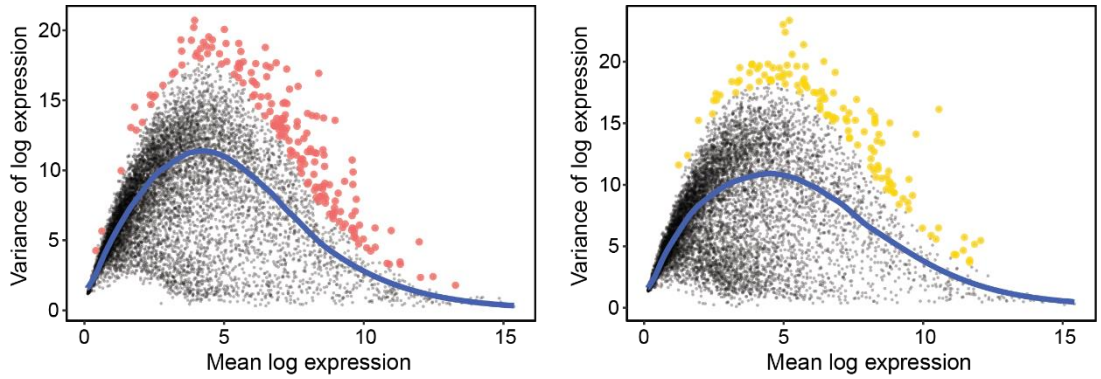


Figure 3.14: Highly variable genes in AA and DT regions.

Scatterplots showing the variance of log2-transformed normalised read counts versus mean log2-transformed normalised read counts, with each black dot representing a gene. Genes identified as highly variable in the AA (red, left) and the DT (yellow, right) are indicated. Blue line represents the estimation of technical variance. The figure is from Dobnikar and Taylor et al. 2018.

To assess the nature of the genes, which were identified as highly variable, I used gene ontology analysis. Highly variable genes within the AA region showed enrichment for "positive regulation of vasculature development" (genes *Pdgfd*, *Hspb6*, *Nras*, *Hk2*, *Efnb2*, *Adm*, *Anxa1*, *F3*, *Rapgef2*, *Myocd*) and "regulation of cell growth" (genes *Cd44*, *Thrb*, *Fn1*, *Rab11a*, *Rtn4*, *Ddx3x*, *Tram1*, *Rbbp7*, *Gja1*, *Sdcbp*, *Sgk1*, *Myocd*, *Efna5*, *Rgs4*) among enriched gene ontology terms (Figure 3.15). Highly variable genes within the DT region showed enrichment for gene ontology term "regulation of cell population proliferation" (genes *Cd44*, *Thrb*, *Itga1*, *Stat1*, *Tes*, *Wisp2*, *Pdgfd*, *Atf3*, *Rtn4*, *Anxa2*, *Dsp*, *Gpx1*, *Nfkbia*, *Fth1*, *Rgs5*, *Efnb2*, *Nupr1*, *Ctnnb1*, *Asph*, *Cdh13*, *Anxa1*, *Gja1*, *Cnbp*, *Hipk1*, *Rps6kb1*). In addition, a number of genes identified as highly variable were involved in processes which play an important role in VSMC biology in disease, such as proliferation and migration. While gene ontology terms "regulation of cell proliferation" and "regulation of cell migration" were not significantly enriched among the highly variable genes in the AA or DT regions, a number of highly variable genes were associated with these gene ontology terms. Forty of the genes identified as highly variable were associated with the gene ontology term "regulation of cell proliferation", including *Rgs5*, *Gja1*, *Pdgfd*, *Irf1*, *Anxa1*, *Anxa2*, *Myocd*, *Fn1* and *Nfkbia*. 21 of the highly variable genes in either region mapped to the go term "regulation of cell migration", including *Pdgfd*, *Anxa1*, *Anxa3*, *Myocd*, *Fn1*, *Adamts1*, *Gja1* and *Postn* among others. Expression profiles for some of the highly variable genes are shown in Figure 3.16.

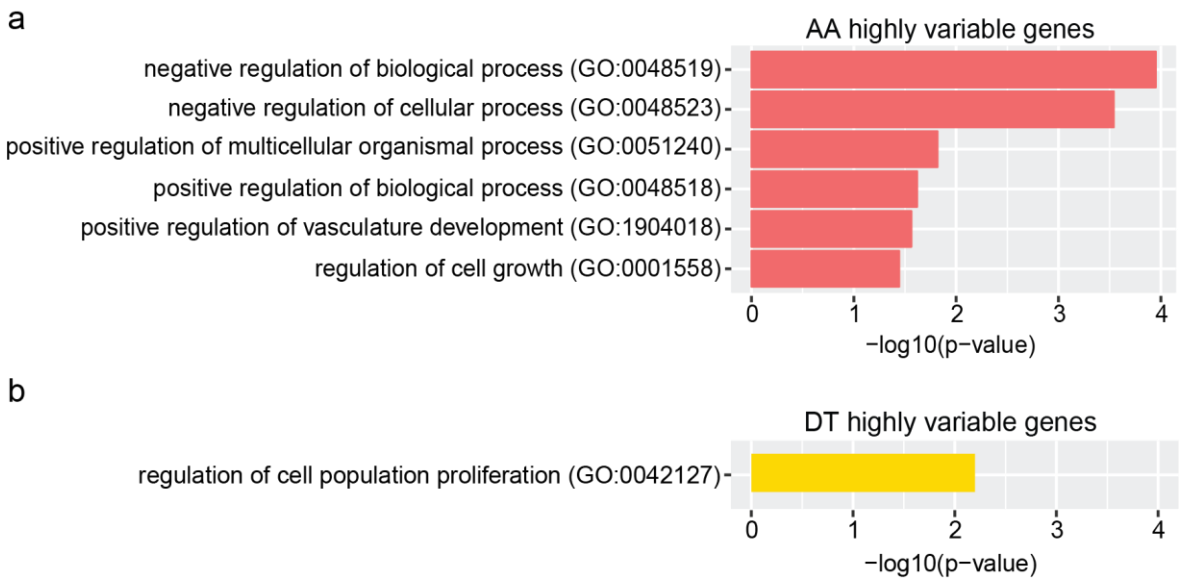


Figure 3.15: Gene ontology analysis of highly variable genes in the AA and DT regions.

Enrichment for gene ontology terms (biological process) among genes identified as highly variable in the AA (a) and DT (b) regions. Bar plots show the significantly enriched gene ontology terms, ranked by their adjusted p-values (details in Methods).

Several of the genes identified as highly variable within the AA or DT regions were previously investigated in the context of cardiovascular disease. For example, *Anxa1* has been observed to be expressed at lower levels in VSMCs isolated from asymptomatic compared with symptomatic human plaques (Viiri et al. 2013). *Pdgfd* is an activator of the PDGF receptor β and overexpression of *Pdgfd* in transgenic mice was observed to lead to increased VSMC proliferation and vascular remodelling (Pontén et al. 2005). *Rgs5* was observed to promote atherosclerotic plaque formation and its expression was found to be increased during arterial remodelling (Arnold et al. 2014). *Rgs5* is also involved in the regulation of blood pressure and VSMC contraction (Gunaje et al. 2011). Overall, this analysis shows that VSMCs from the same vascular region show heterogeneous expression of a range of genes related to biological functions, which are important for VSMC biology in healthy arteries as well as in disease.

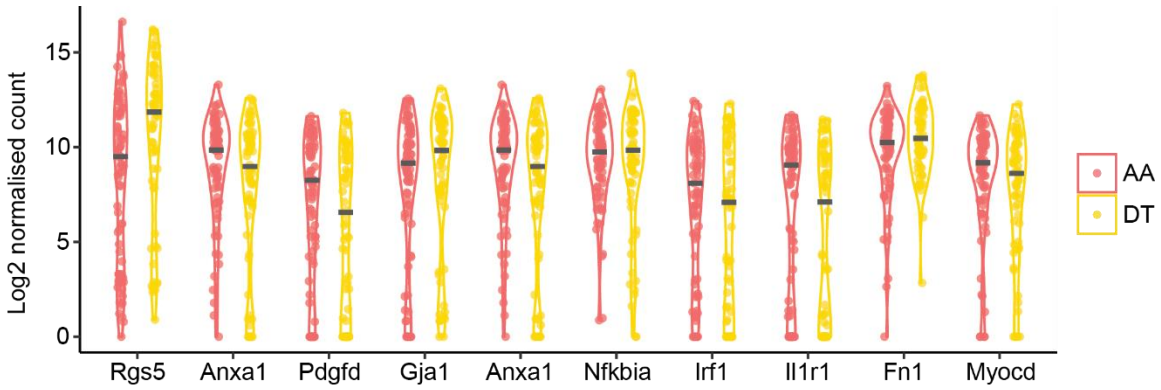


Figure 3.16: Expression levels of highly variable genes in profiled VSMCs.

Violin plots showing the distribution of log2-transformed normalised read counts of selected highly variable genes across the cells from the AA (red) and the DT (yellow). The median is marked with a grey line.

3.2.5 Highly variable gene expression in independent VSMC profiles

Highly variable gene analysis in Section 3.2.4 identified individual genes which showed heterogeneity in expression levels among VSMCs. I was next interested in examining whether subsets of highly variable genes showed coordinated variability between individual VSMCs. However, the number of cells profiled using the Fluidigm C1 platform was too low to detect clear co-expression patterns. Instead, I examined the expression patterns of highly variable genes in independent VSMC profiles. To this end, I used two scRNA-seq datasets of thousands of aortic cells generated using the 10X Genomics Chromium platform in collaboration with A.L. Taylor. Both datasets were generated using aortas of mice on a VSMC lineage traced background (details in Methods). The first dataset was generated only from medial cells FACS-sorted for VSMC-lineage label expression. The second dataset was generated from the whole aorta and contains adventitial and endothelial cells alongside VSMCs.

3.2.6 Processing of the 10X Genomics Chromium VSMC transcriptomes

Sequencing data of single-cell transcriptomes generated using the 10X Genomics Chromium platform was processed through the 10X Genomics pipeline (details in Methods). I then filtered the transcriptomes based on the number of UMIs detected per cell, the number of genes detected and the proportion of mitochondrial reads (Figure 3.17, details in Methods)

and normalised the profiles of the remaining 5686 cells. I then used principal component analysis to identify the major sources of variability between the cells. To reduce noise and dimensionality, clustering analysis was based on the results of principal component analysis, rather than on the expression levels of individual genes. It is desired to exclude non-informative principal components from the analysis to reduce noise and overfitting, however care needs to be taken that important aspects of heterogeneity are not excluded from the analysis. The standard deviation of identified principal components and the clusters of cells they delineate are shown in Figure 3.18. The first 15 principal components, all of which explained significantly more variance between the cells than would be expected at random, were included in the final clustering analysis (details in Methods). In total, 9 clusters were identified. Clusters 0, 1, 2, 3, 5 and 6 expressed VSMC marker genes, such as *Myh11*, and all contained cells from both the whole aorta and the VSMC-only datasets (Figure 3.19). Cells within clusters 4 and 7 expressed adventitial marker gene *Pdgfra* or the endothelial marker gene *Cdh5* and were assigned adventitial and endothelial identities respectively (Figure 3.19).

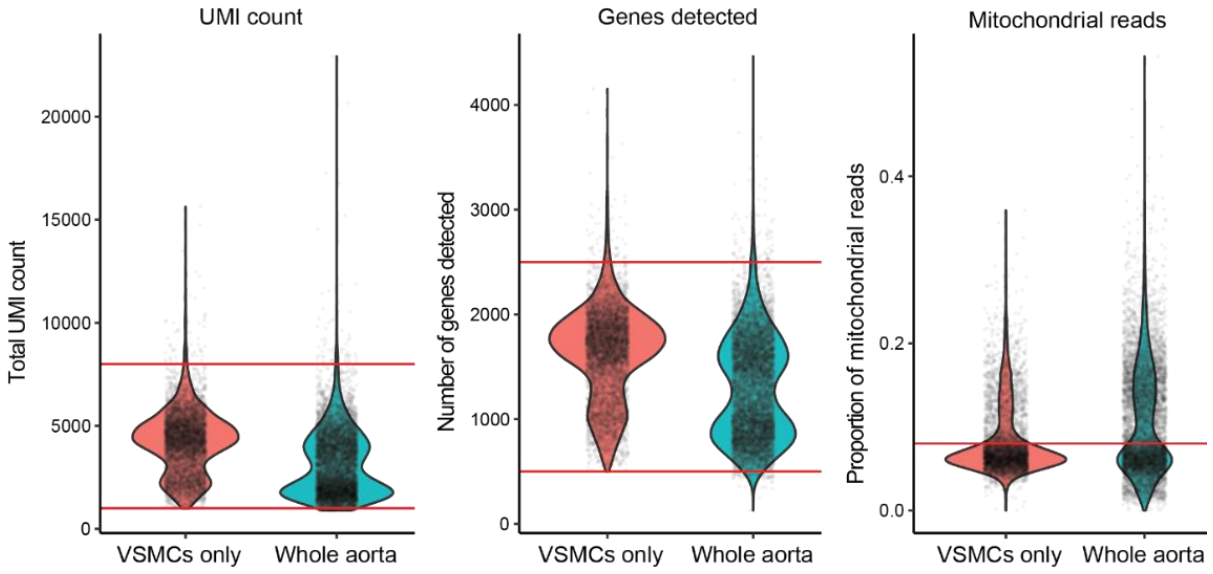


Figure 3.17: Quality control of the VSMCs only (red) and whole aorta (blue) 10X Genomics Chromium datasets.

Violin plots show the distribution of values for total UMI count, number of genes detected and proportion of mitochondrial reads. Black dots represent individual cells and red lines indicate the quality control thresholds. Cells were considered to be of good quality if total UMI count was between 1000 and 8000, number of genes detected was between 500 and 2500 and the percentage of mitochondrial reads was below 8%.

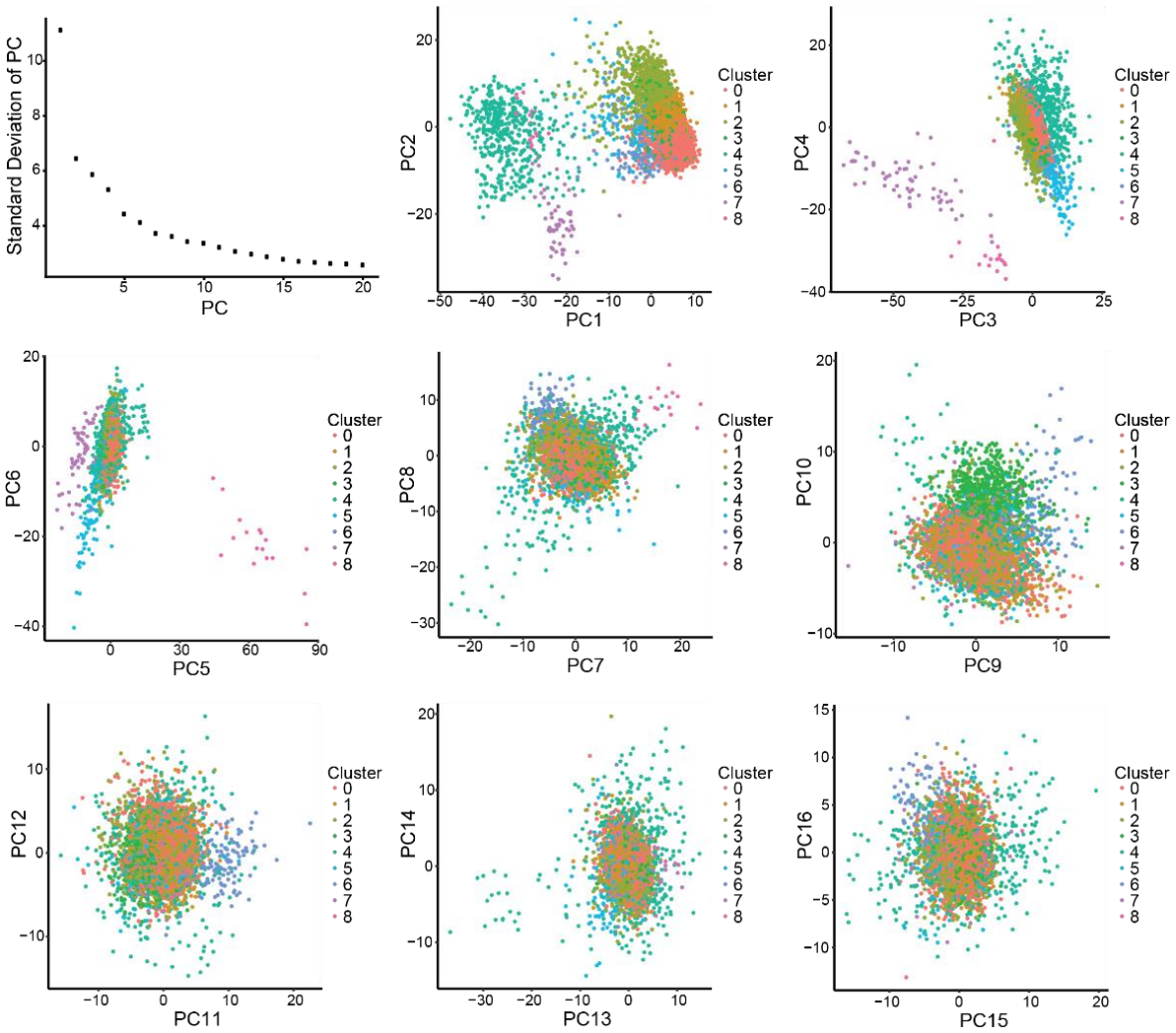


Figure 3.18: Principal component analysis and selection of principal components for further analysis. Top left plot shows the standard deviation of principal components 1-20. The higher the standard deviation, the more variance is explained by that component. The other plots show the values of principal components 1-16 for profiled cells. Each cell is represented by a circle and is coloured-coded by cluster membership. PC1 separated VSMCs (clusters 0, 1, 2, 3, 5 and 6) from other profiled cells, PC2 separated the endothelial cells (cluster 7), P3 further separated endothelial cells as well profiles likely originating from resident immune cells (cluster 8), while PCs 4 and 5 further separated cluster 8.

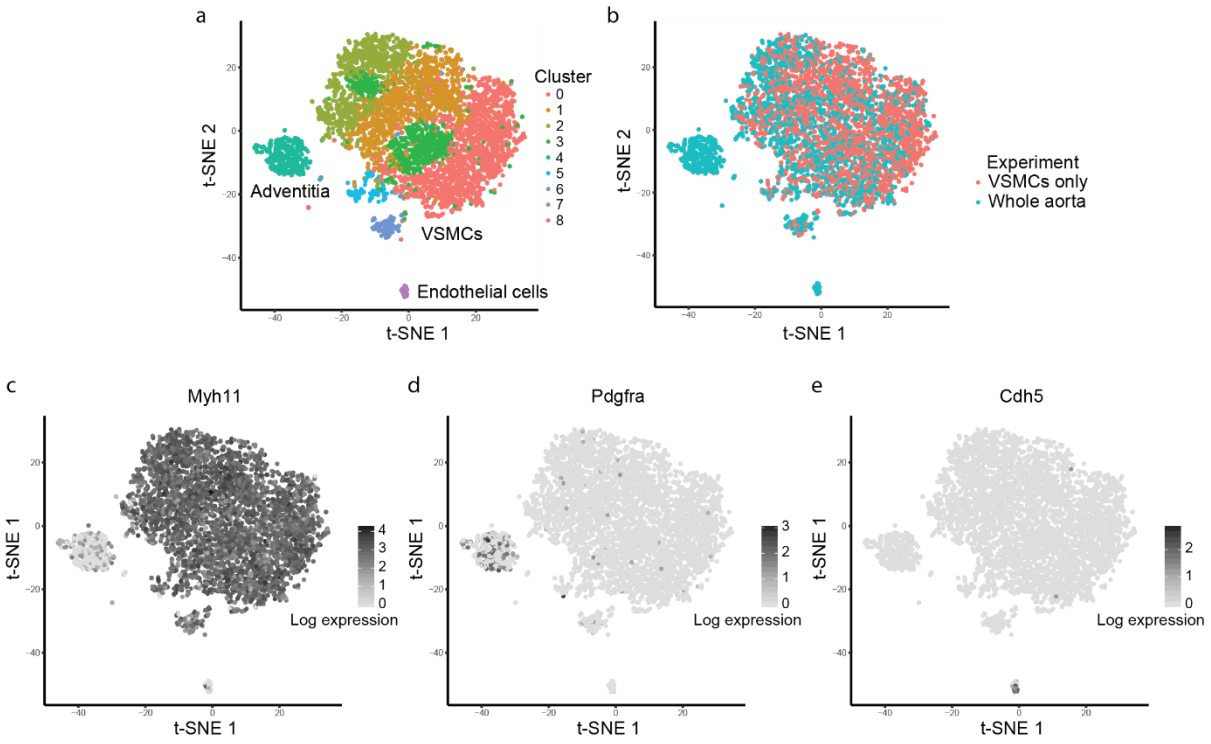


Figure 3.19: Clustering analysis and identification of cell types within the VSMC-only and whole aorta 10X Genomics Chromium datasets.

a) Results of clustering analysis (details in Methods) with adventitial, endothelial and VSMC clusters indicated on a t-SNE plot. b) Distribution of cells from the VSMC-only and whole aorta datasets, which confirmed assignemnt of VSMC cluster identities. Cells from the VSMCs-only dataset are shown in red and whole aorta dataset cells in blue. (c-d) Expression levels of VSMC marker *Myh11*, adventitial marker *Pdgfra* and endothelial marker *Cdh5* are colour coded on a t-SNE plot, with darker gray representing higher expression levels.

3.2.7 Highly variable genes show co-expression in VSMCs profiled using the 10X Genomics Chromium platform

Next I investigated whether highly variable genes identified in the Fluidigm C1 dataset showed differential expression in specific clusters of VSMCs. I performed differential expression analysis for each VSMC cluster relative to the remaining VSMCs (log2 fold change > 0.5 and adjusted p-value < 0.05, details in Methods) and examined the overlap of upregulated and downregulated genes in each cluster with previously identified highly variable genes (Table 3.1).

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 5	Cluster 6
Total upregulated genes	0	0	64	19	3	62
Of which highly variable genes	0	0	17	1	0	3
Total downregulated genes	55	3	1	0	0	42
Of which highly variable genes	13	0	1	0	0	5

Table 3.1: Table shows the numbers of upregulated and downregulated genes within each of the VSMC clusters (details in Methods), which overlap with highly variable genes identified in either the AA or DT region.

One of the most highly variable genes in the Fluidigm C1 dataset *Rgs5*, was found to be upregulated in cluster 6 and downregulated in cluster 2 (Figure 3.20a). Several genes were identified as differentially upregulated or downregulated in Cluster 6, suggesting that its expression profiles differed more substantially than the remaining VSMC clusters (Table 3.1). Clusters 1 and 5 did not differentially express any of the highly variable genes and the only highly variable gene upregulated in cluster 3 was *Gnb1* (Table 3.1). VSMCs in cluster 2 expressed 17 highly variable genes at higher levels and 13 highly variable genes were downregulated in cluster 0 (Table 3.1). Notably, 12 of the highly variable genes upregulated in cluster 2 were also downregulated in cluster 0 (Figure 3.20b). Expression patterns of the most strongly upregulated genes in cluster 2 (*Nfkbia*, *Wsb1* and *Ppp1r15a*) and downregulated genes in cluster 0 (*Nfkbia*, *Gadd45b* and *Ppp1r15a*) are shown in Figure 3.20c. Observed differences in expression levels of highly variable genes between clusters 0 and 2 are not explained by quality control parameters, such as the total normalised log-transformed expression levels or the number of genes detected per cell (Figure 3.20d). Overall, this analysis suggests that a subset of highly variable genes is co-expressed in VSMCs from healthy arteries.

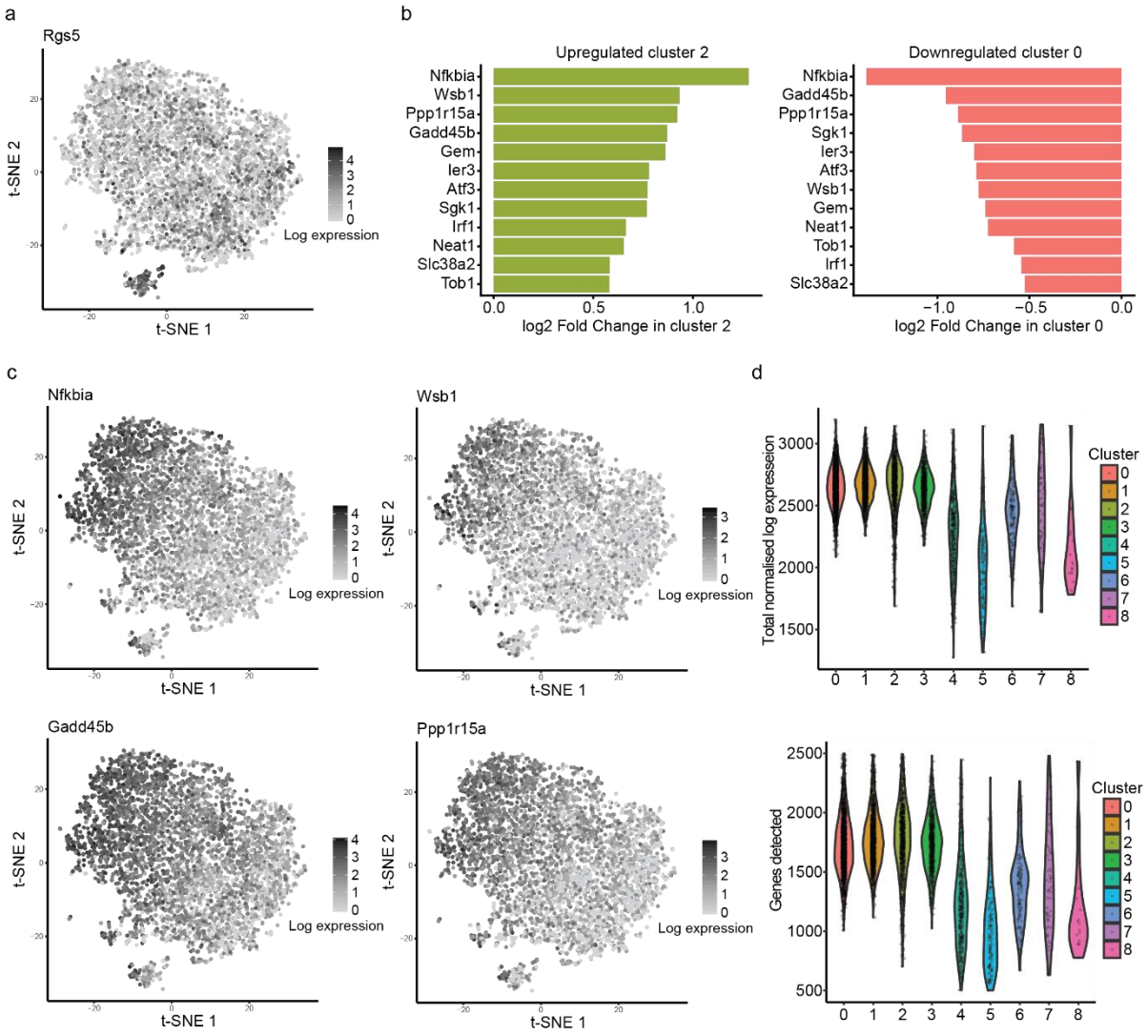


Figure 3.20: Highly variable gene expression in VSMC profiled as part of the 10X Genomics Chromium whole aorta and VSMCs-only datasets.

a, c) Expression levels of highly variable genes Rgs5 (a), Nfkbia, Wsb1, Gadd45b and Ppp1r15a (c) are colour coded on a t-SNE plot, with darker gray representing higher expression levels. b) Barplots showing the highly variable genes, which are upregulated in cluster 2 (left, green) and downregulated in cluster 0 (right, left). The log2 fold change of each gene within clusters 2 and 0 is shown. d) Violin plots representing the quality control parameters of total normalised log expression and the number of genes detected per cell, shown separately for each cluster, as defined in Figure 3.19a. Black dots represent individual cells.

3.2.8 Application of the random forest model to independent VSMC profiles

I was next interested in how generalizable the findings from the random forest analysis were to independent datasets. While the random forest model achieved high accuracy for predicting regional identity of aortic VSMCs within the Fluidigm C1 dataset, I wanted to explore how versatile the model was for predicting regional identity in sparser datasets. I used the scRNA-seq datasets of aortic cells generated using the 10X Genomics Chromium described in Section 3.2.6, to explore the predictive power of the random forest classifier. It is important to note that the AA and DT regional identity of the cells in these datasets was not known in advance, therefore they could not be used as a formal validation of the random forest classifier. Instead, I applied the classifier to these datasets to learn whether it has any predicative power based on expression levels of known differentially expressed genes.

I first checked the expression of random forest predictor genes in VSMCs profiled using the 10X Genomics Chromium platform. While all 30 predictor genes were expressed at appreciable levels in the Fluidigm C1 dataset, expression was much sparser in the 10X Genomics Chromium dataset, with 11 out of 30 genes expressed in less than 1% of VSMCs. The median percentage of VSMCs expressing a particular predictor gene was 56% for Fluidigm C1 cells and 4% for 10X VSMCs (Figure 3.21a). Additionally, the number of random forest predictor genes expressed per cell was drastically lower in the 10X Genomics Chromium dataset compared with the Fluidigm C1 dataset (Figure 3.21b). Such sparsity in random forest predictor genes' expression is expected to affect the classifier performance. This was particularly problematic for DT cells, as there were only 9 predictor genes, which were upregulated in the DT region according to bulk RNA-seq, as opposed to 21 AA predictors.

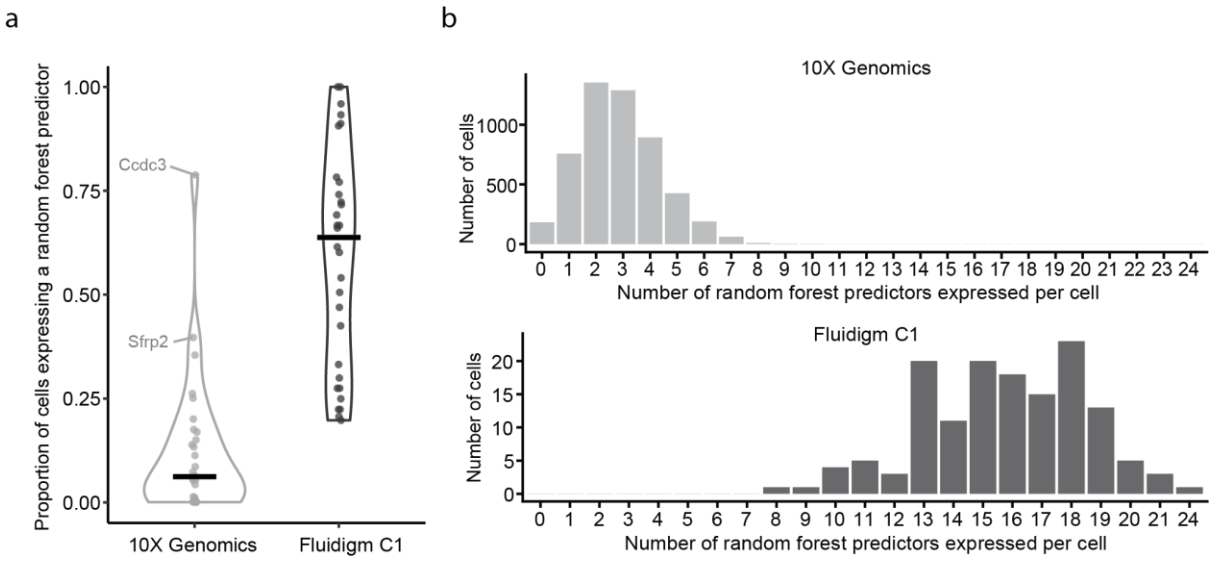


Figure 3.21: Random forest predictor genes are expressed sparsly in VSMCs-only and whole aorta 10X Genomics Chromium datasets.

a) Violin plot showing the proportion of cells that express a random forest predictior, with dots representing individual predictor genes. Median levels for 10X Genomics Chromium (light grey) and Fluidigm C1 (dark grey) are indicated with a black bar. b) Barplots showing the distribution of the number of random forest predictors expressed in an individual cell in the 10X Genomics Chromium (top) and Fluidigm C1 (bottom) datasets.

The random forest classifier predicted that 690 VSMCs were of DT origin and 4479 VSMCs from the AA. The outcome was likely biased towards AA predictions due to a higher number of random forest predictor genes identified as upregulated in the AA with bulk RNA-seq. As an initial check of the regional predictions, I examined the expression levels of random forest predictor genes *Hand2*, *Sfrp2*, *Hoxa7* and *Ccdc3* (Figure 3.22a). As expected, *Hand2* and *Sfrp2* were overall expressed at higher levels in VSMCs predicted to be of the AA origin, while *Hoxa7* and *Ccdc3* showed higher expression levels in VSMCs predicted to originate from the DT region. Since *Hand2*, *Sfrp2*, *Hoxa7* and *Ccdc3* were part of the random forest model, the fact that higher expression of these genes was observed in the appropriate region was not verification of whether the regional identity predictions were sensible. To partially overcome the limitation of not knowing the regional identity of VSMCs in advance, I examined the expression levels of all genes, which were identified as differentially expressed in either the AA or DT regions according to bulk RNA-seq. Figure 3.22b shows that overall single cells with predicted AA identity showed higher expression levels of genes upregulated in the AA region

than DT-predicted cells. Similarly, cells predicted to originate from the DT showed overall higher expression levels of genes upregulated in the DT region.

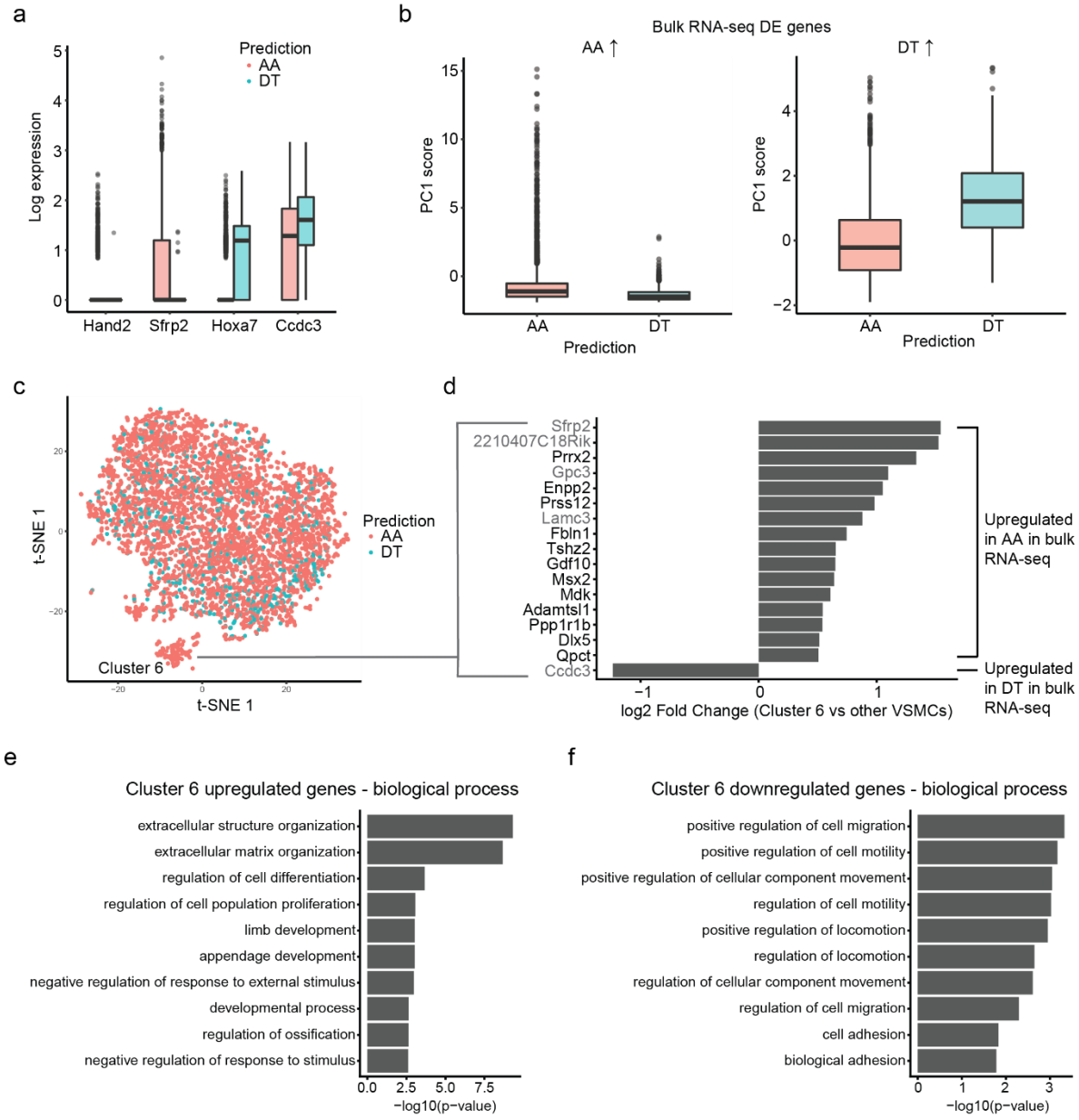


Figure 3.22: Random forest model applied to 10X Genomics Chromium datasets.

a) Boxplots showing the expression levels of AA-upregulated random forest predictors *Hand2* and *Sfrp2* and DT-upregulated predictors *Hoxa7* and *Ccdc3* in 10X profiled VSMCs, which were predicted to originate from either the AA or DT region. b) Boxplots showing the PC1 score in VSMCs segregated based on their predicted regional identity. Principal component analysis is based on all genes identified as upregulated in either AA (left) or DT (right) using bulk RNA-seq. Higher PC1 score represents higher overall expression in a cell (details in Methods). c) t-SNE plot colour coded by the random forest

prediction for each VSMC. AA predictions are shown in red and DT predictions in blue. d) Barplot showing the log2 fold change in expression between cluster 6 and other VSMCs for genes differentially expressed in cluster 6 as well as in bulk RNA-seq. Genes included in the random forest model are indicated in grey. e, f) Barplots showing the top 10 overrepresented gene ontology terms ranked by adjusted p-value, which were found to be significantly enriched among the genes upregulated (e) or downregulated (f) in cluster 6 relative to other VSMCs.

The majority of VSMC clusters contained cells which were predicted as originating from both the AA and the DT regions. Cluster 6, however, contained only cells predicted to originate from the AA region (Figure 3.22c). Despite the fact that more VSMCs were predicted as originating from the AA than the DT region, observing 0% of DT predictions in cluster 6 and 16% of DT predictions in other VSMCs would not be expected at random (Fisher's exact test, p = 4e-9, details in Methods). Differential gene expression analysis discussed in Section 3.2.7 identified 62 genes as upregulated and 42 genes as downregulated in cluster 6 relative to other VSMC clusters. Of the 62 genes identified as upregulated in cluster 6, 16 were upregulated in the AA bulk RNA-seq profiles, while none of the 62 genes were expressed at significantly higher levels in the DT bulk RNA-seq profiles. Additionally, one of the genes downregulated in cluster 6 was found to be upregulated in the DT region in bulk RNA-seq (Figure 3.22d). Importantly, 11 of the 16 AA genes upregulated in cluster 6 and in bulk AA profiles were not part of the random forest model, which supported the assignment of AA identities to VSMCs in this cluster by the classifier (Figure 3.22d). Gene ontology analysis of genes upregulated in cluster 6 showed overrepresentation of genes involved in extracellular matrix organisation, regulation of cell population proliferation and regulation of ossification, among other functions (Figure 3.22e). Downregulated genes showed overrepresentation for genes involved in positive regulation of cell migration and cell adhesion, as well as other gene ontology terms (Figure 3.22f). These functions are relevant for VSMC biology, suggesting that cluster 6 may have a functional role within the AA region. Since VSMCs within cluster 6 comprise only 3% percent of all VSMCs profiled using the 10X Genomics Chromium platform, this subpopulation of cells was likely to rare to have been detected in the Fluidigm C1 dataset, where only 79 AA cells were profiled.

In conclusion, this analysis suggested that the random forest model may be able to identify the AA or DT affiliation of cells in the 10X Genomics Chromium dataset, however the sparsity of the data is likely a limiting factor for the performance of the model. Overall, further

validation where the regional identities of VSMCs are known in advance is required to form concrete conclusions about the performance of the model in this setting. Two separate 10X Genomics Chromium datasets of the AA and DT VSMCs would also enable the validation of the observation that VSMCs in cluster 6 appear to originate only from the neural crest-derived AA region. Interestingly, a similar cluster of VSMCs was observed after carotid ligation injury in carotid arteries, which are also derived from the neural crest (Section 5.2.2).

3.3 Discussion

In this chapter I used single-cell transcriptomics to probe the heterogeneity of VSMCs. In particular, I focused on the regional heterogeneity between parts of the aorta with distinct developmental origins, as well as on the heterogeneity within an aortic region. All profiled VSMCs shared common features, such as expression of VSMC marker genes. However, consistent differences in gene expression profiles were observed between individual VSMCs originating from the AA or DT aortic regions. Moreover, I observed significant cell-to-cell heterogeneity with respect to genes involved in VSMC biology and disease within a vascular bed. This suggests that VSMC heterogeneity in the healthy state may contribute to differential responses to stimuli of VSMCs.

3.3.1 Regional heterogeneity of VSMCs

Several studies have described the regional differences in susceptibility to vascular disease, such as atherosclerosis and vascular calcification (Haimovici & Maier 1971; DeBakey & Glaeser 2000; Leroux-Berger et al. 2011). In this section I studied the transcriptional differences between the atherosclerosis-susceptible AA and atherosclerosis-resistant DT regions, in which the VSMC populations are of distinct embryonic origins (Trigueros-Motos et al. 2013; Majesky 2007). Transcriptional variation between the AA and DT regions has been previously investigated using microarrays (Trigueros-Motos et al. 2013; Van Assche et al. 2011). Bulk RNA-seq analysis replicated the key findings of higher levels of *Hox* gene expression in VSMCs from the DT region. Among the genes identified as upregulated in the DT region according to bulk RNA-seq and previous microarray experiments was *Hoxa9*, which has been previously reported to reciprocally inhibit the proinflammatory transcription factor *Nfkb1* in VSMCs (Trigueros-Motos et al. 2013). Trigueros-Motos et al. (2013) further reported higher levels of

Nfkb1 activity in the AA, which provides a possible explanation of higher susceptibility for atherosclerosis through increased inflammation. In agreement with this, gene ontology analysis of AA upregulated genes according to bulk RNA-seq revealed enrichment for genes involved in the inflammatory response.

Distinct embryonic origins of the neuroectoderm-derived AA and mesoderm-derived DT VSMCs are one of the factors affecting the transcriptional differences between the two regions. However, there are other factors which could be underpinning the observed differences. For example, environmental factors including the local hemodynamic forces and the levels of mechanical stress experienced in a given vascular bed may play a role. For instance, atherosclerotic plaques frequently develop at branching sites or in areas experiencing high mechanical forces (Cunningham & Gotlieb 2005; Chiu & Chien 2011). In this analysis it was not possible to separate the confounding effects of differing local environments and distinct embryonic origins. To isolate the effects of embryonic origin on the susceptibility to atherosclerosis, grafts from the atherosclerosis-susceptible canine abdominal aorta have been transplanted into the jugular vein in a previous study (Haimovici & Maier 1971). The grafts retained higher susceptibility for atherosclerosis in the lower pressure environment of the jugular vein, suggesting an intrinsic difference in disease susceptibility (Haimovici & Maier 1971). More specifically, the transcriptional differences between VSMCs from the AA and DT regions have previously been addressed in the context of environmental factors. Advances in *in vitro* programming of pluripotent stem cells have enabled *in vitro* differentiation of VSMCs of the neuroectodermal or the mesodermal lineages (Cheung et al. 2012). Higher expression levels of several *Hox* genes, including *Hoxa9*, have been observed in VSMCs of mesodermal lineage compared to neuroectodermal VSMCs, both of which have been derived from human pluripotent stem cells *in vitro* (Cheung et al. 2012; Trigueros-Motos et al. 2013). This finding suggests that differences in expression between the two regions are not entirely due to environmental factors, and indicates that observed transcriptional differences between the VSMCs from the AA and DT regions may apply in human VSMCs.

Bulk RNA-seq analysis has provided a refined list of differentially expressed genes between the two regions. However, population-level profiling approaches are unable to determine whether the observed differences are uniformly present in individual cells or whether they are due to differential composition of VSMCs in the two regions. The work in this chapter,

enabled by recent advances in single-cell transcriptomics, has looked at regional heterogeneity of aortic VSMCs at the single-cell level. The fact that it was possible to construct an accurate classifier suggests that the transcriptional signatures of the AA and DT regions are present at the level of individual cells, as reliable classification would not be possible otherwise.

The application of the random forest model to aortic VSMCs profiled using the 10X Genomics Chromium platform indicates that the random forest has some discerning power between the AA and DT VSMCs in sparse datasets. This suggests that such an approach could be useful for assigning regional identities to VSMCs within larger datasets such as the mouse single-cell atlas (The Tabula Muris Consortium 2018), where AA and DT regions were not profiled separately. Interestingly, the application of the random forest classifier to the whole aorta and VSMC-only datasets profiled using the 10X Genomics Chromium platform highlighted a subset of VSMCs, which were predicted to originate exclusively from the AA region and expressed *Rgs5* at higher levels than average for VSMCs. However, further validation in settings where the AA and DT identities of VSMCs are known in advance is required to test the performance of the model in sparser datasets, and to validate whether the subset of cells with increased *Rgs5* expression is present exclusively in the AA region.

3.3.2 Highly variable genes

High variability in levels of VSMC marker proteins between VSMCs resident in the same vascular bed has been previously observed during response to injury (Frid et al. 1994; Rensen et al. 2007). Rensen et al. (2007) proposed that VSMCs exist in a spectrum of phenotypes ranging from contractile to synthetic within healthy tissue. I hypothesised that variability in the healthy state may underpin this differential response of VSMCs to stimulus. Despite the uniform expression of regional signatures in single cells, I observed significant cell-to-cell heterogeneity within each vascular bed. Several genes identified as highly variable within a vascular bed are known to be involved in VSMC biology and disease. Interestingly, there was a low overlap between genes identified as highly variable and those identified as differentially expressed between the AA and DT regions using bulk RNA-seq. This suggests that cell-to-cell heterogeneity within a vascular bed is largely driven by factors independent of the regional identity.

The approach taken to identify highly variable genes in this chapter relies on the estimation of the background technical variation level and identifying genes with significantly higher variability. This approach does not exclude genes which may show true cell-to-cell variation at the time of profiling, but where such variation does not necessarily have functional implications. Such examples of stochastic variation include transcriptional bursts (Raj & van Oudenaarden 2008). Additionally, variable expression of a gene does not necessarily reflect variable activity levels of the corresponding protein product. There is potential for post-transcriptional regulation, which is not captured with the single-cell RNA-seq approach. Since the variability of expression of individual genes could be driven by factors which do not influence the functional activity levels of the resulting protein, it would be informative to look for coordinated variability within a set of highly variable genes. The numbers of cells profiled using the Fluidigm C1 method were too low to detect co-expression patterns within highly variable genes. However, examining the VSMCs profiled using the 10X Genomics Chromium platform revealed that a subset of highly variable genes are co-expressed.

Heterogeneity between single VSMCs at the transcriptional level is in agreement with previous observations that several G-protein coupled receptor genes are heterogeneously expressed in different vascular regions as well as among aortic VSMCs (Kaur et al. 2017). This study was based on single-cell RT-qPCR and further investigated the differences between aortic VSMCs from healthy vessels and atherosclerotic arteries. The authors observed that the transcriptional profiles of a subset of healthy VSMCs matched the profiles of atherosclerotic VSMCs. This subset of healthy VSMCs showed upregulation of a number of G-protein coupled receptor genes, including *Gprc5b*, and was reported to be localised at the inner curvature of the AA, which is at increased risk of plaque development due to disturbed blood flow in the region (Kaur et al. 2017; Chiu & Chien 2011). No clear segregation of VSMCs with analogous expression profiles was observed in either the 10X Genomics Chromium or the Fluidigm C1 datasets, possibly due to lower sensitivity of the method compared to the RT-qPCR approach taken by Kaur et al. (2017). *Gprc5b* was only detected at low levels in a small numbers of cells, and its expression levels may be too low to reliably detect it with scRNA-seq.

3.3.3 Conclusion

In conclusion, the work in this chapter delineated the transcriptional heterogeneity of VSMCs between different regions, as well as between individual VSMCs from the same vascular bed. I have shown that reliable regional classification of individual VSMC transcriptomes is possible, and identified a putative population of cells present only in the AA by applying the classifier to an independent dataset. The analysis of highly variable genes within each vascular bed revealed significant cell-to-cell heterogeneity of many disease-relevant genes. Furthermore, a subset of highly variable genes was co-expressed within the 10X dataset. In the future, it would be interesting to investigate the spatial positioning of VSMCs expressing highly variable genes at different levels. Single-molecule fluorescence *in situ* hybridisation or immunostaining for a subset of co-expressed highly variable genes could reveal whether there are any areas of the aorta, where consistently higher signal is observed. In particular, it would be informative to explore whether areas of disturbed flow, such as the branch sites, express increased levels of disease-relevant highly variable genes in healthy arteries. This may aid in the investigation of the early transcriptional changes in VSMCs, which are involved in atherosclerosis and provide direction for future investigations into early targeting of VSMCs involved in disease.

4 Rare subpopulation of SCA1-positive VSMCs detected in healthy arteries

4.1 Introduction

VSMCs possess a remarkable capacity to switch from a contractile muscle phenotype to a proliferative synthetic state in response to injury or inflammation (Rensen et al. 2007). However, recent studies have challenged the assumption that the majority of VSMCs generate the phenotypically modulated VSMC populations under these conditions. For example, it has been found that only very few VSMCs undergo extensive proliferation in atherosclerosis, which results in clonal or oligoclonal VSMC patches in atherosclerotic plaques (Chappell et al. 2016; Jacobsen et al. 2017).

There have been previous reports of progenitor populations within the vasculature, which have been proposed to differentiate towards VSMCs and to contribute to atherosclerotic plaque formation and blood vessel calcification (Hu et al. 2004; Passman et al. 2008; Kramann et al. 2016). However, several studies showed that VSMCs within the atherosclerotic plaque originated from VSMCs, which expressed contractile differentiation markers prior to atherosclerosis plaque development (Nemenoff et al. 2011; Gomez et al. 2013; Feil et al. 2014; Shankman et al. 2015; Chappell et al. 2016; Jacobsen et al. 2017).

Jointly, these findings have suggested that VSMCs are functionally heterogeneous with respect to their response in atherosclerosis. In collaboration with H.F. Jørgensen's group, we pursued the hypothesis that the differential VSMC response to injury and inflammation may be underpinned by the heterogeneity of VSMCs already present in healthy arteries. To characterise this heterogeneity, we combined scRNA-seq with a VSMC-specific lineage labelling system and profiled the transcriptomes of single VSMC-lineage cells isolated from healthy arteries. The analysis revealed a rare subpopulation of VSMCs expressing the progenitor marker SCA1, which expressed transcriptional signatures characteristic of the synthetic VSMC state.

4.2 Results

4.2.1 Detection of SCA1-positive cells of VSMC lineage in healthy arteries

Analysis of the transcriptomes of *ex vivo* VSMCs from the AA and DT regions revealed that one of the DT cells co-expressed VSMC marker genes as well as the *Ly6a* transcript (*Ly6a/Sca1*) at high levels, which encodes the SCA1 protein (Figure 4.1a). Subsequent flow cytometry analysis of medial aortic cells stained using the anti-SCA1 antibody showed that 0.5-1% of medial cells were SCA1-positive (Figure 4.1b, collaboration with H.F. Jørgensen). SCA1 is a progenitor marker (Holmes & Stanford 2007) and was previously found to be expressed in progenitor cell populations within the vasculature (Hu et al. 2004; Passman et al. 2008). This prompted us to explore the subpopulation of medial SCA1-positive cells further.

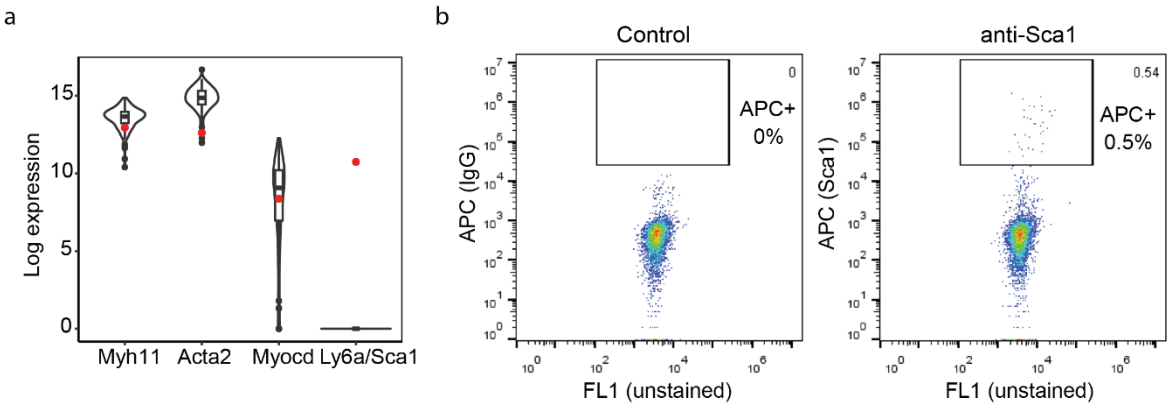


Figure 4.1: *Ly6a/Sca1* transcript and SCA1 protein is expressed in a small number of medial cells.
a) Violin plots showing log2-transformed normalised read counts of *Myh11*, *Acta2*, *Myocd* and *Ly6a/Sca1* in *ex vivo* VSMCs from the AA and DT regions profiled using the Fluidigm C1 platform. The cell co-expressing VSMC marker genes and *Ly6a/Sca1* is shown in red. b) Flow cytometry analysis of SCA1 protein expression in medial cells. Medial cells were stained either with allophycocyanin (APC) conjugated to IgG control (left) or the SCA1 antibody (right). Flow cytometry was carried out by H.F. Jørgensen and all panels are from Dobnikar and Taylor et al. 2018.

We next used a VSMC-specific lineage tracing system to establish whether SCA1-positive medial cells are of the VSMC lineage. We used *Myh11*-CreERt2 transgenic animals in combination with the eYFP single-colour or the R26-Confetti multi-colour reporter (Chappell

et al. 2016) to induce permanent fluorescent protein expression in VSMCs after tamoxifen injection (Figure 4.2, details in Methods). As *Myh11* is a specific marker of VSMCs, tamoxifen exposure lead to Cre expression and subsequent recombination at the R26-Confetti/eYFP locus in VSMC.

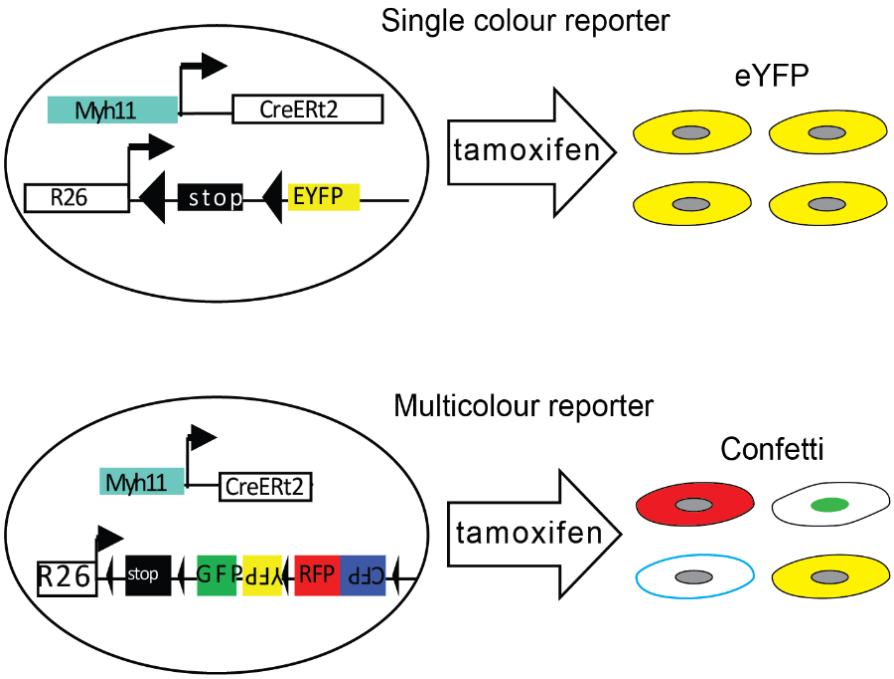


Figure 4.2: Schematic representation of the single-colour (top) and multi-colour (bottom) VSMC lineage labelling approaches.

Upon *Myh11* expression Cre-mediated recombination at the Rosa26 locus occurs. This results in stable expression of eYFP in the single-colour reporter or of one of the fluorescent proteins (GFP, YFP, RFP or CFP) in the multi-colour reporter. Resulting fluorescent label is passed on during cell proliferation and is retained if expression of *Myh11* is subsequently lost, for example during phenotypic switching. Figure adapted from Dobnikar and Taylor et al. 2018.

Flow cytometry analysis of medial cells from the aortas of mice on single-colour VSMC lineage tracing background revealed that 7.4% of medial SCA1-positive cells express the VSMC lineage label (Figure 4.3, collaboration with A.L. Taylor and H.F. Jørgensen). Since the VSMC lineage labelling efficiency has been estimated to be at 40-90% for the eYFP reporter, the proportion of SCA1-positive cells of VSMC lineage in the media may be higher.

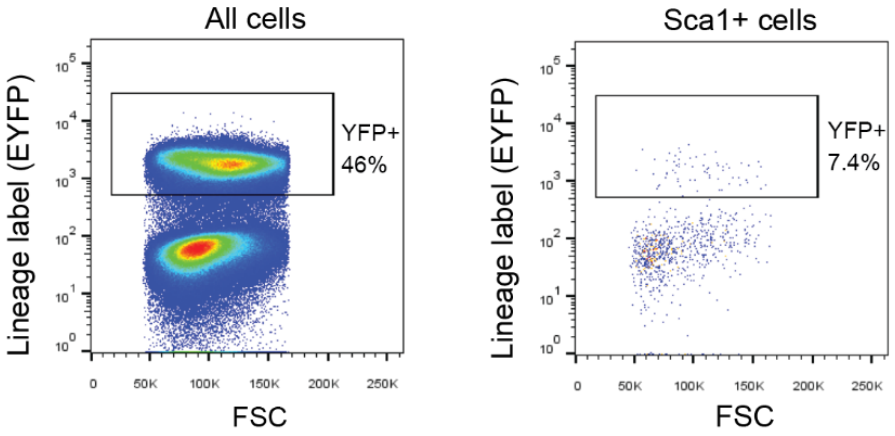


Figure 4.3: A proportion of medial SCA1-positive cells express the VSMC lineage label.
Scatterplots show the forward scatter (FSC) of analysed medial aortic cells from mice on single-colour VSMC lineage tracing background against their eYFP VSMC lineage label expression. All medial cells (left) or SCA1-positive medial cells only (right) are shown. Rectangular boxes represent the eYFP-positive cells. Flow cytometry analysis was carried out by A.L. Taylor and H.F. Jørgensen and the figure was adapted from Dobnikar and Taylor et al. 2018.

To further validate the existence of a VSMC-lineage SCA1-positive subpopulation within the medial layer we analysed the medial cells of transgenic Sca1-GFP mice (Ma et al. 2002). Flow cytometry analysis showed that 0.2-1.6% of medial cells were GFP-positive (Figure 4.4a, collaboration with A.L Taylor), which is in agreement with SCA1 immunostaining analysis (Figure 4.1b). To verify that GFP-positive medial cells have a VSMC identity we sorted GFP-positive medial as well as GFP-positive adventitial cells and immunostained them for the contractile VSMC marker ACTA2. As expected, none of the adventitial GFP-positive cells expressed ACTA2, whereas 25-86% of medial GFP-positive stained positive for ACTA2 (Figure 4.4b, collaboration with A.L Taylor). Overall, this approach has provided complementary evidence for the existence of a rare medial subpopulation of VSMCs, which express SCA1.

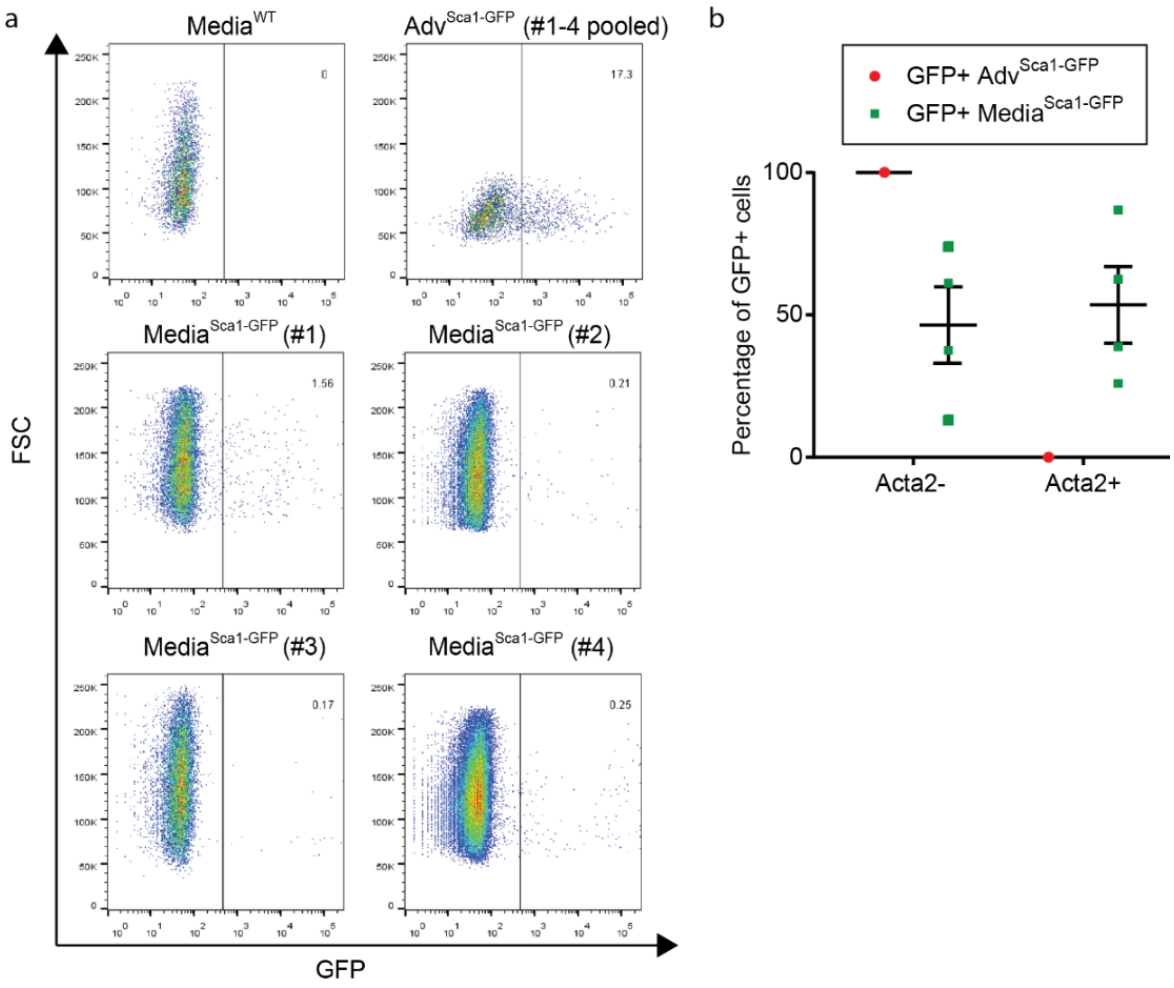


Figure 4.4: A subset of medial cells from transgenic Sca1-GFP animals co-express GFP and VSMC marker ACTA2.

a) Medial cells from wild-type as well as transgenic Sca1-GFP animals were analysed for the expression of GFP using flow cytometry. None of the wild-type medial cells were positive for GFP, whereas 0.2-1.6% of medial cells from Sca1-GFP animals expressed GFP across four experiments. The average GFP expression in adventitial cells pooled from four animals is shown for reference. b) Plot showing the percentage of sorted GFP-positive medial and adventitial cells, which stained positive for ACTA2. GFP-positive adventitial samples are represented by red circles and GFP-positive medial samples by green squares. Flow cytometry analysis and ACTA2 immunostaining was carried out by A.L. Taylor and the panels are from Dobnikar and Taylor et al. 2018.

To investigate whether the transcriptional signatures of SCA1-positive VSMCs showed increased potential for phenotypic switching, we next profiled cells from the whole aorta as well as only the cells positive for the VSMC lineage label (L+) using the 10X Genomics Chromium platform (Zheng et al. 2017, collaboration with A.L. Taylor). This approach allowed

us to profile thousands of cells and enabled us to capture the rare SCA1-positive subpopulation of VSMCs. The transcriptomes of profiled cells in the whole aorta and L+ only datasets have been processed and analysed jointly as described in Chapter 3, Section 3.2.6. This analysis identified three main populations of cells; the VSMCs, adventitial and endothelial cells. I observed several *Ly6a/Sca1* expressing cells within the VSMC cluster of the whole aorta dataset (Figure 4.5a), which further confirmed the existence of SCA1-positive VSMCs. Within the L+ only dataset I observed a total of five cells expressing low levels of *Ly6a/Sca1* (Figure 4.5b). Three of these cells were located in the core VSMC cluster with the remaining two cells showing different transcriptional profiles. One of these cells was located within the endothelial cluster otherwise composed of endothelial cells profiled as part of the whole aorta dataset. Plasticity between the mesenchymal and endothelial cells has been proposed to occur both *in vitro* and *in vivo* (Sainz et al. 2006; Ubil et al. 2014; Evrard et al. 2016; Chen et al. 2016), however the VSMC lineage cell located in the endothelial cluster is more likely to be a contamination from the neighbouring endothelial layer. This could either be as a result of imperfect flow cytometry sorting or due to unspecific VSMC-lineage label expression. Microscopy analysis of 72 aortic sections carried out by A.L. Taylor showed a total of seven VSMC lineage-positive cells in the adventitial layer and one VSMC lineage-positive cell in the endothelial layer. This shows that the VSMC lineage label generally shows high specificity, however there are rare cases of lineage-labelled cells located in the neighbouring adventitial and endothelial layers.

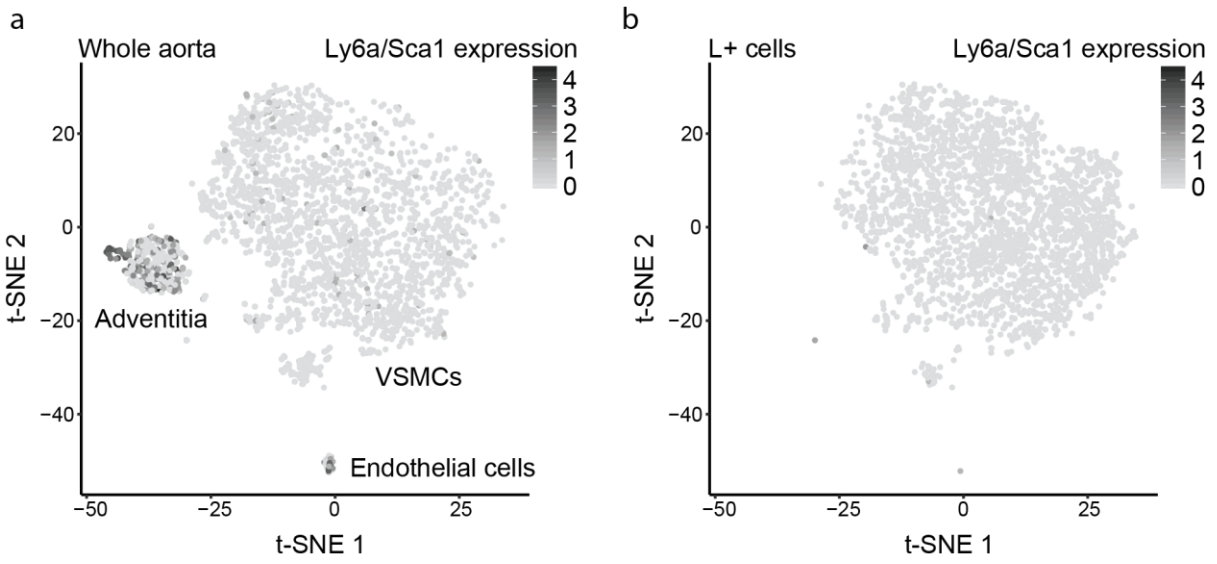


Figure 4.5: *Ly6a/Sca1* transcript is expressed in a subset of VSMCs.

Expression levels of *Ly6a/Sca1* in the whole aorta (a) and the VSMC lineage-labelled (L+) cells only (b) datasets generated using the 10X Genomics Chromium platform are colour coded on t-SNE plots, with darker gray representing higher expression levels. VSMC, adventitial and endothelial clusters were defined in Chapter 3, Figure 3.19.

Overall we have detected a rare subpopulation of VSMCs, which expressed *Ly6a/Sca1*. However, due to the rarity of these cells, the numbers of *Ly6a/*Sca1-expressing L+ cells captured with the unbiased 10X Genomics Chromium platform were too low to robustly investigate their characteristics and heterogeneity. Additionally, the low number of transcripts detected per cell in these experiments may pose challenges in detecting differences in expression, particularly for lowly expressed genes (Figure 3.17).

4.2.2 Targeted profiling of the medial SCA1-positive subpopulation

To obtain sufficient numbers of cells for in-depth analysis of the characteristics of SCA1-positive VSMCs, we decided to use a targeted approach. We selected cells based on the expression of the VSMC lineage label and/or SCA1 using flow cytometry and used the Smart-seq2 protocol (Picelli et al. 2014) to profile their transcriptomes. Typically, aortas from 5-6 mice needed to be pooled to obtain 100 SCA1-positive and VSMC-lineage positive cells for scRNA-seq analysis. Since the number of SCA1-positive VSMCs recovered from primary tissue was the limiting factor in this experiment, we decided to use a lower throughput approach,

which can generate higher coverage transcriptomes. We used the Smart-seq2 protocol, rather than the Fluidigm C1 technology, as it can be combined with index-sorting by flow cytometry. Information on whether each individual cell expressed VSMC lineage label/SCA1 could be used later to guide the data analysis. In addition, this protocol enabled different subpopulations of cells to be combined on a single 96-well plate during index-sorting, which reduced the impact of batch effects on differences between subpopulations.

We isolated the following three subsets of cells from the medial layer of aortas, in which VSMCs were labelled using the multi-coloured VSMC lineage tracing strategy: (1) cells expressing the VSMC lineage label and SCA1 protein (S+L+); (2) cells that expressed SCA1 protein, but were negative for the VSMC lineage label (S+L-); as well as (3) cells that were VSMC lineage-positive and SCA1-negative (S-L+). To reduce the chances of profiling doublets of cells, expression of only one of the four fluorescent labels was required in index-sorted cells. After preparation of cDNA libraries using the Smart-seq2 protocol (Picelli et al. 2014), the samples were sequenced and processed through the Babraham Institute pipeline (Figure 4.6, details in Methods, experimental work including flow cytometry analysis and Smart-seq2 profiling was a collaboration with J. Chappell, A.L. Taylor and H.F. Jørgensen).

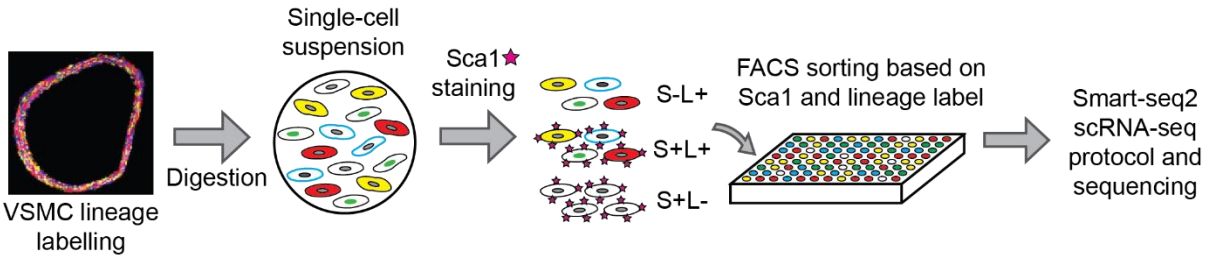


Figure 4.6: Schematic representation of the approach used for profiling medial VSMC-lineage (L) and/or SCA1 (S) positive subpopulations of cells.

Medial cells from mice on multi-colour VSMC lineage tracing background were dissociated to a single-cell suspension. After SCA1 staining, flow cytometry index-sorting was used to isolate individual cells. Their transcriptomes were profiled using the Smart-seq2 protocol (Picelli et al. 2014). Microscopy image was prepared by A.L. Taylor and the schematic is adapted from Dobnikar and Taylor et al. 2018.

4.2.3 Processing and quality control of Smart-seq2 transcriptional profiles

In total, we profiled 187 single cells in three experiments, of which 109 were S+L+, 40 were S-L+ and 38 were S+L-. To exclude low-quality cells, I filtered the single-cell transcriptomes based on the total reads detected (>100,000), total number of genes detected (>1500) and the percentage of reads mapping to ERCC controls (< 30%, Figure 4.7). In this analysis, I decided to retain the cells with high total read count and high number of genes detected. This was because expression of only one of the four fluorescent proteins from the multi-colour lineage labelling system was required during flow cytometry index-sorting, which reduced the chances of profiling doublets. At the same time, the subpopulations of cells profiled in this experiment could show large differences in gene expression profiles and such filters could exclude interesting cells from the analysis. In total, 155 cells passed quality control (92 S+L+, 36 S-L+ and 27 S+L-).

Genes which were expressed below the mean expression level of 1 count per cell were filtered out of further analysis to reduce noise. Transcriptional profiles were then normalised using the pooling normalisation method (Lun et al. 2016, details in Methods) to account for different sequencing depths of individual cells.

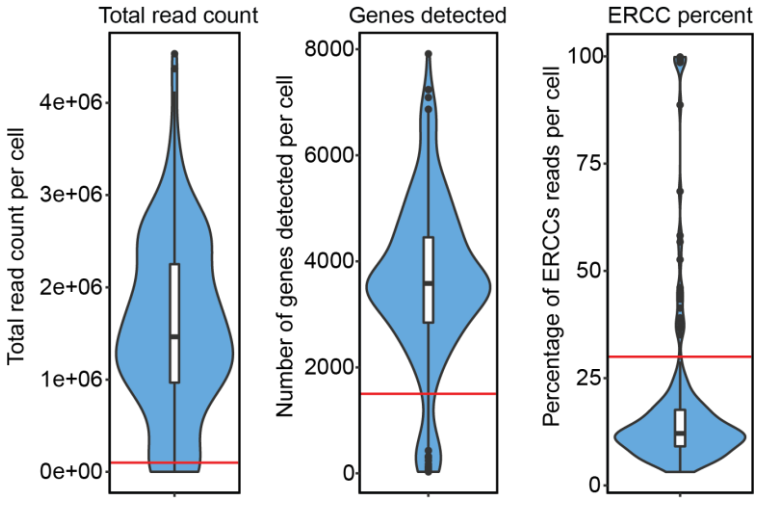


Figure 4.7: Quality control of scRNA-seq profiles generated using the Smart-seq2 protocol.
Violin plots show the distributions of total read count, number of genes detected, and the percentage of reads mapping to ERCC controls. Red lines indicate the quality control filter thresholds. The figure is from Dobnikar and Taylor et al. 2018.

I next assessed the expression of *Myh11* and *Ly6a/Sca1* in the profiled subpopulations of cells. I observed that all S-L+ cells expressed *Myh11*, as did most of the S+L+ cells (Figure 4.8). The majority of S+L- cells expressed *Myh11* at low levels, with a handful of cells showing high *Myh11* expression (Figure 4.8). VSMC lineage labelling efficiency of the multi-colour reporter system has been estimated to be 70-95% in our study and in a previous study from the Jørgensen lab (Chappell et al. 2016), so there is a possibility that the S+L- cells showing high *Myh11* expression were of the VSMC lineage, but did not express the VSMC lineage label. The majority of S+L- cells expressed *Ly6a/Sca1*, as did several S+L+ cells, although overall, the *Ly6a/Sca1* expression levels in this category were lower. A small number of the S-L+ cells showed low levels of *Ly6a/Sca1*, but none of these cells expressed high levels of *Ly6a/Sca1* (Figure 4.8a). Overall, *Ly6a/Sca1* mRNA expression was detected in only 34% of the cells which were identified as Sca1-positive during flow cytometry index-sorting. This may be due to dropout events of *Ly6a/Sca1*, which would be particularly probable in cells where *Ly6a/Sca1* expression is low. Alternatively, although the profiled cells expressed SCA1 protein, the mRNA transcript may not have been present at the time of profiling. *Ly6a/Sca1* showed a higher percentage of zero read counts among the profiled cells than expected according to a linear relationship between the percentage of zeroes and mean expression levels on a log scale (Figure 4.8b). This likely reflects the fact that Sca1 is expressed at differing levels among profiled subpopulations of cells. However, the percentage of zero counts of *Ly6a/Sca1* did not stand out as unusually high and therefore suggests that *Ly6a/Sca1* does not show unexpectedly high dropout levels compared with other profiled genes (Figure 4.8b).

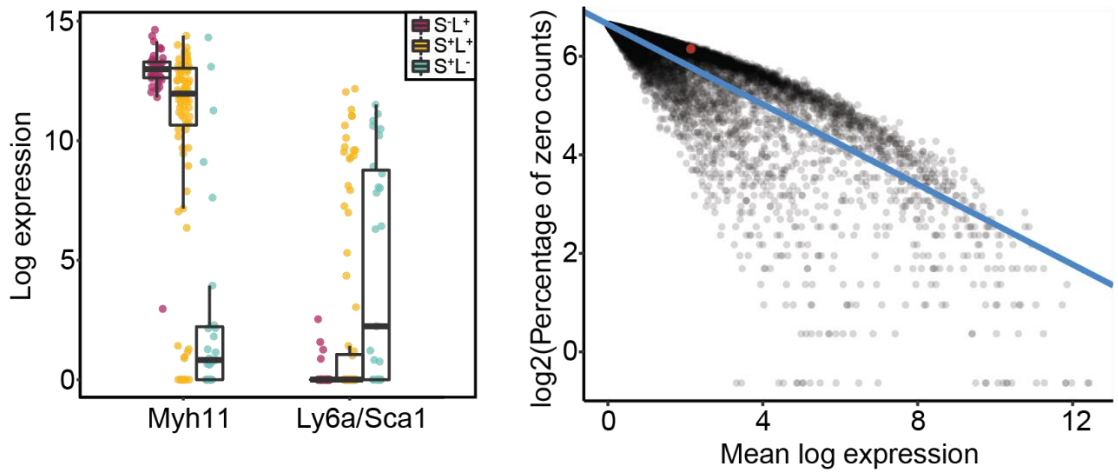


Figure 4.8: *Myh11* and *Ly6a/Sca1* expression levels in profiled cells.

a) Boxplots show the log2-transformed normalised read counts of *Myh11* and *Ly6a/Sca1* transcripts across profiled S+L+ (orange), S+L- (cyan) and S-L+ (magenta) cells. Dots represent individual cells and the median expression level is indicated by the thick black line. The first and third quartiles are represented by the box boundaries and the whiskers show 1.5 of the interquartile range. b) Scatterplot shows the relationship between the $\log_2$ percentage of zero read counts and the mean $\log_2$ expression levels of profiled genes. Individual genes are represented by black dots and *Ly6a/Sca1* is highlighted in red. Blue line shows the results of linear regression (R^2 = 0.78). The analysis is based on the feature selection approach developed in the M3Drop and scmap tools (Tallulah S Andrews & Hemberg 2019; Kiselev et al. 2018). Panel a is from Dobnikar and Taylor et al. 2018.

4.2.4 Heterogeneity of the medial SCA1-positive subpopulation of cells

To further explore the transcriptional differences between the profiled subpopulations of medial cells, I performed principal component analysis (PCA, Figure 4.9a). PCA based on 500 most variable genes resulted in a main cluster of cells expressing high levels of VSMC marker genes, including *Myh11* (Figure 4.9b). The main cluster contained most of the S-L+ cells, a proportion of S+L+ cells, as well as a small number of S+L- cells. S+L- cells located within the main cluster expressed *Myh11* and the lack of VSMC lineage label expression may be due to incomplete recombination of the multi-colour reporter. The majority of the remaining S+L- cells clustered at the opposite end of the principal component analysis plot, showing low *Myh11* expression and higher *Ly6a/Sca1* expression levels (Figure 4.9b and c). The S+L+ subpopulation showed a large degree of heterogeneity, with cells located both in the main cluster of cells as well as spreading out towards the S+L- cluster (Figure 4.9a).

As the cells presented in this Chapter were profiled in three independent experiments, I next examined whether there were any differences between the experiments that may be driving the observed heterogeneity. All three subpopulations of cells were profiled in the first experiment, S+L+ and S-L+ were included in the second, and S+L+ and S+L- in the third experiment. Figure 4.9d shows the distribution of profiled subpopulations of cells along the first two principal components.

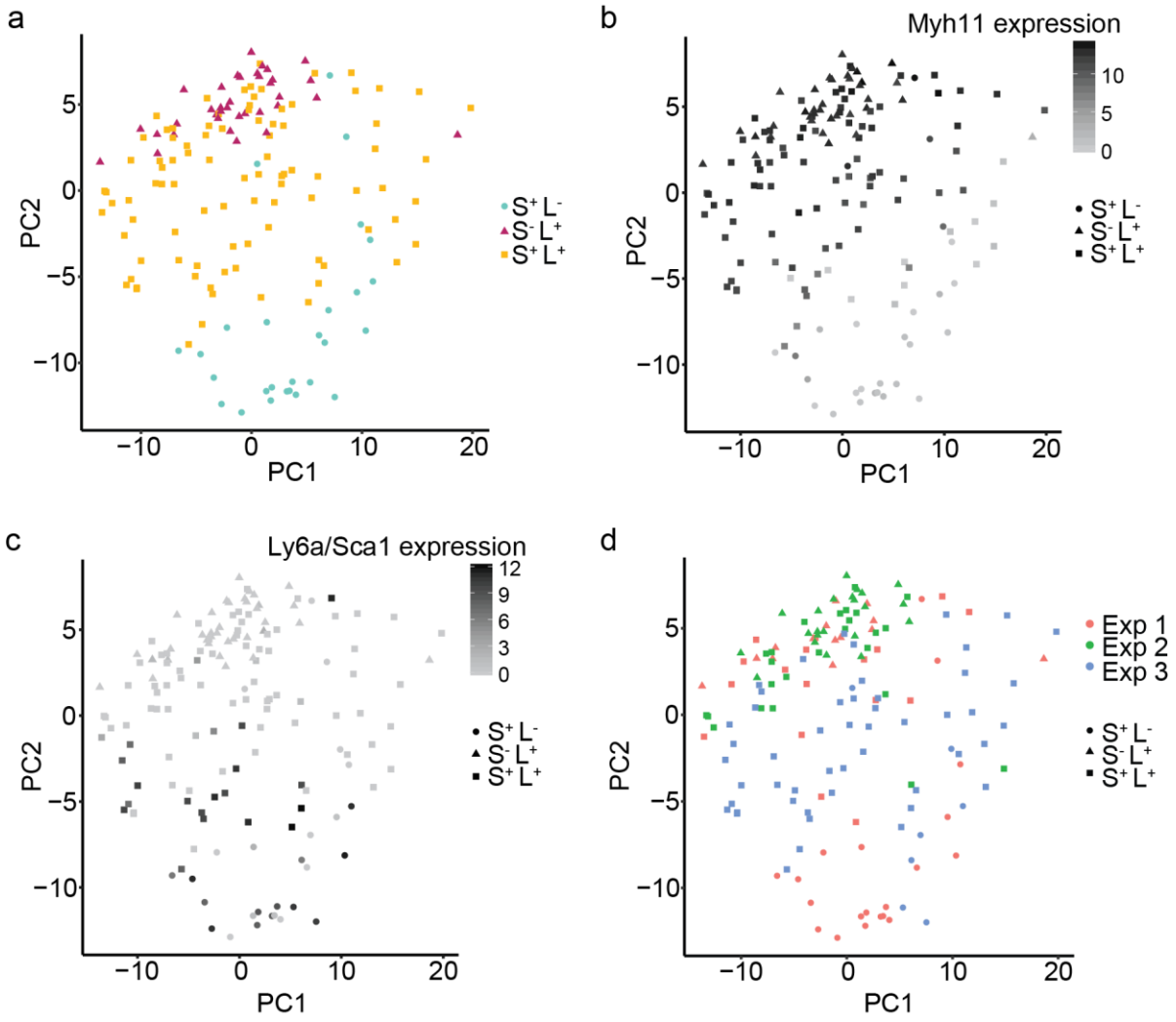


Figure 4.9: Principal component analysis of profiled subpopulations of cells.

Scatterplots show the PC1 and PC2 values for individual cells, based on the 500 most variable genes. PC1 explained 10% of the variance and PC2 explained 6%. a) Cells are colour-coded based on the presence of the VSMC lineage label (L) and/or SCA1 (S), as identified by flow cytometry. S+L- cells are represented by cyan circles, S-L+ cells by red triangles and S+L+ cells by yellow squares. b, c) Log2-transformed normalised expression levels of *Myh11* (b) and *Ly6a/Sca1* (c) are colour coded on the PCA

plot, with darker grey representing higher expression levels. d) Cells profiled in experiment 1 (red), experiment 2 (green) and experiment 3 (blue) are indicated on the PCA plot. Experiment 1 contained cells from all three subpopulations, experiment 2 contained cells from the S-L+ and S+L+ subpopulations, and experiment 3 contained cells from the S+L- and S+L+ subpopulations. Panel a is adapted from Dobnikar and Taylor et al. 2018.

Overall, PCA highlighted the heterogeneity of the S+L+ subpopulation of medial cells. I next examined which genes were expressed heterogeneously among the S+L+ subpopulation. I used an approach which examines the relationship between the variance and mean log expression of genes, and decomposes the variance into the technical and biological components (Lun et al. 2016, similar to Chapter 3, section 3.2.4). To estimate the technical component of variation I used the ERCC controls, which were added to profiled single cells. I first fitted a parametric trend to the variance versus mean expression levels of normalised log-transformed counts of the ERCC controls. I then subtracted the estimated technical variation form the total variance of each gene, and identified a gene as highly variable if the resulting estimated biological variance was significantly greater than zero (adjusted p-value < 0.05, details in Methods, Lun et al. 2016). This approach identified 424 genes as highly variable among the S+L+ cells (Figure 4.10a).

Figure 4.10a shows that while most of the ERCCs agreed with the background technical variance estimation, two of the ERCC controls showed higher variability than expected at their mean expression level, and overlapped with a portion of the genes identified as highly variable. This suggested that the identification criteria for highly variable genes may not have been stringent enough, as no biological variability is expected for ERCC controls. To address this problem, I modified the approach for the identification of highly variable genes. I randomly selected 90% of the S+L+ cells at a time and preformed the highly variable gene expression analysis as described above. I repeated this approach 1000 times and collected a distribution of p-values for individual genes as determined in each iteration. I then used a method implemented by M. Spivakov (details in Methods) to combine the collected p-values. Performing iterative analysis on a subset of cells increased the robustness of the approach, as a gene was required to be consistently highly variable across the majority of the iterations. This excluded the genes showing large variation in a small number of cells, which was more likely to be observed due to technical factors. The limitation of this approach was that the

heterogeneity due to rare subpopulations of cells may have been missed. This approach resulted in a more stringent set of 52 highly variable genes (Figure 4.10b). Figure 4.10c shows the expression profiles of the top 6 highly variable genes.

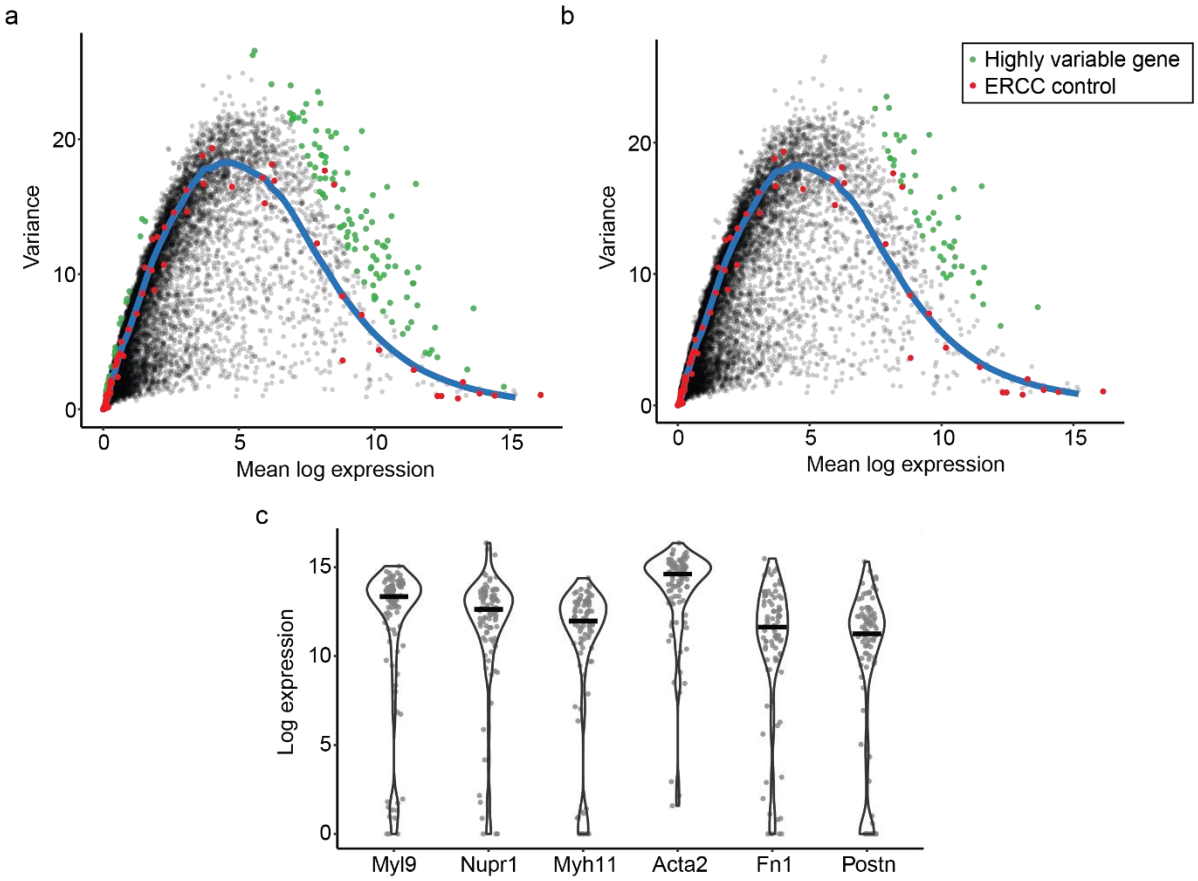


Figure 4.10: Highly variable genes among the S+L+ cells.

a, b) Scatterplots show the mean log expression of genes versus their variance. The blue line represents the estimated technical component of variance at a given mean log expression level. ERCC controls are highlighted in red and the identified highly variable genes are shown in green. a) Highly variable genes were identified as such if their biological component of variance was significantly greater than zero (adjusted p-value < 0.05, details in Methods). b) Iterative validation of highly variable genes was performed 1000 times with 90% of the cells randomly selected at each iteration. p-values for a given gene from individual iterations were then combined using a method devised by M. Spivakov (Dobnikar and Taylor et al. 2018). Genes were then identified as highly variable if their adjusted p-value was below 0.05 (details in Methods). c) Violin plots showing the log2-transformed normalised expression levels of the most highly variable genes. Dots represent individual cells and the black bars indicate the median expression levels. Panels b and c are adapted from Dobnikar and Taylor et al. 2018.

I next investigated whether any of the identified highly variable genes show coordinated variability across the S+L+ cells. Co-expression analysis based on the Spearman's rank correlation coefficient (Lun et al. 2016) revealed high levels of correlation between identified highly variable genes. I identified 281 pairs of positively and 25 pairs of negatively correlated genes, with all of the highly variable genes identified as significantly correlated with at least one other highly variable gene. *Myh11*, *Acta2* and *Tagln* showed the highest number of correlations with other highly variable genes (23, 25 and 25 respectively), with the majority of the genes correlated with all three of these VSMC contractile markers. This indicated that there may be a contractile VSMC signature which is expressed heterogeneously among the profiled cells.

To assess the overall expression levels of the contractile signature, I performed PCA based on the 29 genes correlated with either *Myh11*, *Acta2* or *Tagln*. I used the PC1 scores of each cell, oriented such that higher PC1 score generally aligned with higher expression of the genes, as a measure of the summarised expression levels of this signature in individual cells. Figure 4.11a shows that overall the PC1 score of individual cells correlated strongly with the total sum of normalised read counts of the genes identified as co-expressed with *Myh11*, *Acta2* and *Tagln* (R^2 = 0.96). As expected, PC1 scores were higher in the main cluster of cells, which contained the contractile S-L+ subpopulation (Figure 4.11b).

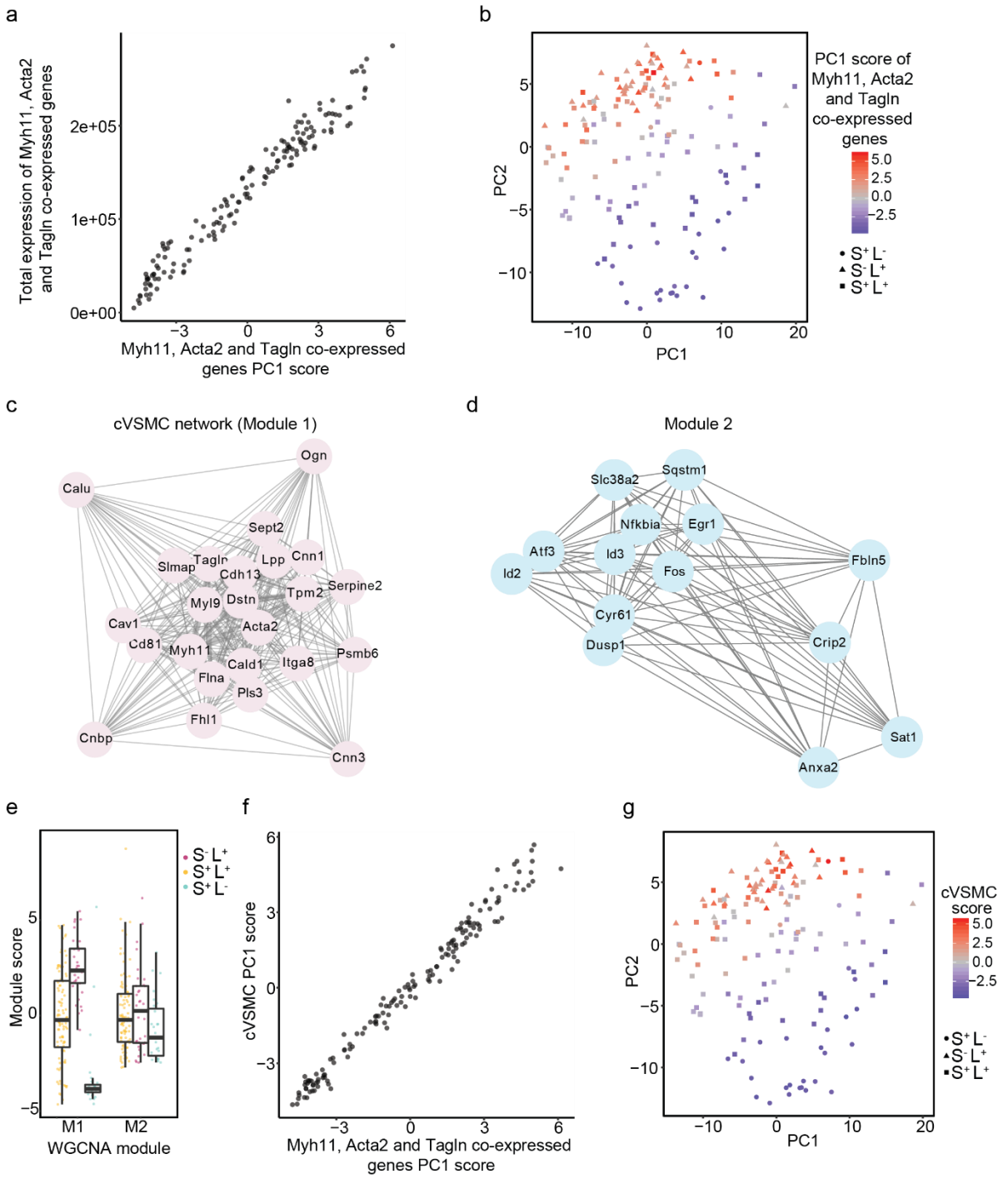


Figure 4.11: Highly variable genes identified in S+L+ cells show high levels of co-expression and contain a contractile VSMC signature.

a) Scatterplot showing the relationship between the total expression levels of genes identified as co-expressed with *Myh11*, *Acta2* and *Tagln* and the PC1 scores from principal component analysis based on these genes. Black points represent individual cells. b, g) Principal component analysis of profiled subpopulations of cells, based on 500 most variable genes. Scatterplots show the PC1 and PC2 values for individual cells. Cells are colour coded based on their PC1 score summarising the expression levels of *Myh11*, *Acta2* and *Tagln* co-expressed genes (b) or their cVSMC score (g), with blue indicating low

and red indicating high overall expression levels. c, d) Network graphs representing the identified networks of co-expressed genes, which were identified using WGCNA among the 52 highly variable genes detected in S+L+ cells. Correlation strength of gene pairs is encoded by the edge thickness. e) Boxplot showing the PC1 scores of WGCNA modules 1 and 2 for the profiled S+L+ (yellow), S+L- (cyan) and S-L+ (red) subpopulations of cells. The median is represented by the thick black line and the first and third quartiles by the box boundaries. Whiskers show 1.5 of the interquartile range. f) Scatterplot showing the relationship between the PC1 scores summarising the overall expression levels of *Myh11*, *Acta2* and *Tagln* co-expressed genes and the cVSMC network, with back points representing individual cells. Panels c, d, e and g are adapted from Dobnikar and Taylor et al. 2018 and WGCNA network analysis was a collaboration with M. Spivakov.

High levels of correlation between highly variable genes suggested that substructure may be present among the profiled S+L+ cells. To obtain a more comprehensive picture of the highly variable gene co-variability we used network analysis (implemented by M. Spivakov). Network analysis performed using the WGCNA approach (Langfelder & Horvath 2008) revealed two modules of genes (Figure 4.11c and d, details in Methods). Module 1 contained several contractile marker genes, such as *Myh11*, *Acta2*, *Tagln* and *Cnn1*, and showed strong overlap with the genes previously found to be correlated with VSMC markers *Myh11*, *Acta2* and *Tagln*. In total, 20 out of 24 of module 1 genes overlapped with VSMC marker correlated genes. We therefore termed this module the contractile VSMC (cVSMC) network.

I next assessed the overall expression levels of the identified gene modules across the profiled cells. I performed PCA based on the genes contained within each of the modules and used the PC1 scores positively aligned with the total expression levels to summarise the overall expression in each cell. Figure 4.11e shows that PC1 scores of module 2 did not vary considerably across the profiled subpopulations of cells. In contrast, the cVSMC network was expressed highly in the S-L+ population, at low levels in S+L- cells and heterogeneously in S+L+ cells (Figure 4.11e). Overall, there was a strong correlation between the PC1 scores across cells between the genes correlated with *Myh11*, *Acta2* and *Tagln* and the cVSMC network (Figure 4.11f, $R^2 = 0.98$).

4.2.5 Transcriptional signatures of SCA1-positive VSMCs expressing low levels of contractile markers

We were next interested in the transcriptional differences between the S+L+ cells expressing low or high levels of the contractile VSMC signature. Figure 4.11g shows that the S+L+ cells, which clustered away from the main contractile cluster scored particularly low for the cVSMC network expression. To delineate their transcriptional differences, we used negative binomial regression to find the genes which correlated positively and negatively with the cVSMC scores of individual S+L+ cells. This approach identified 312 positively and 303 negatively correlated genes (Figure 4.12a, adjusted p-value < 0.05, collaboration with M. Spivakov, details in Methods).

The genes which correlated positively with the cVSMC score (cVSMCpos) showed high expression levels in the S-L+ population, low levels in the S+L- cells and variable levels in S+L+ cells (Figure 4.12b). Gene ontology analysis identified “muscle system process”, “muscle contraction”, “regulation of muscle system process” as the top three overrepresented gene ontology terms, which is in agreement with cVSMCpos representing a differentiated contractile VSMC signature. In contrast, the top three overrepresented gene ontology terms among the genes correlated negatively with the cVSMC network (cVSMCneg) were “positive regulation of cell migration”, “angiogenesis” and “wound healing”, which are processes associated with the synthetic state of VSMCs. This signature was expressed at particularly high levels in a subset of S+L+ cells (Figure 4.12b). This suggested that there may be a subset of VSMC-lineage cells in healthy arteries which show a less differentiated and a more activated transcriptional signature characteristic of synthetic VSMCs. In agreement with this, the cVSMCneg signature contained markers of the VSMC synthetic phenotype, such as *Col8a1, Vcam1* and *Spp1* (Allahverdian et al. 2018). The expression of these genes was sparse but overall higher in the S+L+ cells, which expressed lower levels of the cVSMC signature (Figure 4.12c).

Taken together, the results presented in this chapter suggest that there is a rare subpopulation of VSMCs resident in healthy arteries which has a less differentiated transcriptional profile and shows some characteristics of a synthetic VSMC phenotype.

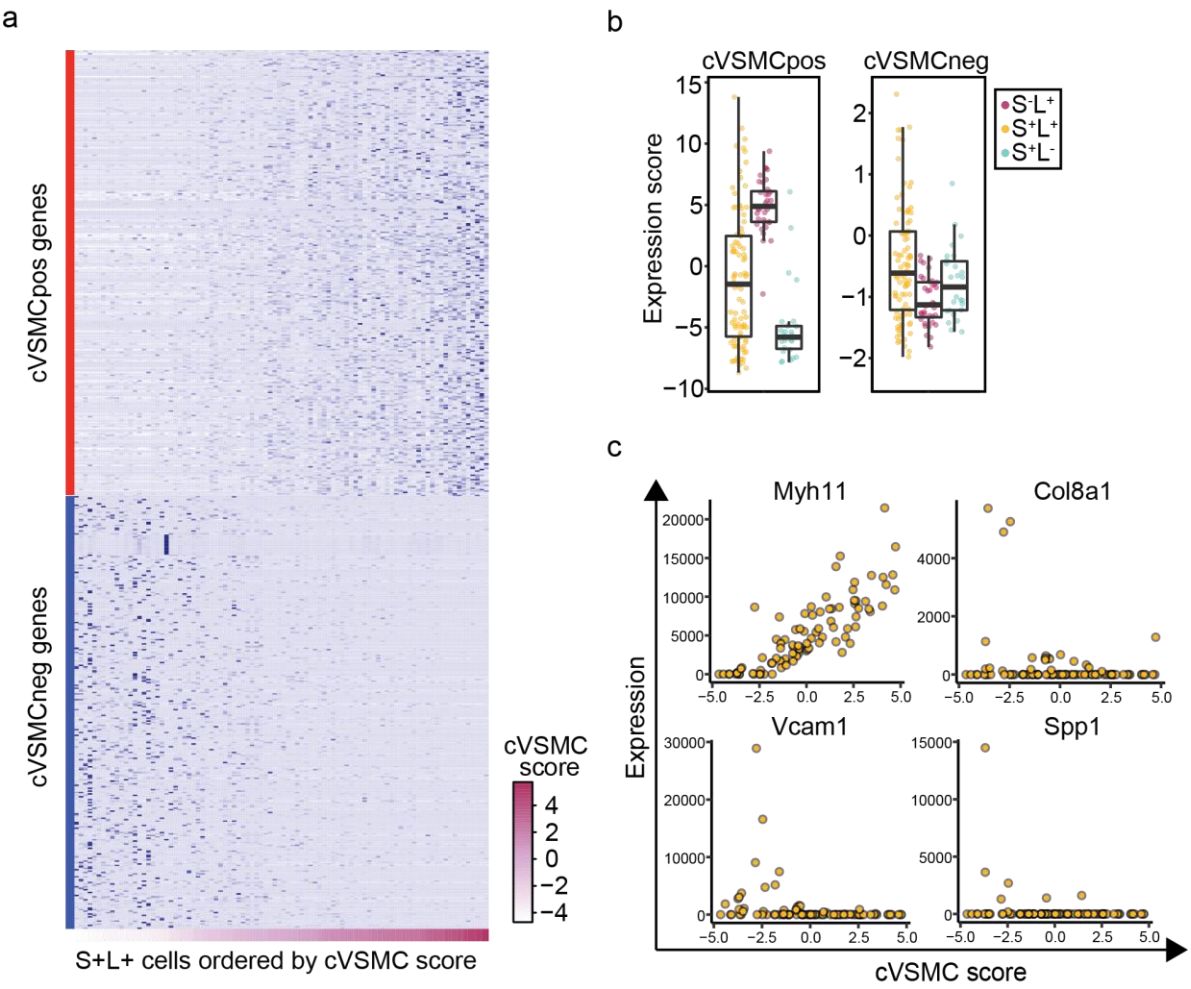


Figure 4.12: A subset of S+L+ cells shows characteristics of a synthetic VSMC phenotype.

a) Heatmap showing the expression levels of cVSMCpos genes (red block) and cVSMCneg genes (blue block) in S+L+ cells, with darker blue representing higher expression levels. Cells are ordered according to their cVSMC score, which is colour coded from white (low) to purple (high). Identification of genes correlated positively and negatively with the cVSMC scores in S+L+ cells was a collaboration with M. Spivakov. b) Boxplots show the PC1 scores summarising the expression levels of cVSMCpos genes (left) and cVSMCneg genes (right) in profiled S+L+ (yellow), S-L+ (red) and S+L- (cyan) subpopulations of cells. The median score is represented by a thick black line, the first and third quartiles by box edges and the 1.5 interquartile range by the whiskers. c) Scatterplots showing the relationship between the cVSMC score and the normalised expression levels of *Myh11*, *Col8a1*, *Vcam1* and *Spp1* in S+L+ cells. Yellow dots represent individual S+L+ cells. Figure adapted from Dobnikar and Taylor et al. 2018.

4.3 Discussion

In this chapter I explored whether VSMC heterogeneity in healthy artery walls may underpin the observed functional heterogeneity among VSMCs in atherosclerosis and vascular injury. I detected a rare subpopulation of medial cells which expressed the progenitor marker SCA1. Lineage tracing has enabled us to confirm the VSMC-lineage identity of a subset of medial SCA1+ cells and we have combined it with scRNA-seq to profile gene expression in individual medial cells from healthy mouse arteries. I found that the SCA1-positive VSMC subpopulation is heterogeneous, with some of these cells resembling contractile VSMCs and others showing a less differentiated transcriptional profile, devoid of the expression of classic VSMC markers and additionally showing some characteristics of the synthetic VSMC state. The existence of VSMCs expressing transcriptional signatures characteristic of the synthetic VSMC state in healthy arteries is in agreement with a previous hypothesis that VSMCs exist on a spectrum between the contractile and synthetic phenotypes (Rensen et al. 2007).

4.3.1 SCA1-positive cells within the adventitial and endothelial layers

SCA1-positive cells have been previously observed to reside in the endothelial and adventitial vascular layers (Psaltis & Simari 2015) and consistent with this, we have observed high levels of *Ly6a/Sca1* expression among the adventitial and endothelial clusters of cells profiled using the 10X Genomics Chromium platform. SCA1-positive subpopulations from both the endothelial and adventitial layers were previously observed to have progenitor potential (Psaltis & Simari 2015). For example, endothelial cells positive for SCA1 among other markers have been shown to have higher potential for clonal expansion than conventional endothelial cells, and were capable of forming functional blood vessels in vivo (Fang et al. 2012; Naito et al. 2012).

Adventitial SCA1-positive cells have been shown to be capable of differentiating towards vascular smooth muscle cells *in vitro* and to contribute to neointima formation *in vivo* (Hu et al. 2004; Passman et al. 2008; Kramann et al. 2016). Transcriptional profiling of adventitial SCA1-positive cells (marked by Gli expression) using single-cell qPCR showed that these cells shared some of the transcriptional signatures with S+L+ cells and some of these adventitial cells expressed the VSMC marker gene *Tagln* (Kramann et al. 2016). There were, however, some differences in the transcriptional profiles the adventitial cells and S+L+ cells profiled in

our study. For example, many of the profiled adventitial cells expressed *Cd29, Vegfr2* and *Sox2* (Kramann et al. 2016), which were not expressed by S+L+ cells. *Myh11* was not profiled by Kramann et al. (2016), therefore it cannot be concluded whether or not these adventitial cells would be likely to express the VSMC lineage label and thus potentially overlap with the S+L+ population. However, S+L+ cells presented in this chapter appear to be generally distinct from adventitial SCA1-positive cells, as highlighted by the flow cytometry and ACTA2 staining experiments using Sca1-GFP transgenic mice. The GFP-positive cells we isolated from the medial layer had high levels of ACTA2 expression by immunostaining, in contrast to GFP-positive cells isolated from the adventitia, which stained negative for ACTA2.

Plasticity between adventitial cells and VSMCs has also been suggested in the reverse direction, with mature VSMCs described as capable of giving rise to a proportion of SCA1-positive adventitial cells. Majesky et al. (2017) reported that mature VSMCs migrate to the adventitial side of the medial-adventitial layer interface *in vivo*. These VSMC-derived adventitial SCA1-positive cells were then observed to expand following vascular injury (Majesky et al. 2017). Using lineage tracing approaches, the authors observed the transition of VSMCs towards adventitial SCA1-positive cells within healthy arteries and suggested this process helps to maintain the vascular progenitor niche within the vasculature (Majesky et al. 2017). The extent to which these VSMC-derived adventitial SCA1-positive cells resemble the activated S+L+ VSMCs isolated from the medial layer in this project remains to be explored. Although we have manually removed the adventitial and endothelial layers during dissection, contamination of cells from the neighbouring layers, particularly from cells located at the borders, cannot be ruled out. Therefore, given that the VSMC-derived adventitial SCA1-positive cells were observed to localise on the adventitial side of the adventitial-medial border and they would be expected to retain the VSMC lineage label expression, these cells may have been included in our study. However, examination of VSMC lineage label expression in the adventitial and endothelial layers carried out by A.L. Taylor, showed only very rare expression of the VSMC lineage label in the adventitia.

4.3.2 SCA1-positive VSMCs within the medial layer

A previous study has identified and isolated a subpopulation of cells within the medial layer of the aorta, which had progenitor potential and expressed SCA1 (Sainz et al. 2006). Using flow

cytometry the authors estimated that the SCA1+ subpopulation of cells represented 6% of all medial cells. They observed that after culturing these cells in the presence of TGF-β1 and PDGF-BB, these cells acquired a VSMC phenotype whereas an endothelial phenotype was acquired after culture with VEGF (Sainz et al. 2006). This SCA1-positive medial subpopulation was identified on the basis of reduced Hoechst staining, which is characterised by the presence of the ABCG2 transporter. A small number of S+L+ cells profiled in our study (6/92) expressed *Abcg2*, suggesting that there may be some overlap between the previously identified medial SCA1+ cells and the S+L+ cells profiled in this project. At the transcriptional level we observed expression of the endothelial cell marker *Vcam1* in several S+L+ cells and it would be interesting to attempt similar *in vitro* culture experiments with VEGF to assess whether S+L+ cells have a potential to differentiate towards endothelial cells. The rarity of S+L+ cells, however, presents technical challenges in conducting such experiments, as it would be difficult to isolate sufficient numbers of S+L+ cells for successful *in vitro* culture.

4.3.3 SCA1-positive cells of VSMC lineage within atherosclerotic plaques

Interestingly, SCA1-positive cells of VSMC lineage were observed previously within atherosclerotic lesions, where they were suggested to mark a mesenchymal stem cell-like population (Shankman et al. 2015). These SCA1-positive cells downregulated the expression of VSMC markers and VSMC-specific lineage tracing has been used to confirm that these cells originated from VSMCs (Shankman et al. 2015). These findings suggested that SCA1-positive VSMCs may have relevance in disease and prompted us to further investigate S+L+ cells in atherosclerosis (see next chapter).

A recent study of VSMCs in healthy and atherosclerotic vessels has confirmed our observation of a small proportion of VSMCs in healthy arteries expressing transcriptional signatures characteristic of phenotypically switched VSMCs (Wirka et al. 2019). Wirka and colleagues profiled the transcriptomes of lineage-traced VSMCs in healthy aortic root as well as in atherosclerotic plaques and observed that 1.3% of healthy VSMCs clustered with phenotypically modulated VSMCs from the atherosclerotic plaques, some of which expressed *Ly6a/Sca1* (Wirka et al. 2019).

4.3.4 Conclusions

Overall in this chapter I have identified and characterised a rare subpopulation of VSMCs, which expressed *Ly6a/Sca1*. The scRNA-seq approach has been invaluable in the analysis of medial cell heterogeneity, as it enabled the identification of a rare subpopulations of cells, which could not have been detected with bulk RNA-seq methods. Combining scRNA-seq with VSMC-specific lineage labelling has enabled us to confirm that a proportion of SCA1-positive medial cells were of the VSMC lineage and to ensure that profiled cells were not a contamination from the neighbouring adventitial and endothelial layers. scRNA-seq has further enabled the characterisation of the transcriptional signatures of SCA1-positive VSMCs, which indicated that VSMCs showing a less differentiated transcriptional profile and expressing transcriptional signatures characteristic of the synthetic VSMC state are present in healthy arteries. These observations motivated us to further investigate the expression of SCA1 in model systems of VSMC phenotypic switching, which will be discussed in the next chapter.

5 SCA1-positive VSMCs in model systems of VSMC phenotypic switching

5.1 Introduction

VSMCs play an important role in the development of atherosclerotic plaques and in neointima formation following vascular injury (Bennett et al. 2016). Previous studies observed that 30-70% of cells within atherosclerotic plaques originate from VSMCs (Shankman et al. 2015; Chappell et al. 2016). Despite extensive accumulation of VSMCs within lesions, multicolour lineage tracing studies suggested that only a very small proportion of VSMCs proliferate and expand during atherosclerotic plaque and neointima formation (Chappell et al. 2016; Jacobsen et al. 2017). A possible explanation for this observation is that VSMCs are functionally heterogeneous.

In Chapter 4, I presented evidence for a rare subset of SCA1-positive VSMC, which expressed transcriptional signatures characteristic of the activated synthetic state of VSMCs. Here, I hypothesised that these SCA1-positive VSMCs may be more primed for expansion in atherosclerosis and injury. SCA1-positive cells of the VSMC lineage have been previously observed within atherosclerotic plaques (Shankman et al. 2015) and the aim of this chapter is to investigate whether the transcriptional profiles of *Ly6a/Sca1*-expressing VSMCs within plaques share similarities to healthy S+L+ cells, which expressed transcriptional signatures characteristic of synthetic VSMCs. Additionally, I investigated how the transcriptional profiles of SCA1-positive VSMCs detected in healthy arteries compare with phenotypically modulated VSMCs induced *in vivo* and *in vitro*.

To investigate the transcriptional signatures of VSMCs during phenotypic switching I used three separate mouse models, in which VSMC phenotypic switching has been extensively documented. *In vitro* culture is a simple but frequently used model system of VSMC phenotypic switching, in which some of the hallmarks of the process, such as downregulation of VSMC marker protein expression and proliferation, are replicated (Chamley-Campbell et al. 1979; Rensen et al. 2007). The ApoE$^{-/-}$ mouse atherosclerosis model system provided a disease-relevant setting for the investigation of VSMC phenotypic switching (Getz & Reardon

2012). Additionally, the carotid ligation injury model (Kumar & Lindner 1997) enabled an *in vivo* investigation of VSMC phenotypic switching in a more acute setting compared with the mouse atherosclerosis model. Together, these model systems provided a range of environments for the study of VSMC phenotypic switching and additionally enabled the use of VSMC-specific genetic lineage tracing, which would not be possible in human atherosclerotic plaques. The analysis presented in this chapter revealed that SCA1 and *Ly6a/Sca1* upregulation marks VSMC phenotypic switching and that the transcriptional signatures of SCA1-positive VSMCs in healthy arteries are similar to those observed in model systems of VSMC phenotypic switching.

5.2 Results

5.2.1 SCA1 is upregulated in VSMCs cultured *in vitro*

To investigate the gene expression changes that occur in VSMCs during phenotypic switching in culture, I used a publicly available single-cell transcriptomics dataset of cultured VSMCs, which was generated on the Fluidigm C1 platform (GSE79436, Adhikari et al. 2015). I compared the transcriptomes of cultured VSMCs with our profiles of *ex vivo* VSMCs from the AA and DT regions of the aorta, which were generated using the same platform. The *ex vivo* and cultured VSMC transcriptomes were processed in an analogous way and the read counts were normalised jointly to account for differences in sequencing depth (details in Methods). PCA (originally presented in Section 3.2.1) resulted in separate *ex vivo* and cultured VSMC clusters. VSMC contractile markers, such as *Myh11*, were downregulated in cultured VSMC, as expected. Interestingly, I observed increased levels of *Ly6a/Sca1* expression among cultured VSMCs (Figure 5.1).

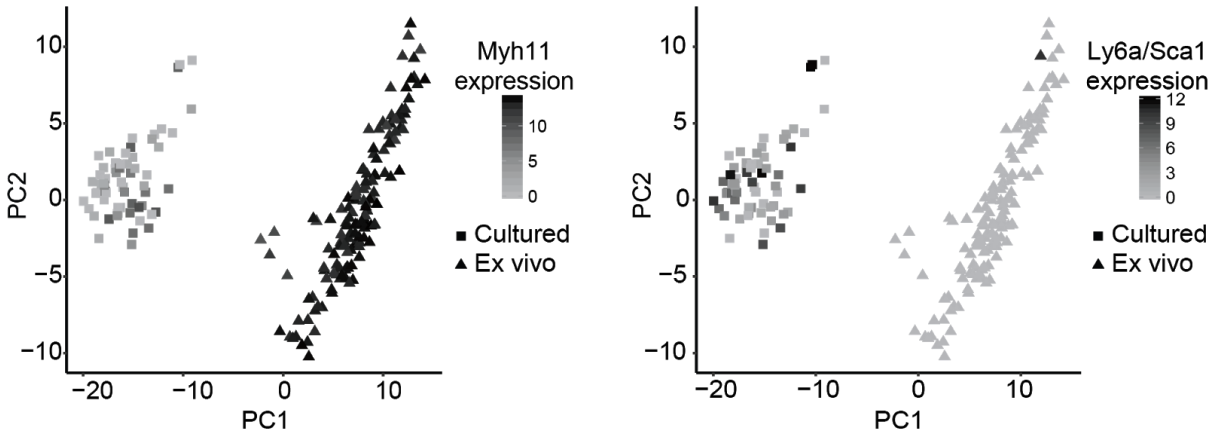


Figure 5.1: *Ly6a/Sca1* expression levels are increased in cultured compared with *ex vivo* VSMCs.
PCA of cultured (squares) and *ex vivo* (triangles) VSMC transcriptomes with expression levels of *Myh11* (left) and *Ly6a/Sca1* (right) colour coded from low (grey) to high (black). PC1 explained 25% of the variance and PC2 explained 4%. Transcriptional profiles of cultured VSMCs are publicly available at the Gene Expression Omnibus (GSE79436, Adhikari et al. 2015). *Ly6a/Sca1* panel was adapted from Dobnikar and Taylor et al. 2018.

Our next question was whether contractile SCA1-negative VSMCs were capable of upregulating SCA1 after stimulus, or whether pre-existing SCA1-positive VSMCs selectively expanded in culture. To address this we cultured VSMCs isolated from Sca1-GFP animals (Ma et al. 2002) and observed that sorted GFP-negative medial cells upregulated GFP after 10 days in culture (Figure 5.2, collaboration with A.L. Taylor). In contrast, we did not observe any GFP-positive cells after culture of VSMCs from wild-type animals, as expected. Overall, these experiments indicated that mature VSMCs have the capacity to upregulate SCA1 during phenotypic switching in culture.

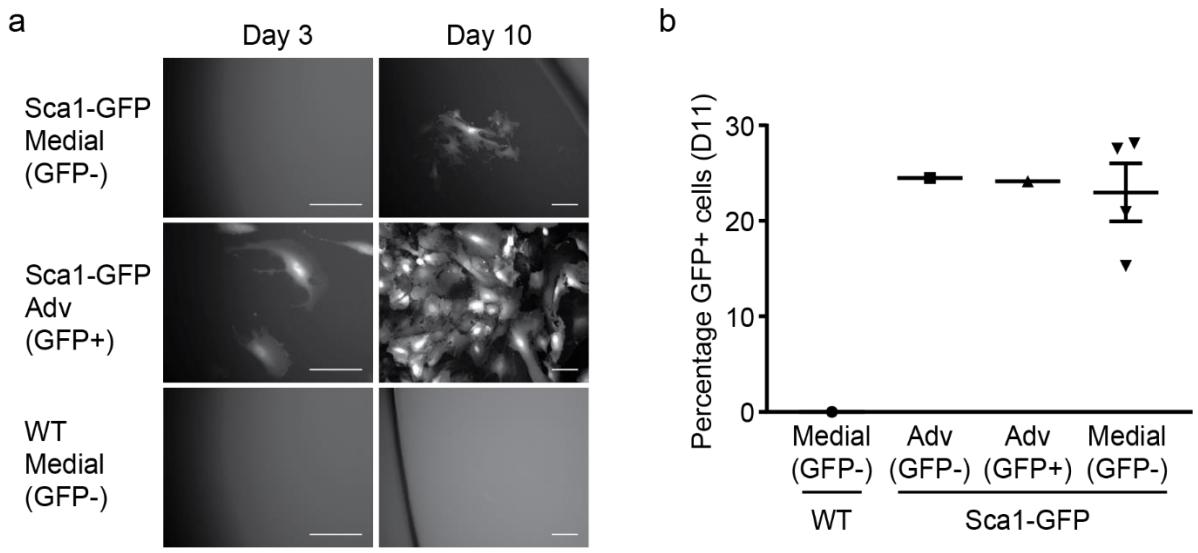


Figure 5.2: SCA1 is upregulated in sorted GFP-negative medial cells during culture.

Medial GFP-negative (n=4), adventitial GFP-negative (pooled tissue from four animals) and adventitial GFP-positive (pooled tissue from four animals) cells from Sca1-GFP reporter transgenic animals were sorted using flow cytometry and cultured *in vitro*. As a negative control, medial GFP-negative cells from a wild type mouse were also isolated and cultured (n=1). a) Images show the GFP expression in isolated cell populations after 3 and 10 days in culture. b) Quantitation of the percentages of GFP-positive cells among isolated populations after 11 days of culture. Collaboration with A.L. Taylor and the figure is from Dobnikar and Taylor et al. 2018.

We next investigated whether VSMCs also upregulated SCA1 *in vivo* using lineage-labelled animals. We analysed the proportion of S+L+ VSMCs using flow cytometry after varying time periods between tamoxifen injection and flow cytometry analysis in single-colour lineage-labelled animals (collaboration with A.L. Taylor and M. Spivakov). Logistic regression analysis suggested that there was a small but significant increase in the proportion of S+L+ VSMCs with longer time periods between the induction of lineage labelling and analysis (Figure 5.3a), and this effect did not appear to originate from the confounding factor of varying animal age (Figure 5.3b and c). This analysis therefore suggests that VSMCs upregulate SCA1 *in vivo*, although the increase in the proportion of S+L+ VSMC over time is small.

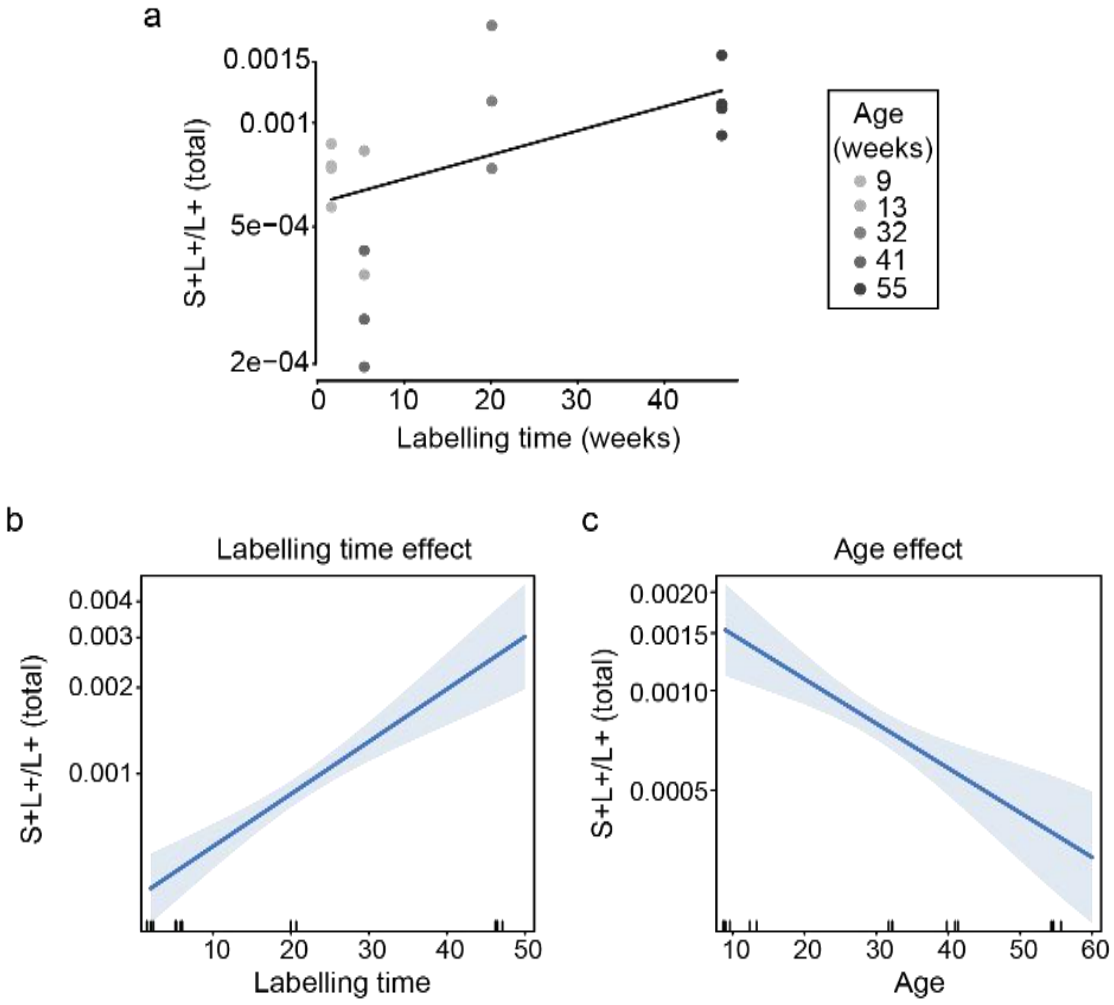


Figure 5.3: The proportion of S+L+ VSMCs increases with increased time period between lineage labelling and analysis.

a) Scatterplot showing the relationship between the proportion of S+L+ VSMCs, and the time period between tamoxifen injection and flow cytometry analysis. The age of individual animals is colour-coded, with darker grey dots representing higher age. The trendline was estimated using logistic regression, with the proportion of S+L+ VSMCs as the response variable and the labelling time period as the explanatory variable (logit-link regression coefficient = 0.016 (mean) ± 0.005 (95% confidence interval), Student's distribution p-value = 2.56e-10). b, c) Bivariate logistic regression analysis using the labelling time period and animal age as explanatory variables and the proportion of S+L+ VSMCs as the response variable. Blue lines represent the trendline and shaded areas represent 95% confidence intervals. b) Effect of labelling time, logit-link regression coefficient = 0.042 ± 0.014, Student's distribution p-value = 2.72e-9. c) Effect of age, logit-link regression coefficient = -0.028 ± 0.015, Student's distribution p-value = 3.07e-5). Partial correlation analysis was used as an alternative approach to determine the influence of animal age on the relationship between labelling time and proportion of S+L+ VSMCs (*pcor* function of R package *ppcor*). While controlling for animal age, there

was a positive and significant correlation between the proportion of S+L+ VSMCs and labelling time (partial correlation = 0.66, p-value = 0.0069). The correlation between animal age and the proportion of S+L+ VSMCs while controlling for labelling time was negative (partial correlation = -0.46). Collaboration with A.L. Taylor (experimental work) and M. Spivakov (statistical analysis). The figure is adapted from Dobnikar and Taylor et al. 2018.

I next investigated whether the overall transcriptional signatures of cultured VSMCs resembled those of the S+L+ cells identified in healthy arteries, which expressed high levels of the cVSMCneg transcriptional signature (described in Chapter 4). Figure 5.4 shows that the overall expression level of the cVSMCneg signature was markedly higher in cultured VSMCs compared with *ex vivo* VSMCs. This observation prompted us to investigate SCA1 expression levels and the transcriptional signatures of VSMCs in models, where VSMC phenotypic switching is observed *in vivo*.

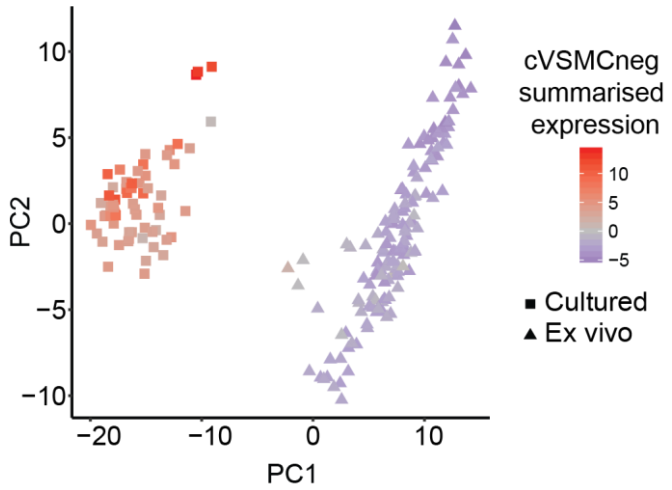


Figure 5.4: Cultured VSMCs express transcriptional signatures of activated S+L+ cells from healthy arteries.

PCA of cultured and *ex vivo* VSMC transcriptomes with the summarised expression levels of the cVSMCneg response signature colour coded from low (blue) to high (red) overall expression levels. To summarise the overall gene expression levels, the PC1 score of PCA based on the genes within the cVSMCneg signatures has been used (details in Methods). Transcriptional profiles of cultured VSMCs are publicly available at the Gene Expression Omnibus (GSE79436, Adhikari et al. 2015)

5.2.2 SCA1 is upregulated in VSMCs after carotid ligation injury

Next we investigated whether SCA1 expression is upregulated in VSMCs after carotid ligation injury, which is an *in vivo* model of VSMC phenotypic switching (Kumar & Lindner 1997). In no injury control carotid arteries, we observed a very low proportion of VSMCs positive for both the VSMC lineage label and SCA1 (Figure 5.5a), as expected based on previous analysis of aortic VSMCs (Chapter 4). After carotid ligation, the proportion of lineage labelled cells which also stained positive for SCA1 expression, increased to over 25% on average (Figure 5.5b and c, collaboration with J.L. Harman).

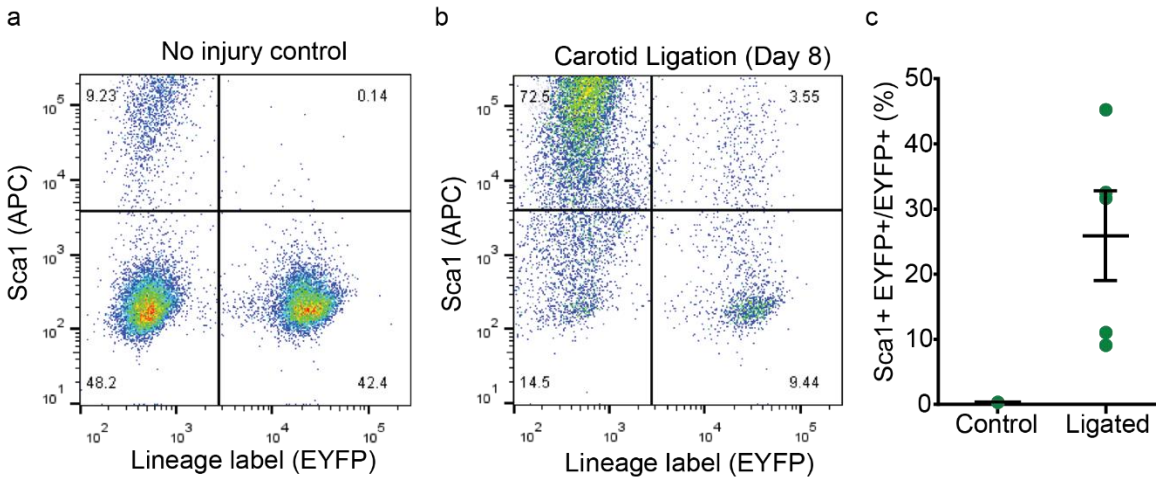


Figure 5.5: SCA1 is upregulated in VSMCs 8 days after carotid ligation injury.

a,b) Flow cytometry scatter plots showing the expression of the VSMC lineage label (eYFP) on the x axis and the expression of SCA1 on the y axis. Whole carotid arteries in no injury control (a) and 8 days after carotid ligation (b) were analysed. c) Percentages of eYFP-positive cells, which stained positive for expression of SCA1 in no injury controls and 8 days after carotid ligation are shown (5 animals per group). Collaboration with J.L. Harman and the figure is from Dobnikar and Taylor et al. 2018.

We were next interested in the overall transcriptional changes in VSMCs after carotid ligation injury. We used the 10X Genomics Chromium platform to profile the transcriptomes of individual VSMC-lineage labelled cells 7 days after carotid ligation (collaboration with A.L. Taylor and J. Chappell). In order to capture as many cells involved in phenotypic switching as possible, we initially used a reporter inserted into the *Mki67* locus to enrich for the cells involved in the response to injury. For this experiment we combined the single-colour eYFP

VSMC lineage tracing system with the Ki67-RFP reporter allele (Basak et al. 2014, details in Methods). Using FACS, we isolated eYFP lineage labelled cells and additionally enriched for RFP-positive cells, which also expressed the eYFP lineage label. cDNA libraries were then prepared from sorted cells using the 10X Genomics Chromium platform. After sequencing, the reads were processed using the 10X Genomics *cellranger* pipeline, which recovered a total of 1335 cells (as detailed in Methods).

Quality control of captured single-cell transcriptomes showed that the majority of profiled cells were of good quality. The remaining cells, which were not profiled to sufficient depth or showed a high proportion of mitochondrial reads, were removed from further analysis (Figure 5.6). To pass quality control, over 5000 UMI counts and over 2000 genes had to be detected in a cell. Additionally, cells were required to contain less than 6% of mitochondrial reads in order to exclude the cells that may have burst during processing. Cells with particularly high total UMI count or high number of detected genes were not removed from this dataset as larger differences in these parameters were expected between VSMCs undergoing phenotypic switching and contractile VSMCs. Removing such cells could therefore lead to loss of interesting cell populations. Additionally, the number of cells captured in this experiment was relatively low for the 10X Genomics Chromium platform, which reduced the chances of profiling doublets (Zheng et al. 2017). Overall 1126 out of 1335 cells passed quality control and the transcriptomes of good quality cells were normalised (details in Methods).

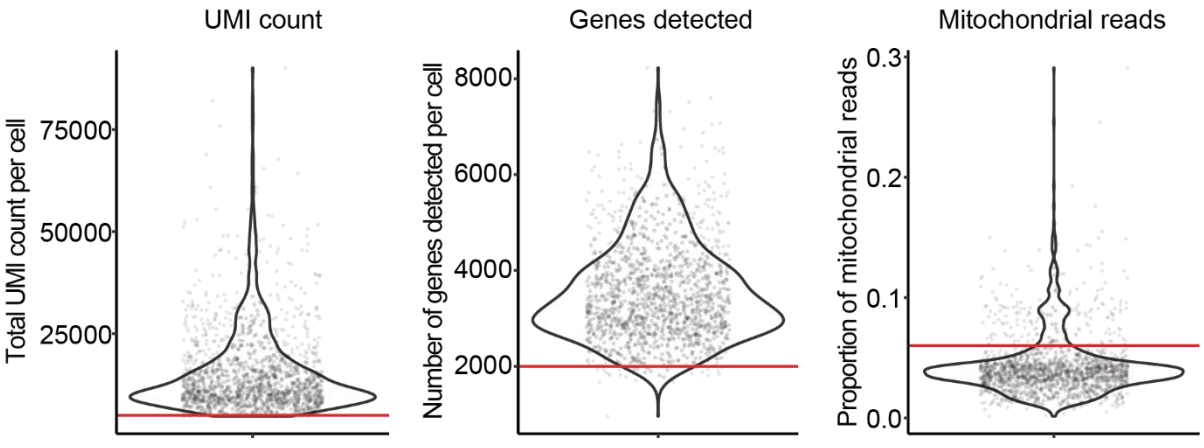


Figure 5.6: Quality control of VSMCs profiled 7 days after carotid ligation injury.

Violin plots show the distribution of total UMI count, the number of genes detected and the proportion of mitochondrial reads across profiled VSMCs. Red lines represent the filtering thresholds and grey dots represent individual cells. Cells with total UMI count over 5000, over 2000 detected genes and less that 6% of mitochondrial reads passed quality control.

To understand the heterogeneity among the profiled cells I next performed principal component analysis. The first principal component separated the cells which expressed *Ly6a/Sca1* as well as cells showing increased *Mki67* expression from the cells expressing the contractile VSMC marker *Myh11* (Figure 5.7). The second principal component further differentiated the cells, which expressed the proliferative marker *Mki67* (Figure 5.7).

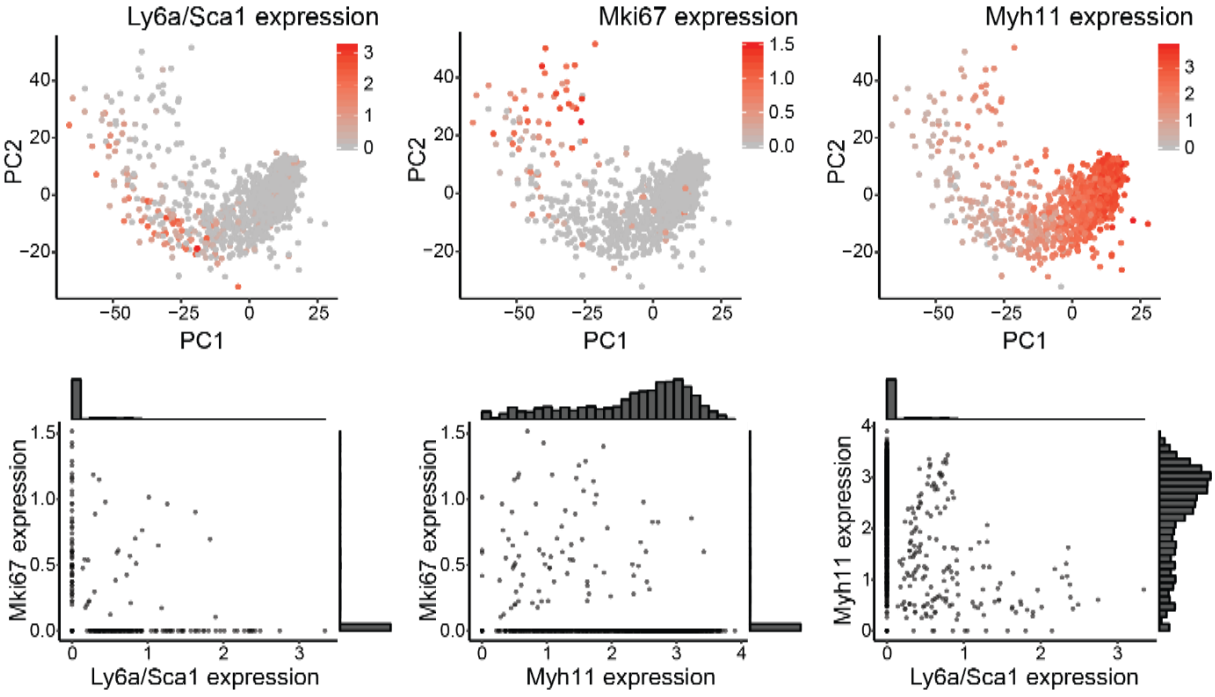


Figure 5.7: Principal components 1 and 2 separate cells showing lower expression levels of the contractile marker *Myh11* and higher expression levels of *Ly6a/Sca1* and *Mki67*.

Top: the log2-transformed normalised expression levels of *Ly6a/Sca1* (left), *Mki67* (middle) and *Myh11* (right) are colour-coded from grey (low) to red (high) on the PCA plot. The first two principal components, which explained the most variance are shown. PC1 explained 25% and PC2 explained 10% of the variance. Bottom: scatterplots showing the pairwise gene expression relationships between *Ly6a/Sca1*, *Myh11* and *Mki67*. Histograms show the distribution of cells along the x and y axes.

To explore the profiled subpopulations of VSMCs and identify the overall differences between the cells I performed clustering analysis. Clustering was based on the first 11 principal components, all of which explained significantly more variance than expected at random (details in Methods). While further principal components also explained significantly more variance than expected at random, there was a sharp drop in significance levels after the 11th principal component and examination of further principal components did not reveal that they highlighted biologically meaningful differences among profiled cells. Overall, clustering analysis revealed 7 clusters of cells (Figure 5.8).

Clusters 0, 1, 3 and 5 expressed high levels of contractile VSMC markers, such as *Myh11* and *Acta2* (Figure 5.9). These VSMCs may represent the contractile cells from the non-remodelled parts of the carotid artery, as the entire artery was included in scRNA-seq profiling. Cluster 4 contained the majority of *Mki67*-positive cells and the majority of *Ly6a/Sca1*-positive cells were located in cluster 2 (Figure 5.10).

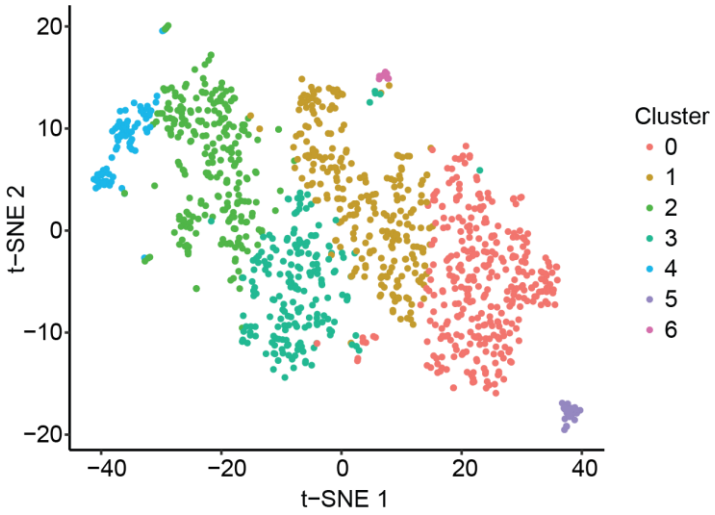


Figure 5.8: Clustering analysis of profiled VSMCs after carotid ligation injury.

Clusters of cells were identified using the graph-based clustering method from the Seurat package (Butler et al. 2018) and the clustering analysis was based on the top 11 principal components. In total, 7 clusters of cells were identified and are colour coded on a t-SNE representation of profiled cells.

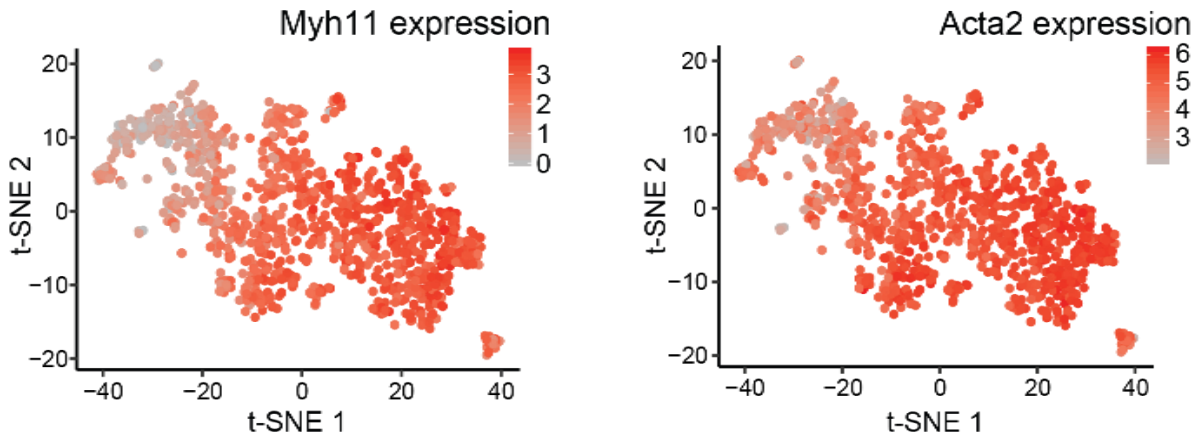


Figure 5.9: Expression levels of VSMC marker genes *Myh11* and *Acta2* in VSMCs profiled following carotid ligation injury.

t-SNE plot showing the expression levels of *Myh11* and *Acta2* in profiled cells following carotid ligation injury. The expression levels are colour-coded in individual cells with grey representing low and red high log2-transformed normalised expression levels.

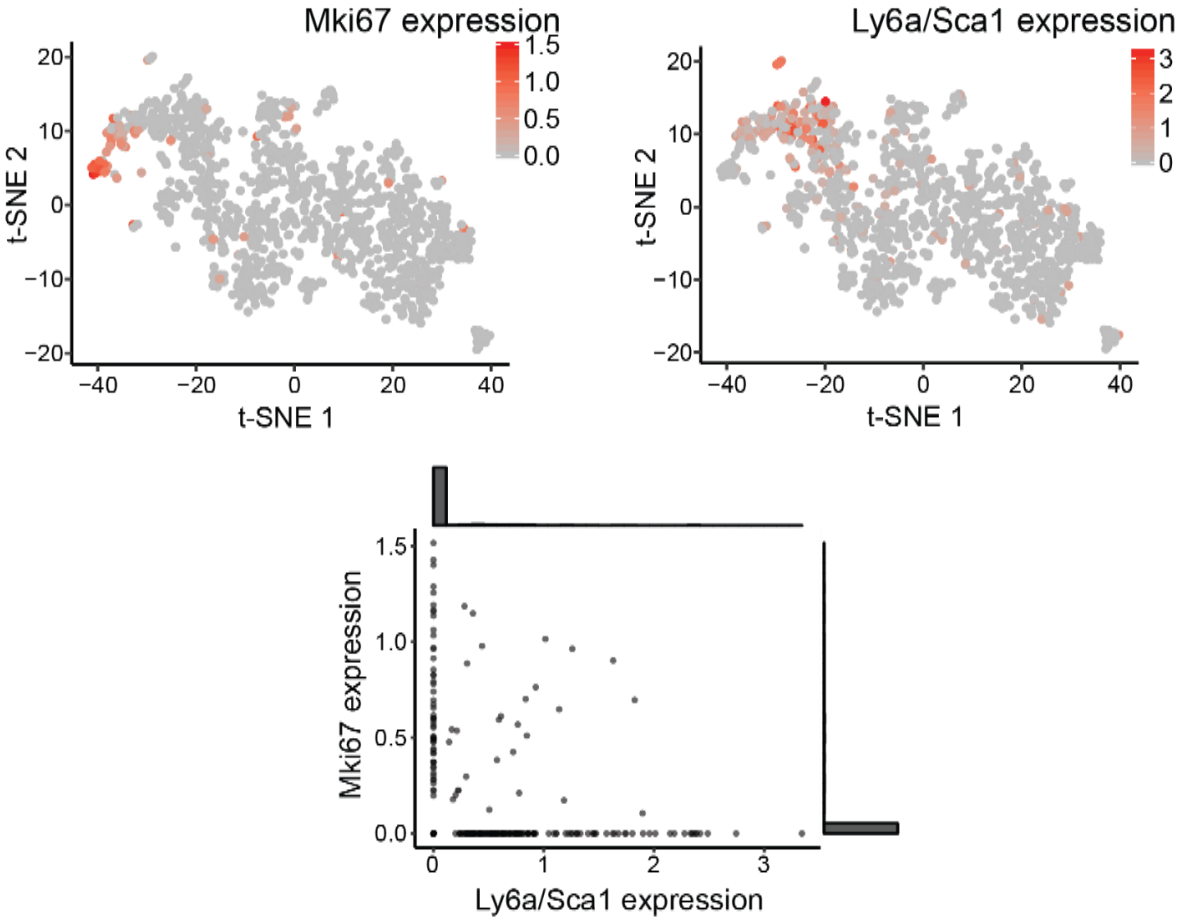


Figure 5.10: *Mki67* and *Ly6a/Sca1* expression levels among the cells profiled after carotid ligation injury.

Top: t-SNE plot show the expression levels of *Mki67* and *Ly6a/Sca1* in profiled cells following carotid ligation injury. The expression levels are colour-coded in individual cells with grey representing low and red high log2-transformed normalised expression levels. Bottom: scatterplot showing the relationship between expression levels if *Ly6a/Sca1* and *Mki67* in profiled cells. Histograms show the distribution of cells along the x and y axes.

To obtain a more comprehensive view of the transcriptional differences between the clusters, I performed differential gene expression analysis between each of the identified clusters and the remaining cells. Cluster 0 showed particularly high levels of expression of VSMC contractile markers *Myh11* and *Acta2*, which were identified as significantly upregulated in this cluster (Figure 5.11). Cluster 1 expressed higher levels of several genes which we observed to form module 2 among highly variable genes identified in healthy S+L+ cells (Section 4.2.4). Examples of these genes include *Egr1*, *Fos* and *Nfkbia*, indicating that heterogeneity observed among

VSMCs in the healthy arteries was also present following carotid ligation surgery (Figure 5.11). Additionally, cluster 1 VSMCs expressed higher level of *Klf4*, which has previously been shown to play a key role in VSMC phenotypic switching (Shankman et al. 2015). This suggests that VSMCs in cluster 1 may be in a primed state for the early stages of response to injury compared with other contractile VSMCs. Cluster 3 did not express a distinct signature, with only 5 genes found to be upregulated in this cluster. Similarly, no genes were found to be significantly upregulated in the smallest identified cluster 6 (Figure 5.11).

Interestingly, cluster 5 expressed transcriptional signatures resembling the AA-only cluster identified during the application of random forest analysis to the 10X Genomics Chromium datasets of VSMCs from healthy aortas. The gene with highest log2 fold change in expression in cluster 5 was *Rgs5* and additionally this cluster upregulated several other genes, which were also observed to be upregulated in the AA only cluster (Section 3.2.8), including *Sfrp2* and *Prxx2* (Figure 5.11).

Cluster 2 showed upregulation of several known markers of VSMC phenotypic switching, such as *Spp1*, *Col8a1*, *Vcam1* and *Mgp*. *Ly6a/Sca1* was also upregulated in cluster 2 (Figure 5.11). Cluster 4 showed higher expression levels of the proliferative marker *Mki67*, which we used to enrich for VSMCs undergoing phenotypic switching (Figure 5.11). Additionally this cluster showed upregulation of the proliferative markers *Pcna* and *Mcm7* (Juríková et al. 2016) and likely represented actively proliferating VSMCs.

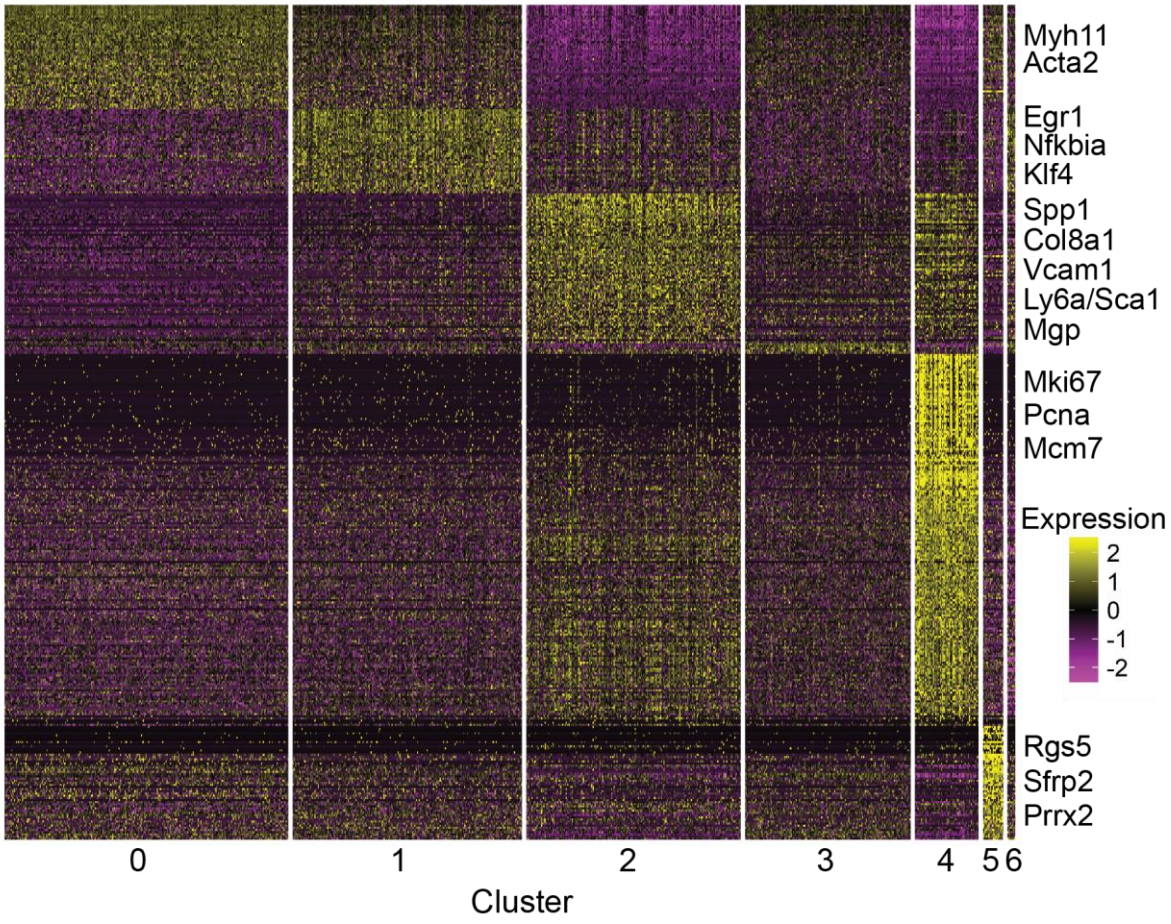


Figure 5.11: Transcriptional signatures of identified clusters of VSMCs following carotid ligation injury.

Heatmap showing the scaled expression levels of genes, identified as significantly upregulated in at least one cluster of cells (adjusted p-value < 0.05 and log2 fold change > 0.5, details in Methods). All differentially expressed genes are shown and their scaled expression levels are colour-coded from purple (low) to yellow (high). Cells are shown in columns and are grouped by cluster memberships and genes are shown in rows. Selected genes are indicated on the right.

Evaluation of the quality control metrics in individual clusters showed that VSMCs displaying signs of phenotypic switching in clusters 2 and 4 had higher total UMI counts and higher numbers of detected genes in individual cells, as would be expected for activated VSMCs (Figure 5.12). Thus, removing the cells with high UMI counts and large numbers of detected genes during quality control would have removed a significant proportion of cells of interest.

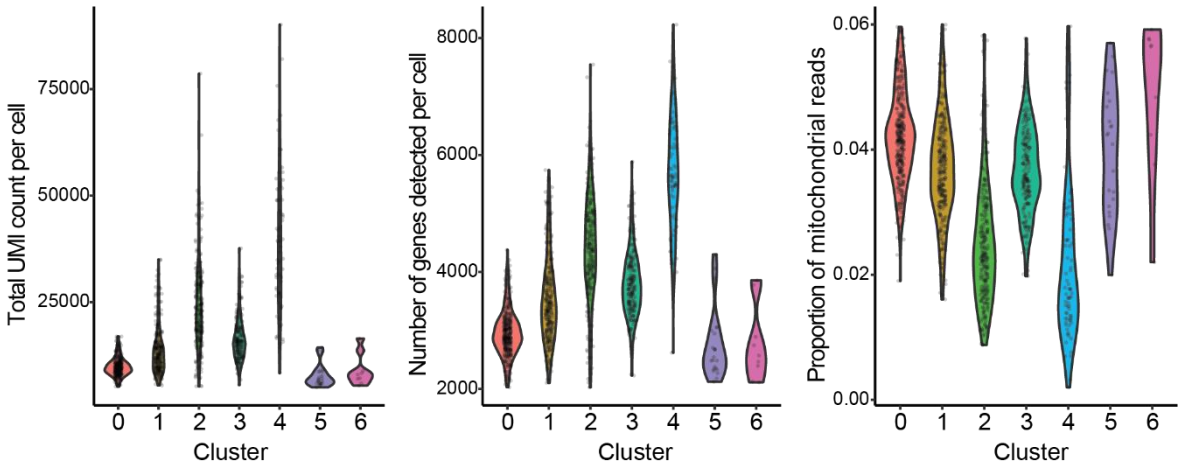


Figure 5.12: Quality control metrics in identified clusters of cells following vascular injury.
Violin plots show the distributions of quality control metrics of total UMI count, number of genes detected and the proportion of mitochondrial reads per cell in identified clusters of cells. Individual cells are represented as black dots and only the cells, which passed the quality control thresholds, are included.

I was next interested in understanding whether the cells which showed transcriptional profiles characteristic of VSMCs undergoing phenotypic switching resembled the S+L+ cells expressing high levels of the cVSMCneg signature in healthy arteries. I visualised the summarised expression levels of the cVSMCneg and cVSMCpos transcriptional signatures, which were identified as genes which correlated negatively or positively with the cVSMC contractile network in healthy S+L+ cells (described in Chapter 4). I used the PC1 values derived from PCA based on the cVSMCneg or cVSMCpos signatures to summarise the expression levels of these two sets of genes in VSMCs following carotid ligation (details in Methods).

The *Ly6a/Sca1*-positive cluster of cells (cluster 2) expressed particularly high levels of the cVSMCneg signature, which contained genes characteristic of the synthetic state of VSMCs (Figure 5.13). In addition, this subset of cells expressed low levels of the contractile cVSMCpos signature (Figure 5.13). This suggested that *Ly6a/Sca1*-positive VSMCs after carotid ligation injury shared transcriptional similarities with activated S+L+ cells previously observed in healthy arteries.

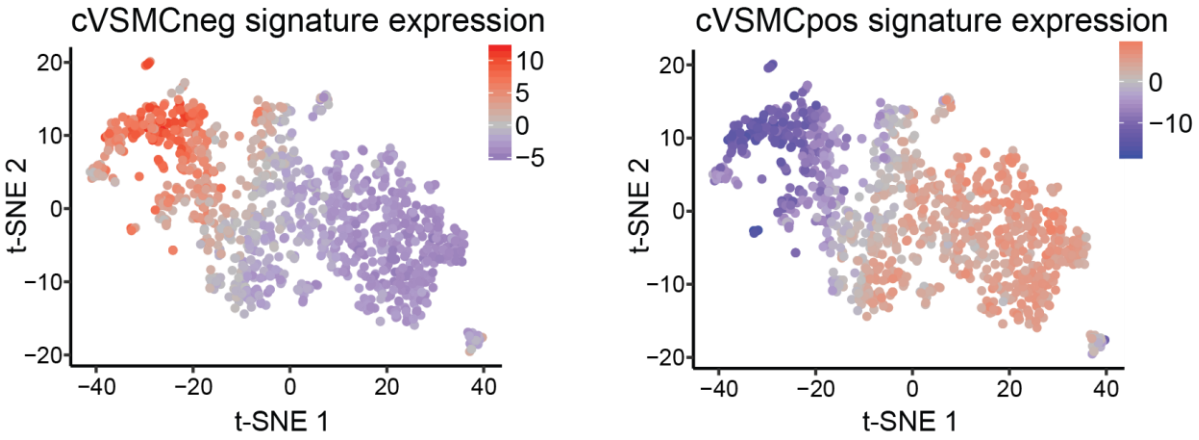


Figure 5.13: Summarised expression levels of the cVSMCneg and cVSMCpos signatures among cells profiled after carotid ligation injury.

t-SNE plots show the summarised expression levels of the cVSMCneg (left) and cVSMCpos (right) transcriptional signatures (details in Methods). Summarised expression levels are encoded from blue (low) to red (high) in individual cells, which are represented by dots.

VSMCs in cluster 2, which contained *Ly6a/Sca1*-positive cells, also expressed higher levels of *Col8a1* and *Spp1*, which are characteristic of the synthetic VSMC state (Figure 5.14). Overall, this analysis suggested that *Ly6a/Sca1*-positive VSMCs were among the phenotypically modulated subpopulation of cells following carotid ligation injury.

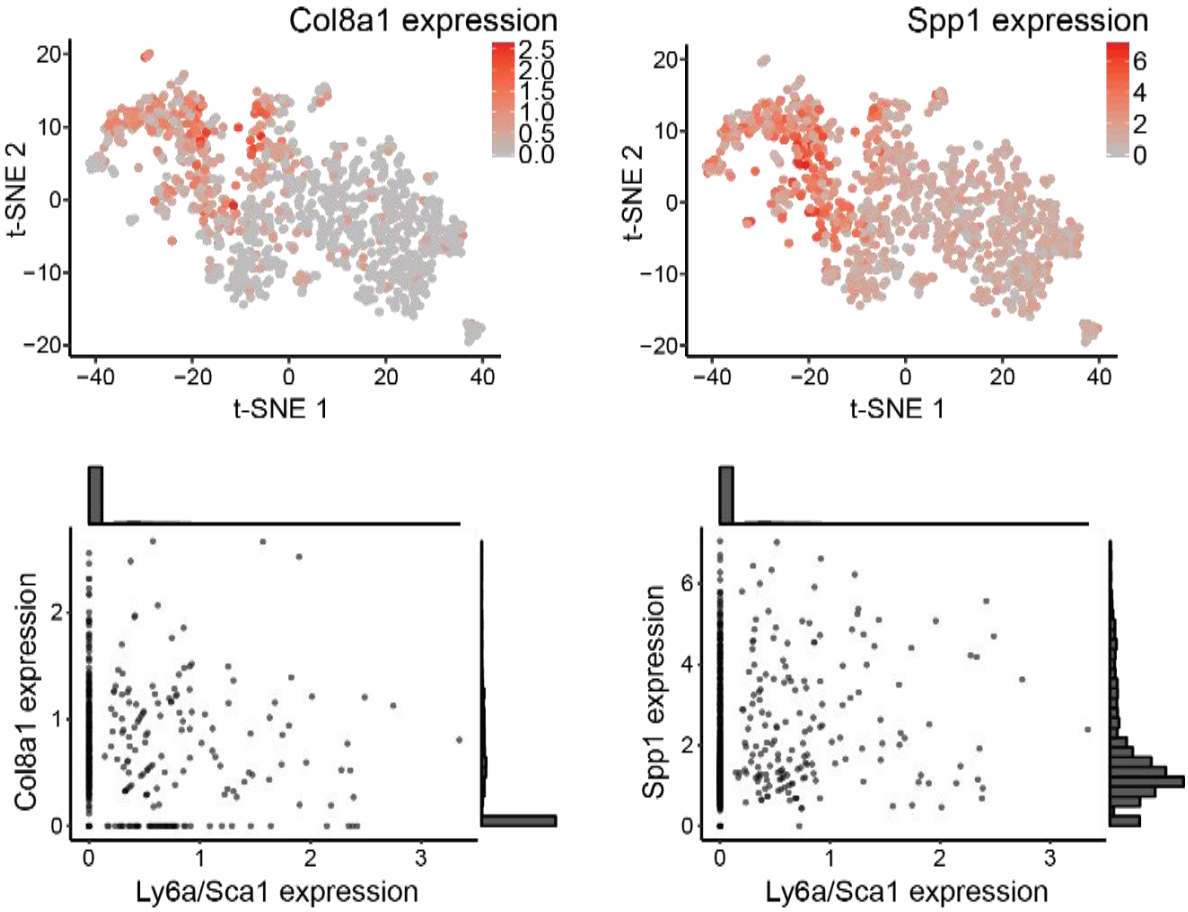


Figure 5.14: *Col8a1* and *Spp1*, which are characteristic of the synthetic state of VSMCs, are expressed among the *Ly6a/Sca1*-positive cells.

Top: t-SNE representations show the expression levels of *Col8a1* and *Spp1* in profiled VSMCs following carotid ligation injury. The log2-transformed normalised expression levels are colour-coded in individual cells with grey representing low and red high expression levels. Bottom: scatterplots show the relationship between the expression levels of *Ly6a/Sca*1, *Col8a1* and *Spp1*. Histograms show the distributions of cells along the x and y axes.

5.2.3 Transcriptional signatures of VSMCs within the atherosclerotic plaque

I next investigated the transcriptional signatures of VSMCs undergoing phenotypic switching in atherosclerotic plaques. The inflammatory environment of the atherosclerotic plaque has been described to drive VSMC phenotypic modulation towards a range of different phenotypes, such as calcifying-like cells, macrophage-like cells and mesenchymal stem cell-like cells, which have been observed to express SCA1 (Shankman et al. 2015; Bennett et al. 2016; Durham et al. 2018). To understand the transcriptional changes which occur in VSMCs

undergoing phenotypic modulation to these states, we profiled the lineage-labelled VSMCs isolated from atherosclerotic plaques of ApoE$^{-/-}$ mice, which were fed a high fat diet for either 14 or 18 weeks (jointly in collaboration with H.F. Jørgensen).

In total we profiled 3346 VSMCs from 14-week plaques and 2757 VSMCs from 18-week plaques. The quality of profiled cells was assessed based on their total UMI count, the number of genes detected and the proportion of mitochondrial reads. The distribution of the total UMI count among the profiled cells was bimodal and closer examination of the relationship between the total UMI count and the proportion of mitochondrial reads per cell showed that there was a group of cells with relatively low UMI counts and a high proportion of mitochondrial reads (Figure 5.15a and b). The lower threshold for acceptable total UMI count was therefore set at 5000 UMI counts and cells were required to express less than 9% of mitochondrial reads to pass quality control. Cells with particularly high total UMI count (over 20000) and number of genes detected (over 5000) were removed from this analysis, as the higher number of captured cells increased the chances of profiling doublets (Figure 5.15a).

The transcriptomes of good-quality VSMCs were then normalised (details in Methods). I next performed principal component analysis to reduce the dimensionality of the dataset prior to clustering. Subsequent clustering analysis and t-SNE visualisation initially identified 7 clusters of cells (Figure 5.16a, details in Methods). When I performed differential gene expression analysis among the identified clusters I noticed that several mitochondrial genes were upregulated in cluster 4, which contained VSMCs from both 14- and 18-week atherosclerotic plaque stages (Figure 5.16b). Inspection of the quality control metrics for individual clusters showed that cluster 4 displayed lower total UMI counts and a higher proportion of mitochondrial reads than the other identified clusters (Figure 5.16c). This indicated that VSMCs in cluster 4 were of poorer quality and were excluded from further analysis. Clustering analysis and t-SNE visualisation was then repeated with remaining VSMCs (Figure 5.16d). Although there was some variability in the proportions of VSMCs from different stages of high-fat diet between the clusters, all clusters contain cells from both the 14- and 18-week stages (Figure 5.16e).

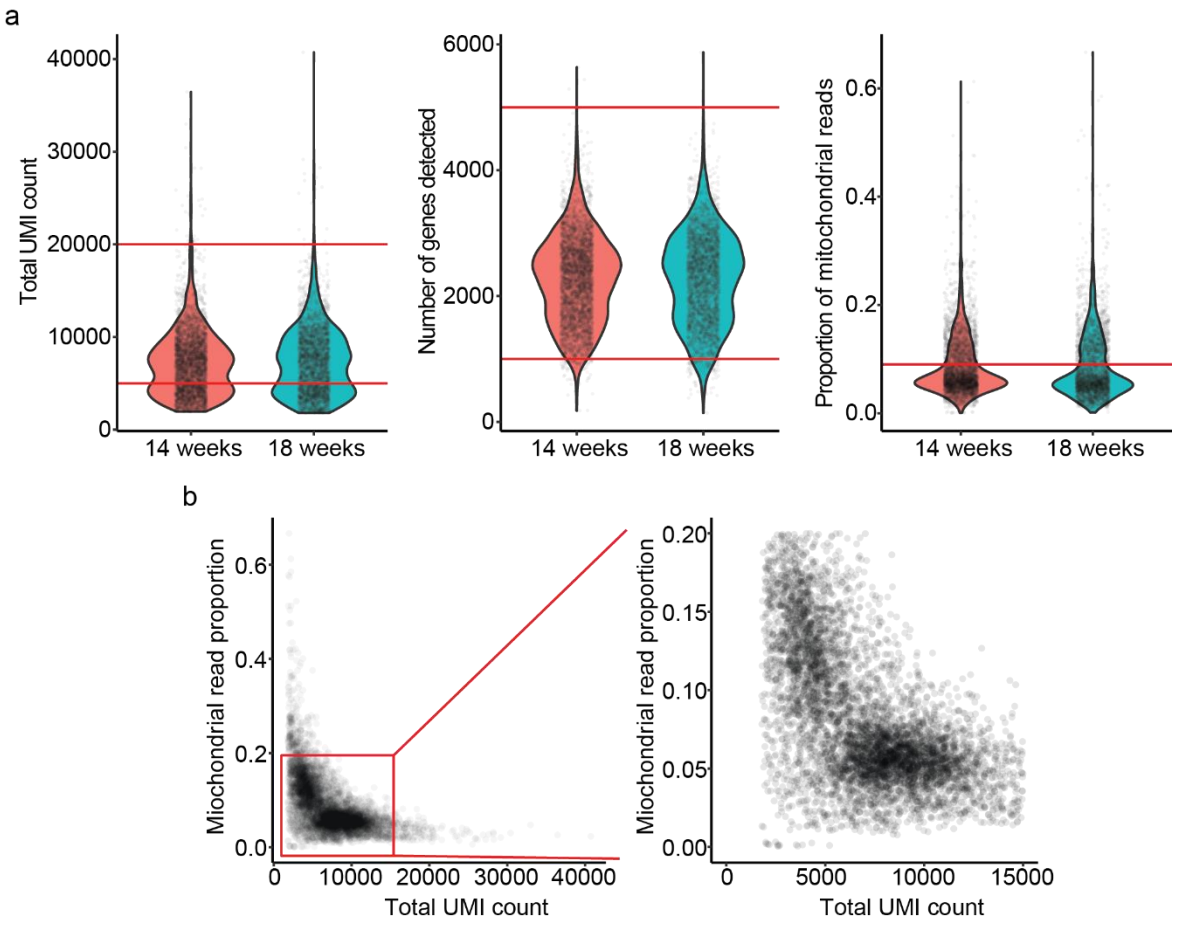


Figure 5.15: Quality control of single-cell transcriptomes of VSMCs from atherosclerotic plaques.

a) Violin plots showing the distribution of the quality control metrics of total UMI count, number of genes detected and the proportion of mitochondrial reads per cell. Individual cells are represented by black dots and the red lines represent the cut offs for good quality cells. b) Scatterplots showing the relationship between the total UMI count and the proportion of mitochondrial reads in profiled cells. All cells are included on the right, while the left scatterplot shows a zoomed-in view between 0 and 15000 UMIs and 0 to 0.2 proportion of mitochondrial reads.

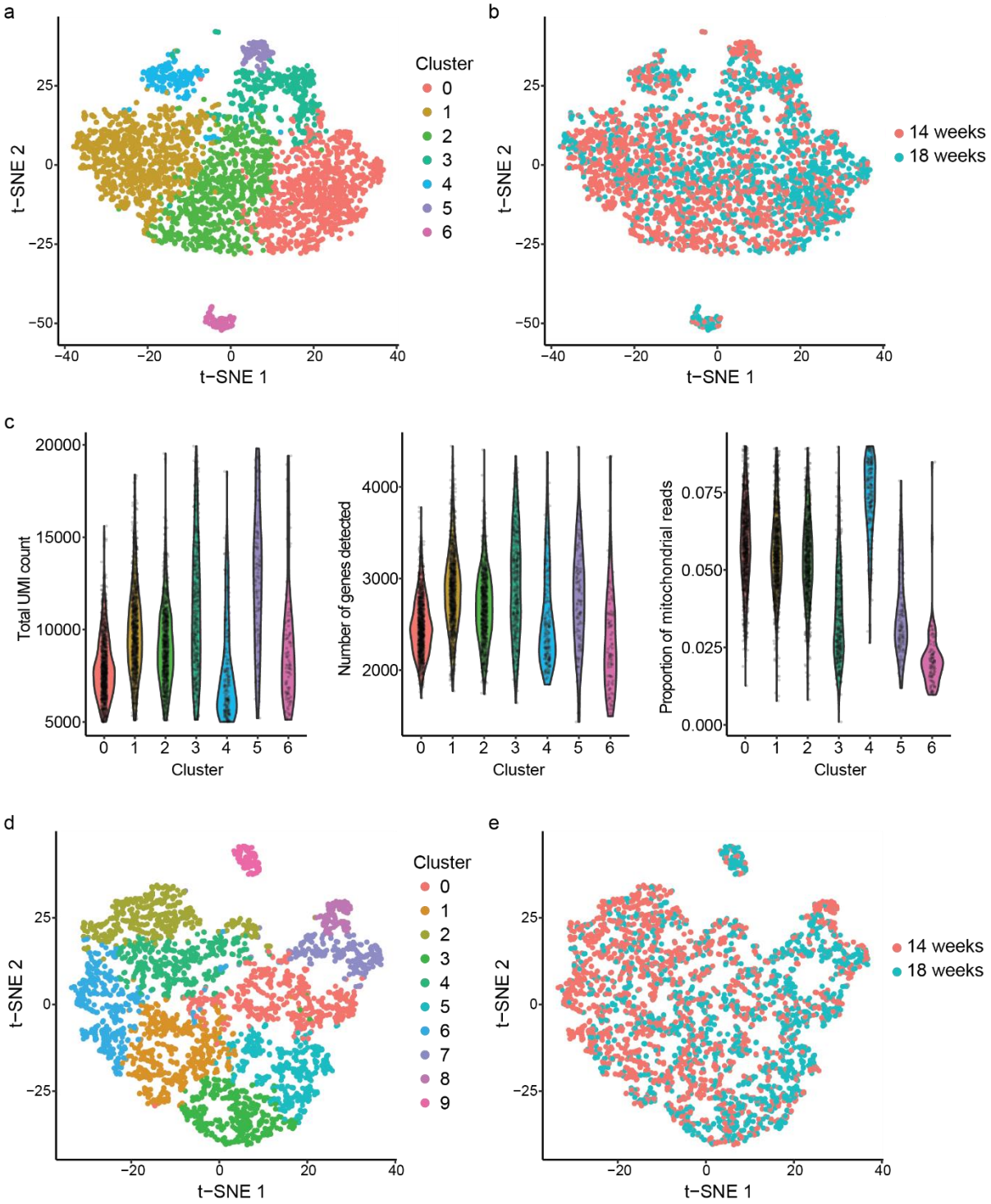


Figure 5.16: Clustering analysis of VSMCs from atherosclerotic arteries.

a, b, d, e) t-SNE representation of profiled VSMCs. Individual cells are represented by dots, which are colour coded according to cluster membership (a, d) or by the duration of high fat diet prior to analysis (b, e). c) Violin plots showing the distribution of total UMI count, the number of genes detected and the proportion of mitochondrial reads within identified clusters of VSMCs in panel a. Clusters of cells were identified using the graph-based clustering method from the Seurat package (Butler et al. 2018). Panel d is adapted from Dobnikar and Taylor et al. 2018.

Differential gene expression analysis between the identified clusters of VSMCs showed varying levels of transcriptional differences between the clusters. Cluster 9 was transcriptionally the most distinct cluster and showed increased expression levels of several macrophage marker genes, such as *Cd68* and *Lgals3* (Figure 5.17). There have been several previous reports of VSMC lineage-labelled cells within the plaque co-expressing macrophage markers (Feil et al. 2014; Shankman et al. 2015; Chappell et al. 2016; Albarrán-Juárez et al. 2016). Differential gene expression analysis suggests that the transcriptional profiles of VSMC-derived macrophages are substantially different from other plaque VSMCs, beyond the expression of a handful of macrophage markers. It is, however, not possible to rule out that observed macrophage-like cells could be a contamination from bone marrow-derived macrophages. Possible contamination is supported by the fact that *Ptprc*, which encodes the bone marrow marker Cd45, was identified as upregulated in cluster 9 (Figure 5.17), although VSMC lineage-labelled cells have also been observed to express *Ptprc* within the plaque previously (Albarrán-Juárez et al. 2016). *Myh11* and *Acta2* transcripts were detected at low levels in cluster 9 (Figure 5.18a), however, this could be explained through background ambient mRNA contamination (Macosko et al. 2015), which may be particularly likely for genes expressed at high levels by the majority of profiled cells. Due to possible technical factors it is therefore not possible to conclusively describe the origin of observed macrophage-like cells.

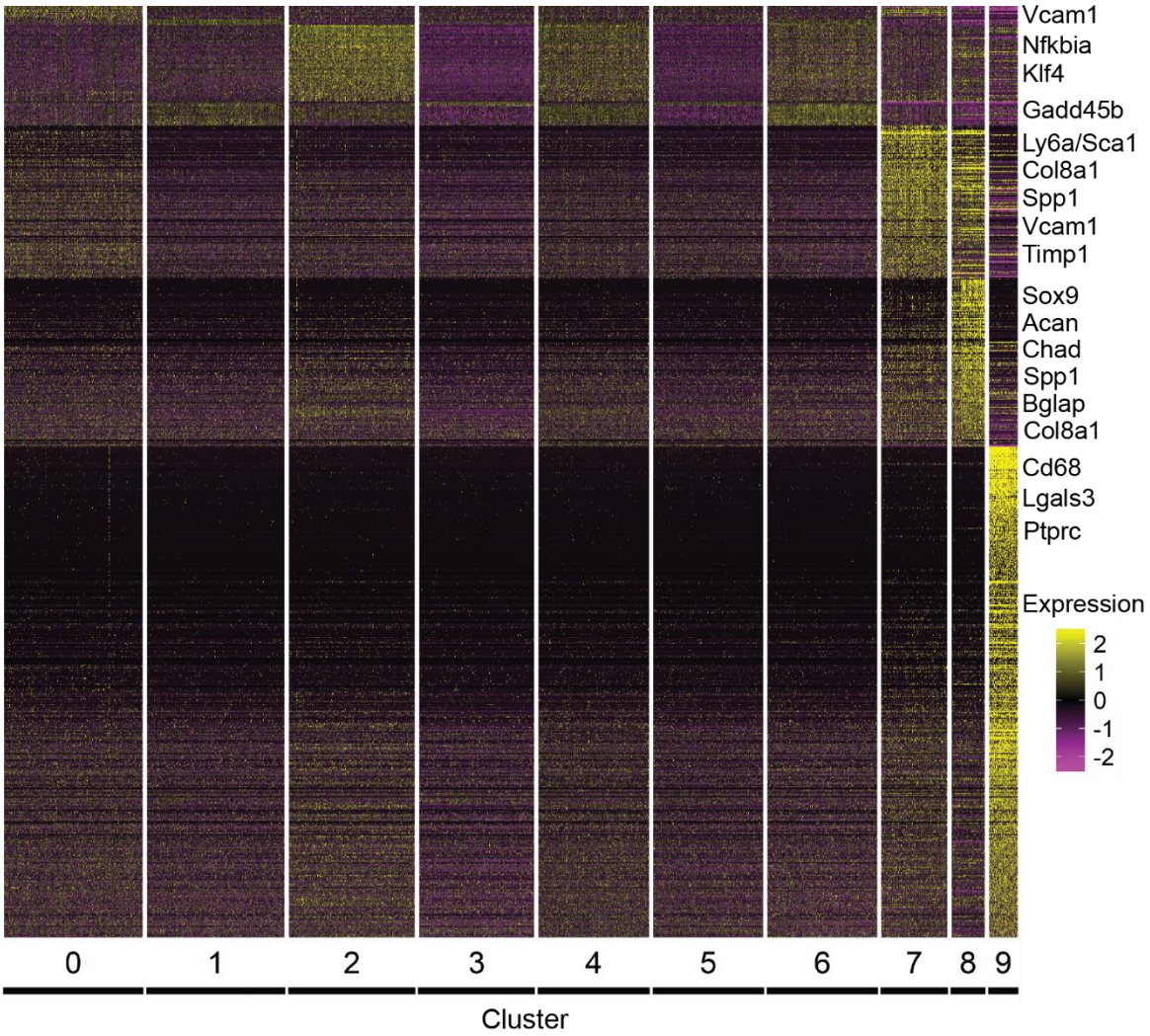


Figure 5.17: Differential gene expression among identified clusters of VSMCs from atherosclerotic arteries.

Heatmap showing the scaled expression levels of genes, identified as significantly upregulated in at least one cluster of cells (adjusted p-value < 0.05 and log2 fold change > 0.5, details in Methods). Cells are grouped by cluster and represented in columns and the differentially expressed genes are shown in rows. Selected upregulated genes are indicated on the right. Scaled expression levels are colour coded from purple (low) to yellow (high).

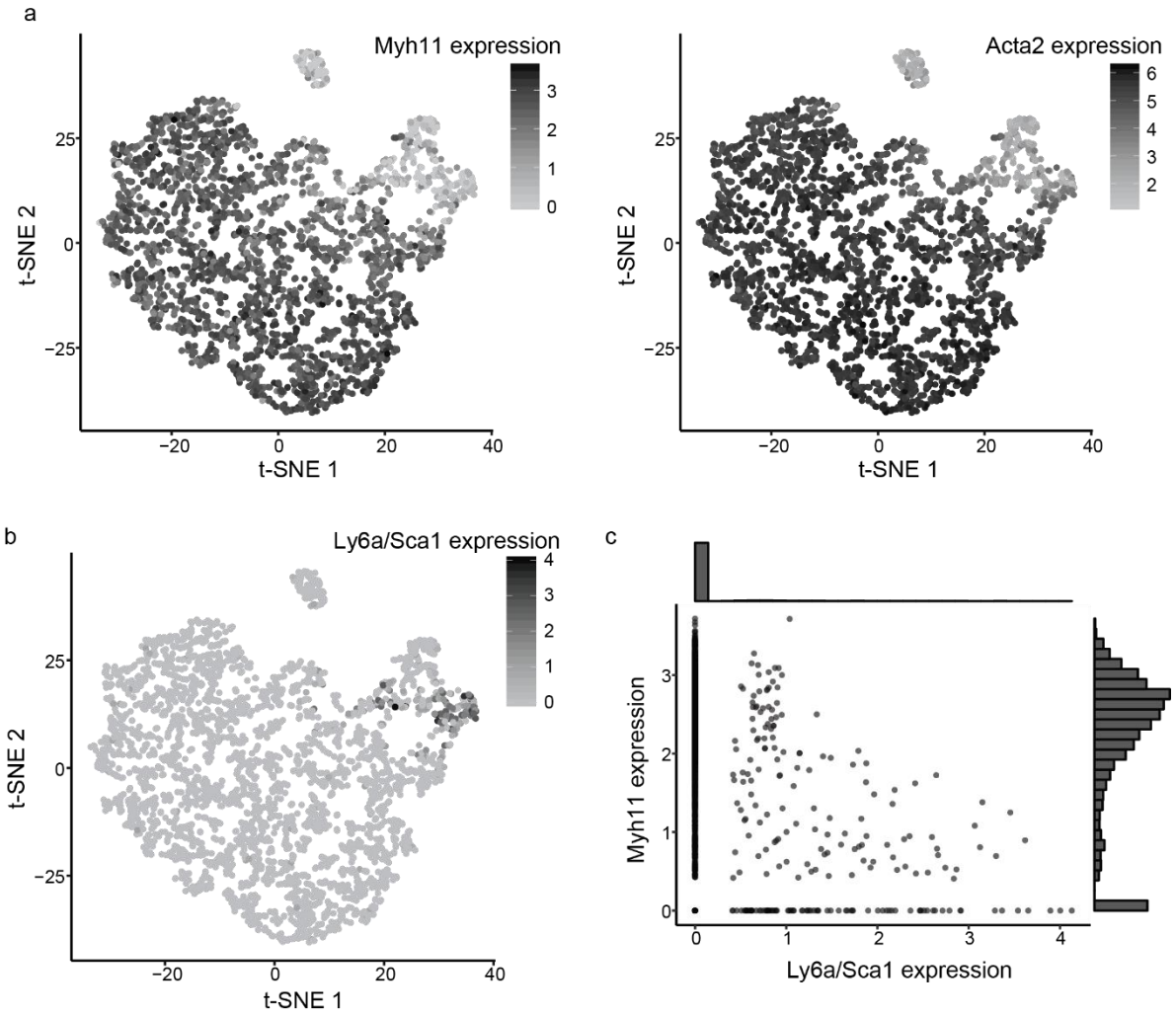


Figure 5.18: Contractile VSMC marker and *Ly6a/Sca1* expression in profiled cells from atherosclerotic arteries.

t-SNE visualisations of profiled VSMCs with individual cells colour-coded for the log2-transfromed, normalised expression levels of *Myh11* and *Acta2* (a) and *Ly6a/Sca1* (b), with light grey representing low and dark grey high expression levels. c) Scatterplot showing the relationship between expression levels of *Myh11* and *Ly6a/Sca1*. Histograms show the distribution of cells along the x and y axes. Panel b is adapted from Dobnikar and Taylor et al. 2018.

VSMCs in cluster 7 showed upregulation of *Ly6a/Sca1* (Figure 5.18b) as well as several markers of VSMC phenotypic switching, such as *Vcam1*, *Col8a1* and *Spp1* (Figure 5.17). Cluster 8 shared some of the transcriptional signatures of cluster 7 and additionally upregulated several genes which mark osteochondrocytes, such as *Sox9*, *Acan* and *Bglap* (Figure 5.17). This cluster of VSMCs may therefore be involved in arterial calcification. Clusters 0-6 expressed high levels

of VSMC contractile markers such as *Myh11* and *Acta2* (Figure 5.18a). Differential gene expression analysis revealed some differences among the contractile clusters of VSMCs. Clusters 2, 4 and 6 showed similar transcriptional signatures, with increased expression of several genes including *Nfkbia* and *Klf4*. Cluster 0 showed a small increase in expression of several genes marking phenotypic activation of VSMCs, such as *Vcam1* and *Mgp* (Figure 5.17). These and other genes marking VSMC phenotypic switching were also more strongly upregulated in the adjacent cluster 7. Overall, there was a degree of overlap in the transcriptional signatures of clusters 0-6 (Figure 5.17), and not all of these clusters showed large transcriptional differences. As such, identified clusters likely do not represent distinct subpopulations of VSMCs and could be merged. Nevertheless, this analysis did reveal interesting heterogeneity within contractile VSMCs, with several biologically relevant genes identified as differentially expressed in the identified clusters of cells.

I next investigated the similarity of VSMCs showing transcriptional signatures characteristic of phenotypic switching in atherosclerotic plaques to the activated healthy S+L+ cells. Cells in clusters 7, 8 and 9 all expressed low levels of the cVSMCpos signature (Figure 5.19), which positively correlated with the contractile cVSMC network in healthy S+L+ cells (described in Chapter 4). At the same time, the overall expression levels of the negatively correlated cVSMCneg signature were higher in clusters 7, 8 and 9 than in the contractile clusters 0-6 (Figure 5.19). This suggested that *Ly6a/Sca1*-positive VSMCs, as well as other clusters showing signatures of phenotypic modulation shared similarities with healthy S+L+ cells, which expressed high levels of the cVSMCneg signature.

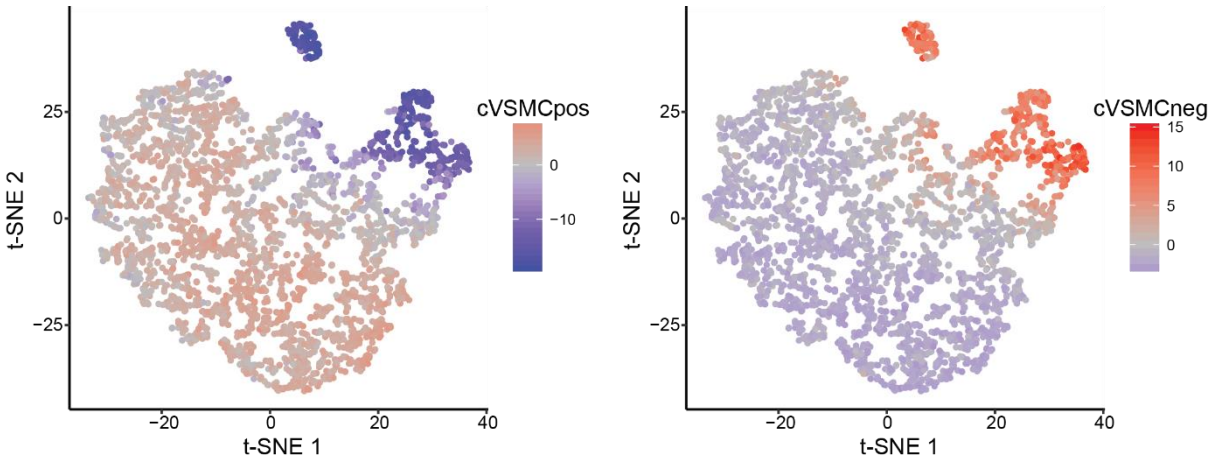


Figure 5.19: Summarised expression levels of the cVSMCpos and cVSMCneg signatures in plaque VSMCs.

t-SNE representations of VSMCs isolated from atherosclerotic arteries are colour-coded for the summarised expression levels of the cVSMCpos (left) and cVSMCneg (right) signatures, which were originally identified in healthy S+L+ cells. PC1 values positively aligned with the total expression levels of cVSMCpos and cVSMCneg genes were used to summarise the overall expression levels (details in Methods), with red representing high and blue low expression. The figure is adapted from Dobnikar and Taylor et al. 2018.

5.2.4 Alternative marker genes of *Ly6a/Sca1*-expressing VSMCs in models of phenotypic switching

So far in this chapter I have shown that *Ly6a/Sca1*-positive VSMCs are present in model systems of VSMC phenotypic switching and that the transcriptional signatures of *Ly6a/Sca1*-positive VSMCs following *in vitro* culture, carotid ligation injury and atherosclerosis share similarities with healthy S+L+ cells, which express high levels of the cVSMCneg signature. SCA1 protein, however, has no known human orthologue, which presents challenges for identifying a similar population of VSMCs in human atherosclerotic plaques. Knowing the wider transcriptional signatures of VSMCs undergoing phenotypic switching in mice is valuable for investigating potential analogous VSMC subpopulations in atherosclerotic plaques from other organisms using transcriptomics approaches. To further facilitate the isolation of the desired subpopulations of VSMCs in other organisms I next aimed to identify a set of alternative

marker genes of *Ly6a/Sca1*-positive VSMCs in model systems of VSMC phenotypic switching. To investigate which genes would be suitable candidates for alternative markers of *Ly6a/Sca1*-positive VSMCs, I used co-expression analysis to identify a set of genes, which showed a similar pattern of expression to *Ly6a/Sca1*. This analysis identified a set of 708 genes as co-expressed with *Ly6a/Sca1* in VSMCs following carotid ligation injury (adjusted p-value < 0.01, details in Methods). For a gene considered to be a suitable alternative marker of *Ly6a/Sca1*-positive VSMCs, it was also required to be upregulated in cluster 2 of the carotid ligation injury dataset, which contained *Ly6a/Sca1*-positive VSMCs. Additionally, I required suitable alternative markers of the *Ly6a/Sca1*-positive VSMCs to be present in the cVSMCneg signature. This approach resulted in a set of 17 genes, which shared expression patterns with *Ly6a/Sca1* (Figure 5.20). The majority of these genes were clearly expressed at higher levels in the *Ly6a/Sca1* cluster 2 post carotid ligation injury (Figure 5.20).

I next checked the expression levels of these genes in VSMCs isolated from atherosclerotic plaques of ApoE$^{-/-}$ mice to ensure that potential alternative markers of *Ly6a/Sca1*-positive VSMCs were representative across different model systems of VSMC phenotypic switching. In the atherosclerotic plaque dataset *Ly6a/Sca1* was expressed in cluster 7 and while the majority of the tested genes were expressed at higher levels in this cluster, only *Fbln2* and *Slco2a1* were specific to VSMCs in this cluster. The majority of other genes were also upregulated in clusters 8 or 9 (Figure 5.20).

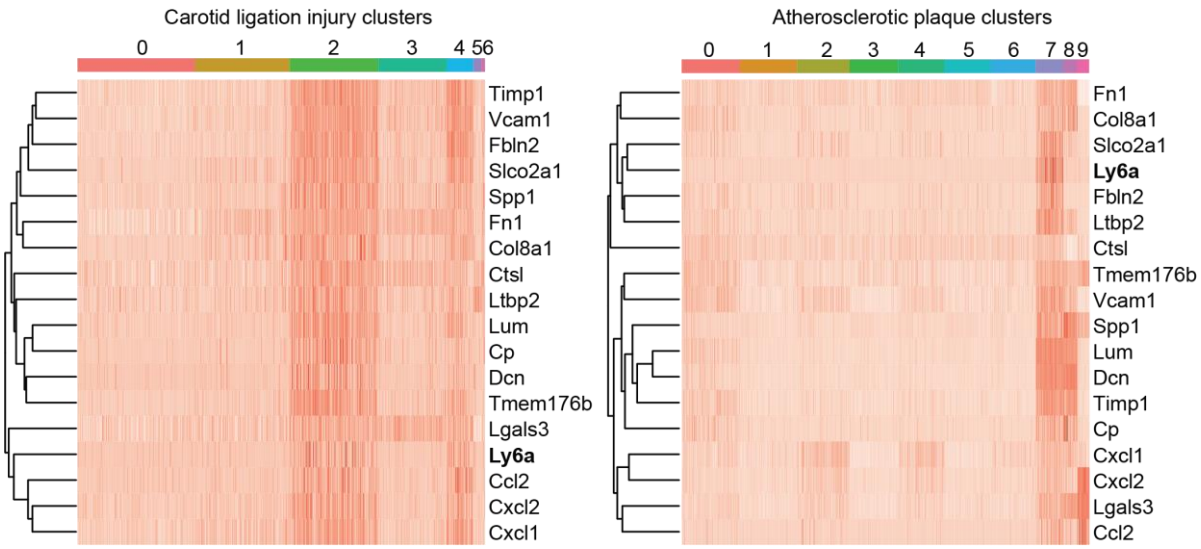


Figure 5.20: Identification of alternative markers of *Ly6a/Sca1*-positive VSMCs.

Heatmaps showing the expression levels of genes, which were identified as correlated with *Ly6a/Sca1* in the injury dataset, were upregulated in cluster 2 of post injury VSMCs and were part of the cVSMCneg signature. The expression levels in the carotid ligation injury (left) and the atherosclerotic plaque (right) datasets are colour-coded with darker red representing higer expression. Genes are ordered in rows according to the results of hierarchical clustering and cells are grouped by cluster membership in columns.

Closer examination of *Ly6a/Sca1*, *Fbln2* and *Slco2a1* expression patterns showed that these genes were generally expressed in the same subpopulations of VSMCs in model systems of VSMC phenotypic switching. However, in the carotid ligation injury model, *Fbln2* and *Slco2a1* were expressed in a larger proportion of cluster 2 VSMCs than *Ly6a/Sca1*, while in cultured VSMCs *Fbln2* marked a larger proportion of cells than both *Ly6a/Sca1* and *Slco2a1* (Figure 5.21). There was a tighter overlap in the expression of *Ly6a/Sca1*, *Fbln2* and *Slco2a1* within atherosclerotic plaques (Figure 5.21). However, a small proportion of VSMCs located in the contractile VSMC clusters in all three models of VSMC phenotypic switching also expressed *Ly6a/Sca1*, *Fbln2* or *Slco2a1*. An alternative and targeted approach for isolation of *Ly6a/Sca1*-positive VSMCs would therefore require selecting for reduced expression levels of contractile markers alongside higher expression of *Fbln2* or *Slco2a1*. Taken together, *Fbln2* or *Slco2a1* may be suitable alternative markers of *Ly6a/Sca1*-positive VSMCs for the investigation and isolation of an analogous population of VSMCs. However, further testing of expression and specificity at the protein-level is still required.

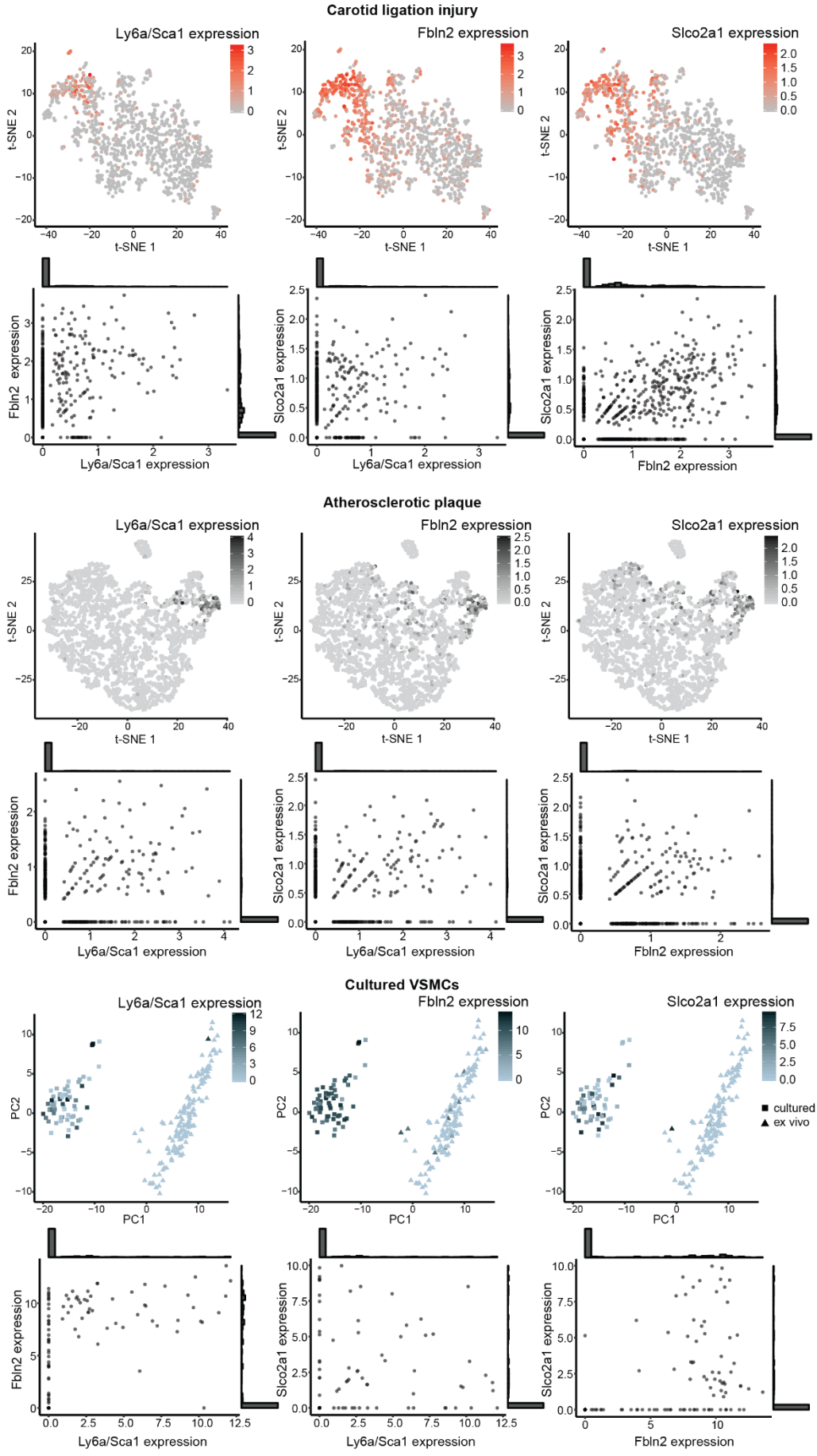

Carotid ligation injury
Ly6a/Sca1 expression
Fbln2 expression
Slco2a1 expression
t-SNE 1
t-SNE 2
Fbln2 expression
Slco2a1 expression
Ly6a/Sca1 expression
Atherosclerotic plaque
Cultured VSMCs
PC1
PC2
cultured
ex vivo

Figure 5.21: Expression levels of *Fbln2* and *Slco2a1* overlap with *Ly6a/Sca1* expression in model systems of VSMC phenotypic switching.

The expression levels of *Ly6a/Sca1*, *Fbln2* and *Slco2a1* are visualised after carotid ligation injury (top), in the atherosclerotic plaque (middle) and in cultured VSMCs (bottom). Expression levels are colour-coded on t-SNE or PCA visualisations and pairwise gene expression relationships between of *Ly6a/Sca1*, *Fbln2* and *Slco2a1* are shown in scatterplots. Histograms show the distributions of cells along the x and y axes of shown scatterplots.

5.3 Discussion

In this chapter I characterised the transcriptional signatures of phenotypically switching VSMCs in three model systems of phenotypic switching. The proportion of VSMCs which expressed *Ly6a/Sca1* was higher in all three model systems of phenotypic switching compared with healthy artery walls. Closer examination of the transcriptional signatures of *Ly6a/Sca1*-expressing VSMCs following phenotypic switching revealed similarities with SCA1-positive VSMCs from healthy arteries, which expressed transcriptional signatures characteristic of synthetic VSMCs. Additionally, I identified a set of putative alternative markers of SCA1-positive VSMCs to facilitate the investigation of a similar population in human atherosclerosis.

5.3.1 VSMC phenotypic switching in culture

I initially used *in vitro* culture as a model system of VSMC phenotypic switching, as single-cell transcriptomes of cultured VSMCs were publicly available (GSE79436, Adhikari et al. 2015). As expected, I observed downregulation of VSMC marker gene expression in cultured VSMCs compared with the *ex vivo* VSMCs we profiled, which is a hallmark of VSMC phenotypic switching. Cultured VSMCs also upregulated *Ly6a/Sca1* and showed higher expression of the cVSMCneg transcriptional signature. *In vitro* culture is a simple model system, with clear differences to the physiological environment, in which VSMC phenotypic switching occurs *in vivo* (Rensen et al. 2007). However, further *in vivo* work presented in Sections 5.2.2 and 5.2.3 showed that these findings were replicated in phenotypically switching VSMCs in the carotid ligation injury or atherosclerosis mouse models.

5.3.2 Induction of VSMC phenotypic switching after carotid ligation injury

VSMC proliferation and accumulation in the neointima is reproducibly and acutely induced following the cessation of blood flow through carotid ligation surgery (Kumar & Lindner 1997). Only a small number of VSMCs were previously observed to expand extensively following carotid ligation injury, similar to the observations of clonality in the mouse atherosclerotic plaque (Chappell et al. 2016). Consistent with *in vitro* culture, we observed increased frequency of *Ly6a/Sca1* transcript expression in VSMCs following carotid ligation injury compared with healthy arteries. In addition, we confirmed that SCA1 is upregulated in VSMCs at the protein level following carotid ligation using flow cytometry. Similarly to cultured VSMCs*, Ly6a/Sca1*-expressing VSMCs following carotid ligation injury also expressed high levels of the cVSMCneg signature originally identified in healthy S+L+ cells.

VSMCs showing signs of phenotypic switching expressed lower levels of contractile VSMC genes than other profiled cells. Previous studies have reported extensive downregulation of VSMC contractile genes following vascular injury (Regan et al. 2000; Herring et al. 2017). Herring et al. (2017) reported a significant decrease in VSMC marker gene expression as early as 3 days following carotid ligation injury and Regan et al. (2000) observed nearly complete downregulation of VSMC contractile genes in the neointima 7 days after vascular wire injury, which is the same time point used for scRNA-seq analysis presented in this chapter. I observed downregulation of VSMC contractile marker genes in the transcriptomes of VSMCs showing characteristics of phenotypic switching, however, the majority of profiled cells did express contractile markers, such as *Myh11* and *Acta2*. The discrepancy in observed contractile gene expression levels may be due to different models of vascular injury, as well as the fact that the entire carotid artery was profiled in our study, and VSMCs expressing contractile genes may have originated from the non-remodelled parts of the artery. Herring et al. (2017) also investigated transcriptional changes in carotid arteries following carotid ligation and observed upregulation of a number of inflammatory cytokines, such as *Il6, Il1b* and *Ccl2.* Herring et al. (2017) profiled gene expression changes in the entire carotid artery and not specifically in VSMCs. *Ccl2* was among the upregulated genes in cluster 2 of our dataset, however *Il6* and *Il1b* were only expressed very sparsely and at low expression levels. This observation suggests that VSMCs partly contribute to the observed upregulation of inflammatory cytokines following injury. Additionally, *Spp1* was observed to be upregulated after injury (Herring et al.

2017), which is consistent with higher expression levels of *Spp1* observed among phenotypically modulated VSMCs presented in this chapter.

Carotid ligation injury model (Kumar & Lindner 1997) provided a setting for investigation of VSMC phenotypic switching in an alternative and more acute environment to the atherosclerotic plaque. Interestingly, I observed a lower level of diversity of VSMC phenotypic modulation after injury compared with atherosclerosis. The phenotypically switching VSMCs in the carotid ligation dataset showed the transcriptional signatures of both the calcifying and the *Ly6a/Sca1*-positive clusters identified in the atherosclerotic plaque. This may be either due to environmental differences or due to the shorter timeframe of lesion development, with 7 days passing from carotid ligation to scRNA-seq analysis in case of injury and 14-18 weeks of high-fat diet duration prior to the analysis of VSMCs within atherosclerotic plaques.

5.3.3 VSMCs are phenotypically modulated towards a variety of phenotypes in atherosclerosis

Single-cell transcriptional profiling of lineage-labelled VSMCs from the atherosclerotic plaques revealed distinct subpopulations of phenotypically modulated VSMCs, which expressed low levels of contractile VSMC marker genes. This finding is in agreement with previous reports of VSMC modulation towards a range of different phenotypes, such as macrophage-like cells and calcifying cells (Naik et al. 2012; Nguyen et al. 2013; Feil et al. 2014; Shankman et al. 2015; Chappell et al. 2016; Jacobsen et al. 2017). In addition, VSMC-derived mesenchymal stem cell-like cells, characterised by the expression of *Ly6a/Sca1* and *Eng*, have been observed previously within atherosclerotic plaques by Shankman et al. (2015). I detected a cluster of *Ly6a/Sca1*-positive VSMCs showing signs of phenotypic switching within atherosclerotic plaques, which is in also agreement with a recent report confirming expression of *Ly6a/Sca1* among phenotypically modulated VSMCs within the plaque (Wirka et al. 2019). *Eng* was, however, expressed sparsely in all detected clusters of cells and was not specific to the *Ly6a/Sca1* cluster in our dataset.

VSMC involvement in arterial calcification through transition to an osteochondrogenic state has been established in recent years (Naik et al. 2012; Nguyen et al. 2013; Durham et al. 2018). VSMCs transitioning towards a calcifying phenotype were found to be marked by upregulation

of *Sox9*, *Spp1* (encodes osteopontin)*, Bglap* (encodes osteocalcin) and *Runx2* among others (Durham et al. 2018). In agreement with this, I detected a cluster of VSMCs isolated from the atherosclerotic plaques, which showed transcriptional signatures characteristic of VSMCs involved in vascular calcification, including differential expression of *Sox9*, *Spp1* and *Bglap*.

I also observed a cluster of plaque cells expressing transcriptional signatures characteristic of macrophages. VSMC-lineage cells expressing macrophage markers have been documented in several studies in mice (Feil et al. 2014; Shankman et al. 2015; Chappell et al. 2016; Albarrán-Juárez et al. 2016). Additionally, *in vitro* cholesterol loading experiments have suggested that VSMCs can upregulate macrophage markers, as well as a selection of other genes characteristic of the macrophage phenotype (Rong et al. 2003; Vengrenyuk et al. 2015). Macrophage-like cells detected in our study showed extensive transcriptional differences compared with contractile VSMC as well as other phenotypically modulated VSMC subpopulations. This may suggest that VSMCs are phenotypically modulated towards a macrophage state in atherosclerosis, and that transcriptional changes are global rather than confined to the the expression of a handful of macrophage markers. All cells included in the atherosclerotic plaque datasets were selected for VSMC lineage label expression by FACS, however possible contamination from bone-marrow derived macrophages cannot be ruled out. Observed macrophage-like cells expressed low levels of *Ptprc*, which encodes the bone marrow-derived cell marker Cd45. However, low expression levels of *Ptprc* have been previously observed in VSMC lineage-labelled cells within the plaque (Albarrán-Juárez et al. 2016), therefore expression of *Ptprc* may not necessarily imply contamination. Observed macrophage-like cells did express low levels of VSMC markers, such as *Myh11* and *Acta2*, however this may be due to background contamination from cell-free mRNA released from damaged cells (Young & Behjati 2018). The fact that the majority of profiled VSMCs expressed these genes at high levels would make the presence of cell-free *Myh11* and *Acta2* transcripts particularly likely. A method for removing ambient mRNA from droplet-based scRNA-seq datasets has been developed (Young & Behjati 2018), however its requirement for prior specification of cluster markers could result in incorrect removal of contractile VSMC markers from the phenotypically switching populations. Interestingly, VSMC-derived macrophages were not captured in a recent study which characterised mouse atherosclerotic plaques using scRNA-seq, although low level lipid uptake was observed in plaque VSMCs (Wirka et al. 2019). Wirka and colleagues reported that *Lgals3* was the only marker typically used to study macrophages which was upregulated in phenotypically switching VSMCs. This is in agreement

with our data, where I observed higher *Lgals3* expression in non-macrophage phenotypically switching VSMCs in atherosclerosis and following injury. Overall, it is difficult to draw conclusions about the origins of macrophage-like cells observed in this chapter. Further studies will be required to reconcile the discrepancies in the literature on whether VSMCs transition to a macrophage-like state within the plaque, and if so, what is the extent of such a transition. For example, the list of significantly upregulated genes in the macrophage-like cluster 9 could be used to inform further immunostaining experiments in the plaque, where co-expression of the VSMC lineage label and previously untested cluster 9 markers could be investigated.

In our analysis, we included VSMCs isolated from atherosclerotic arteries of mice fed a high-fat diet for either 14 or 18 weeks. These time points represent an intermediate and advanced stage of atherosclerosis in the $ApoE^{-/-}$ mouse model (Nakashima et al. 1994). We did not observe large differences between the samples, however, with all of the clusters containing cells from both stages. During dissection we observed a large amount of heterogeneity in the sizes of atherosclerotic plaques within a single animal as well as between animals, which were fed a high fat diet for the same duration. This may explain the fact that significant differences were not present between VSMCs isolated from 14 or 18 week stage plaques. We did observe some differences in the proportions of VSMCs from each plaque stage present in certain clusters. However further replicates at each time point would be required to substantiate these observations.

Studies of the clonality of VSMCs within the atherosclerotic plaque have shown that a single VSMC is capable of giving rise to the full range of phenotypes VSMCs adopt within atherosclerotic plaques (Feil et al. 2014; Chappell et al. 2016; Jacobsen et al. 2017; Misra et al. 2018). One of these studies has investigated VSMC localisation within atherosclerotic plaques throughout their early development (Misra et al. 2018). Misra and colleagues observed that VSMCs first populated the plaque cap, in which they were highly proliferative. They suggested that a currently unknown progenitor expressing contractile VSMC markers may initially form the cap and later enter the plaque core where VSMCs would be further modulated towards other phenotypes. Wirka et al. (2019) however observed that *Lum*, which marked phenotypically switching VSMCs in their study, was expressed in VSMCs located in the

plaque cap, suggesting that VSMCs in the plaque cap may at least in part show characteristics of phenotypically modulated VSMCs.

In light of the observations of the clonality of atherosclerotic plaques, the presence of phenotypically switched VSMCs in the plaque and rare S+L+ VSMCs in healthy arteries, which share aspects of transcriptional signatures with phenotypically switched VSMCs, it is a compelling hypothesis that activated S+L+ cells undergo selective expansion in atherosclerosis and following injury. At present, there is however no direct evidence to evaluate this hypothesis. Further experimental validation with *Myh11* and *Sca1* dual lineage tracing is required to establish whether it is the healthy S+L+ VSMCs that expand after stimulus.

5.3.4 Alternative markers of *Ly6a/Sca1*-positive VSMCs

In order to extend the findings from this chapter to the clinically relevant setting of the human atherosclerotic plaque, alternative markers of SCA1-positive VSMCs are required, as there is no known human orthologue of SCA1. Knowing the wider transcriptional signatures of *Ly6a/Sca1*-positive VSMCs in healthy as well as atherosclerotic arteries may enable identification of analogous subpopulations of VSMCs in human in transcriptome-wide studies. Transcriptional signatures of phenotypically switching VSMCs in human plaques have recently been established and have been broadly matched to mouse plaque VSMCs (Wirka et al. 2019). *Ly6a/Sca1* was expressed in a subset of phenotypically switching VSMCs both in our dataset as well as in the recent study by Wirka and colleagues (2019), where it appeared to show less advanced signatures of phenotypic modulation to VSMCs expressing calcifying signatures. *Ly6a/Sca1*-expressing VSMCs may therefore represent an intermediate state along phenotypic modulation of VSMCs. Identifying specific alternative markers of *Ly6a/Sca1*-positive VSMCs within the plaque may therefore be valuable to study the heterogeneity of phenotypically switching VSMCs in human. In this chapter I have identified *Fbln2* and *Slco2a1* as putative alternative markers of *Ly6a/Sca1*-positive VSMCs. These genes were expressed in equivalent subpopulations of VSMCs as *Ly6a/Sca1* following in vitro culture, carotid ligation injury as well as in atherosclerosis. Experimental validation of co-expression with SCA1 at the protein level is ongoing in collaboration with A.L. Taylor.

5.3.5 Conclusion

Overall in this chapter I presented evidence for the existence of *Ly6a/Sca1*-positive VSMCs following induction of phenotypic switching in three separate model systems. I observed a higher proportion of *Ly6a/Sca1*-positive VSMCs in all three models of phenotypic switching compared with healthy arteries. Phenotypically switching VSMCs showed similar transcriptional profiles to healthy S+L+ VSMCs, which expressed high levels of the cVSMCneg signature. These findings, as well as the observation that *Ly6a/Sca1*-positive VSMCs in atherosclerosis appear to represent an intermediate state of phenotypic switching, are in support of the hypothesis that *Ly6a/Sca1*-positive VSMCs may be responsible for clonal expansion of VSMCs in atherosclerosis and following injury. However, further validation using dual Sca1 and Myh11 lineage tracing systems is required to test this hypothesis experimentally.

Recent work by Wirka and colleagues (2019) has confirmed our observation of *Ly6a/Sca1*-positive VSMCs within the atherosclerotic plaque and additionally identified a population of phenotypically modulated VSMCs in human atherosclerotic plaques. To overcome the lack of human SCA1 orthologue, I have identified two alternative markers of *Ly6a/Sca1*-positive VSMCs, which may aid in further understanding of VSMC heterogeneity in human plaques. Despite the challenges in translating the findings from this chapter to human VSMCs due to lack of a SCA1 orthologue, the use of mouse models has been invaluable in this study. Genetic VSMC-specific lineage tracing systems, which could not have been used in human, have enabled us to identify the transcriptional signatures of phenotypically switching VSMCs, which have lost the expression of VSMC marker genes.

6 General Discussion

In this thesis I explored the heterogeneity of VSMCs and their transcriptional profiles in disease. I found that VSMCs located in the atherosclerosis-prone AA and atherosclerosis-resistant DT regions have distinct transcriptional signatures at the single-cell level and additionally identified the heterogeneous expression of several disease-relevant genes within both vascular beds. I also identified and characterised a rare subset of VSMCs in healthy arteries which expressed the progenitor marker SCA1 and showed that these cells express transcriptional signatures characteristic of the synthetic state of VSMCs. Examination of the VSMC response to *in vitro* culture and injury suggested that SCA1 upregulation marks the process of VSMC phenotypic switching. Furthermore, the transcriptional profiles of *Ly6a/Sca1*-positive VSMCs in atherosclerotic plaques were similar to those of SCA1-positive VSMCs detected in healthy arteries. I therefore hypothesise that SCA1-positive VSMCs may be the rare subset of cells which undergoes clonal expansion of VSMCs in atherosclerosis, which would have clinical implications for specific targeting of clonally expanding VSMCs and earlier detection of the disease in the future.

VSMCs were previously proposed to exist on a spectrum of contractile and synthetic states (reviewed in Rensen et al. 2007). In line with this hypothesis, I observed diverse VSMC transcriptional profiles both in healthy arteries and after response to stimulus. In healthy arteries the phenotypic spectrum was heavily biased towards the contractile state, with the vast majority of VSMCs expressing the contractile transcriptional signature. S+L+ VSMCs, however, showed progressively higher expression levels of transcriptional signatures associated with the synthetic state. Following stimulus in the form of vascular injury or induction of atherosclerosis, the proportion of VSMCs expressing signatures characteristic of phenotypic switching increased, even though the majority of VSMCs in profiled arteries retained the expression of contractile VSMC genes. Previous studies have described distinct subpopulations of VSMC-derived cells within atherosclerotic plaques (Feil et al. 2014; Shankman et al. 2015; Bennett et al. 2016; Durham et al. 2018). However, our data suggested that the transcriptional signatures of *Ly6a/Sca1*-positive VSMCs and calcifying-like VSMCs share a degree of similarity. A very recent study by Wirka et al. (2019) observed a continuous axis of phenotypic modulation of VSMCs within the plaque, which suggests that there may be

a continuity in the transcriptional profiles of phenotypically modulated VSMCs within the plaque rather than distinct subpopulations of cells.

The findings presented in this thesis have shed light on the transcriptional signatures of VSMCs located within the atherosclerotic plaques. Atherosclerosis is a complex disease and many different cell types are involved in plaque formation. scRNA-seq approaches are particularly well suited to the exploration of cellular heterogeneity within a complex environment, where population-level analysis would obscure the diversity of cells involved. The power of scRNA-seq for studying the atherosclerotic plaque is illustrated by recent studies of macrophages, which have revealed heterogeneity among this population in the plaque (Winkels et al. 2018; Cochain et al. 2018). Characterisation of the transcriptional profiles of identified macrophage subpopulations also shed light on the putative functions of different macrophage subpopulations identified within the plaque (Winkels et al. 2018; Cochain et al. 2018). Increased understanding of VSMC and macrophage heterogeneity in atherosclerosis highlights that the composition of atherosclerotic plaques may be even more complex than previously thought, with not only several cell types involved but there also being significant heterogeneity within each cell type. Further studies will be required to understand the role of different subpopulations of VSMCs in the plaque and to understand how different subpopulations of cells interact with one another. If it is confirmed that the diversity of VSMCs in the plaque is generated from SCA1-positive VSMCs, it will be important to understand how potential elimination of plaque-accumulating VSMCs would impact disease progression, given that VSMCs are currently thought to possess both protective and destabilising properties within the plaque (Bennett et al. 2016; Basatemur et al. 2019).

Our study used dissociated tissue for analysis, however recent advances in spatial transcriptomics have made high-resolution measurements possible within intact tissue samples at transcriptome-wide or near transcriptome-wide coverage (Eng et al. 2019; Vickovic et al. 2019). The spatial architecture of the atherosclerotic plaque is disrupted during dissociation of tissue to a single-cell suspension, which leads to loss of spatial positioning information for individual cells. Availability of spatial information from the atherosclerotic plaques would enable a detailed characterisation of plaque architecture and characterisation of any subpopulations of cells, which may localise or co-localise in specific areas of the plaque. Additionally, applying spatial transcriptomics to healthy arteries may shed light on whether

VSMCs showing signs of phenotypic switching in healthy arteries localise in areas of increased atherosclerotic risk, such as near branch sites. Another advantage of using spatial transcriptomics as a complementary approach to scRNA-seq would be that any artefacts induced during tissue dissociation could be eliminated. A recent study observed that dissociation of tissue could induce a stress response in cells, which included upregulation of several genes identified as heterogeneously expressed in several analyses in this thesis, such as *Egr1*, *Nfkbia*, *Fos* and *Atf3* (O'flanagan et al. 2019). O'flanagan et al. (2019) further observed that the precise effects of dissociation-induced stress response on transcriptional profiles varied between different cell types. Since VSMCs were not included in the analysis there is still the need to understand the specific effects that the tissue dissociation approach used in our study had on the transcriptomes of profiled VSMCs. However, I cannot rule out the possibility that some of the observed heterogeneity may have originated from a dissociation-induced stress response. Any dissociation-induced gene expression changes may also impact future investigations of early transcriptional changes during VSMC phenotypic switching using scRNA-seq. Dissociation-induced transcriptional changes would likely be similar to early changes induced by *in vitro* culture, which is a model system for VSMC phenotypic switching. Spatial transcriptomics approaches, which would enable investigation of gene expression directly in primary tissues, would therefore provide a valuable alternative in such studies.

It is currently not possible to determine whether healthy VSMCs showing signs of phenotypic switching in healthy arteries are primed for response to stimulus and subsequently expand, or whether phenotypic switching is a homeostatic process, which happens to a lesser extent in healthy than stimulated arteries and does not impact on clonal expansion. To validate whether SCA1-positive VSMCs in healthy arteries expand in atherosclerosis and after vascular injury, dual VSMC/Sca1 lineage tracing systems would be required. A tamoxifen-inducible Sca1-Cre tracing approach for tracking SCA1-positive cells has been developed (Vagnozzi et al. 2018). However, since the Cre approaches used in Myh11-Cre and Sca1-Cre lineage tracing systems are not independent, it would not be possible to use a combined Sca1-Cre and Myh11-Cre tracing system for such validation. This is because the induction of Cre expression under the control of either the *Myh11* or *Sca1* promoter would in both cases trigger the recombination at a Cre-reporter allele. A potential way around this problem would be to combine the Sca1-Cre (Vagnozzi et al. 2018) and Myh11-Dre (available from Shanghai Model Organisms) mouse lines, where the Cre-*loxp* and Dre-*rox* recombination systems would act

independently upon tamoxifen induction (Anastassiadis et al. 2009). This approach would require characterisation of the specificity and efficiency of the Myh11-Dre system and subsequent establishment of a dual Myh11-Dre/Sca1-Cre mouse line. Such a model could be used in conjunction with the Ai66 reporter, which requires recombination at both the *rox* and *loxp* sites to induce expression of the fluorescent reporter (Rosa26-*rox*-stop-*rox*-*loxp*-stop-*loxp*-tdTomato) (Madisen et al. 2015).

In order to avoid the extensive crossing required to establish the dual lineage tracing system, an alternative approach could be taken to substantiate the hypothesis that SCA1-positive VSMCs may expand after stimulus. Independent animals traced through the Myh11-Cre or Sca1-Cre systems coupled with fluorescent reporters could be used in conjunction with the carotid ligation model, which would not require the establishment of a Sca1-Cre/ApoE$^{-/-}$ model. This experiment would involve scRNA-seq profiling of four samples; the entire carotid arteries from both Myh11-Cre and Sca1-Cre mice serving as reference points, as well as isolated lineage-labelled cells from each of the Myh11-Cre and Sca1-Cre mice. Joint analysis of profiled cells could then reveal whether the phenotypically switching VSMC population is found within both the Myh11-traced and Sca1-traced samples. However, for unambiguous validation, dual lineage tracing experiments would likely be required in the future.

Myh11-Cre used throughout chapters 4 and 5 of this thesis was inserted into the Y chromosome (Wirth et al. 2008). A limitation of this work is therefore that only male mice were studied. However, there are some differences in the presentation and risk factors of atherosclerosis between men and women (Han et al. 2008; Yahagi et al. 2015) and inclusion of both sexes is important for reducing the sex bias of research findings and potential future clinical implications. Studies have shown that there are widespread gene expression differences between female and male mice (Yang et al. 2006) and that sex bias in biomedical research may result in poorer clinical outcomes for women (Beery & Zucker 2011). The tamoxifen-inducible Myh11-Dre mouse line mentioned above is not linked through the Y chromosome, and the use of this model would enable investigation of whether the findings in this thesis can be extended to female mice and to generate future findings by investigating both sexes. Due to the Y-linked Myh11-Cre, only male mice were also used in the recent study by Wirka et al. (2019), which confirmed our observations of a population of VSMCs in healthy arteries showing characteristics of phenotypic switching, as well as of *Ly6a/Sca1*-positive

VSMCs with similar transcriptional signatures within atherosclerotic plaques. Wirka and colleagues have extended their observations to human VSMCs by jointly analysing their transcriptomes with those of mouse VSMCs, which enabled the lineage-traced mouse VSMCs to act as a guide in interpreting human atherosclerotic cells. One of the four human atherosclerotic samples used by the authors was female, however they have not reported on whether the conclusions were supported by each individual sample (Wirka et al. 2019). Inclusion of more female samples and a specific comparison between male and female samples would have enabled an assessment of whether the findings applied to both sexes. A similar strategy of joint data analysis to that employed by Wirka et al (2019) could also be used to test the existence of *Ly6a/Sca1*-positive VSMCs in healthy as well as diseased arteries in unlabelled female mice in absence of the Myh11-Dre system in the future.

Advances in scRNA-seq and lineage tracing approaches have enabled investigations of cellular heterogeneity and plasticity in many different tissues. Investigations at the single-cell level are highlighting the extent of cellular heterogeneity within tissues. It is becoming apparent that numerous cell types, which were previously thought to be relatively homogeneous, are in fact heterogeneous and are composed of many different subtypes of cells (Zeisel et al. 2015; Villani et al. 2017; Chen et al. 2017; Papalexi & Satija 2018; Guo et al. 2019), which is consistent with observed heterogeneity in VSMCs. In addition, it is becoming clear that plasticity of resident differentiated cells in response to injury is common to many different tissues (Merrell & Stanger 2016). De-differentiation of mature cells under such circumstances is thought to have the function of tissue repair (Merrell & Stanger 2016). Transitioning of contractile VSMCs to the synthetic state can be seen as de-differentiation and the observation of VSMC-lineage cells expressing transcriptional signatures associated with synthetic VSMCs in healthy arteries suggests that this process may also occur during tissue homeostasis, albeit to a lesser extent.

Overall the findings presented in this thesis highlight that the heterogeneity of VSMCs at the single-cell level is a widespread phenomenon, both in healthy arteries as well as in disease. Additionally, the transcriptional signatures expressed by VSMCs appear to be dynamic both in healthy arteries and following stimulus, suggesting VSMC plasticity may be observed both during homeostasis and after injury or in disease. Future investigations will further the understanding of the role that the heterogeneity and plasticity of VSMCs play in cardiovascular disease.

References

Aaronson, P.I., Ward, J.P.T. & Connelly, M.J., 2012. *The Cardiovascular System at a Glance*, Wiley.

Adhikari, N. et al., 2015. Guidelines for the Isolation and Characterization of Murine Vascular Smooth Muscle Cells. A Report from the International Society of Cardiovascular Translational Research. *Journal of Cardiovascular Translational Research*, 8(3), pp.158–163.

Albarrán-Juárez, J. et al., 2016. Lineage tracing of cells involved in atherosclerosis. *Atherosclerosis*, 251, pp.445–453.

Allahverdian, S. et al., 2018. Smooth muscle cell fate and plasticity in atherosclerosis. *Cardiovascular Research*, 114(4), pp.540–550.

Anastassiadis, K. et al., 2009. Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice. *Disease models & mechanisms*, 2(9–10), pp.508–515.

Anders, S., Pyl, P.T. & Huber, W., 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), pp.166–169.

Andrews, T.S. & Hemberg, M., 2019. False signals induced by single-cell imputation. *F1000Research*, 7, p.1740.

Andrews, T.S. & Hemberg, M., 2019. M3Drop: dropout-based feature selection for scRNASeq. *Bioinformatics*, 35(16), pp.2865–2867.

Arnold, C. et al., 2014. RGS5 promotes arterial growth during arteriogenesis. *EMBO molecular medicine*, 6(8), pp.1075–1089.

Van Assche, T. et al., 2011. Transcription Profiles of Aortic Smooth Muscle Cells from Atherosclerosis-Prone and -Resistant Regions in Young Apolipoprotein E-Deficient Mice before Plaque Development. *Journal of Vascular Research*, 48(1), pp.31–42.

Bailey, B. et al., 2012. Sca-1 knockout impairs myocardial and cardiac progenitor cell function. *Circulation research*, 111(6), pp.750–760.

Basak, O. et al., 2014. Mapping early fate determination in Lgr5+ crypt stem cells using a novel Ki67-RFP allele. *The EMBO journal*, 33(18), pp.2057–2068.

Basatemur, G.L. et al., 2019. Vascular smooth muscle cells in atherosclerosis. *Nature Reviews Cardiology*, pp.1–18.

Becht, E. et al., 2019. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*, 37(1), pp.38–44.

Beery, A.K. & Zucker, I., 2011. Sex bias in neuroscience and biomedical research. *Neuroscience & Biobehavioral Reviews*, 35(3), pp.565–572.

Benditt, E.P. & Benditt, J.M., 1973. Evidence for a Monoclonal Origin of Human Atherosclerotic Plaques. *PNAS*, 70(6), pp.1753–1756.

Bennett, M.R., Sinha, S. & Owens, G.K., 2016. Vascular Smooth Muscle Cells in

Atherosclerosis. *Circulation Research*, 118(4), pp.692–702.

Bentzon, J.F. et al., 2006. Smooth muscle cells in atherosclerosis originate from the local vessel wall and not circulating progenitor cells in ApoE knockout mice. *Arteriosclerosis, thrombosis, and vascular biology*, 26(12), pp.2696–2702.

Bochaton-Piallat, M.L. et al., 1996. Phenotypic heterogeneity of rat arterial smooth muscle cell clones. Implications for the development of experimental intimal thickening. *Arteriosclerosis, thrombosis, and vascular biology*, 16(6), pp.815–820.

Bonyadi, M. et al., 2003. Mesenchymal progenitor self-renewal deficiency leads to age-dependent osteoporosis in Sca-1/Ly-6A null mice. *PNAS*, 100(10), pp.5840–5845.

Bradfute, S.B., Graubert, T.A. & Goodell, M.A., 2005. Roles of Sca-1 in hematopoietic stem/progenitor cell function. *Experimental Hematology*, 33(7), pp.836–843.

Brasier, A.R., 2010. The nuclear factor- B-interleukin-6 signalling pathway mediating vascular inflammation. *Cardiovascular Research*, 86(2), pp.211–218.

Brennecke, P. et al., 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nature Methods*, 10(11), pp.1093–1095.

Butcher, M.J. et al., 2011. Flow Cytometry Analysis of Immune Cells Within Murine Aortas. *Journal of Visualized Experiments*, (53), p.e2848.

Butler, A. et al., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, 36(5), pp.411–420.

Cai, Y. et al., 2015. Role of cAMP-Phosphodiesterase 1C Signaling in Regulating Growth Factor Receptor Stability, Vascular Smooth Muscle Cell Growth, Migration, and Neointimal Hyperplasia. *Circulation Research*, 116(7), pp.1120–1132.

Carmeliet, P., 2000. Mechanisms of angiogenesis and arteriogenesis. *Nature Medicine*, 6(4), pp.389–395.

Chakraborty, R. et al., 2019. Promoters to Study Vascular Smooth Muscle. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 39(4), pp.603–612.

Chamley-Campbell, J., Campbell, G.R. & Ross, R., 1979. The smooth muscle cell in culture. *Physiological reviews*, 59(1), pp.1–61.

Chappell, J. et al., 2016. Extensive Proliferation of a Subset of Differentiated, yet Plastic, Medial Vascular Smooth Muscle Cells Contributes to Neointimal Formation in Mouse Injury and Atherosclerosis Models. *Circulation research*, 119(12), pp.1313–1323.

Chen, Q. et al., 2016. Endothelial cells are progenitors of cardiac pericytes and vascular smooth muscle cells. *Nature Communications*, 7(1), p.12422.

Chen, R. et al., 2017. Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. *Cell Reports*, 18(13), pp.3227–3241.

Cheung, C. et al., 2012. Generation of human vascular smooth muscle subtypes provides insight into embryological origin–dependent disease susceptibility. *Nature Biotechnology*, 30(2), pp.165–173.

Chiu, J.-J. & Chien, S., 2011. Effects of disturbed flow on vascular endothelium:

pathophysiological basis and clinical perspectives. *Physiological reviews*, 91(1), pp.327–387.

Cochain, C. et al., 2018. Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis. *Circulation Research*, 122(12), pp.1661–1674.

Coifman, R.R. et al., 2005. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *PNAS*, 102(21), pp.7426–7431.

Cordes, K.R. et al., 2009. miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature*, 460(7256), pp.705–710.

Coward, K. & Wells, D., 2013. *Textbook of Clinical Embryology*, Cambridge University Press.

Crow, M. et al., 2018. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nature Communications*, 9, p.884.

Cunningham, K.S. & Gotlieb, A.I., 2005. The role of shear stress in the pathogenesis of atherosclerosis. *Laboratory Investigation*, 85(1), pp.9–23.

Davis-Dusenbery, B.N. et al., 2011. Micromanaging Vascular Smooth Muscle Cell Differentiation and Phenotypic Modulation. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 31(11), pp.2370–2377.

DeBakey, M.E. & Glaeser, D.H., 2000. Patterns of atherosclerosis: effect of risk factors on recurrence and survival—analysis of 11,890 cases with more than 25-year follow-up. *The American Journal of Cardiology*, 85(9), pp.1045–1053.

van Dijk, D. et al., 2018. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell*, 174(3), p.716–729.e27.

Dingemans, K.P. et al., 2000. Extracellular matrix of the human aortic media: An ultrastructural histochemical and immunohistochemical study of the adult aortic media. *The Anatomical Record*, 258(1), pp.1–14.

Dobin, A. et al., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), pp.15–21.

Dobnikar, L. et al., 2018. Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. *Nature Communications*, 9(1), p.4567.

Duò, A., Robinson, M.D. & Soneson, C., 2018. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Research*, 7, p.1141.

Durham, A.L. et al., 2018. Role of smooth muscle cells in vascular calcification: implications in atherosclerosis and arterial stiffness. *Cardiovascular Research*, 114(4), pp.590–600.

Eling, N. et al., 2018. Correcting the Mean-Variance Dependency for Differential Variability Testing Using Single-Cell RNA Sequencing Data. *Cell Systems*, 7(3), p.284–294.e12.

Eng, C.-H.L. et al., 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, 568(7751), pp.235–239.

Evanko, S.P., Angello, J.C. & Wight, T.N., 1999. Formation of hyaluronan- and versican-rich pericellular matrix is required for proliferation and migration of vascular smooth muscle

cells. *Arteriosclerosis, thrombosis, and vascular biology*, 19(4), pp.1004–1013.

Evrard, S.M. et al., 2016. Endothelial to mesenchymal transition is common in atherosclerotic lesions and is associated with plaque instability. *Nature Communications*, 7(1), p.11853.

Fang, S. et al., 2012. Generation of Functional Blood Vessels from a Single c-kit+ Adult Vascular Endothelial Stem Cell. *PLoS Biology*, 10(10), p.e1001407.

Feil, R. et al., 1996. Ligand-activated site-specific recombination in mice. *PNAS*, 93(20), pp.10887–10890.

Feil, R. et al., 1997. Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains. *Biochemical and Biophysical Research Communications*, 237(3), pp.752–757.

Feil, S. et al., 2014. Transdifferentiation of Vascular Smooth Muscle Cells to Macrophage-Like Cells During Atherogenesis. *Circulation Research*, 115(7), pp.662–667.

Ferguson, J.E., Kelley, R.W. & Patterson, C., 2005. Mechanisms of endothelial differentiation in embryonic vasculogenesis. *Arteriosclerosis, thrombosis, and vascular biology*, 25(11), pp.2246–2254.

Fernandes, K.J.L. et al., 2004. A dermal niche for multipotent adult skin-derived precursor cells. *Nature Cell Biology*, 6(11), pp.1082–1093.

Finak, G. et al., 2015. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology*, 16(1), p.278.

Frid, M.G. et al., 1997. Smooth muscle cells isolated from discrete compartments of the mature vascular media exhibit unique phenotypes and distinct growth capabilities. *Circulation research*, 81(6), pp.940–52.

Frid, M.G., Moiseeva, E.P. & Stenmark, K.R., 1994. Multiple phenotypically distinct smooth muscle cell populations exist in the adult and developing bovine pulmonary arterial media in vivo. *Circulation research*, 75(4), pp.669–681.

Fu, G.K. et al., 2011. Counting individual DNA molecules by the stochastic attachment of diverse labels. *PNAS*, 108(22), pp.9026–9031.

Gadson, P.F. et al., 1997. Differential Response of Mesoderm- and Neural Crest-Derived Smooth Muscle to TGF-β1: Regulation of c-myb and α1 (I) Procollagen Genes. *Experimental Cell Research*, 230(2), pp.169–180.

Galley, H.F. & Webster, N.R., 2004. Physiology of the endothelium. *British Journal of Anaesthesia*, 93(1), pp.105–113.

Gardeux, V. et al., 2017. ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data. *Bioinformatics*, 33(19), pp.3123–3125.

Getz, G.S. & Reardon, C.A., 2012. Animal models of atherosclerosis. *Arteriosclerosis, thrombosis, and vascular biology*, 32(5), pp.1104–1115.

Gimbrone, M.A. & García-Cardeña, G., 2016. Endothelial Cell Dysfunction and the Pathobiology of Atherosclerosis. *Circulation Research*, 118(4), pp.620–636.

Gomez, D. et al., 2013. Detection of histone modifications at specific gene loci in single cells in histological sections. *Nature Methods*, 10(2), pp.171–177.

Gomez, D., Swiatlowska, P. & Owens, G.K., 2015. Epigenetic Control of Smooth Muscle Cell Identity and Lineage Memory. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 35(12), pp.2508–2516.

Greif, D.M. et al., 2012. Radial construction of an arterial wall. *Developmental cell*, 23(3), pp.482–493.

Grün, D. et al., 2015. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature*, 525(7568), pp.251–255.

Gunaje, J.J. et al., 2011. PDGF-dependent regulation of regulator of G protein signaling-5 expression and vascular smooth muscle cell functionality. *American Journal of Physiology - Cell Physiology*, 301(2), pp.C478–C489.

Guo, M. et al., 2019. Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. *Nature Communications*, 10(1), p.37.

Haghverdi, L. et al., 2018. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, 36(5), pp.421–427.

Haghverdi, L. et al., 2016. Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods*, 13(10), pp.845–848.

Haghverdi, L., Buettner, F. & Theis, F.J., 2015. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*, 31(18), pp.2989–2998.

Haimovici, H. & Maier, N., 1971. Experimental canine atherosclerosis in autogenous abdominal aortic grafts implanted into the jugular vein. *Atherosclerosis*, 13(3), pp.375–384.

Han, S.H. et al., 2008. Sex differences in atheroma burden and endothelial function in patients with early coronary atherosclerosis. *European Heart Journal*, 29(11), pp.1359–1369.

Hao, H. et al., 2002. Heterogeneity of smooth muscle cell populations cultured from pig coronary artery. *Arteriosclerosis, thrombosis, and vascular biology*, 22(7), pp.1093–1099.

Hashimshony, T. et al., 2012. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell reports*, 2(3), pp.666–673.

Hashimshony, T. et al., 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biology*, 17(1), p.77.

Hedin, U. et al., 1988. Diverse effects of fibronectin and laminin on phenotypic properties of cultured arterial smooth muscle cells. *The Journal of Cell Biology*, 107(1), pp.307–319.

Hellstrom, M. et al., 1999. Role of PDGF-B and PDGFR-beta in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse. *Development*, 126, pp.3047–3055.

Herring, B. et al., 2014. Previously differentiated medial vascular smooth muscle cells contribute to neointima formation following vascular injury. *Vascular Cell*, 6(1), p.21.

Herring, B.P. et al., 2017. Inflammation and vascular smooth muscle cell dedifferentiation following carotid artery ligation. *Physiological Genomics*, 49(3), pp.115–126.

Holmes, C. & Stanford, W.L., 2007. Concise Review: Stem Cell Antigen-1: Expression, Function, and Enigma. *Stem Cells*, 25, pp.1339–1347.

Hu, Y. et al., 2004. Abundant progenitor cells in the adventitia contribute to atherosclerosis of vein grafts in ApoE-deficient mice. *Journal of Clinical Investigation*, 113(9), pp.1258–1265.

Hug, H. & Schuler, R., 2003. Measurement of the Number of Molecules of a Single mRNA Species in a Complex mRNA Preparation. *Journal of Theoretical Biology*, 221(4), pp.615–624.

Hyvärinen, A. & Oja, E., 2000. Independent component analysis: algorithms and applications. *Neural networks*, 13(4–5), pp.411–430.

Ignarro, L.J. et al., 1987. Endothelium-derived relaxing factor produced and released from artery and vein is nitric oxide. *PNAS*, 84(24), pp.9265–9269.

Ishibashi, S. et al., 1993. Hypercholesterolemia in low density lipoprotein receptor knockout mice and its reversal by adenovirus-mediated gene delivery. *Journal of Clinical Investigation*, 92(2), pp.883–893.

Islam, S. et al., 2011. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research*, 21(7), pp.1160–1167.

Islam, S. et al., 2014. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods*, 11(2), pp.163–166.

Ito, C.Y. et al., 2003. Hematopoietic stem cell and progenitor defects in Sca-1/Ly-6A-null mice. *Blood*, 101(2), pp.517–523.

Iwata, H. et al., 2010. Bone Marrow-Derived Cells Contribute to Vascular Inflammation but Do Not Differentiate Into Smooth Muscle Cell Lineages. *Circulation*, 122(20), pp.2048–2057.

Jacobsen, K. et al., 2017. Diverse cellular architecture of atherosclerotic plaque derives from clonal expansion of a few medial SMCs. *JCI Insight*, 2(19), p.e95890.

Jaitin, D.A. et al., 2014. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science*, 343(6172), pp.776–779.

Ji, Z. & Ji, H., 2016. TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Research*, 44(13), p.e117.

Jiang, X. et al., 2000. Fate of the mammalian cardiac neural crest. *Development*, 127(8), pp.1607–1616.

Juríková, M. et al., 2016. Ki67, PCNA, and MCM proteins: Markers of proliferation in the diagnosis of breast cancer. *Acta Histochemica*, 118(5), pp.544–552.

Kaur, H. et al., 2017. Single-cell profiling reveals heterogeneity and functional patterning of GPCR expression in the vascular system. *Nature Communications*, 8, p.15700.

Kharchenko, P. V, Silberstein, L. & Scadden, D.T., 2014. Bayesian approach to single-cell

differential expression analysis. *Nature methods*, 11(7), pp.740–742.

Kiselev, V.Y. et al., 2017. SC3: consensus clustering of single-cell RNA-seq data. *Nature Methods*, 14(5), pp.483–486.

Kiselev, V.Y., Andrews, T.S. & Hemberg, M., 2019. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nature Reviews Genetics*, 20(5), pp.273–282.

Kiselev, V.Y., Yiu, A. & Hemberg, M., 2018. scmap: projection of single-cell RNA-seq data across data sets. *Nature Methods*, 15, pp.359–362.

Klein, A.M. et al., 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 161(5), pp.1187–1201.

Klein, D. et al., 2011. Vascular wall-resident CD44+ multipotent stem cells give rise to pericytes and smooth muscle cells and contribute to new vessel maturation. *PLoS ONE*, 6(5), p.e20540.

Kolodziejczyk, A.A. et al., 2015. The technology and biology of single-cell RNA sequencing. *Molecular cell*, 58(4), pp.610–620.

Kramann, R. et al., 2016. Adventitial MSC-like Cells Are Progenitors of Vascular Smooth Muscle Cells and Drive Vascular Calcification in Chronic Kidney Disease. *Cell Stem Cell*, 19(5), pp.628–642.

Kumar, A. & Lindner, V., 1997. Remodeling With Neointima Formation in the Mouse Carotid Artery After Cessation of Blood Flow. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 17(10), pp.2238–2244.

Langfelder, P. & Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1), p.559.

Leroux-Berger, M. et al., 2011. Pathologic calcification of adult vascular smooth muscle cells differs on their crest or mesodermal embryonic origin. *Journal of Bone and Mineral Research*, 26(7), pp.1543–1553.

Lesauskaite, V. et al., 2001. Smooth muscle cells of the media in the dilatative pathology of ascending thoracic aorta: Morphology, immunoreactivity for osteopontin, matrix metalloproteinases, and their inhibitors. *Human Pathology*, 32(9), pp.1003–1011.

Levine, J.H. et al., 2015. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell*, 162(1), pp.184–197.

Li, B. et al., 2010. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, 26(4), pp.493–500.

Li, S. et al., 2001. Innate Diversity of Adult Human Arterial Smooth Muscle Cells. *Circulation Research*, 89(6), pp.517–525.

Li, W.V. & Li, J.J., 2019. A statistical simulator scDesign for rational scRNA-seq experimental design. *Bioinformatics*, 35(14), pp.i41–i50.

Li, W.V. & Li, J.J., 2018. An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nature Communications*, 9(1), p.997.

Libby, P., Ridker, P.M. & Hansson, G.K., 2011. Progress and challenges in translating the

biology of atherosclerosis. *Nature*, 473(7347), pp.317–325.

Licht, C., 2010. *New methods for generating significance levels from multiply-imputed data*. Doctoral thesis, University of Bamberg.

Le Lièvre, C.S. & Le Douarin, N.M., 1975. Mesenchymal derivatives of the neural crest: analysis of chimaeric quail and chick embryos. *Development*, 34(1), pp.125–154.

Lin, P., Troup, M. & Ho, J.W.K., 2017. CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biology*, 18(1), p.59.

Lindner, V., Fingerle, J. & Reidy, M.A., 1993. Mouse model of arterial injury. *Circulation research*, 73(5), pp.792–796.

Liu, Y. et al., 2005. Kruppel-like factor 4 abrogates myocardin-induced activation of smooth muscle gene expression. *The Journal of biological chemistry*, 280(10), pp.9719–9727.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), pp.550–571.

Lun, A.T.L. et al., 2016. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biology*, 17(1), p.75.

Lun, A.T.L., McCarthy, D.J. & Marioni, J.C., 2016. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research*, 5, p.2122.

Ma, X. et al., 2002. The Ly-6A (Sca-1) GFP Transgene is Expressed in all Adult Mouse Hematopoietic Stem Cells. *Stem Cells*, 20(6), pp.514–521.

Maaten, L. Van Der & Hinton, G., 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9, pp.2579–2605.

Macosko, E.Z. et al., 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, 161(5), pp.1202–1214.

Madisen, L. et al., 2015. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron*, 85(5), p.942.

Maegdefessel, L., Rayner, K.J. & Leeper, N.J., 2015. MicroRNA Regulation of Vascular Smooth Muscle Function and Phenotype. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 35(1), pp.2–6.

Maiellaro, K. & Taylor, W., 2007. The role of the adventitia in vascular inflammation. *Cardiovascular Research*, 75(4), pp.640–648.

Majesky, M.W., 2007. Developmental Basis of Vascular Smooth Muscle Diversity. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 27(6), pp.1248–1258.

Majesky, M.W. et al., 2017. Differentiated Smooth Muscle Cells Generate a Subpopulation of Resident Vascular Progenitor Cells in the Adventitia Regulated by Klf4. *Circulation Research*, 120(2), pp.296–311.

Matsuura, K. et al., 2004. Adult cardiac Sca-1-positive cells differentiate into beating cardiomyocytes. *The Journal of biological chemistry*, 279(12), pp.11384–11391.

McCarthy, D.J. et al., 2017. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*, 33(8), pp.1179–1186.

McDonald, O.G. et al., 2006. Control of SRF binding to CArG box chromatin regulates smooth muscle gene expression in vivo. *Journal of Clinical Investigation*, 116(1), pp.36–48.

McInnes, L., Healy, J. & Melville, J., 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 (preprint)*.

Merrell, A.J. & Stanger, B.Z., 2016. Adult cell plasticity in vivo: de-differentiation and transdifferentiation are back in style. *Nature Reviews Molecular Cell Biology*, 17(7), pp.413–425.

Mi, H. et al., 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, 47(D1), pp.D419–D426.

Miano, J.M. et al., 1994. Smooth muscle myosin heavy chain exclusively marks the smooth muscle lineage during mouse embryogenesis. *Circulation research*, 75(5), pp.803–12.

Misra, A. et al., 2018. Integrin beta3 regulates clonality and fate of smooth muscle-derived atherosclerotic plaque cells. *Nature Communications*, 9(1), p.2073.

Moiseeva, E., 2001. Adhesion receptors of vascular smooth muscle cells and their functions. *Cardiovascular Research*, 52(3), pp.372–386.

Monahan-Earley, R., Dvorak, A.M. & Aird, W.C., 2013. Evolutionary origins of the blood vascular system and endothelium. *Journal of thrombosis and haemostasis*, 11(Suppl 1), pp.46–66.

Mortazavi, A. et al., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7), pp.621–628.

Murry, C.E. et al., 1997. Monoclonality of smooth muscle cells in human atherosclerosis. *The American journal of pathology*, 151(3), pp.697–705.

Naik, V. et al., 2012. Sources of cells that contribute to atherosclerotic intimal calcification: an in vivo genetic fate mapping study. *Cardiovascular Research*, 94(3), pp.545–554.

Naito, H. et al., 2012. Identification and characterization of a resident vascular stem/progenitor cell population in preexisting blood vessels. *The EMBO Journal*, 31(4), pp.842–855.

Nakamura, T., Colbert, M.C. & Robbins, J., 2006. Neural crest cells retain multipotential characteristics in the developing valves and label the cardiac conduction system. *Circulation research*, 98(12), pp.1547–54.

Nakashima, Y. et al., 1994. ApoE-deficient mice develop lesions of all phases of atherosclerosis throughout the arterial tree. *Arteriosclerosis and Thrombosis: A Journal of Vascular Biology*, 14(1), pp.133–140.

Nemenoff, R.A. et al., 2011. SDF-1α induction in mature smooth muscle cells by inactivation of PTEN is a critical mediator of exacerbated injury-induced neointima formation. *Arteriosclerosis, thrombosis, and vascular biology*, 31(6), pp.1300–1308.

Nguyen, A.T. et al., 2013. Smooth muscle cell plasticity: fact or fiction? *Circulation research*, 112(1), pp.17–22.

Nguyen, N., Naik, V. & Speer, M.Y., 2013. Diabetes mellitus accelerates cartilaginous

metaplasia and calcification in atherosclerotic vessels of LDLr mutant mice. *Cardiovascular Pathology*, 22(2), pp.167–175.

O'flanagan, C.H. et al., 2019. Dissociation of solid tumour tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *bioRxiv (preprint)*.

Oh, H. et al., 2003. Cardiac progenitor cells from adult myocardium: Homing, differentiation, and fusion after infarction. *PNAS*, 100(21), pp.12313–12318.

Orr, A.W. et al., 2010. Complex regulation and function of the inflammatory smooth muscle cell phenotype in atherosclerosis. *Journal of vascular research*, 47(2), pp.168–180.

Owens, A.P. et al., 2010. Angiotensin II Induces a Region-Specific Hyperplasia of the Ascending Aorta Through Regulation of Inhibitor of Differentiation 3. *Circulation Research*, 106(3), pp.611–619.

Owens, G.K., 1995. Regulation of differentiation of vascular smooth muscle cells. *Physiological Reviews*, 75(3), pp.487–517.

Owens, G.K., Kumar, M.S. & Wamhoff, B.R., 2004. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiological reviews*, 84(3), pp.767–801.

Palotie, A. et al., 1983. Components of subendothelial aorta basement membrane. Immunohistochemical localization and role in cell attachment. *Laboratory investigation*, 49(3), pp.362–370.

Papalexi, E. & Satija, R., 2018. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nature Reviews Immunology*, 18(1), pp.35–45.

Passman, J.N. et al., 2008. A sonic hedgehog signaling domain in the arterial adventitia supports resident Sca1+ smooth muscle progenitor cells. *PNAS*, 105(27), pp.9349–9354.

Patel, A.P. et al., 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, 344(6190), pp.1396–1401.

Paul, F. et al., 2015. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell*, 163(7), pp.1663–1677.

Pfaltzgraff, E.R. & Bader, D.M., 2015. Heterogeneity in vascular smooth muscle cell embryonic origin in relation to adult structure, physiology, and disease. *Developmental Dynamics*, 244(3), pp.410–416.

Picelli, S. et al., 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols*, 9(1), pp.171–181.

Picelli, S. et al., 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods*, 10(11), pp.1096–1098.

Plump, A.S. et al., 1992. Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous recombination in ES cells. *Cell*, 71(2), pp.343–353.

Pontén, A. et al., 2005. Platelet-derived growth factor D induces cardiac fibrosis and proliferation of vascular smooth muscle cells in heart-specific transgenic mice.

Circulation research, 97(10), pp.1036–1045.

Pouget, C. et al., 2006. Somite-derived cells replace ventral aortic hemangioblasts and provide aortic smooth muscle cells of the trunk. *Development*, 133(6), pp.1013–22.

Psaltis, P.J. & Simari, R.D., 2015. Vascular Wall Progenitor Cells in Health and Disease. *Circulation Research*, 116(8), pp.1392–1412.

Qiao, L. et al., 2014. Endothelial Fate Mapping in Mice With Pulmonary Hypertension. *Circulation*, 129(6), pp.692–703.

Raj, A. & van Oudenaarden, A., 2008. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(2), pp.216–226.

Rajendran, P. et al., 2013. The vascular endothelium and human diseases. *International journal of biological sciences*, 9(10), pp.1057–1069.

Ramsköld, D. et al., 2012. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*, 30(8), pp.777–782.

Regan, C.P. et al., 2000. Molecular mechanisms of decreased smooth muscle differentiation marker expression after vascular injury. *The Journal of clinical investigation*, 106(9), pp.1139–1147.

Regev, A. et al., 2017. The Human Cell Atlas. *eLife*, 6.

Rensen, S.S.M., Doevendans, P.A.F.M. & van Eys, G.J.J.M., 2007. Regulation and characteristics of vascular smooth muscle cell phenotypic diversity. *Netherlands heart journal*, 15(3), pp.100–108.

Risso, D. et al., 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, 32(9), pp.896–902.

Ritchie, M.E. et al., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), p.e47.

Rodrigues, R.M. et al., 2014. Human Skin-Derived Stem Cells as a Novel Cell Source for In Vitro Hepatotoxicity Screening of Pharmaceuticals. *Stem Cells and Development*, 23(1), pp.44–55.

Rong, J.X. et al., 2003. Transdifferentiation of mouse aortic smooth muscle cells to a macrophage-like state after cholesterol loading. *PNAS*, 100(23), pp.13531–13536.

Roostalu, U. et al., 2018. Distinct Cellular Mechanisms Underlie Smooth Muscle Turnover in Vascular Development and Repair. *Circulation Research*, 122(2), pp.267–281.

Ross, R. & Glomset, J.A., 1973. Atherosclerosis and the Arterial Smooth Muscle Cell. *Science*, 180(4093), pp.1332–1339.

Sainz, J. et al., 2006. Isolation of Side Population Progenitor Cells From Healthy Arteries of Adult Mice. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 26(2), pp.281–286.

Sata, M. et al., 2002. Hematopoietic stem cells differentiate into vascular cells that participate in the pathogenesis of atherosclerosis. *Nat Med*, 8(4), pp.403–409.

Schmidt, A., Brixius, K. & Bloch, W., 2007. Endothelial precursor cell migration during vasculogenesis. *Circulation research*, 101(2), pp.125–36.

Schwartz, S.M. & Murry, C.E., 1998. Proliferation and the Monoclonal Origins of Atherosclerotic Lesions. *Annual Review of Medicine*, 49(1), pp.437–460.

Shankman, L.S. et al., 2015. KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nature Medicine*, 21(6), pp.628–637.

Shannon, P. et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), pp.2498–504.

Sinha, S., Iyer, D. & Granata, A., 2014. Embryonic origins of human vascular smooth muscle cells: implications for in vitro modeling and clinical application. *Cellular and Molecular Life Sciences*, 71(12), pp.2271–2288.

Soneson, C. & Robinson, M.D., 2018. Bias, robustness and scalability in single-cell differential expression analysis. *Nature Methods*, 15(4), pp.255–261.

Spangrude, G.J., Heimfeld, S. & Weissman, I.L., 1988. Purification and characterization of mouse hematopoietic stem cells. *Science*, 241(4861), pp.58–62.

Stanford, W.L. et al., 1997. Altered proliferative response by T lymphocytes of Ly-6A (Sca-1) null mice. *The Journal of experimental medicine*, 186(5), pp.705–17.

Stegemann, J.P., Hong, H. & Nerem, R.M., 2005. Mechanical, biochemical, and extracellular matrix effects on vascular smooth muscle cell phenotype. *Journal of Applied Physiology*, 98(6), pp.2321–2327.

Stegle, O., Teichmann, S.A. & Marioni, J.C., 2015. Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews Genetics*, 16(3), pp.133–145.

Steinbach, S.K. et al., 2011. Directed differentiation of skin-derived precursors into functional vascular smooth muscle cells. *Arteriosclerosis, thrombosis, and vascular biology*, 31(12), pp.2938–2948.

Stenmark, K.R. et al., 2013. The adventitia: essential regulator of vascular wall structure and function. *Annual review of physiology*, 75, pp.23–47.

Street, K. et al., 2018. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19(1), p.477.

Svensson, V. et al., 2017. Power analysis of single-cell RNA-sequencing experiments. *Nature Methods*, 14(4), pp.381–387.

Svensson, V., Vento-Tormo, R. & Teichmann, S.A., 2018. Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols*, 13(4), pp.599–604.

Tang, F. et al., 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6(5), pp.377–382.

Tang, Z. et al., 2012. Differentiation of multipotent vascular stem cells contributes to vascular diseases. *Nature communications*, 3, p.875.

The Tabula Muris Consortium, 2018. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, 562(7727), pp.367–372.

Thyberg, J. & Hultgårdh-Nilsson, A., 1994. Fibronectin and the basement membrane

components laminin and collagen type IV influence the phenotypic properties of subcultured rat aortic smooth muscle cells differently. *Cell and tissue research*, 276(2), pp.263–71.

Toma, J.G. et al., 2001. Isolation of multipotent adult stem cells from the dermis of mammalian skin. *Nature Cell Biology*, 3(9), pp.778–784.

Topouzis, S. & Majesky, M.W., 1996. 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. *Developmental Biology*, 178(2), pp.430–445.

Tran, P.-K. et al., 2004. Increased Intimal Hyperplasia and Smooth Muscle Cell Proliferation in Transgenic Mice With Heparan Sulfate–Deficient Perlecan. *Circulation Research*, 94(4), pp.550–558.

Trapnell, C. et al., 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, 32(4), pp.381–386.

Trapnell, C. et al., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5), pp.511–515.

Trapnell, C., Pachter, L. & Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), pp.1105–11.

Trigueros-Motos, L. et al., 2013. Embryological-Origin-Dependent Differences in Homeobox Expression in Adult Aorta: Role in Regional Phenotypic Variability and Regulation of NF-kB Activity. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 33(6), pp.1248–1256.

Trivedi, C.M., Patel, R.C. & Patel, C. V., 2007. Homeobox gene HOXA9 inhibits nuclear factor-kappa B dependent activation of endothelium. *Atherosclerosis*, 195(2), pp.e50–e60.

Ubil, E. et al., 2014. Mesenchymal–endothelial transition contributes to cardiac neovascularization. *Nature*, 514(7524), pp.585–590.

Vagnozzi, R.J. et al., 2018. Genetic Lineage Tracing of Sca-1+ Cells Reveals Endothelial but Not Myogenic Contribution to the Murine Heart. *Circulation*, 138(25), pp.2931–2939.

Vallejos, C.A. et al., 2017. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nature Methods*, 14(6), pp.565–571.

Vallejos, C.A., Marioni, J.C. & Richardson, S., 2015. BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. *PLOS Computational Biology*, 11(6), p.e1004333.

Vengrenyuk, Y. et al., 2015. Cholesterol Loading Reprograms the MicroRNA-143/145–Myocardin Axis to Convert Aortic Smooth Muscle Cells to a Dysfunctional Macrophage-Like Phenotype. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 35(3), pp.535–546.

Vickovic, S. et al., 2019. High-density spatial transcriptomics arrays for in situ tissue profiling. *bioRxiv (preprint)*.

Viiri, L.E. et al., 2013. Smooth muscle cells in human atherosclerosis: Proteomic profiling reveals differences in expression of Annexin A1 and mitochondrial proteins in carotid disease. *Journal of Molecular and Cellular Cardiology*, 54, pp.65–72.

Villani, A.-C. et al., 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, 356(6335), p.eaah4573.

Wagner, A., Regev, A. & Yosef, N., 2016. Revealing the vectors of cellular identity with single-cell genomics. *Nature Biotechnology*, 34(11), pp.1145–1160.

Wang, G. et al., 2015. Origin and differentiation of vascular smooth muscle cells. *The Journal of Physiology*, 593(14), pp.3013–3030.

Wang, X. et al., 2006. The Role of the Sca-1+/CD31− Cardiac Progenitor Cell Population in Postinfarction Left Ventricular Remodeling. *Stem Cells*, 24(7), pp.1779–1788.

Wang, Z. et al., 2003. Myocardin is a master regulator of smooth muscle gene expression. *PNAS*, 100(12), pp.7129–7134.

Wasteson, P. et al., 2008. Developmental origin of smooth muscle cells in the descending aorta in mice. *Development*, 135(10), pp.1823–32.

Williams, J.K. & Heistad, D.D., 1996. Structure and function of vasa vasorum. *Trends in Cardiovascular Medicine*, 6(2), pp.53–57.

Winkels, H. et al., 2018. Atlas of the Immune Cell Repertoire in Mouse Atherosclerosis Defined by Single-Cell RNA-Sequencing and Mass Cytometry. *Circulation research*, 122(12), pp.1675–1688.

Wirka, R.C. et al., 2019. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nature Medicine*, 25(8), pp.1280–1289.

Wirth, A. et al., 2008. G12-G13–LARG–mediated signaling in vascular smooth muscle is required for salt-induced hypertension. *Nature Medicine*, 14(1), pp.64–68.

Wolf, F.A., Angerer, P. & Theis, F.J., 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1), p.15.

Wolinsky, H. & Glagov, S., 1967. A Lamellar Unit of Aortic Medial Structure and Function in Mammals. *Circulation Research*, 20(1), pp.99–111.

Yahagi, K. et al., 2015. Sex differences in coronary artery disease: Pathological observations. *Atherosclerosis*, 239(1), pp.260–267.

Yang, X. et al., 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome research*, 16(8), pp.995–1004.

Yao, Q. et al., 2014. Sonic hedgehog mediates a novel pathway of PDGF-BB–dependent vessel maturation. *Blood*, 123(15), pp.2429–2437.

Yoshida, T. et al., 2003. Myocardin Is a Key Regulator of CArG-Dependent Transcription of Multiple Smooth Muscle Marker Genes. *Circulation Research*, 92(8), pp.856–864.

Yoshida, T. et al., 2007. Platelet-derived growth factor-BB represses smooth muscle cell marker genes via changes in binding of MKL factors and histone deacetylases to their promoters. *American Journal of Physiology - Cell Physiology*, 292(2), pp.C886–C895.

Yoshida, T. et al., 2013. Smooth Muscle–Selective Inhibition of Nuclear Factor-κB Attenuates Smooth Muscle Phenotypic Switching and Neointima Formation Following Vascular

Injury. *Journal of the American Heart Association*, 2(3), p.e000230.

Yoshida, T., Kaestner, K.H. & Owens, G.K., 2008. Conditional Deletion of Krüppel-Like Factor 4 Delays Downregulation of Smooth Muscle Cell Differentiation Markers but Accelerates Neointimal Formation Following Vascular Injury. *Circulation Research*, 102(12), pp.1548–1557.

Young, M.D. & Behjati, S., 2018. SoupX removes ambient RNA contamination from droplet based single cell RNA sequencing data. *bioRxiv (preprint)*.

Yu, G. et al., 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*, 16(5), pp.284–287.

Zeisel, A. et al., 2015. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*, 347(6226), pp.1138–42.

Zhang, L. et al., 2018. Cardiac Sca-1+ Cells Are Not Intrinsic Stem Cells for Myocardial Development, Renewal, and Repair. *Circulation*, 138(25), pp.2919–2930.

Zhang, S. et al., 1992. Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. *Science* , 258(5081), pp.468–471.

Zhao, S. et al., 2014. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS ONE*, 9(1).

Zhao, Y., Vanhoutte, P.M. & Leung, S.W.S., 2015. Vascular nitric oxide: Beyond eNOS. *Journal of Pharmacological Sciences*, 129(2), pp.83–94.

Zheng, G.X.Y. et al., 2017. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8, p.14049.

Ziegenhain, C. et al., 2017. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular Cell*, 65(4), p.631–643.e4.