THE UNIVERSITY
of EDINBURGH

A comprehensive accounting of lncRNA dynamics within vascular smooth muscle cell pathological transitions

Matthew A. Bennett

Submitted for the degree of Doctor of Philosophy

University of Edinburgh 2021

Declaration

I declare that the work done in the composition of this thesis is entirely my own, except where stated in the text. This work has not been submitted for any other degree, and to the best of my knowledge contains no material published or written by any other person, except where stated in the text.

Matthew Bennett

September 2021

Acknowledgements

I am very grateful to my primary supervisory team of Dr. Julie Rodor and Professor Andrew H. Baker whose constant oversight, patience and direction allowed this thesis to come to fruition whilst pushing my own scientific and personal development. Their faith in my ability to succeed was instrumental to getting to this point, especially in what was a strange year to write a thesis.

Thanks also to the wider Baker group and the other PhD students for their support. Particularly to Vladislav Miscianov, Gregor Aitchison, Amira Mahmoud, Margaret Ballantyne and Jean Iyinikkel who handled the bulk of my laboratory training along with Gregor Aitchison and Kathryn Newton for constant - often behind the scenes - technical support. The CVS community as a whole provided an encouraging and inspiring home for me throughout my PhD. I thank Professors John Mullins, Matt Bailey and Tom Freeman for initially letting me in and to the British Heart Foundation for funding my time in Edinburgh.

Finally, thanks to my partner Sandra who first planted the seed of this whole endeavour.

Abstract

Vascular smooth muscle cells (VSMCs) provide vital contractile force within blood vessel walls, yet also propagate widespread cardiovascular pathologies with high mortality rates through pathological activities. The targeting of such phenotypes in VSMCs has been a commonly-touted strategy for decades yet we still have no viable option to implement this. Recent studies have established that VSMC phenotypes are driven, in part, by the diverse effects of long non-coding RNAs (lncRNAs) on gene expression. This class of largely uncharacterised gene regulators may offer a wealth of novel targets to be used to target VSMCs. However, their characterisation in VSMCs in pathological states is hampered by incomplete lncRNA representation in reference annotation.

In this thesis, we address this by assembling non-reference transcripts in RNA sequencing datasets describing saphenous vein VSMCs stimulated *in vitro* with cytokines and growth factors or arterial VSMCs stimulated with mechanical stress. We also utilised VSMCs isolated from atherosclerotic plaques. All transcripts were subject to a rigorous lncRNA prediction pipeline to provide an expanded VSMC transcriptome with an unprecedented level of detail on the lncRNAs associated with VSMC pathological states.

We found substantially improved coverage of lncRNAs responding to pro-mitogenic stimuli, with non-reference lncRNAs contributing 21–32% per dataset. We also demonstrate non-reference lncRNAs were biased towards enriched expression within VSMCs, suggesting extra lncRNAs highlighted by our pipeline have particular relevance to VSMC-specific processes. They were also biased towards transcription from enhancer sites suggesting they coordinate the regulation of neighbouring protein-coding genes. Both VSMC-enriched and enhancer-transcribed lncRNAs were large components of lncRNAs responding to pathological stimuli, yet without novel transcript discovery 33–46% of these lncRNAs would remain hidden. In parallel to this analysis, we mined the expanded VSMC annotation to initially explore functionality in a small cohort of uncharacterised lncRNAs within the saphenous vein VSMC *in vitro* model.

In our final round of analysis, we hypothesised that many lncRNAs may be involved in directing early transcriptional changes leading up to proliferation – and so constitute targets that may be particularly high value through acting upstream of multiple mitogenic or pathogenic pathways. We therefore used our expanded VSMC annotation as a foundation to perform a deeper analysis of lncRNA activity within RNAseq samples obtained from the first 24 hours of stimulation in the saphenous vein VSMC *in vitro* model, aiming to identify lncRNAs influencing initial transcriptional changes prior to observable cell division.

We noted an enrichment of lncRNA induction – particularly those which were VSMC-enriched or enhancer-transcribed – within an early phase of SVSMC stimuli response prior to proliferation. Transcription factor mRNA dynamics also localised to earlier phases whilst cell cycle mRNAs were overwhelmingly induced after 8 hours. This suggests the involvement of lncRNAs in an early phase of gene regulation sets the VSMC on a path towards later proliferation. To predict lncRNAs with functional impact in the 4 hour regulatory phase, we looked for evidence of their *cis*-regulation of nearby genes. Genes located near differentially expressed intergenic lncRNAs were 1.51x more likely to be differentially expressed within the four hour window than those located elsewhere in the genome. This effect was also identified for transcription factors and was particularly potent for genes around enhancer-transcribed lncRNAs (2.26x) but weakened when considering over longer time periods (1.16x). Together this suggests a focus of lncRNA-dependent *cis*-regulation activity in the first four hours after stimulation that could lead to wider downstream impact on VSMC pathological states. We finish by identifying a cohort of uncharacterised lncRNAs regulated in the initial four hour phase that have strong correlations in expression with transcription factors or other genes that explicitly link to vasculoproliferative pathology.

Overall, our comprehensive VSMC lncRNA repertoire provides much needed clarity on the activity of lncRNAs within VSMC pathological states. The approach we outline allows proper prioritisation of candidates for characterisation and exemplifies a strategy to broaden our knowledge of lncRNA across a range of disease states.

Lay Summary

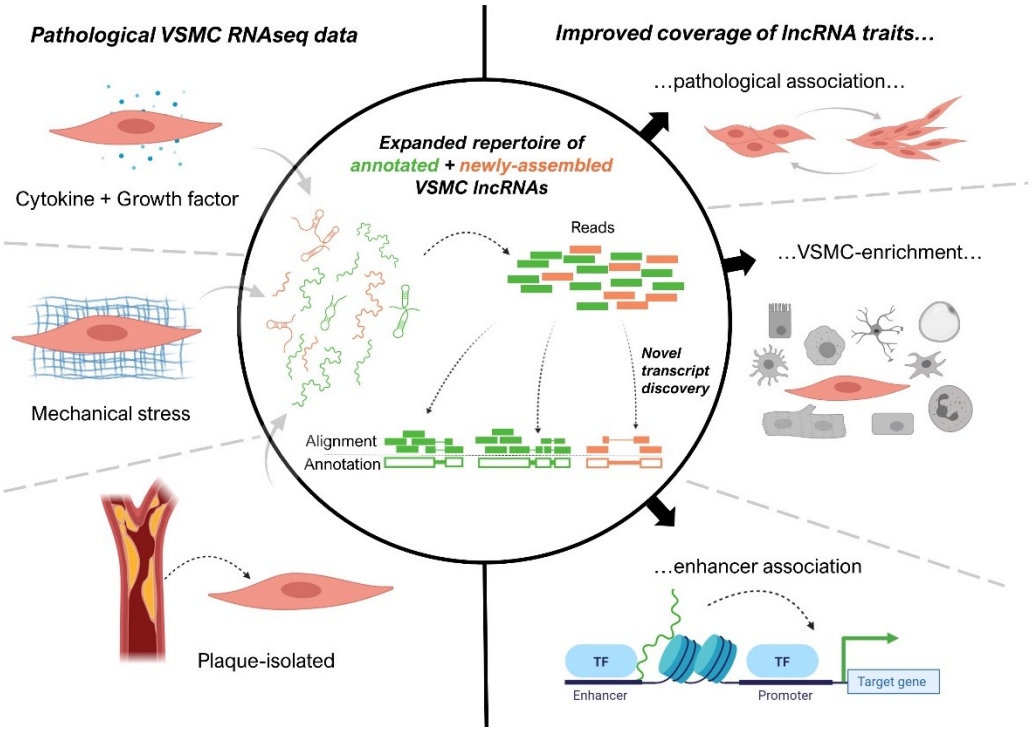
Vascular smooth muscle cells (VSMCs) make up the bulk of the wall of all large blood vessels (arteries and veins) where they assemble into a thick muscular inner layer. Here, they contract and relax to properly deliver blood throughout the body but also act as central contributors to the most life-threatening and widespread blood vessel diseases. VSMCs cause such diseases when they lose muscular traits and instead travel towards the inner surface of the blood vessel wall where they begin to divide. This lowers the ability of vessels to contract and reduces space for blood flow, eventually causing blockages and heart attacks (the number one most common global cause of death). To combat this, we want to understand how disease-causing signals (for example cholesterol from a high fat diet absorbed into the blood vessel wall) change VSMCs from being a crucial muscular component of blood vessel walls to disease-causing agents.

We now know that all cellular processes, including VSMC contribution to disease, are controlled by an internal machinery which directs their movement, growth and ability to divide. This machinery is created inside the cell according to a blueprint consisting of a molecule called DNA. Typically, elements of this blueprint are copied into another molecule called RNA which is then used to make proteins. These proteins are the structural components that are used to build a cell as well as components of the internal cellular machinery. However, recent advances in our ability to detect RNA have shown that some of these “blueprint copies” are not used to make proteins at all. Such RNAs are referred to as “non-coding” RNAs with several now known to be integral components of cellular machinery – much like proteins. This discovery initiated a wave of studies which hope to separate those RNAs which are made into protein or those which are non-coding and may instead have some other function.

In the field of VSMC biology, one particular type of non-coding RNA (called long non-coding RNAs or lncRNAs) have now been found to control their ability to show disease-causing traits. We may be able to use gene therapy to control such lncRNAs in patients and reduce the consequences of vascular disease. The field is therefore attempting to identify all such lncRNAs and many studies have been geared towards building catalogues of lncRNAs. However, despite their central role in vascular diseases, very few of these have focused on

VSMCs which has meant many lncRNAs of interest are likely unaccounted for. In this thesis we build a catalogue of lncRNAs in disease-causing VSMCs for the first time, showing that many of likely importance to processes leading to VSMC-related disease were previously missing from most previous studies. Ultimately, we hope to provide a resource for further studies of particular lncRNAs that could help stop disease-causing VSMCs. We also provide an example for others to use to build more informative lncRNA catalogues in other disease settings.

Graphical Abstract



Publications

Bennett. M. *et al.,* Novel Transcript Discovery Expands the Repertoire of Pathologically-Associated, Long Non-Coding RNAs in Vascular Smooth Muscle Cells, Int. J. Mol. Sci., 22(3), 1484 (2021)

Bennett. M. *et al.,* Endothelial function and dysfunction in the cardiovascular system: the long non-coding road, Cardiovasc. Res., 115, 1692–1704 (2019)

Presentations

"Novel transcript discovery expands the repertoire of pathologically-associated, long non-coding RNAs in vascular smooth muscle cells" *Poster presentation at Keystone: Non-Coding RNAs: Biology and Applications Online, May 2021*

"Deep Seq Exploration: Hidden LncRNAs in the Depths of Pathological VSMCs" *3MT Competition at CVS Symposium, University of Edinburgh, July 2020*

"Novel transcript discovery expands the repertoire of vascular smooth muscle cell long noncoding RNAs associated with pathology, enhancers and cell-specificity" *Poster presentation at RNA 2020, Online, May 2020 (including travel fellowship award)*

List of Abbreviations

AdvSca1	adventitial stem cell antigen 1
AOSMC	aortic smooth muscle cell
AngII	angiotensin II
ANOVA	analysis of variance
a-SMA	alpha smooth muscle actin
ASO	antisense oligonucleotide
BH correction	Benjamini-hochberg correction
BRG1	brahma-related gene 1
CAD	coronary artery disease
CAGE	cap analysis of gene expression
CAGEseq	cap analysis of gene expression sequencing
CASMC	coronary artery smooth muscle cell
ChIPseq	Chromatin immunoprecipitation sequencing
Ct	cycle threshold
dCt	delta cycle threshold
DIT	diffuse intimal thickening
dsiRNA	double-stranded interfering RNA
EC	endothelial cell
ECM	extracellular matrix
elncRNA	enhancer-transcribed lncRNA
eQTL	expression-quantitative trait loci
FBS	fetal bovine serum
FC	fold change
FDR	false discovery rate
FPKM	fragments per kilobase per million mapped reads
GEO	gene expression omnibus
GWAS	genome-wide association study
HDAC	histone deacetylase
IEG	immediate early gene
IL-1a	interleukin 1 alpha
IPAH	idiopathic pulmonary artery hypertension
ITGB3	integrin beta 3
KLF	Kruppel-like factor
LDL	low-density lipoprotein
lncRNA	long non-coding RNA
LRT	likelihood ratio test
miRNA	micro RNA
MMP	matrix metalloproteinase
MYH11	myosin heavy chain 11
oxLDL	oxidised low-density lipoprotein
PAH	pulmonary arterial hypertension
PASMCs	pulmonary artery smooth muscle cells
PCG	protein coding gene
PDGF	platelet-derived growth factor
PH	pulmonary hypertension
PLAR	pipeline for annotation of lncRNAs
qRT-PCR	quantitative real time polymerase chain reaction

RISC	RNA-induced silencing complex
RNAseq	RNA sequencing
Sca1	stem cell antigen 1
scRNAseq	single-cell RNA sequencing
Shh	Sonic hedgehog
SMMHC	smooth muscle myosin heavy chain
SNP	single nucleotide polymorphism
SRF	serum response factor

Contents

List of Figures

Chapter 3 contains the following figures within a manuscript:

List of Tables

Chapter 1: Introduction

1.1 VSMC proliferation in homeostasis and disease

1.1.1 VSMCs, Blood Vessels and Cardiovascular Disease

The principal role of vascular smooth muscle cells (VSMCs) is to provide contractile force in the walls of arteries and veins across the entire body to ensure proper circulation. During vascular development VSMCs stabilise the formation of new blood vessels, eventually becoming the bulk component of the middle layer of the vessel media. In fully formed vessels they overwhelmingly exist in a final differentiation state characterised by a contractile phenotype with minimal levels of proliferation. This provides a stable layer of muscle which can constrict or relax in response to changes in blood pressure. However, in contrast to most other cell types, VSMCs possess a high level of plasticity and a ready ability to lose differentiation state. In response to a multiplicity of stimuli they are exposed to in the vasculature, they can lose contractile ability and develop a range of phenotypes including proliferation, migration and extracellular matrix (ECM) production. The transition is reversible[1] and can be conventionally summarised as a move from a passive, quiescent and contractile state towards an active, synthetic and non-contractile state referred to as phenotypic switching. This phenomenon is thought to have evolved to allow vessels to be adaptable, switching between maintaining homeostasis or reacting to new circumstances with growth and repair mechanisms[2].

The balance between VSMC states is a central determinant of tissue remodelling during prevalent and life-threatening cardiovascular diseases such as coronary artery disease (CAD), hypertension as well as aortic and cerebrovascular aneurysms[3,4]. Both the genetic and environmental causes of these cardiovascular diseases are many-fold. Similarly, a high variety exists in such diseases in the range of VSMC phenotypes exhibited after loss of differentiation state[2,5]. However, for diseases like CAD and hypertension, VSMC proliferation in the intimal or medial layers is central to their pathophysiology, hence their characterisation as “vasculoproliferative” diseases. Proliferation has been characterised as the "final common pathway” in such contexts[6], enacted by multiple convergent stimuli such as growth factors, cytokines and biochemical or mechanical stresses. These also induce

other phenotypes which do not cause such diseases in the absence of VSMC proliferation but may sustain, accelerate or worsen the effects of VSMC proliferation. For instance, migration of VSMCs from the media to intima allows their proliferation in this space to produce neointimal expansions – lesions dense with VSMCs and ECM – on the surface of blood vessels[6]. Alternatively, apoptosis of VSMCs can lead to release of pro-inflammatory factors which act as stimuli for VSMC proliferation. Hence, development of therapies aimed directly at VSMC proliferation, or indirectly via supportive VSMC phenotypes is a long-standing aim to reduce cardiovascular disease burden.

Vessel wall remodelling can take years to decades to build up before significant downstream consequences are incurred, for example through reductions in the blood supply to surrounding tissue (ischaemia and hypoxia) or blood pressure increases that underlie vascular dementia or the hypertrophy of heart tissue[3]. Conversely, more immediate consequences are often particularly fatal, such as the rupture of atherosclerotic lesions and the resulting thrombosis leading to a blockage of blood flow to the heart and myocardial infarction[7]. CAD-related deaths are the number one most common cause of death globally in 2016, whilst deaths linked to strokes were second[8]. Hypertension is also a strong causal risk factor for both. It is therefore imperative to understand the regulatory mechanisms underlying the contribution of VSMC phenotypes to vasculoproliferative diseases. However, the causes and consequences of VSMC proliferation and associated phenotypes are not yet understood sufficiently to therapeutically target them effectively. Indeed, the net effect of VSMC proliferation on clinical outcomes is increasingly recognised as context-dependent, with the potential for both stabilisation or disruption of vessel wall function based on timing or presence of other phenotypes[5,9]. This means that despite decades of study on their foundational role in vascular remodelling, therapies which can directly modify the regulation of VSMCs in the vessel wall are still in desperately short supply or absent for nearly all vasculoproliferative pathologies. In this thesis, a novel cohort of candidate genes relevant to the control of VSMC proliferation in various disease settings are defined and explored.

1.1.2 VSMC role in vessel development

From early embryogenesis VSMCs are essential to stabilise arteries and veins at a nascent stage in development where they consist solely of endothelial cells (ECs). EC secretion of platelet-derived growth factor-BB (PDGF-BB), a key signalling molecule in the lifespan of VSMCs, leads to both recruitment of progenitor cells and their differentiation into VSMCs[10]. Many such progenitors and routes to VSMC differentiation are apparent in the early embryo. Flk1+ or 10T1/2 embryonic stem cells derived from the mouse mesodermal layer can adopt features of differentiated VSMCs such as upregulation of genes for contractility, and adoption of a spindle-like morphology through stimulation with PDGF-BB or transforming growth factor-β (TGF-β) - another key signal for VSMCs, often acting in opposition to PDGF-BB in mature blood vessels[11,12]. PDGF-BB provides another source of mural cells for nascent blood vessels as it is chemotactic for differentiated VSMCs, inducing their migration to the tip of newly formed vessels via a signalling pathway involving Sonic hedgehog (Shh)-dependent phosphorylation of ERK1/2 and PI3K[13]. ECs also communicate to VSMCs in close-contact via Notch signalling. Presentation of the Jag1 ligand on the surface of ECs is recognised by VSMC Notch receptors which then activate tight linkages to the endothelium via expression of integrin αvβ3[14]. Further, the Notch3 receptor is required for expression of VSMC contractile genes in the aortic arch and coronary arteries[15]. These embryonic processes and pathways can be reactivated during repair processes and form the basis of many responses to vascular injury in later mature vessels.

VSMCs are recruited to nascent vessels from distinctly developed embryonic regions, a factor contributing to a well-described trait of VSMCs; their heterogeneity between and within vascular beds. For example, VSMC progenitors for vessels in the immediate vicinity of the mammalian heart originate from at least 4 embryonic regions[5]. Differentiation paths from human pluripotent stem cells to VSMCs via these regions have been replicated *in vitro* by Cheung et al.[16](Figure 1.1). This showed populations of VSMCs from the neuroectoderm or lateral plate mesoderm embryonic regions that form the disease-prone aortic arch, carotid arteries and coronary arteries show an innate capacity for degradation of extracellular matrix (ECM) common to many

tissue remodelling events in these vessels. This manifested as relatively high production of matrix metalloproteinase 9 (MMP9) and low production of TIMP1 – a general inhibitor of MMPs. Conversely, those derived from hSPC-derived paraxial mesoderm showed far less pathogenic capacity and produce the disease-resistant descending aorta. Additionally, VSMCs from the aortic arch and descending aorta possess distinct gene expression profiles, reflecting the differences in disease-susceptibility between these vessel regions. The former show greater capacity for immune response and proliferation whilst the latter emphasise expression of developmental genes (e.g. *Hox* genes)[17,18]. These differences appear not to stem from differences in sizes of VSMC sub-populations but rather are apparent at the level of individual VSMCs isolated from either region. These studies demonstrate a strong genetic component to heterogeneity amongst VSMCs as their tendency to produce a pathogenic phenotype can be initially leveraged during embryogenesis.

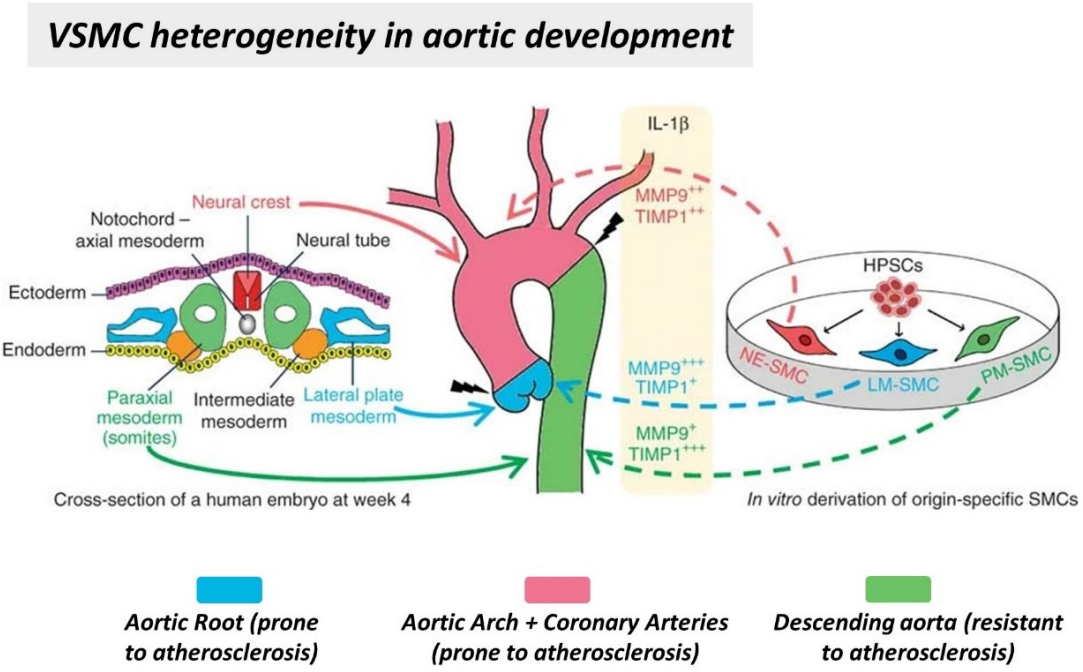


Figure 1.1 VSMC heterogeneity in aortic development. Various distinct embryonic regions provide cells to distinct regions of the aorta. Cheung et al. provide evidence that this spatially-separated heterogeneity in embryonic origin influences propensity towards disease. Adapted from Cheung et al., 2012[16]

Subsequent proliferation and differentiation of progenitors into VSMCs, builds up concentric layers of smooth muscle resulting in the muscularisation of large vessels[10,19]. A case study in the mouse pulmonary artery provides

some mechanistic clues, showing that these layers are formed initially of cells expressing the PDGF receptor PDGFR-β[20](Figure 1.2). This receptor is later downregulated concomitant with an increase in markers associated with VSMC identity such as α-smooth muscle actin as mature, contractile VSMCs are formed. Morphological changes are also apparent as they reduce their longitudinal size and reorient to wrap around the vessel circumferentially in a more uniform manner. With repetitions of this process in multiple consecutive layers the vessel media is formed whilst in the final layer the transition is not completed leaving a less mature layer involved in forming the outer adventitial layer. Aside from PDGF-BB and TGF-β signals, likely derived from the endothelium, other likely pro-differentiation signals for VSMCs are shear stress provided by blood flow[21]. Additionally the ECM components produced by immature VSMCs - particularly collagen and elastin – are not only necessary to provide tensile strength but also repress VSMC chemotaxis and their synthesis of other ECM components often found in pathological contexts[22]. Production of such proteins peaks soon after VSMC recruitment to nascent vessels then rapidly falls as VSMCs mature and reduce synthetic activity[23]. In mature arteries and veins VSMCs exist in the medial layer, separated from the endothelium via a basement membrane and in healthy conditions are tightly controlled to maintain their mature, quiescent state.

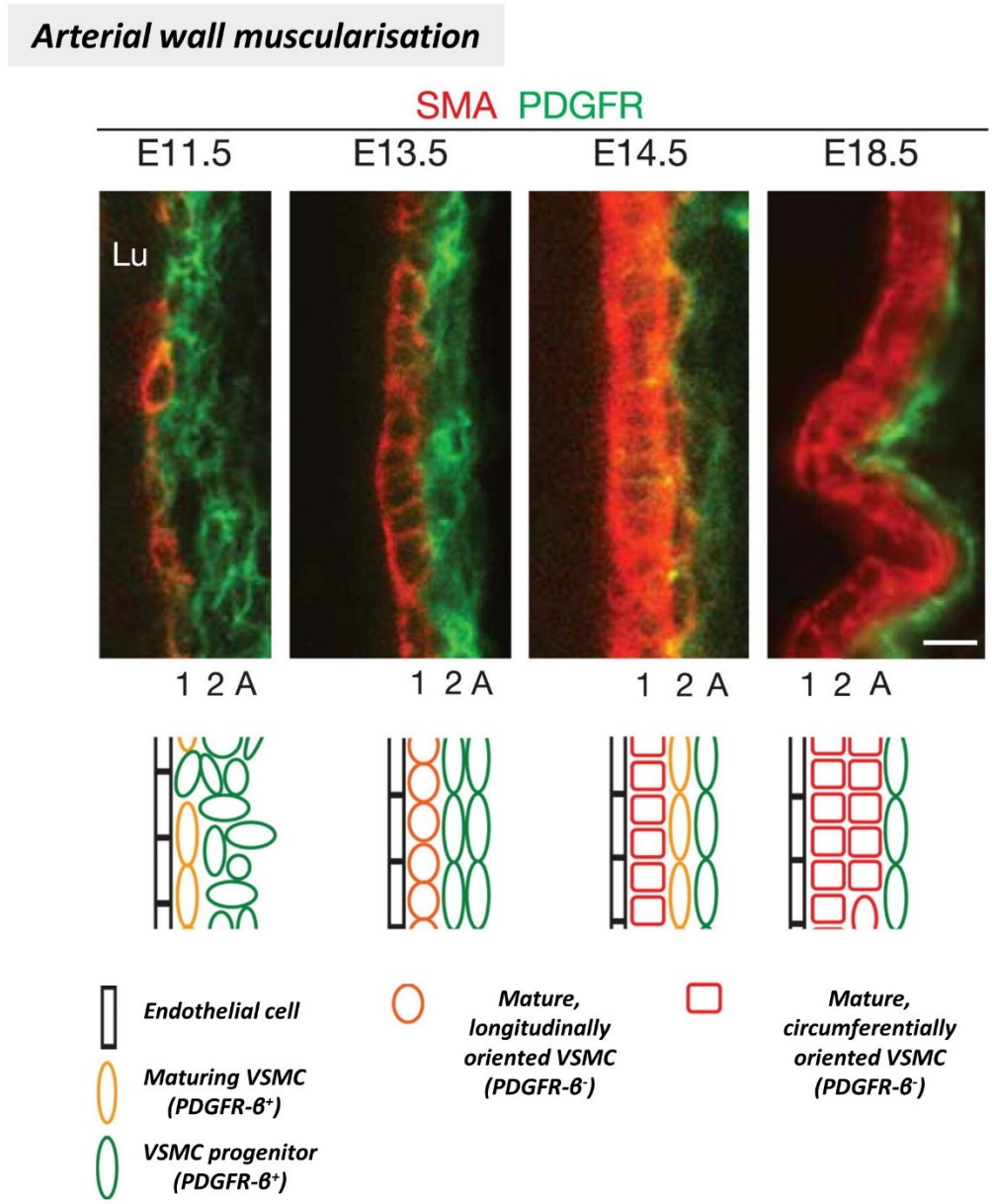


Figure 1.2 Arterial wall muscularisation. Case study of early embryonic development of the pulmonary artery in mice showing evidence for sequential layering and maturation of VSMC progenitors. Adapted from Greif et al., 2012[20]

1.1.3 VSMC proliferation during atherosclerosis

Atherosclerosis is defined as the development of plaques comprised of inflammatory cells, lipids, cellular debris and VSMCs in the intima of arteries. This process occurs over decades in humans and is a root cause of myocardial infarction, stroke, heart failure and angina and so the top cause of death globally[24]. Despite widespread plaque occurrence across the human population, attempts to reduce the associated mortality risk are limited to either

invasive surgical interventions or broad-brush treatments to lower risk factors like blood lipid levels. Halting or reversing plaque growth or even changing plaque morphology therefore requires finding new druggable targets in the most relevant cell types. VSMC phenotypic modulation has long been known to be involved in atherosclerosis progression and is extensively studied in this setting. However, in recent years we have gained a clearer view on the somewhat ambiguous role of VSMCs during atherosclerosis alongside other causative factors, such as a sustained local inflammatory response[5].

The timeline of the role of VSMCs in human atherosclerosis could begin with the formation of diffuse intimal thickenings (DITs), accumulations of structurally organised VSMCs and ECM thought to occur as an adaption to arterial flow pressure. They are visible in the intima, predominantly in arteries prone to the disease, from as early as 36 weeks after birth[25]. Though DITs are absent in animal models of atherosclerosis, so challenging to definitively link to atherosclerosis, their localisation and age-related increase in thickening suggests they are likely to act as the initial base for plaque formation. VSMCs in DITs show low proliferative ability relative to VSMCs in plaques[26] though appear to be dedifferentiated as they are responsible for synthesising the ECM component of this layer. The ability of this negatively charged ECM, comprised mainly of proteoglycans, to bind to negatively charged circulating lipids is thought to allow plasma-derived low-density lipoproteins (LDLs), particularly apolipoprotein B, to accumulate in intimal pools over a potentially decades-long period of time (Figure 1.3).

Uptake of LDLs can stimulate the proliferation and pro-inflammatory activity of VSMCs. However long term exposure, or conversion of LDLs into oxidised LDLs (oxLDLs), induces VSMC apoptosis and cell cycle arrest[27]. In early stages of atherogenesis uptake of oxLDLs by VSMCs as well as macrophages gives them a distinct lipid-dense “foam cell” morphology which coincides with a cycle of increased oxidative stress and apoptosis[28]. Sustained inflammation then arises as clearance of apoptosing cells through phagocytosis by macrophages or VSMCs is inhibited, in part through effects of oxLDL accumulation[29]. Lack of clearance leads to necrosis of apoptosing cells, with the resulting debris acting as a particularly potent inducer of the secretion of cytokines such as IL-6 and MCP-1. VSMCs and ECs can be stimulated by

other LDL derivates to increase secretion of MCP-1 and adhesion molecules, aiding recruitment of circulating monocytes and their differentiation into macrophages – a process also resulting in loss of endothelial barrier integrity[27,30,31]. The initiation of atherosclerosis through lipid retention is supported by observations that lipid accumulation precedes an increased density of macrophages (derived from monocytes) in early human lesions[32,33]. Further, altering the proteoglycan binding capacity of apolipoprotein B results in lower intimal retention of this lipid and smaller lesion size in mice aortas[34]. Together this apoptosis, necrosis, influx of macrophages, disruption of the endothelial layer and the accompanying release of cytokines and growth factors lays the foundation for a wave of VSMC proliferation to contain the inflamed lipid-rich area.

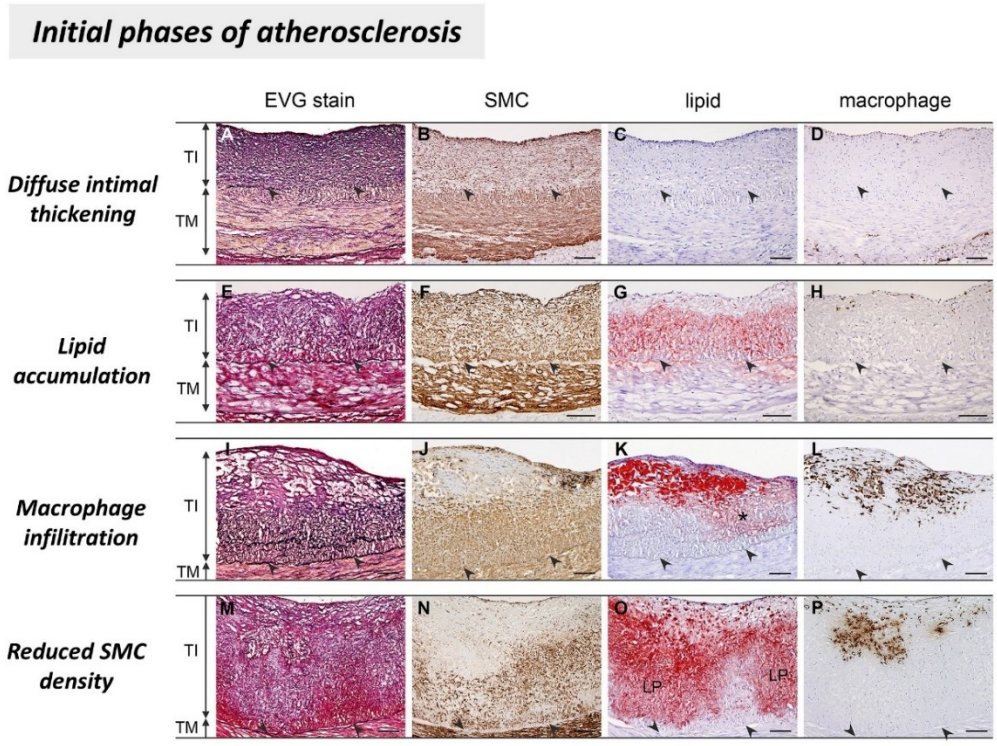


Figure 1.3 Initial phases of atherosclerosis. Each row represents a phase if tissue remodelling prior to atherosclerosis, displayed using EVG stain in left-hand panels, anti-SMA immunostaining in 2nd column, Sudan IV staining in 3rd column and anti-CD68 immunostaining in the right-hand panels. Arrows indicate the elastic lamina basement membrane whilst TI refers to the tunica intima (referred to as the intima in the text) and TM refers to the tunica media (referred to as the media in the text). Adapted from Nakagawa et al., 2018[32]

VSMC proliferation predominantly occurs during formation of advanced atherosclerotic lesions or fibroatheromas (Figure 1.4). These are covered by a

fibrous cap which, in the best cases, stabilises the inner core of inflammatory activity initiated in earlier stages thereby preventing it from destabilising and rupturing the plaque. PDGF-BB, is secreted by ECs, platelets and macrophages and is widely-appreciated as a central driver of the VSMC accumulation that builds advanced atherosclerosis lesions[35,36]. Production of PDGF-BB by macrophages is visible even in early atherosclerotic plaques, still consisting mainly of lipid rich deposits, but this output markedly increases in the advanced plaques with fibrous caps[37]. PDGF-BB can stimulate proliferation in VSMCs through using the ubiquitous ERK1/2 and PI3K/Akt signalling pathways to activate cell cycle components like cyclin D1[38].

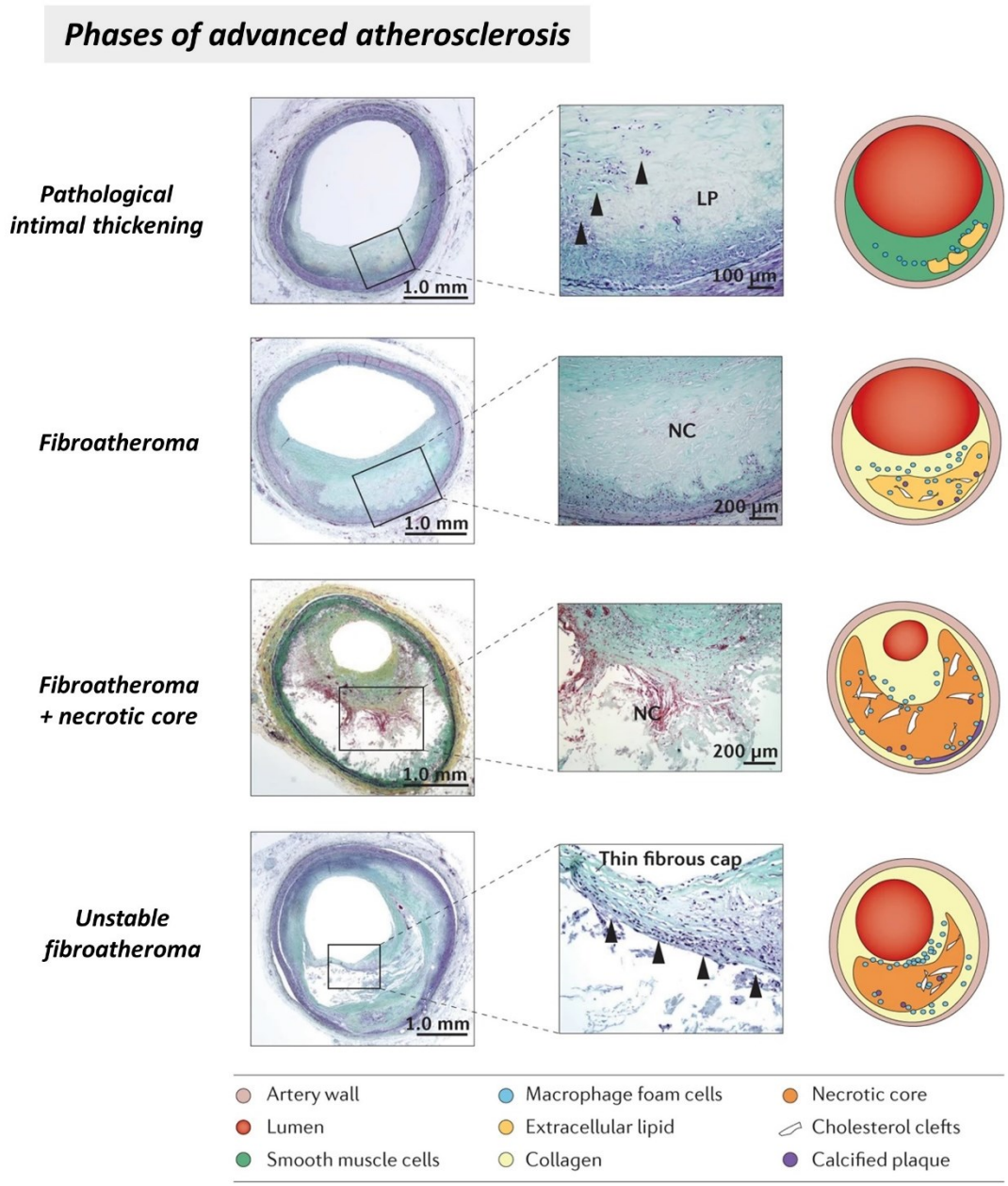


Figure 1.4 Phases of advanced atherosclerosis displayed via histology and representative images. Arrows indicate macrophages, LP indicates lipid pools, NC indicates necrotic cores. Adapted from Yagahi et al., 2016[7]

Pro-inflammatory factors secreted by macrophages, activated ECs or present within apoptotic debris such as IL-6, IL-8, MCP-1, IL-1α and IL-1β can also stimulate VSMC proliferation and migration *in vitro* given the right conditions[39–44] so are also contributors to VSMC intimal accumulation during progression of atherosclerosis. Additionally, PDGF-BB and pro-inflammatory signalling pathways can together boost proliferative activity synergistically. For example, IL-1β has been shown to play a supportive role in PDGF-BB mediated proliferation through downregulating expression of cell cycle kinase

inhibitors p27 and p21[45]. In VSMCs, treatment with both factors together causes close physical association of PDGFRβ and the IL-1 receptor at the cell membrane leading to sustained phosphorylation of PDGFRβ and extended Akt signalling[46]. VSMC proliferation is not thought to be implemented directly by NF-κB, a central mediator of transcriptional activity typically activated by pro-inflammatory factors. However, the activation of NF-κB appears to support VSMC migration, proliferation and survival by inducing anti-apoptotic, ECM degrading and pro-inflammatory factors[38]. Conversely, PDGF-BB plays a supportive role in IL-1β signalling as activation of the ERK pathway can in turn sustain NF-κB activation for longer[47] which can thus accentuate pathogenic VSMC behaviour. For example, the expression of the matrix degrading enzyme, MMP9 by VSMCs is unaffected by PDGF-BB alone and is upregulated by IL-1α. However, PDGF-BB treatment can potentiate IL-1α-induced MMP9 expression in an NF-κB dependent manner - also implying wider support of other NF-κB-targeted genes by PDG-BB[48,49]. As both PDGF-BB and pro-inflammatory factors are generated together within atherosclerotic lesions, VSMC proliferation is therefore likely driven not only by signalling cascades initiated by PDGF-BB or pro-inflammatory factors alone but also by synergistic effects resulting from the interaction of these cascades.

Our understanding of how these pathways are used in VSMCs during plaque progression has been aided by the development of more sophisticated tools to track cell subpopulations in the blood vessel wall. Initial evidence that clonal patches of cells exist in human plaques was established nearly 5 decades ago using patterns of X-chromosome inactivation to track VSMC progeny[50]. This method was later finessed to show proliferation of a tiny number or even a single medial VSMC can be responsible for these clonal patches and also identified them within DITs in healthy arteries suggesting that this clonal expansion may begin in early life[51]. The monoclonal nature of atherosclerosis distinguishes it from other more polyclonal events like VSMC proliferation during vessel development[52] or tissue remodelling during wound healing.

Lineage tracing techniques allowing permanent fluorescent labelling of medial VSMCs and their progeny in healthy arteries prior to induction of atherosclerosis have recently expanded on these early findings. In mice, tracing of VSMC clones from a mature differentiated phenotype in the media,

into advanced plaques shows they not only provide cap cells, which express high levels of ACTA2, but a set of ACTA2^{-} cells localised in the core that stain positive for oxLDL and express macrophage markers LGASL3 and CD68[53,54]. This is consistent with earlier *in vitro* observations that cholesterol-loaded VSMCs upregulate macrophage markers[55]. Another VSMC tracing effort set these findings in a temporal context, showing that a single VSMC is capable of producing all clonal cap cells which express contractile markers as well as PDGFR-β and are hyperproliferative (Figure 1.5 – top panel)[52]. These cap cells were also identified to provide VSMC-derived cells in the plaque core which downregulate contractile genes and are ~5-fold less proliferative. The ability to sequence the transcriptome of individual VSMCs during plaque development using single cell RNA sequencing (scRNAseq) allowed Wirka et al. to add further detail, indicating VSMCs follow a linear progression from the contractile phenotype to a modulated phenotype which can then progress to cells expressing macrophage markers (Figure 1.5 – bottom panel)[56]. Notably, there was no detectable divergence from this path to obtain a macrophage-like transcriptome. Rather, the transcriptional changes underlying this progression point towards a reprogramming of VSMCs towards a fibroblast-like or "fibromyocyte" phenotype visible in both mouse and human atherosclerosis. VSMCs expressing macrophage markers appear to be an extreme form of this progression and do not show overall transcriptional similarity to macrophages.

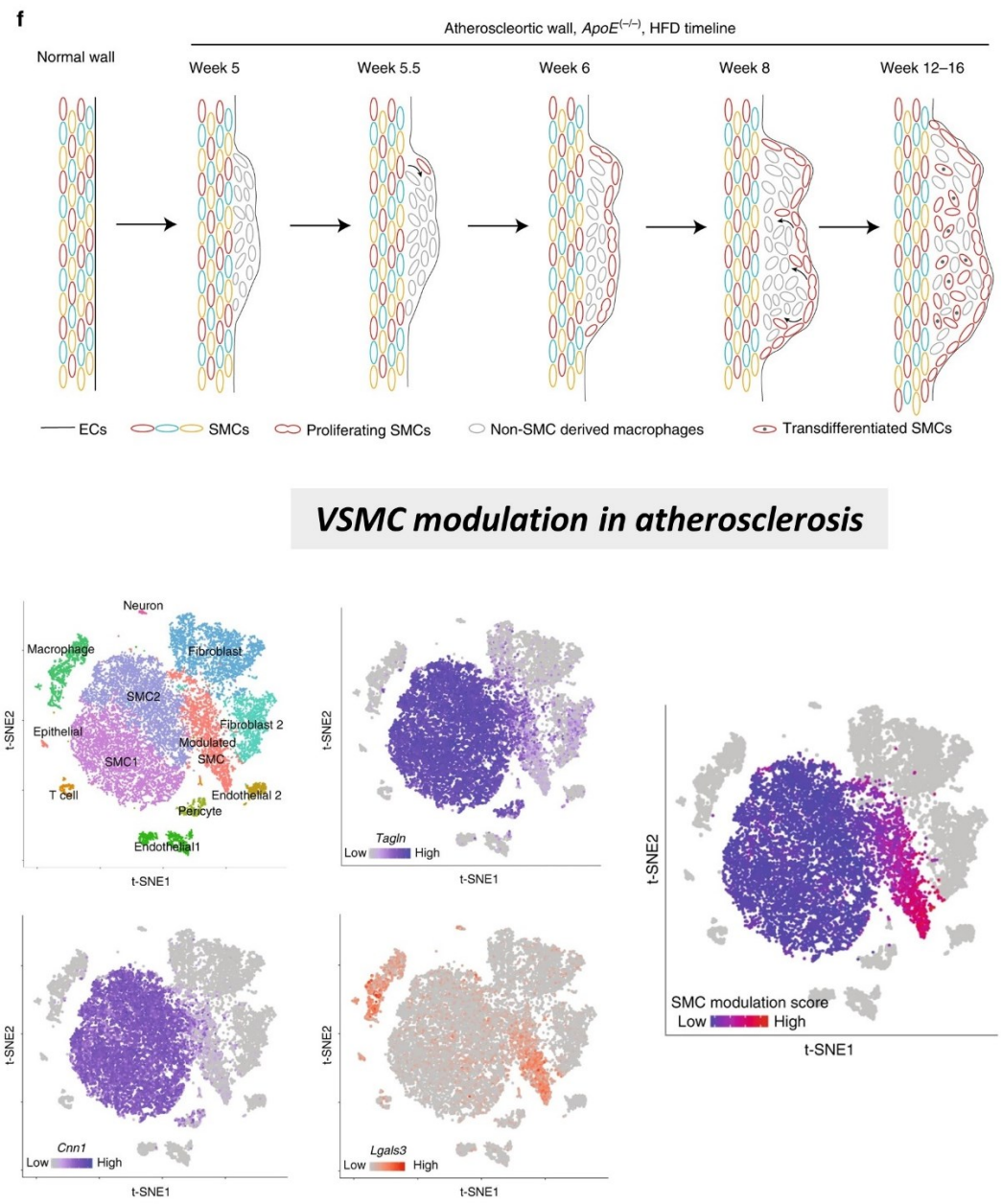


Figure 1.5 Recent insights into VSMC dynamics in atherosclerosis. Top panel adapted from Misra et al., 2018[52] shows schematic model of how VSMCs expand clonally after leaving the media to create the plaque cap before subsequently entering the plaque core. Bottom panel adapted from Wirka et al., 2019[56] shows single cells isolated from plaques and clustered based on transcriptional similarities via scRNAseq. Each panel represents this cell clustering coloured by relative expression strength of VSMC markers TAGLN (SM22α) or CNN1, the macrophage marker LGALS3 as well as the bespoke score for VSMC modulation defined by the authors.

Tracing VSMC populations also advanced the idea of altering the balance of their phenotypes within plaques to reduce chance of rupture. VSMC-specific

knockout of KLF4, a core component of VSMC dedifferentiation, identified KLF4 as a driver of macrophage-markers as well as proliferative and apoptotic phenotypes during mouse plaque growth. Accordingly, this led to >50% reduced lesion size and increased the density of VSMCs expressing contractile markers within caps and overall cap size within these lesions[54]. Conversely, a VSMC-specific knockout of TCF21 in mice reduced the fibrous cap area relative to the LGALS3^{+} core as well as the arterial proportion of fibromyoctes relative to non-modulated VSMCs[56]. TCF21 then, appears to promote the proliferation of VMSCs with a fibromyocyte phenotype allowing proper containment of the lipid-rich core, whilst KLF4 activity may push these VSMCs to further along the phenotypic modulation axis and into the lipid-rich plaque core. Other factors identified using lineage tracing studies have also determined factors controlling the initial transfer of medial VSMCs into the plaque. Oct4, a pro-pluripotency factor induced by KLF4, is required during mouse plaque progression and when knocked out greatly reduces VSMC density in the fibrous cap[57]. The only visible phenotypic effect from Oct4 knockout in VSMC *in vitro* is loss of migration, implicating it as a possible driver of medial to intima transfer. Alternatively, integrin β3 (ITGB3), a transmembrane protein interacting with both ECM and the cytoskeleton, showed biased expression to cap rather than core VSMCs and suppresses toll-like receptor (TLR)4 signalling – used to generate innate immune responses. This desensitised these cells to pro-inflammatory factors whilst also reducing oxLDL uptake and macrophage marker expression[52]. Further, in myeloid cells ITGB3 appeared to maintain the monoclonality of VSMCs in plaques and cap formation, possibly through suppressing the *en masse* proliferation and migration of multiple medial VSMCs.

The ability to track alterations in VSMC dynamics through perturbing known factors in dedifferentiation, pluripotency and cell communication have obvious, immediate implications for therapies aiming to adjust the balance of VSMC phenotypic modulation. These findings provide evidence that a single medial progenitor can provide all plaque VSMC-derived cells, of both beneficial and disruptive phenotypes, and that these are a major cellular component of the lesion. Multiple competing hypotheses exist to attempt to define this population of initial causative VSMCs in the vessel wall. Evidence that resident stem cells

in the adventitia are competent for differentiation into VSMCs in atherosclerotic lesions was initially provided by Hu et al. in 2004[58]. These cells, identified through possession of many markers of stem cells including stem cell antigen 1 (Sca1), could be seeded on to the outer adventitia of irradiated vessels and be observed to migrate through to the intima in a mouse vessel injury model *in vivo*. A group of Sca1$^+$ adventitial (AdvSca1) progenitors was later identified to arise in 2-3 week old mice at the medial border, possessing a high level of Shh signalling[59]. AdvSca1 cells with Shh activity were subsequently traced (via Shh component, Gli1) from localisation in the adventitia through to neointimal formations (accumulations of VSMCs and ECM similar to DITs) after wire injury in mice[60]. In these neointima, >50% of mature VSMCs were derived from Gli1+ AdvSca1 cells showing that resident progenitors in the adventitia have the capacity to contribute to VSMC mass after injury. However, other tracings of AdvSca1 cells, labelled directly with Sca1, have failed to reproduce this contribution to the neointima post-wire injury, instead seeing influx of AdvSca1-derived VSMCs to the media only, and requiring a more severe trans-sectional injury of the vessel wall[61]. Therefore, though the potential of AdvSca1 cells to provide VSMCs during vessel remodelling is proven, this may be a particularly context-dependent event. The mechanisms, prevalence and clinical relevance of adventitial-derived VSMCs are therefore still an open question.

Another Sca1$^+$ population with potential as a source of pathological VSMCs has been identified outside of the adventitia, within medial VSMCs in mouse aortas[17]. Using scRNAseq, Sca1 was found amongst a set of genes with particularly high variation across populations of medial VSMCs. Expression of Sca1 was strongly negatively correlated with a set of genes that together provide a contractile expression signature in differentiated VSMCs. The Sca1$^+$ population is <1% of the entire medial population of VSMCs though increased in size in response to *in vitro* culturing, *in vivo* aging and carotid ligation injury. 8 days post-injury, Sca1$^+$ medial cells could expand to provide 10-45% of arterial VSMCs[17]. Within mouse plaque-derived VSMCs, lineage traced using SMMHC labelling, Sca1$^+$ cells represented only a proportion of dedifferentiated VSMCs and did not overlap with those expressing macrophage markers. Sca1 may therefore mark out VSMCs in an intermediate state between contractile and dedifferentiated VSMCs which readily expand during vasculoproliferative

disease. Neointimal and plaque VSMCs are therefore shown to derive from medial cells expressing VSMC marker genes in some studies[17,52–54,56], yet from AdvSca1 cells in others[58,60]. Both models could potentially act in concert, either through AdvSca1 cells being a precursor to medial Sca1$^+$ VSMCs or both providing independent routes to intimal VSMCs. However, the finding that Sca1 is expressed in medial VSMCs provides initial evidence that the two populations could, at least in part, be of the same lineage. A potential model which unifies both sets of studies would need to reconcile differences between the potential VSMC progenitor populations. For instance, Gli1-tracked VSMC progenitors express distinct markers from medial Sca1$^+$ VSMCs[17,60]. Untangling the contributions and clinical relevance of both adventitial and medial progenitors of intimal VSMCs will continue to be an area of high interest in the coming years.

This provides an overview of the role of VSMC across various stages of atherosclerosis. Despite the advances outlined, we still have no therapy to address VSMCs directly in this context. However, work in this area has provided a basis to understand VSMC dynamics in distinct proliferative contexts such as those seen after interventions undertaken to reduce risk from atherosclerosis.

1.1.4 VSMC proliferation after surgical interventions

A commonly used strategy to intervene in atherosclerosis of the coronary or peripheral arteries is to circumvent the diseased arteries using other non-diseased blood vessels as grafts to allow revascularisation. Saphenous veins are regularly used as grafts as they are relatively convenient to surgically extract, of a large enough diameter to meet arterial blood flow demands and lengthy enough to provide sufficient material for grafting. Despite this, the consensus is that 10-20% of saphenous vein grafts (SVGs) used to bypass coronary arteries lose patency in the first year alone and this increases to >60% in the 10-20 year period[62]. This is much higher than the 10–20-year failure rate seen when using alternative arterial grafts, particularly the internal mammary arteries (<5%) and radial arteries (<20%). The increased failure rate in SVGs relative to arterial grafts is due to an initial risk of thrombosis in the

first months after surgery and then in subsequent years inward vascular remodelling[9]. This process shares much in common with the initial formation of atherosclerosis but occurs in an accelerated time frame.

Current treatments attempt to boost graft patency through use of statins and aspirin that target the identified risk factors cholesterol and platelet levels respectively[63]. Improved "no-touch" surgical techniques which limit graft damage during harvesting of the vein may lead to more substantiative changes in future SVG failure rates, with 83% grafts recorded as patent after 16 years in small scale trials of <30 patients[64]. However, the sheer number of SVG procedures undertaken means that without targeted treatments to effectively reduce adverse remodelling, the failure rate will remain an issue. Gene therapy is a strategy which holds high potential as grafts can be treated *ex vivo*, post-harvesting and immediately prior to surgical implantation[65]. This contrasts with other scenarios where a therapeutic may have to be administered to a tissue *in situ*, likely requiring a more sophisticated targeted delivery system to ensure the tissue of interest receives proper dosage. As the vein graft procedure incorporates a period in which the harvested graft is prepared for implantation, this provides a window of opportunity to expose the graft to therapeutics (Figure 1.6). This could allow manipulation of VSMCs in the early phases of graft remodelling which are known to determine the long-term functionality of the graft. Notably, this idea has been previously pursued in a phase 3 clinical trial which utlised decoy oligonucleotides to prevent activity of the pro-proliferative E2F in pre-implant veins[66]. Despite widespread evidence showing that the etiology of vein graft failure in the months to years post implantation is related to an early phase of VSMC proliferation – no beneficial effects were seen in this trial. This demonstrated that we still have much to decipher in the early influences on graft remodelling and any link with pro-proliferative influences. Exploration of the molecular mechanisms underlying the VSMC contribution to vein graft failure therefore remains a high priority, to both stimulate and inform future therapeutic developments.

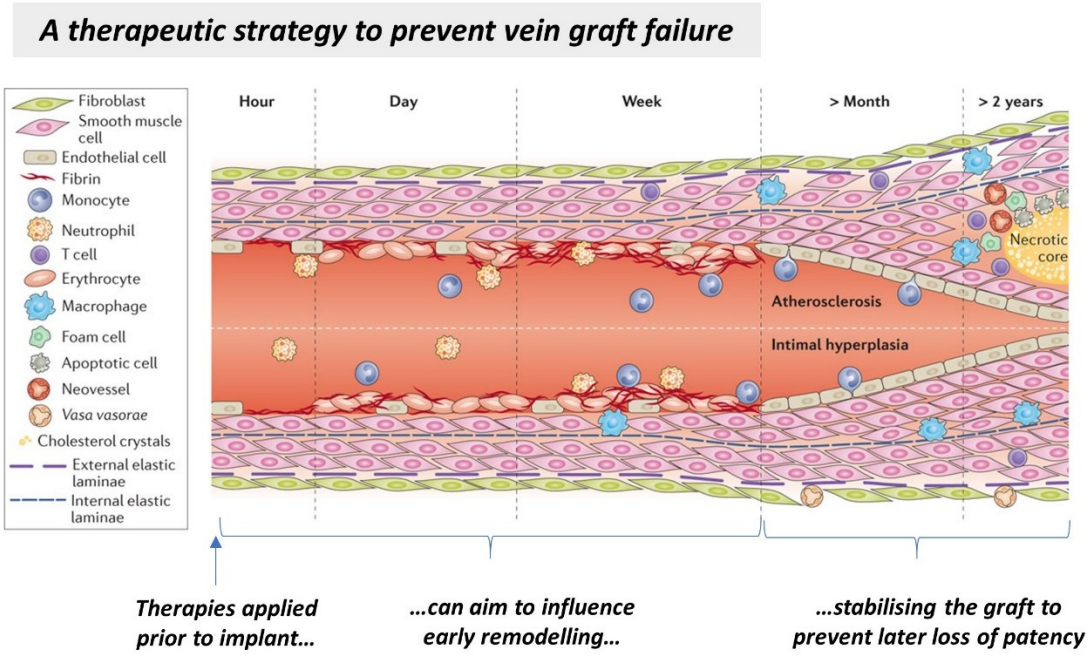


Figure 1.6 The opportunity for therapeutic intervention prior to onset of negative remodelling in vein graft failure. Preventing formation of a VSMC-rich neointima in the first weeks post-implantation may be possible by pre-treatment of a graft with gene therapy e.g. to reduce expression of pro-proliferative regulators in VSMCs. Effects from such a therapy would likely have to be sustained over a time period of 1-2 weeks minimum and not bring a halt to beneficial remodelling required for muscularisation of the graft. Adapted from de Vries et al. 2016[9].

Targeting VSMC proliferation with gene therapy may be a way to counter the uncontrolled intimal thickening in the later phases of SVG failure, rather than earlier failures largely due to thrombosis. In the later phase, intimal thickening can proceed towards occlusion of the entire lumen whilst providing a highly atherogenic environment which often leads to the development of new plaques. Compared to “native” atherosclerosis this process occurs in a time-frame of years rather than decades and the resulting plaques show greater signs of instability[7]. VSMC proliferation again plays a context-dependent role with both beneficial and detrimental effects to patency during remodelling of the graft. Neointimal hyperplasia is already apparent in many SVGs pre-implantation[67] but growth of such lesions also occurs in the weeks and months after surgery stimulated by loss of endothelium, platelet attachment and macrophage infiltration at the vessel wall surface. Successful grafts limit this intimal thickening to muscularise the vein and allow it to adapt to arterial blood pressure levels without effecting lumen diameter. Outward remodelling of the media and adventitia aids long term patency, allowing luminal dilation to adapt

to increased shear stress[68,69]. In failed SVGs, neointimal hyperplasia growth continues unabated, often with minimal outward remodelling and stiffening of the graft wall.

Steps taken during the surgical procedure to prepare SVGs are an initial cause of SVG intimal thickening. In *ex vivo* culture, surgically prepared veins show an increase in intimal thickening relative to unprepared veins which could be attributed to loss of the endothelium and increased induction of oxidative stress[70]. Endothelial loss – initiated during the initial harvesting of veins – is exacerbated by surgical preparation steps to ensure the vein is graft-worthy and results in loss of endothelial cell nitric oxide production which has anti-proliferative, anti-thrombotic and anti-inflammatory effects. Surgical handling involves high pressure distention of the vein, which initiates expression of adhesion molecules and TLRs in the graft which, as in native atherosclerosis, induce the recruitment of circulating monocytes and platelets[71]. The magnitude of pressure required to damage the endothelial layer is regularly exceeded during the vein preparation and positively correlates with the amount of endothelial cell loss and neointimal hyperplasia seen in the weeks after procedure[72]. "No touch" techniques to harvest grafts may alleviate this damage by removing the graft whilst keeping the surrounding tissue including the adventitial vasa vasorum. Retention of these microvessels may be particularly important for the stability of venous grafts, as the increased number of endothelial cells facilitate greater nitric oxide production and preserve the graft blood supply for longer over initial post-surgery phase[73]. In rodent models, vein grafts lose their endothelium by 1-7 days with concurrent platelet and monocyte attachment prior to regeneration within ~30 days - no such injury occurs in arterial grafts[74,75]. The phase of loss and regeneration of endothelium sows later vessel wall disruption by providing an early opportunity for leukocyte and platelet attachment and secretion of signals for VSMC proliferation and dedifferentiation.

Rodent models of early vein graft remodelling show that an initial wave of VSMC apoptosis in the first days is overtaken in the first week by a sharp increase in mostly medial and adventitial proliferation which then sharply declines[75,76]. A VSMC-dense neointimal hyperplasia is present after 14 days and can continue thickening up to at least 70 days without loss of luminal area.

Many SVGs also show signs of intimal hyperplasia prior to implantation which stain heavily for TGF-β likely an indication of extensive ECM secreton[67]. In SVG patients an initial increase in the graft lumen diameter occurs in the first month, correlating with the initial shear stress level after implantation[68]. The extent of this early outward remodelling is a strong predictor of graft patency in subsequent years[69]. Mechanical factors like shear stress can stimulate SVG adaption to a higher pressure environment through inducing VSMC proliferation, acting via the PI3K pathway to repress activity of the cell cycle inhibitor p27[77]. Patients with a single nucleotide polymorphism (SNP) that can boost expression of p27 have a greater increase in lumen diameter in the first month and a greater 5-year graft patency rate[78]. Thus, though VSMC proliferation in the graft is required to effect outward remodelling, a greater ability to limit this proliferation - for instance through an SNP leading to elevated p27 expression - may aid long-term patency. Graft remodelling immediately after the first month of implantation appears minimal with no change in wall thickness but a gradual reduction in luminal diameter over the rest of the first year[79]. Proliferating cells over this time period in human graft lesions are few in number (1-2% of all cells) but mainly VSMCs, endothelial cells (from infiltrating microvessels) monocytes or macrophages[80]. The first month post-implantation, influenced by VSMC proliferation and apoptosis, is therefore a key remodelling phase that influences later graft patency and a strong target for gene therapy (Figure 1.7).

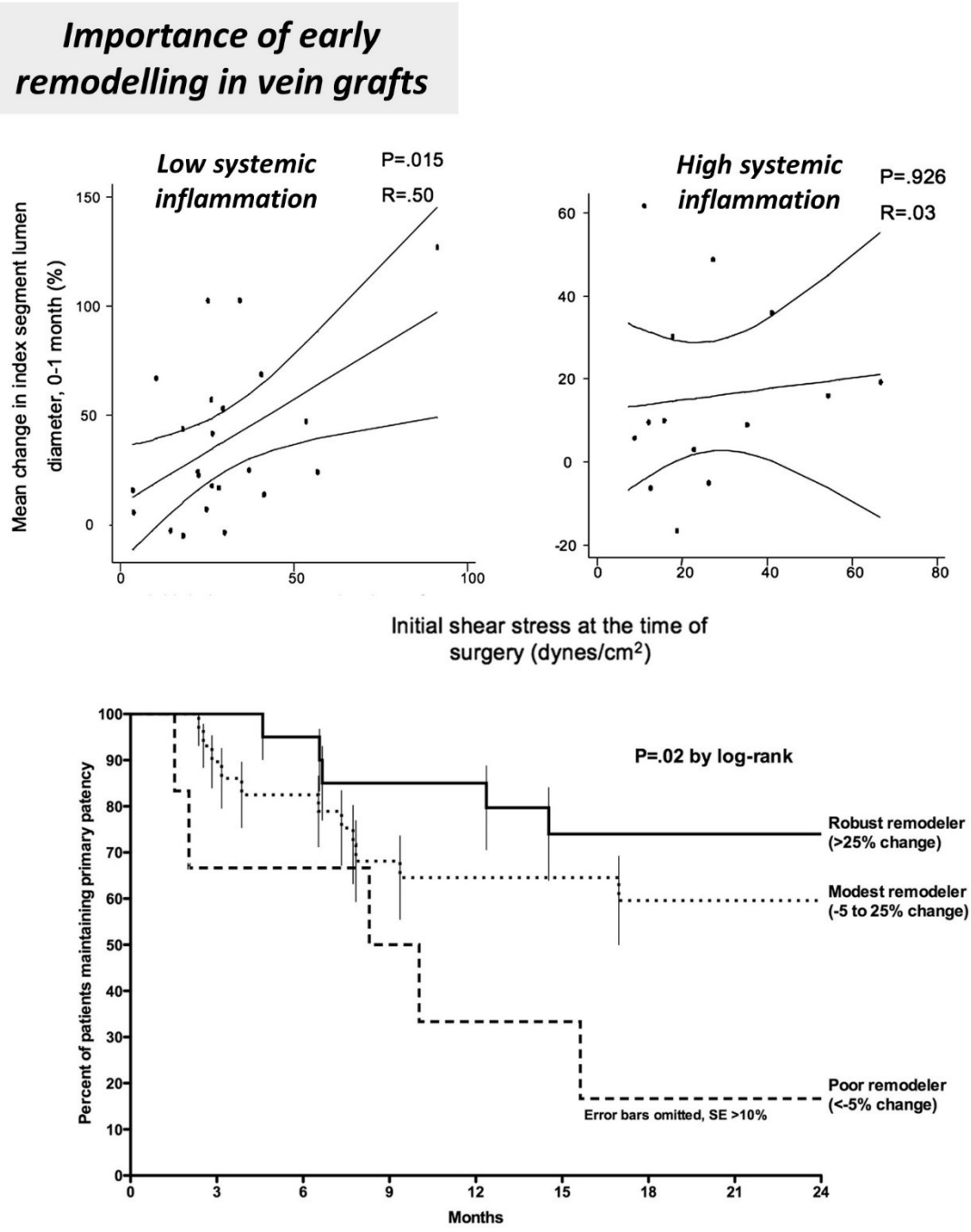


Figure 1.7 Importance of early remodelling in vein grafts. Top panels adapted from Owens et al., 2008[81] using measurements obtained through imaging of patients undergoing lower extremity bypass surgery. Correlations between 1 month changes in lumen diameter of vein grafts and shear stress in the graft at time of implant in cases of high or low systemic inflammation (via hsCRP level) are shown. Bottom panel, adapted from Gasper et al., 2013[69], shows patients separated into 3 groups by their magnitude of 1 month remodelling in lower extremity vein grafts and the percentage of each group that maintains graft patency over a 2 year period.

In the worst vein graft scenarios, outward remodelling may be cancelled out, or exceeded by inward remodelling stimulated by an inflammatory, pro-

mitogenic environment. The correlation of early luminal dilation and shear stress is not apparent in patients with high levels of systemic inflammation – a high proportion of those requiring a SVG procedure[81]. As described in the setting of atherosclerosis, platelets and leukocytes attach to blood vessels with a disrupted endothelial layer, secreting growth factors and cytokines which increase the proliferative and migratory stimuli received by VSMCs. As increased shear stress also induces endothelial cells to secrete cytokines and growth factors such as IL-1α and PDGF-BB[82], a lack of graft adaption to higher blood flow is likely to lead to expansion of VSMCs in the intima, even after regeneration of the endothelium.

Other influences on vein graft failure appear tied to differences in venous and arterial VSMC. Saphenous vein VSMCs (svSMCs) demonstrate an inherently more proliferative, more invasive (greater ECM degradation) and more apoptosis-resistant phenotype *in vitro* as compared to VSMCs derived from the alternative, more reliable graft source – the internal mammary artery[83,84]. This manifests in greater activation of ERK1/2 and Akt signalling in response to culture serum and growth factors. Lineage tracing of mature VSMCs via SMMHC labelling indicates venous and arterial VSMCs provide the bulk of neointimal hyperplasia cells at anastomosis sites but only venous VSMCs contribute to middle lesions[85]. Endothelial cells which have undergone phenotypic changes to a mesenchymal state (via the endothelial-mesenchymal transition) have also been proposed as a source of VSMCs in the SVG neointima[86] though this may be due to non-specific labelling technique[85]. An inherently greater proliferative capacity in svSMCs relative to arterial SMCs is therefore a likely factor in the accelerated rate of intimal thickening and atherosclerosis seen in failing vein grafts in the period after the first two years of implantation[87].

Loss of endothelium is a driving feature of VSMC proliferation and initiating factor causing vein graft failure. This is also true of stent implantation – another typically-used surgical intervention to counter atherosclerosis[88]. Again, in this case the endothelium damaged by an acute surgical event, this time the expansion of a scaffolding device within the occluded, stenosed vessel to attempt to restore and maintain vessel patency. Expansion of the stented vessel wall into the lumen through neointimal hyperplasia formation within and

over the stent struts is referred to as restenosis. Thrombotic events in the years after stent implantation are associated with the incidence of restenosis in the vessel[89] hence reducing VSMC proliferation is vital. Current state-of-the-art stent designs achieved over the last two decades have provided traits such as biocompatibility, release of anti-proliferative drugs (drug-eluting stents) and ultrathin construction to achieve improvements in incidence of subsequent cardiovascular events[90,91]. Improvements in clinical outcomes are such that this area is no longer a high priority target for gene therapy[92] so this pathology will not be detailed further here.

1.1.5 VSMC proliferation during pulmonary hypertension

A quite distinct vessel wall remodelling context in which the contribution of phenotypically modulated VSMCs appears key is pulmonary hypertension (PH). This umbrella term describes a range of chronic remodelling events in arteries, veins and smaller vessels that result in elevated blood pressure in the pulmonary vasculature. Such events are risk factors for development of wider CVDs and are often fatal. Pulmonary arterial hypertension (PAH) in particular has a mortality rate of ~50% within 5 years of diagnosis, a rate that has little improved despite therapeutic advances and incremental improvements in our understanding of the initiation and progression of the pathology[93]. The underlying blood pressure elevation is brought on by chronic increased medial thickness and vessel stiffness but specific pathological steps in human PH settings have been challenging to ascertain due to difficulty in obtaining patient lung tissue, the distinct nature of human PH compared to animal models and a variety of suspected root causes[94,95]. For instance in PAH, various routes to pathogenesis include infectious disease, hereditary factors, congenital heart defects and as yet poorly-defined (characterised as idiopathic PAH or IPAH). Common to all these routes to PH is vessel wall remodelling likely brought about through a phase of hyperproliferative, apoptosis-resistant phenotypes shown by VSMCs and ECs. This produces the hallmark remodelling events – a medial layer dense with hypertrophic VSMCs and ECM, alongside EC-rich lesions (characterised as plexiform lesions) in the intimal layer. As with the previous described vasculoproliferative diseases, druggable targets are in high

demand and the study of VSMC regulators is a high potential area in which to identify them.

Pioneering studies in the bovine pulmonary artery show that specific subsets of VSMCs may be pre-disposed to a proliferative phenotype and responsible for medial thickening in PH. Neonatal calf arteries show mature ACTA2+, SMMHC+ VSMCs that are segregated by presence or absence of the contractile marker meta-vinculin - with hypoxia-induced proliferation seen nearly exclusively seen in the latter[96]. This is part of the stratification of VSMC populations the neonatal calf model of pulmonary artery development, with the layer adjacent to the lumen and patches within the deeper layers showing an absence of ACTA2 and SMMHC expression(Figure 1.8)[97]. This suggests medial expansion in PH may be due to proliferation within specific patches or layers of immature VSMCs, in contrast to the clonal expansion seen in atherosclerosis and acute injury-induced neointima formation. The contribution of VSMC proliferation to medial thickening appears limited to an early phase of pulmonary hypertension in several animal models[98]. For instance, a 1-week peak is seen in hypoxia-induced hypertension in the neonatal calf model[96] and a 3-week peak, with particularly low numbers of proliferating VSMCs, in the “Sugen” rat model which is morphologically similar to human PH[99]. In addition, staining for markers of proliferation in the pulmonary vasculature of PH patients indicates little to no VSMC proliferative activity in the final stages of disease[100,101]. Therefore, the therapeutic window to block VSMC proliferation in PH remodelling may be quite small. Nonetheless therapies with pro-apoptotic, anti-proliferative effects on responsive layers of VSMCs in the arterial media are desirable to potentially restore vessel functions in PH.

VSMC stratification in the pulmonary artery

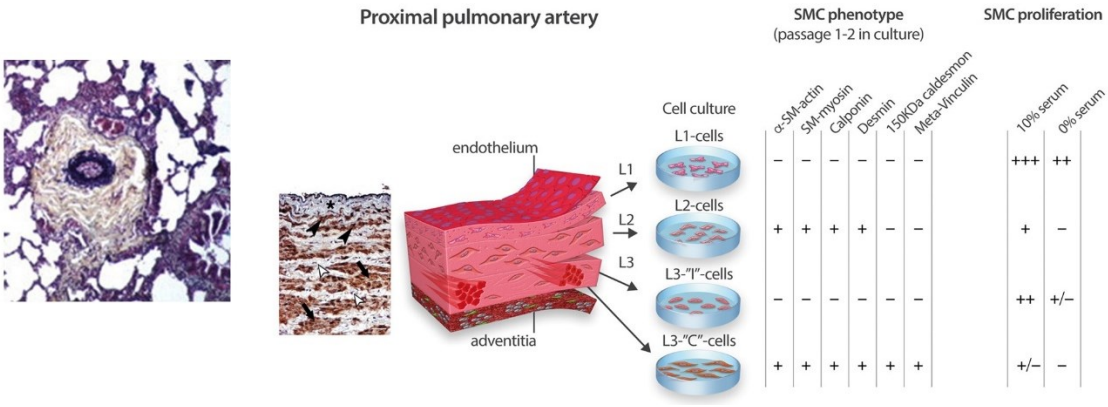


Figure 1.8 VSMC stratification in the pulmonary artery. Left hand panel shows a pulmonary artery from a PAH patient with thickening of all layers of the vessel wall - adapted from Stenmark et al., 2009[94]. Elsewhere the stratification of medial VSMCs in bovine pulmonary arteries is shown in terms of ability to express VSMC markers and proliferate after isolation and culturing *in vitro* – adapted from Stenmark et al., 2018[98]

As with the other described vasculoproliferative pathologies, in PH a disrupted endothelial layer and pro-inflammatory environment are early activators of VSMC proliferation. The initiating causes of endothelial disruption here are suspected to be increased shear stress, oxidative stress and hypoxia. Rat models of PH show the pulmonary artery is particularly susceptible to infiltration of monocytes, dendritic cells and secretion of pro-inflammatory factors in response to hypoxia[102]. This includes chemokines and cytokines such as MCP-1 and IL-6 which as described already are known to activate VSMC proliferation and migration[39,40,44]. Again, the convergence of pro-inflammatory factors and growth factors from an activated endothelial layer is likely to influence VSMC proliferation so could be targeted to redirect pulmonary artery VSMCs (PASMCs) within a PH context. This has been explored somewhat for FoxO1, a transcription factor involved in VSMC differentiation and associated with expression of the cell-cycle inhibitor CDKN1B and repression of cyclins B1 and D1[103]. The FoxO family in general are known to respond to both cytokine and growth factor signalling whilst FoxO1 appears particularly relevant to PH as it was found to be both downregulated and deactivated via phosphorylation in the medial layer of human IPAH and rat PAH models. *In vitro*, stimulation of PASMCs with

cytokines such as TNFα or IL-6 and growth factors such as PDGF-BB or fetal calf serum activated a range of kinases which converged to phosphorylate FoxO1. The centrality of this to PH pathology is highlighted by the targeted knockout of FoxO1 within VSMC which was sufficient to induce PH characteristics such as elevated pressure and remodelling in the right ventricle without hypoxia treatment. Reinstatement of FoxO1 activity has anti-proliferative, pro-apoptotic effects on wild-type PASMCs *in vitro* and could reduce detrimental pulmonary remodelling *in vivo*.

The distinct metabolic environment seen in PH may offer another VSMC-relevant therapeutic avenue. Evidence suggests hypertensive vessels have an increased reliance on glycolytic processes rather than mitochondrial oxidation – a process similar to that seen in tumour cells which is thought to provide greater ability to maintain hyperproliferative, anti-apoptotic phenotypes[104]. This metabolic reprogramming is induced by ECM stiffness which also leads to increased PASMC and EC proliferation[105]. Using siRNAs against a key enzymatic component of the reprogramming - glutaminase, glycolytic processes can be reduced to ameliorate disease progression in a PH rat model. Interestingly, TGF-β involvement is key as it can promote both ECM remodelling and this shift to glycolysis. This may provide some explanation for the mechanism behind the majority of cases of inherited PAH (and a lesser proportion of IPAH) which are linked to defects in the TGF-β receptor bone-morphogenic protein receptor [95]. If these mutations in TGF-β signalling provide a boost in metabolic support for hyperproliferative vascular cells then this could provide a target for a future treatment strategy for a wide range of PH variants.

Altogether, targeting VSMCs to address large vessel PH pathologies still requires more foundational studies to tease out their specific contributing role. For instance, the time-frame of their medial proliferation in which any such intervention may be effective is still under investigation. Despite this, promising initial strategies for VSMC-based interventions from various studies have been described here which will hopefully encourage much-needed further active exploration and cataloguing of novel regulators in pulmonary VSMC

1.2 Molecular control of VSMC pathological states

1.2.1 Control of VSMC maturation

So far various scenarios of vessel wall remodelling have been described in which loss of VSMC differentiation leads to a variety of phenotypes. Of these, a principal target of therapeutic approaches for CVDs is the proliferation of VSMCs, representing a core element of vasculoproliferative diseases and a common final endpoint for various dedifferentiation signals and pathways[6]. Circumstances in which VSMCs undergo dedifferentiation often lead to increased proliferation and vice-versa, though with notable exceptions[106]. In vasculoproliferative diseases VSMC dedifferentiation and proliferation are intertwined. The molecular mechanisms at this intersection of phenotypes are therefore relatively well-studied with the aim of finding routes to re-establish VSMC maturation or reduce VSMC proliferation within the vessel wall.

The identity of contractile, mature, differentiated VSMCs is defined through the expression of marker proteins such as α-smooth muscle actin (α-SMA aka SMA or ACTA2), smooth muscle myosin heavy chain (SMMHC aka MYH11), SM22α (aka TAGLN), h1-calponin and caldesmon. Of these, SMMHC is the most specific for VSMCs with others expressed to a greater extent in other cell types such as fibroblast or cardiac mesenchymal cells. However, their profile together marks out mature, differentiated VSMCs that possess the high cytoplasmic density of myofilaments required for contractility[107]. This profile generally coincides with a low amount of rough endoplasmic reticulum (and thereby protein synthesis), low level of ECM secretion, a distinct metabolic profile and exceedingly low levels of proliferation together representative of a quiescent cell type.

Though a definitive model remains elusive, the regulatory network controlling maturation of VSMC identity has been outlined over the last 20 years (Figure 1.9). An initial success was the identification of a master transcription factor, Myocardin, which is sufficient to induce VSMC identity - defined by expression of ACTA2, SMMHC and SM22α - when transfected into 10T1/2 cells or fibroblast cell lines[108]. Transcription at these and several other VSMC marker gene promoters, is dependent on DNA motifs known as CArG boxes (following the pattern: CC(A/Tx6)GG). Myocardin acts as an activator of myogenic

transcription and binds to these promoters indirectly via the ubiquitously-expressed serum response factor (SRF) which interacts with CArG boxes as a homodimer[109]. The multifunctional tumour-suppressive protein, PTEN also binds SRF and together with Myocardin is required for SRF-CArG interactions in the promoters of several VSMC marker genes[110]. Myocardin, SRF and PTEN are also expressed across various other cell types so specific activation of VSMC marker genes within VSMCs cannot always be explained by these factors alone. Further, SRF-CArG binding events also activate transcription of widely utilised genes, including c-fos which is utilised during proliferation, which seemingly contrasts with the phenotype of a quiesced, differentiated VSMC. Notably, the number and precise spacing of CArG boxes is distinct within some VSMC marker gene promoters, appearing to confer some specificity by allowing assembly of multiple Myocardin-SRF complexes[111]. This CArG motif spacing is absent at non-VSMC SRF-bound promoters, including c-fos, however many VSMC marker genes also have only a single CArG box. Because of this, several studies have sought to identify other influences which induce SRF-CArG binding specifically at VSMC marker loci.

Further explanation of VSMC-specific expression is provided by epigenetic factors which control the accessibility of chromatin to DNA-binding proteins. Chromatin structure is dependent on histones, proteins which assemble into nucleosome complexes that act as storage units for transcriptionally inactive regions. An array of post-translational modifications of histones are associated with “condensed” (inactive, nucleosome-bound) or “relaxed” (active, nucleosome-released) chromatin regions[2]. In VSMCs, contractile marker genes are enriched with particular histone modifications that indicate relaxed chromatin in close proximity to CArG boxes within their promoters[112]. This contrasts sharply with opposing condensed chromatin modifications deposited near these promoters in other cell types including ECs. Histones with di-methylation of histone 3 lysine (H3K4diMe), are VSMC-enriched and interact with Myocardin-SRF complexes and may improve SRF-CArG binding and activation of VSMC marker genes. Further modifications enriched at VSMC marker promoters in VSMCs include histone acetylations which typically are associated with chromatin relaxation[2]. The deposition of these acetylations is dependent on SRF-CArG binding. This implicates Myocardin-SRF promoter

binding as a necessary precursor to histone acetylation and chromatin relaxation. This may be implemented via the identified interaction of Myocardin with histone acetyltransferase enzymes such as p300[113].

Differentiated VSMCs also appear to use specific histone variants as components in the nucleosomes surrounding VSMC marker gene promoters. The presence of one such variant, H2.AZ, is strongly correlated with expression of VSMC marker genes as identified through transcriptional profiling of VSMCs at the single cell level[114]. H2A.Z is implicated in recruitment of MED1 a DNA-binding factor known to promote chromatin relaxation. It also recruits, SMAD3, a transcription factor and primary effector of TGF-β stimulated VSMC differentiation, which can bind Myocardin resulting in a synergistic increase in the expression of VSMC contractile marker genes[115]. Intriguingly, SMAD3 is at the centre of a network of TGF-β-induced genes in VSMCs which are linked to incidence of CAD via genome-wide association studies (GWAS – a method identifying enrichment of phenotype-related mutations within a given population)[116]. This implicates SMAD3 - and H2.AZ by association - as critical determinants to the incidence of CAD. This demonstrates the influence epigenetic factors may have on disease outcome. Indeed, both SMAD3 and Myocardin are recruited to VSMC marker genes via a particular corresponding VSMC-specific histone type (H2A.Z) or modification (H3K4diMe).

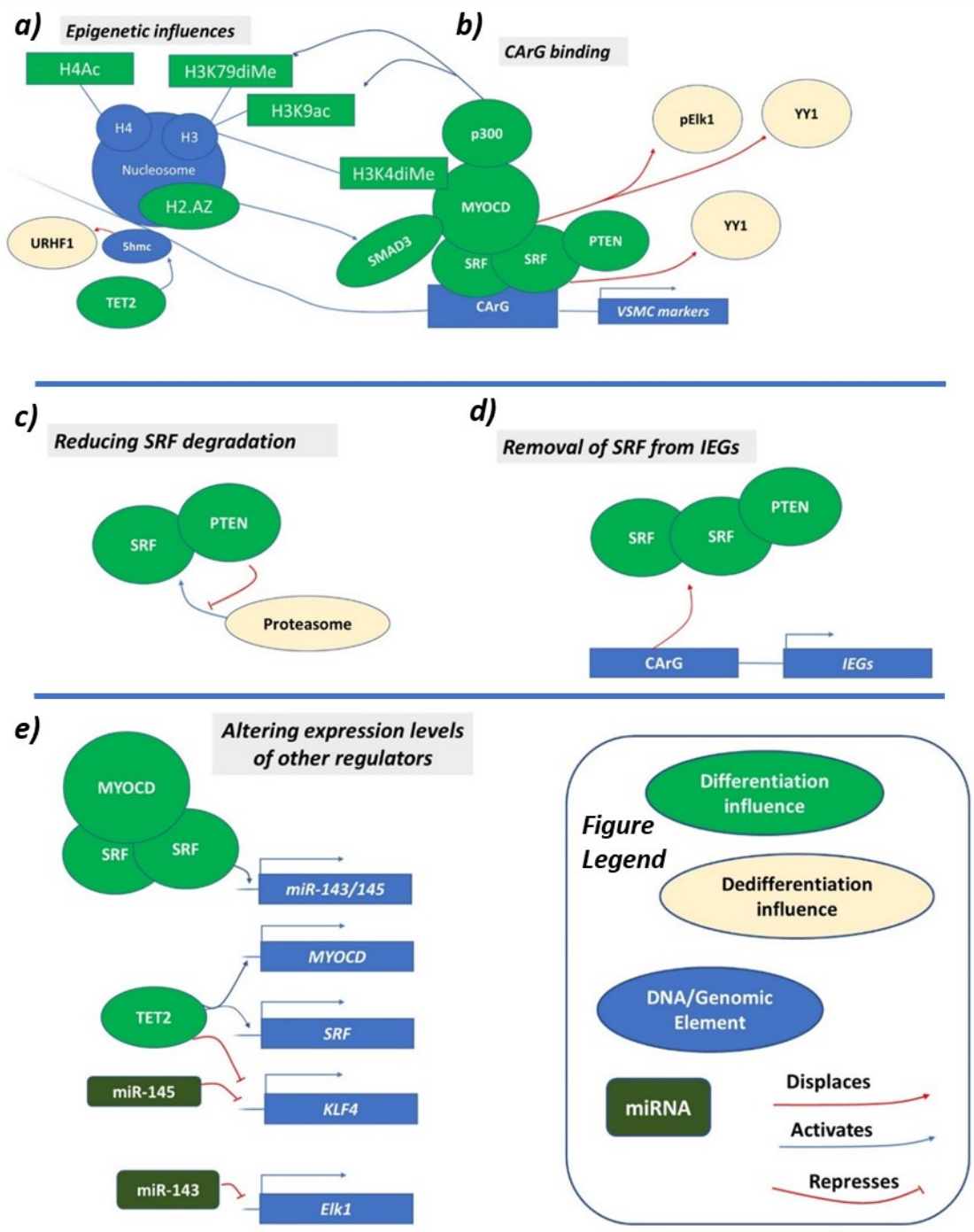


Figure 1.9 A Myocardin-centred view of mechanisms controlling VSMC maturation. a) + b) Formation of a myocardin-containing complex at VSMC marker gene promoters leads to both changes in histone signatures associated with chromatin relaxation as well as displacement of transcription factors that cause dedifferentiation of VSMCs. c) Proposed role of the myocardin co-factor PTEN as a modulator of SRF availability. d) PTEN also has a role in redirecting SRF away from immediate-early gene (IEG) promoters that initialise VSMC dedifferentiation and proliferation (e.g. *c-fos*). e) The capacity of VSMC maturation factors to support expression of other maturation factors and inhibit expression of dedifferentiation factors (KLF4/Elk1).

Promoter-binding events need accessible DNA so chromatin relaxation may be a prerequisite for other VSMC maturation mechanisms. For instance, H3K4diMe and H4ac deposition at VSMC marker promoters occurs independently of SRF-CArG binding suggesting this likely occurs beforehand. Any VSMC-specific mechanisms for depositing such modifications are therefore of high interest to explain how this happens. A recently discovered example is TET2, which has enriched expression in contractile VSMCs and oxidises 5'-methylcytosine DNA bases to 5'-hydroxymethylcytosine. This reaction not only primes DNA for demethylation but also appears to shift histone methylation away from repressive H3K27me3 marks and towards H3K4me3 marks associated with chromatin relaxation[117]. Genome-wide upregulation of 5-hmc marks was shown in this study as a hallmark of contractile VSMCs but TET2 also binds specifically to the Myocardin, SRF and SMMHC promoters. As with Myocardin, overexpression of TET2 is sufficient to activate expression of these markers in fibroblasts cell lines. Identification of other histone modifying enzymes or control mechanisms specific to activation of VSMC maturation is an active area of study with few others known so far.

Aside from activation of CArG-dependent contractile genes, the Myocardin-SRF complex also upregulates expression of another class of regulators involved in ensuring VSMC contractility. RNA products from these genes are not translated into proteins, but instead are processed to create short, 22-25 nucleotides long, single stranded molecules termed micro RNAs (miRNAs). These miRNAs guide the RNA-induced silencing complex (RISC), to specific mRNAs through Watson-Crick base pairing to effect translation control or RNA degradation and modulation of expression levels. Their ability to target numerous mRNAs means they are often central components of transcriptional regulatory networks controlling phenotypes – though their effects are thought to generally be more fine-tuning of expression than large changes. In mature VSMCs, two particularly conserved miRNAs, miR-143 and miR-145, are known components of the Myocardin-SRF network and support contractility in the vessel wall[118,119]. Transcribed as a pair in a VSMC-enriched manner, they target a range of mRNAs that enable the support of Myocardin expression, organisation of actin stress fibres and temperance of Angiotensin II levels in

the circulation to prevent VSMCs from becoming desensitised to this contractile signalling molecule. Loss of this miRNA pair leads to a thinner arterial smooth muscle layer and defects in contractility.

Over the last two decades the field has elucidated several influences on VSMC identity including the Myocardin-SRF complex, the permissive chromatin state created by epigenetic factors and more recently the influence of miRNA regulation. The mechanisms described above represent focal points of the cellular machinery advancing the maturation state of VSMCs in the vessel wall. However, another inherent feature of VSMCs compared to cardiac or skeletal muscle is their duality; their ability to repress expression of these contractile markers and exit quiescence when circumstances require. Reversible activation of a set of factors that oppose and dismantle this pro-differentiation machinery is therefore another component of the VSMC identity. Elucidating the nature of the interaction between these two sets of opposing factors is key to find therapeutic avenues to manipulate VSMCs directly in the vessel wall.

1.2.2 Loss of mature VSMC identity – phenotypic switching

In vivo evidence that transcriptional downregulation of contractile VSMC marker genes was key to the loss of contractility and quiescence seen during VSMC phenotypic switching (or dedifferentiation) was found decades after initial observations of this transition *in vitro* in the early 1970s[1]. In a study by Regan et al., the promoters of the VSMC markers α-SMA, SMMHC and SM22α were variously used to drive expression of the β-galactosidase reporter in transgenic mouse lines. Loss of VSMC maturation was stimulated via removal of the carotid artery endothelium by wire injury. By 7 days this led to markedly reduced mRNA for the contractile markers in the media and intima for all mouse lines. Crucially this was accompanied with loss of β-galactosidase signal, indicating that loss of contractile genes was at least in part through loss of transcription[120]. This did not disprove loss of mRNA stability for VSMC marker genes as another possible explanation, but it sanctioned new avenues of investigation into mechanisms acting to repress their promoters. These often were discovered in parallel with mechanisms activating CArG-dependent

VSMC marker promoters described in the previous section. Though the exact configuration of elements remains to be identified and put in place, we now have a grasp on the repressive influences that mirror pro-contractile VSMC mechanisms. As the Myocardin-SRF complex is a focal point for activation of VSMC contractile genes, it is perhaps not unexpected that the disruption of this pairing and loss of this expression underpins the opposing VSMC dedifferentiation.

One route to disruption of VSMC marker genes is the binding of repressive elements to the SRF homodimer or removal of the homodimer from VSMC marker CArG boxes. An example identified early on was the ETS-domain family member, Elk1, which is phosphorylated (to pElk1) to become capable of binding to the SRF homodimer within 24 hours of response to PDGF-BB[121]. This growth factor is one of the most studied VSMC dedifferentiation signals (Figure 1.10), secreted not only during formation of nascent vessels but also by ECs and myeloid-derived cells in the vessel wall activated by pathogenic stimuli. Oxidised phospholipids are another commonly used VSMC dedifferentiation signal to model VSMC phenotypes in CAD and also induce phosphorylation of Elk1[122]. The pElk1-SRF interaction occurs at the same SRF domain that binds Myocardin, thereby allowing displacement of the VSMC master transcription factor. pElk1 was later co-immunoprecipitated from PDGF-BB-treated VSMCs in complex with Krüppel-like factor (KLF)4, a prominent pluripotency-inducing transcription factor[122]. As this interaction occurs separately to the SRF interaction it suggests existence of an SRF-pElk1-KLF4 complex. As PTEN could not be found in such complexes, the formation of the SRF-pElk1-KLF4 complex likely opposes formation of the SRF-Myocardin-PTEN complex used during VSMC maturation[110]. SRF occupancy of VSMC marker CArG boxes is eliminated, or at least drastically reduced, during KLF4 overexpression as well as treatment with PDGF-BB or oxidised phospholipids. This indicates that pElk1 splits the SRF-Myocardin complex and also displaces it from VSMC marker promoters by the resulting complex[123].

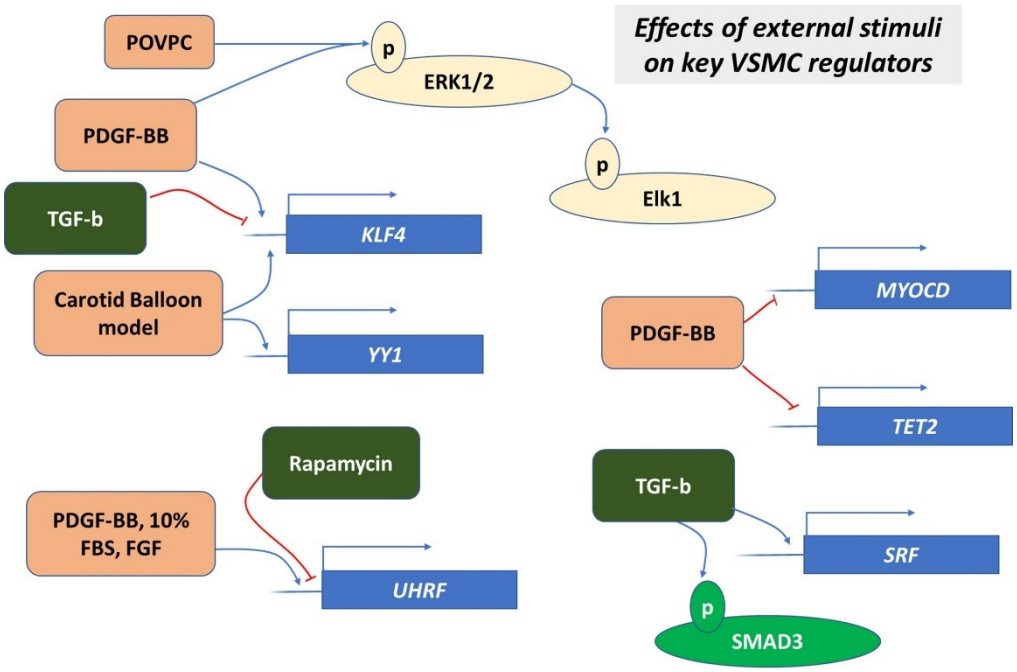


Figure 1.10 Effects of canonical external stimuli on key VSMC regulators. Biochemical stimuli such as POVPC and PDGF-BB activate the ubiquitous ERK1/2 pathway or support expression of transcription factors like KLF4 and UHRF that counter the action of Myocardin complex and TET2 respectively (both of which are also downregulated by PDGF-BB). Models of vessel wall injury such as the carotid balloon model typically have similar effects on such dedifferentiation transcription factors. Conversely, TGFβ and Rapamycin are often used to counter these dedifferentiation signals or develop a more mature VSMC state. Promoters of VSMC maturation are shown in dark green whilst inhibitors are shown in orange.

Other factors disrupt SRF-Myocardin-CArG interactions through direct Myocardin binding. Yin Yang 1 (YY1), known to be induced during vascular injury, inhibits transcription at α-SMA, SM22α and SMMHC promoters by interacting with and competitively displacing Myocardin from SRF homodimers[124]. SRF-CArG binding is also reduced through YY1 blocking a CArG box within the SM22α promoter through direct binding. A similar Myocardin-SRF displacement is performed by Transcription factor 21 (TCF21), which like SMAD3 has an identified link to CAD incidence through GWAS[125]. TCF21 binds to DNA regions near SRF-bound CArG boxes and these two binding events appear to act largely in opposition, with loci generally showing enrichment of one or the other factor exclusively. This suggests TCF21 binding opposes SRF activity and vice-versa. TCF21, similar to YY1 can also bind directly to Myocardin to displace SRF.

The expression levels of Myocardin are also repressed by several factors involved in VSMC dedifferentiation including KLF4, YY1 and TCF21. Conversely, pro-differentiation factors miR-143 and miR-145 are transcribed through Myocardin activity to repress Elk1 and KLF4 expression respectively[126]. This suggests existence of a negative feedback mechanism whereby disrupting SRF-Myocardin complexes, Elk1 and KLF4 can repress Myocardin and the miRNAs thus derepressing their own expression levels. Altogether a basic Myocardin-centred model of VSMC dedifferentiation can be put forward where the removal of SRF from both VSMC marker CArG boxes and Myocardin is enacted by pElk1-KLF4 displacement. This is maintained by activity of factors like YY1 and TCF21 which sequester Myocardin to prevent any further competition with pElk1 for SRF binding. Loss of Myocardin activity and expression provides further support to dedifferentiation as this leads to loss of miR-143 and miR-145 and so derepression of the VSMC dedifferentiation factors (Figure 1.11).

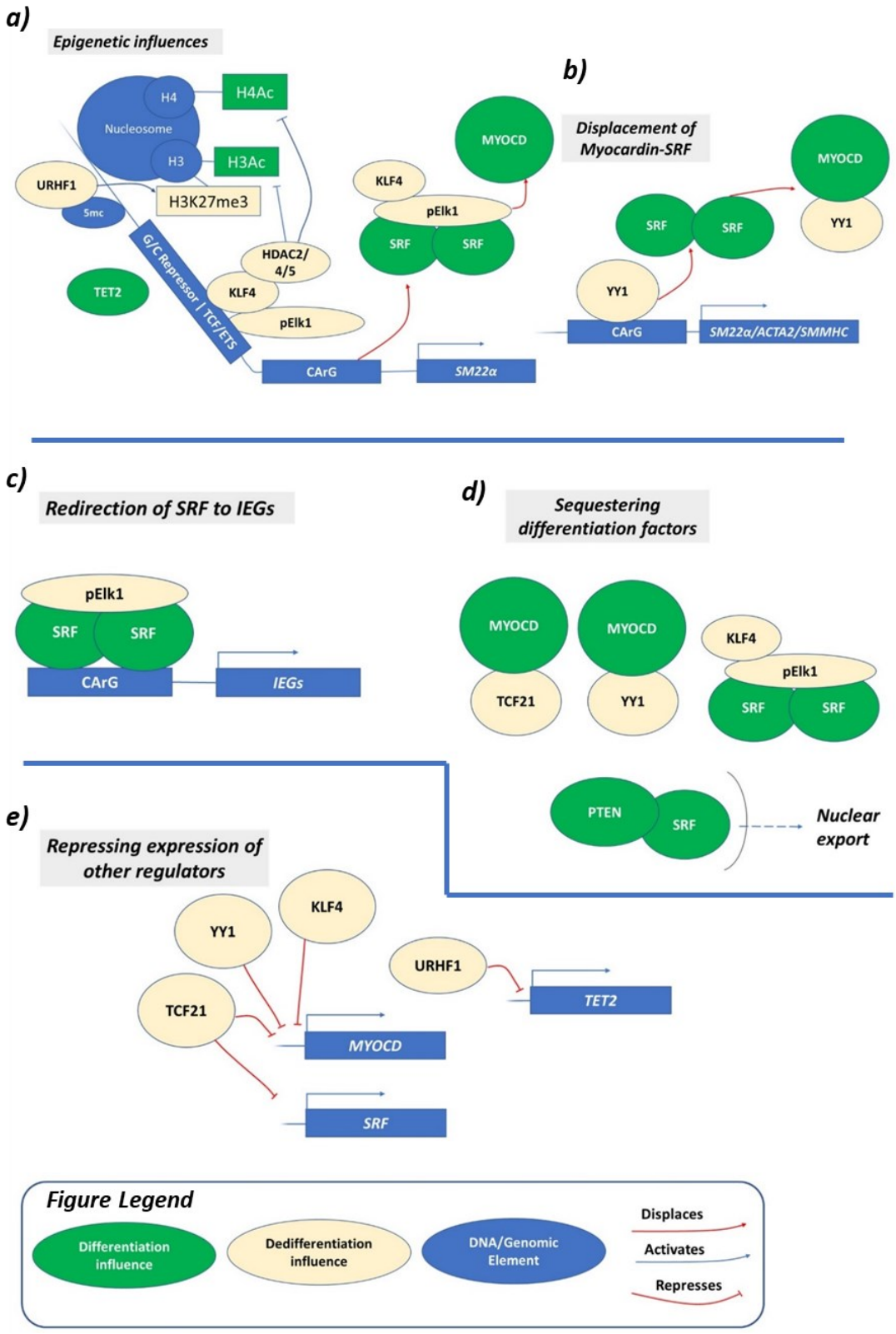


Figure 1.11 A Myocardin-centred view of mechanisms leading to VSMC dedifferentiation. a + b) Displacement of Myocardin complexes by factors such as KLF and phosphorylated Elk1 (pElk1) are thought to lead histone signature changes (from relaxed to condensed chromatin) through loss of Myocardin-interacting epigenetic modifiers and gain of KLF4-interacting deacetylation enzymes. c) pElk1 is thought to redirect the promoter binding activity of SRF towards immediate-early genes (IEGs)

which build the initial VSMC dedifferentiation response and wider associated transcriptional activation. d) The sequestration of Myocardin and SRF is also a key theme in the literature describing loss of VSMC maturation state. e) Many of these same sequestering factors also demonstrate capacity to repress expression of VSMC maturation factors.

Disruption of this pairing also directly contributes to adoption of proliferative VSMC phenotypes as SRF, in combination with pElk1, has another role in activating transcription of immediate early genes (IEGs) such as c-fos[127]. This class of genes constitute an initial transcriptional burst involved in multiple stimuli response pathways include cell cycle activators. Removal of Myocardin from SRF and SRF from contractile gene promoters may free SRF to bind to these other promoters and actively contribute to the dedifferentiation process. One outlined mechanism where SRF could be redirected to IEGs was postulated during characterisation of PTEN as a co-factor. In PDGF-BB-treated VSMCs, this PTEN-SRF complex was observed to translocate out of the nucleus by an unknown mechanism and PTEN also protected SRF from proteasomal degradation[110]. With a resulting lower pool of nuclear SRF, SRF-CArG binding shifted towards IEG promoters rather than contractile gene promoters, suggesting lower nuclear concentrations of SRF may favour the former. This also fits with an observed suppression of SRF transcription levels by the pro-dedifferentiating factor TCF21[125], and on the opposing side the induction of SRF by TGF-β[128] both of which may alter intracellular redistribution of SRF. Additionally, depletion of SRF leads to loss of miR-143 and subsequent release of the PTEN-repressive miR-21[129]. Redirection of SRF from contractile marker promoters therefore represses PTEN expression, which could feed forward to reduce capacity for PTEN-SRF to form on these contractile markers. The switching of SRF between activation of quiescent or synthetic genes could be beneficial to VSMCs in terms of energy expenditure. It may be more efficient for the cells to direct their efforts solely towards either growth and repair, or the upkeep of the cytoskeletal architecture and ion balance required for contractility.

The pattern of histone marks used at VSMC marker gene promoters is altered by dedifferentiation signals and this likely aids the transcriptional repression at these sites. The KLF4-pElk1 complex is known to bind to a conserved G/C-rich repressor element in the promoters of most CArG-

dependent VSMC marker genes. The effects of this interaction were modelled using carotid ligation injury in mouse showing that the resulting recruitment of histone deacetylase enzyme (HDAC)2 removes H3Ac marks associated with chromatin relaxation. Earlier *in vitro* work also demonstrates KLF4-dependent loss of H4ac, recruitment of HDAC2 and HDAC5 and loss of VSMC marker gene accessibility with either KLF4 overexpression[112] or addition of oxidised phospholipids[122]. A HDAC2 interaction has also been identified for TCF21 which may complement the KLF4-based recruitment and deacetylation[130]. Of note the H3K4diMe mark, left at VSMC marker genes specifically within VSMCs, is unaffected by KLF4 overexpression suggesting it could remain in place to allow a route back to maturation for dedifferentiated VSMCs[112]. A mechanism that counters the activity of the VSMC-enriched TET2 epigenetic marker enzyme has recently been ascribed to UHRF1[131]. This factor is one of the few genes involved in epigenetic modification that are upregulated by PDGF-BB. As a target of miR-145 it increases in abundance upon loss of this miRNA with dedifferentiation and represses TET2 expression. This results in loss of the 5-hmc promoter marks suggested to be a key initiator of the VSMC maturation cascade as URHF1 recruits methylase enzymes to leave 5-mc marks. Such marks are subsequently enriched at VSMC marker genes leading to increased H3K27me3 deposition associated with chromatin condensation.

Several of these dedifferentiation factors (PTEN, YY1, TCF21, UHRF1) have been discovered in the last 5 years so the disparate dedifferentiation mechanisms have not yet been integrated into a unified model. Building such a model is still required to address the current absence of therapeutics that could target VSMC dedifferentiation and/or proliferation. The field is also still seeing a steady influx of studies describing new regulators or entirely novel classes of regulators (such as miRNAs) that control this process, several of which underline previously missing layers of mechanistic detail. Cataloguing and understanding of the factors driving VSMC remodelling in an ever more comprehensive manner is therefore still a key aim. This is especially true considering we are still attempting to dissect previous unsuccessful attempts to develop VSMC therapeutics[66] and that distinct tissue remodelling environments will have distinct influences on VSMC dynamic

1.3 Therapeutic opportunities from VSMC lncRNAs

1.3.1 General traits and functional aspects of lncRNAs

The previous section highlights the complexity of VSMC dynamics seen in vasculoproliferative tissue remodelling events. This suggests a regulatory network capable of promoting distinct pathways in particular vessels at particular times in control of VSMC behaviour. The heterogeneity of VSMCs between and within vessel walls also supports this conclusion. To understand such a network requires full identification of the constituent components and definition of their function.

Completion of the human genome project nearly two decades ago marked the end of a period in which the majority of the ~20,000 human protein-coding genes (PCGs) were annotated and confirmed that these genes were a tiny fraction of the genome. The functionality of the remaining fraction has since been a subject of intense debate - fuelled by the advent of RNA sequencing (RNAseq) technology which allowed an unbiased view of transcription across the human genome. Subsequent surveys provided by the ENCODE consortium provided estimates that 75% of the genome could be transcribed and that 80% had some functionality – a number arrived at by considering all loci that could be either transcribed, bound by transcription factors or involved in determining chromatin conformation[132,133]. A counter-argument raised was that this definition of functionality is broad, assuming for instance that a DNA binding event or produced RNA transcript is always indicative of a biological function[134]. A more conservative estimate of ~8% has been produced by considering the percentage of the human genome which is evolutionarily constrained (within eutherian mammals at least) and so likely to perform some function[135]. In terms of transcribed portions of the human genome, the 1-2% fraction encoding proteins is dwarfed by the remaining fraction broadly characterised as non-coding RNAs with often debatable functionality.

Some non-coding RNA fall within the miRNA class, including those already described to regulate aspects of VSMC biology such as miR-143, miR-145 and miR-21, and so have a defined mechanism of dampening mRNA activity via the RISC complex. Other classes include ribosomal RNAs and transfer RNAs which generally make up most of the RNA mass of the cell and have well

established roles in fundamental biology and often ubiquitous housekeeping functions. However, another class contains the remainder of non-coding transcripts of 200bp or more in length - a separate, large constituent of the transcribed genome referred to as long non-coding RNAs (lncRNAs). The number of lncRNA genes is estimated to be significantly greater numbers than PCGs at ~27,000[136] though more conservative estimates show them roughly equal[137]. They make up less of the cellular RNA mass than mRNA as they have a tendency for lesser abundance[134]. Their conservation is also distinct compared to mRNA, with selection pressure seeming to apply only to certain sections of transcript sequence for the more conserved lncRNAs while other lncRNAs are not conserved at all. The conservation constraints results in a high rate of evolutionary turnover of lncRNA sequences and exonic structures[138]. This is demonstrated in a recent survey of lncRNAs across 17 species, revealing that >70% of lncRNAs in each species originated within the last 50 million years – and approximately 1000 human lncRNAs conserved across mammals[139].

Like much of the non-coding genome, the functionality of lncRNAs was initially in doubt. However, an indication that this class has deep significance was already apparent in 1996 when the genomic deletion of the lncRNA *Xist*, showed it to be a master regulator of X chromosome inactivation – required to stabilise gene expression on this chromosome in female mammals[140]. Similarly, deletion of the *H19* lncRNA in 1995 demonstrated a role in preventing maternal expression of the neighbouring insulin-related genes[141]. Both these early examples showed lncRNAs involvement in imprinting (epigenetic silencing) of chromosomal loci - now understood as just one of the many fundamental roles that lncRNAs can perform. Subsequent years brought examples of lncRNA roles across biological contexts including regulating *Hox* genes that lay down developmental axes in the body plan[142], triggering apoptosis as part of the ubiquitous p53 tumour suppressive pathway[143] and - of particular relevance to this thesis - tissue remodelling during CVD[144,145]. A recent extensive study to collate together various human lncRNA gene maps (or annotations) identified evidence of functionality for 69% of lncRNAs through integration of conservation and GWAS data[136].

More definitive answers have begun to be produced recently by studies screening hundreds of lncRNAs simultaneously for phenotypes. In human dermal fibroblasts for example, transfection of a library of antisense-oligonucleotides was used to reduce expression of 194 selected lncRNAs, showing ~30% had either a phenotype related to cell proliferation or morphology and ~10.9% showed a phenotype of a robust molecular change[146]. Alternatively, using a CRISPR inhibition-based method to epigenetically silence ~5,000-16,000 lncRNAs within 7 cell types, Liu et al. showed that ~0.3-6% of lncRNAs were observed to robustly effect cellular growth[147]. High-throughput screens such as these are limited by necessity to observe only a certain number of phenotypic traits. Despite this they do provide initial estimates of widespread functionality. The historic characterisation of lncRNA genes as “junk” DNA is no longer applicable.

As more lncRNAs are studied, their diversity becomes more apparent. This is such that a variety of methods to attempt their classification now exist, for instance using their mechanism, subcellular localisation, genomic location relative to surrounding PCGs or chromatin signatures at their transcriptional start sites (TSS)(Figure 1.12). As relatively few have been characterised with a phenotype or mechanism compared to the tens of thousands annotated[148], our understanding of how they perform their function is ever-growing. We know of a wide range of possible mechanisms now, which can be enacted by lncRNAs in nuclear, cytoplasmic or chromatin-bound subcellular regions. At a fundamental level these include interactions with mRNA, interactions with double-stranded DNA as a “triple helix”, scaffolding interactions with regulatory proteins, hosting of small RNAs (including miRNAs) or blocking of miRNAs.

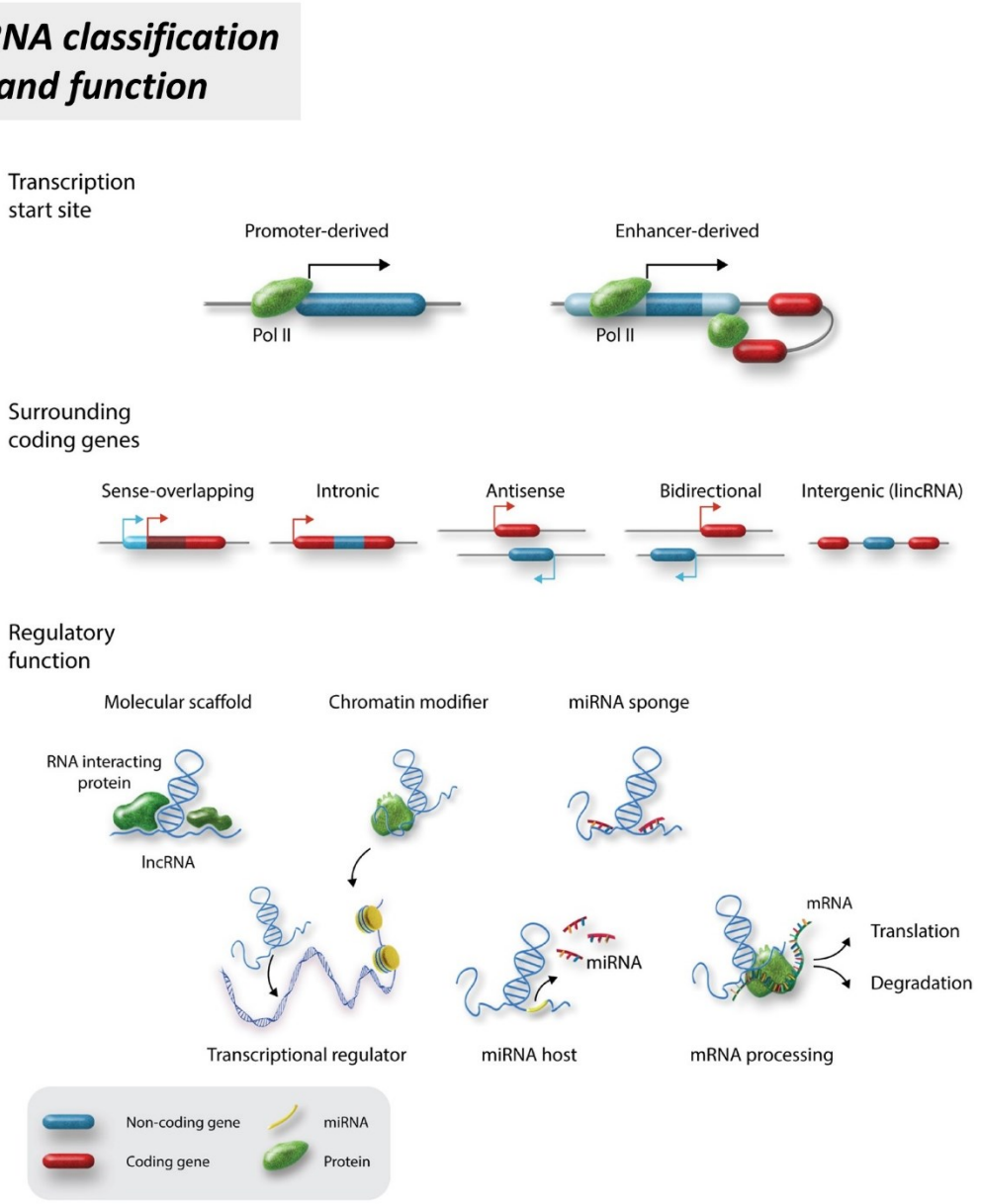


Figure 1.12 LncRNA classification and function. Adapted from Bennett et al., 2019[149]

Recent evidence also suggests many lncRNAs are translated to produce small peptides (micropeptides) generally <100 amino acids – several examples of which have some regulatory function[150]. Their small size, low abundance and often lack of conservation means micropeptides are hard to detect and so the true extent of lncRNA translation remains an open question[150,151]. Examples also exist of both lncRNA and hosted micropeptide demonstrating differing, or even opposing, functions[152] whilst micropeptides can demonstrate distinct biochemical characteristics to PCGs[153]. Such genes may therefore not simply be “typical” PCGs which have remained undetected due to their size. It is therefore likely that a population of micropeptide-

producing long “non-coding” RNAs exist, and so will be reclassified to a more appropriately named class in the near future - exemplifying the complexity often encountered in the lncRNA field.

LncRNAs are also attractive to explore because of their well-described tendency towards specific expression patterns – a feature that could be clinically advantageous. This was initially clear from attempts to reconstruct lncRNA annotation across various human tissue and cell types, showing their tendency towards tissue-specificity outstripped that of mRNA[154] and supporting earlier visualisation of lncRNA spatial separation in the mouse brain[155]. LncRNA expression profiles are also often restricted to specific developmental stages of cell lineages[156,157] so are often described as showing a high “spatio-temporal” specificity of expression. This specificity is often a bias towards higher expression within a given set of cell types rather than exclusive expression within one cell type alone[136]. As such, if describing specificity on a scale from ubiquitous (expressed in all cell types) to cell-type specific (expressed within one cell type only) lncRNAs show a stronger tendency towards the latter than mRNAs. Execution of function by lncRNAs can also occur in a cell-specific manner - the CRISPRi high-throughput screen performed by Lui et al. demonstrated that 89% of functional lncRNAs showed this function in only one of the 7 cell types tested[147]. In keeping with this, many characterised lncRNAs are involved in epigenetic regulation of chromatin states which also occurs in a cell-specific manner. Enhancers – genomic loci which can serve as platforms for transcription factors to bind and propagate changes in chromatin accessibility – are often associated with lncRNA production and such enhancer-transcribed lncRNAs are particularly cell-specific[136]. A specific expression profile is of interest as it implies that the function of such genes may show particular relevance to the biological conditions to which they are restricted[158] and could allow therapeutic targeting of such genes with minimal off-target effects or use to aid diagnosis of particular pathologies. Further, existence of lncRNAs localised mainly to VSMCs from a particular embryonic origin or stimulated with a particular cytokine or pathological stimulus may help explain heterogeneity of VSMC phenotypes in homeostasis and disease.

The sheer number of uncharacterised transcripts, coupled with examples of those driving fundamental biological processes means there is a huge opportunity for discovery of novel regulatory mechanisms and novel therapeutic targets by studying lncRNAs. This could provide much-needed detail and context to our understanding of mechanisms controlling any given cell state transition. Moreover, their specificity in terms of expression and function may also provide an advantage over therapeutics aimed at more ubiquitous targets - such as mRNAs –which could require additional targeting considerations to prevent off-site toxicity. Their relatively recent discovery means no lncRNA therapeutic yet exists, though this may change in the near future with promising translational results in large animal models – for example the targeting of *H19* in a pig model of cardiac hypertrophy[159] – beginning to emerge.

Characterisation of lncRNAs driving pathological VSMC phenotypes is still in a nascent stage. However, several notable examples found so far suggest lncRNA contribution to regulation of VSMC remodelling phenotypes is significant, demonstrating them as intrinsic components of pathways controlling VSMC proliferation, migration and apoptosis - whilst examples exploring their effects on differentiation and pro-inflammatory phenotypes are also identified. This is a product of the widespread aim within the last decade to profile lncRNA activity within VSMCs stimulated *in vitro* with established mitogenic stimuli or *in vivo* models of neointimal hyperplasia formation or hypertension. Several studies also take the approach of identifying lncRNAs which regulate - or are regulated by - known relevant transcription factors (e.g. Myocardin, SRF or TGF-β-driven SMADs). Selection of a novel VSMC lncRNA for study has generally been through identifying those with a particularly large expression change associated with pathology, enriched expression levels in VSMCs, or genomic proximity to a relevant gene of interest. Other VSMC lncRNAs were initially selected for characterisation as they had an established phenotype in another biological context. Of particular interest from a clinical perspective are those which are present in human, mechanistically characterised and have a demonstrable phenotype in an *in vivo* animal model or *ex vivo* human model.

1.3.2 LncRNAs with roles in VSMC identity

Recognition of the specific expression patterns of many lncRNAs has driven the discovery of several with a VSMC-restricted expression profile or some association with VSMC identity control (Figure 1.13 and Table 1.1). An early example was the identification of the lncRNA *SENCR*, with expression largely restricted to the cytoplasm of human umbilical venous endothelial cells and coronary artery VSMCs (CASMCs)[160]. In CASMCs, *SENCR* was shown to reduce migration and promote expression of *Myocardin* and VSMC marker genes including *ACTA2*, *SM22α* and *CNN1* via a mechanism which so far remains unexplored. This same study revealed another lncRNA with restricted expression in vascular cell types - *PEBP1P2*, later shown to have a phenotype of reducing proliferation and migration whilst promoting expression of VSMC marker genes[161]. *PEBP1P2* interacts with and appears to reduce expression of cyclin-dependent kinase 9, a little-explored member of the CDK family of cell cycle regulators. In contrast to *SENCR* and *PEBP1P2*, the lncRNA *CARMN* shows minimal expression in endothelial cells, in keeping with an identified role as host gene for the key VSMC identity miRNAs miR-143 and miR-145[162,163]. An additional, miR-independent role in preventing VSMC proliferation is established for *CARMN* transcripts via an as yet unknown mechanism[164]. Restoration of VSMC identity to regenerate blood vessel wall function is a long-held gene therapy strategy so the potential to use these lncRNAs to support this is of high interest. Indeed, a *PEBP1P2* ortholog could be overexpressed to reduce neointima formation size in rat balloon injury models[161] whilst depletion of the *CARMN* mouse ortholog exacerbates plaque formation and instability[164]. Both examples provide encouraging evidence that manipulation of lncRNAs could be used to maintain VSMC identity in pathological contexts.

These lncRNAs were initially identified in VSMCs using *in vitro* culture conditions, known to inherently induce a degree of VSMC dedifferentiation. Several other approaches to identify lncRNAs controlling VSMC identity have looked for associations with MYOCD. A study by Zhao et al. overexpressed *MYOCD,* aiming to more closely approximate the contractile phenotype seen *in vivo*[165]. This identified *MYOSLID*, a lncRNA directly regulated by the Myocardin-SRF complex that has a similar expression profile to classic VSMC

marker genes in terms of response to PDGF-BB and TGF-β. As with other lncRNAs involved in VSMC identity, *MYOSLID* has an anti-proliferative phenotype that also promotes VSMC marker gene expression. A primarily cytoplasmic lncRNA, *MYOSLID* function appears to be related to the maintenance of actin organisation which allows shuttling of pro-contractile transcription factors from the cytoplasm to the nucleus as well as the activation of SMAD2 phosphorylation within TGF-β signalling pathways. Another lncRNA with a Myocardin link is *HIF1α-AS1*, expression of which is positively regulated by Brahma-related gene (BRG)1 an epigenetic co-factor required for the effect of Myocardin on VSMC marker genes[166,167]. *HIF1α-AS1* was in turn shown to support the expression and pro-apoptotic, anti-proliferative phenotype promoted by BRG1 - though effects on VSMC marker genes and overall mechanism remain unexplored for this lncRNA.

1.3.3 LncRNA with roles in VSMCs and vasculoproliferative pathologies

Alongside their role in stabilising VSMC identity, lncRNAs also have a recognised role in the gain of synthetic, proliferative phenotypes (Figure 1.13 and Table 1.1). An early-discovered example was the lncRNA, *ANRIL*, identified via a high density of CAD-associated SNPs in the chromosomal locus from which it originates[168]. Subsequently found to be expressed widely in VSMCs, ECs and macrophages, *ANRIL* demonstrates an atherogenic phenotype that exacerbates proliferation, migration and oxidative stress in aortic VSMCs (AOSMCs)[144,169] and is anti-apoptotic within leukocytes[170]. A substantial number of the CAD-associated SNPs at this locus fall in intronic regions suggesting that the role of *ANRIL* during plaque formation may be influenced by differential splice patterns[171]. Indeed, though unexplored in VSMCs, *ANRIL* isoforms containing Alu motifs have been shown to recruit both repressive and activating epigenetic complexes to various promoters culminating in the atherogenic phenotype[170]. In contrast, circular isoforms (formed through the splicing together of linear isoforms at their extremities) have no Alu motifs and in CASMCs destabilise ribosome maturation complexes likely via direct binding, so could protect against atherosclerotic

growth through limiting protein translation[172]. An additional role of the *ANRIL* locus is also demonstrated through presence of CAD risk SNPs linked to expression of neighbouring tumour-suppressive genes[116]. The early identification of this locus via a high density of CAD risk SNPs suggests that *ANRIL* is particularly relevant to CAD progression. Indeed, no other lncRNA effecting VSMCs has yet been identified via GWAS for CAD or other vasculoproliferative pathologies. However, the density of CAD-linked SNPs at this locus and extensive studies on various *ANRIL* isoforms also indicates a plethora of regulatory mechanisms may be responsible. This has complicated any strategy to target *ANRIL* thus far and is an example of the challenges often encountered during lncRNA characterisation.

Other lncRNAs driving pathological VSMC phenotypes have been found to directly influence the activity of cell cycle components. *SMILR*, is a pro-proliferative lncRNA that is upregulated in svSMCs after IL-1α and/or PDGF-BB treatment[173]. *SMILR* expression is largely restricted to VSMCs and interacts with the mRNA of the mitotic component CENPF to stabilise it[174]. Further, SMILR is upregulated at least 20 hours prior to detectable cell proliferation which places it upstream of crucial cell cycle pathways – an example of how temporal context can highlight lncRNAs of importance. This example demonstrates firstly that VSMC-enriched lncRNAs are not solely involved in determination of VSMC identity and secondly that they may often act as cell-specific regulators of generic processes such as the cell cycle. Due to lack of an ortholog in a relevant animal model, SMILR knockdown was performed in an *ex vivo* human vein model of neointimal formation and was able to reduce the number of proliferating cells in the media by ~6-fold. Conversely, two ubiquitously-expressed, anti-proliferative lncRNAs have also been shown to control the pro-apoptotic cell cycle inhibitor p53 in VSMCs. The lncRNA *lincRNA-p21*, downregulated in mouse atherosclerosis, is thought to sequester MDM2 in AOSMCs, preventing it from degrading p53 via the ubiquitin-proteasome pathway[175]. Inhibition of this lncRNA aggravated neointimal formation post-carotid wire injury in mice with intimal/medial ratio increased ~3-fold relative to sham injury. The lncRNA *GAS5*, has also been shown to prevent p53 degradation in AOSMCs - forming a complex with both p53 and the co-activator p300[176]. *GAS5* repression reduced this binding whilst

dampening p53 levels and proliferation of VSMCs stimulated with fetal bovine serum (FBS). Other exploration of *GAS5* in AOSMCs has revealed a pro-contractile phenotype and an interaction with β-catenin, a component of the Wnt signalling pathway involved in developmental processes[177]. This interaction may reduce the nuclear translocation of β-catenin and reduce induction of proliferative gene targets such as c-Myc and cyclin D1 demonstrating another possible anti-proliferative role for *GAS5*. These studies also show overexpression of *GAS5* could reduce neointima formation in rat carotid artery injury[176] whilst knockdown led to increased medial width in spontaneously hypertensive rats[177]. Another example of a p53-influencing lncRNA active in VSMCs and hypertension is *TYKRIL*, demonstrated to be upregulated in both pericytes and PASMCs in hypoxic conditions as well as in IPAH patient lung tissue[178]. The proposed role of *TYKRIL* appears to act in direct opposition to *GAS5* through binding p53 and blocking a stabilising interaction with p300, ultimately resulting in loss of p53/p300 nuclear translocation. As with *SMILR*, low conservation of *TYKRIL* necessitated development of an *ex vivo* human model of tissue remodelling - here using lung biopsies from IPAH patients. This revealed that *TYKRIL* knockdown could reduce proliferating cells in lung slices ~4-fold whilst increasing pro-apoptotic cells ~2-fold indicating a beneficial effect on reducing the vessel remodelling brought on by hypertension.

Another group of lncRNAs involved in VSMC proliferation are those linked to TGF-β signalling or the downstream SMAD effector transcription factors. Study of *GAS5* in 10T1/2 VSMC progenitors showed it is enriched amongst lncRNAs bound to the TGF-β effectors SMAD3 and SMAD4[179]. However, in contrast to other identified GAS5 interactions the SMAD3 interaction appears inhibitory, sequestering it from binding to VSMC marker gene promoters. Upon TGF-β stimulation *GAS5* is downregulated and exported from the nucleus whilst SMAD3 is imported. Intriguingly, this anti-differentiation role for *GAS5* in VSMC progenitors contrasts somewhat with the pro-contractile, anti-proliferative role in AOSMCs identified by Wang et al.[177]. This requires further investigation but may mean *GAS5* blocks VSMC accumulation along several axes during vessel homeostasis - reducing maturation of progenitors whilst also blocking proliferation and maintaining the pro-contractile identity of existing VSMCs.

The lncRNA *CRNDE* was also identified to interact with SMAD3 in VSMCs though in this case the interaction stabilises and supports the proliferative activity AOSMCs cultured in FBS[180]. This interaction indicates *CRNDE* is likely to be involved in TGF-β signalling though the nature of this is yet to be determined. As *CRNDE* also supports PDGF-BB mediated proliferation and downregulation of VSMC marker genes[181] one hypothesis is that it could act to reconfigure SMAD3 function from pro-contractile to an alternate pro-mitogenic role also described for this transcription factor.

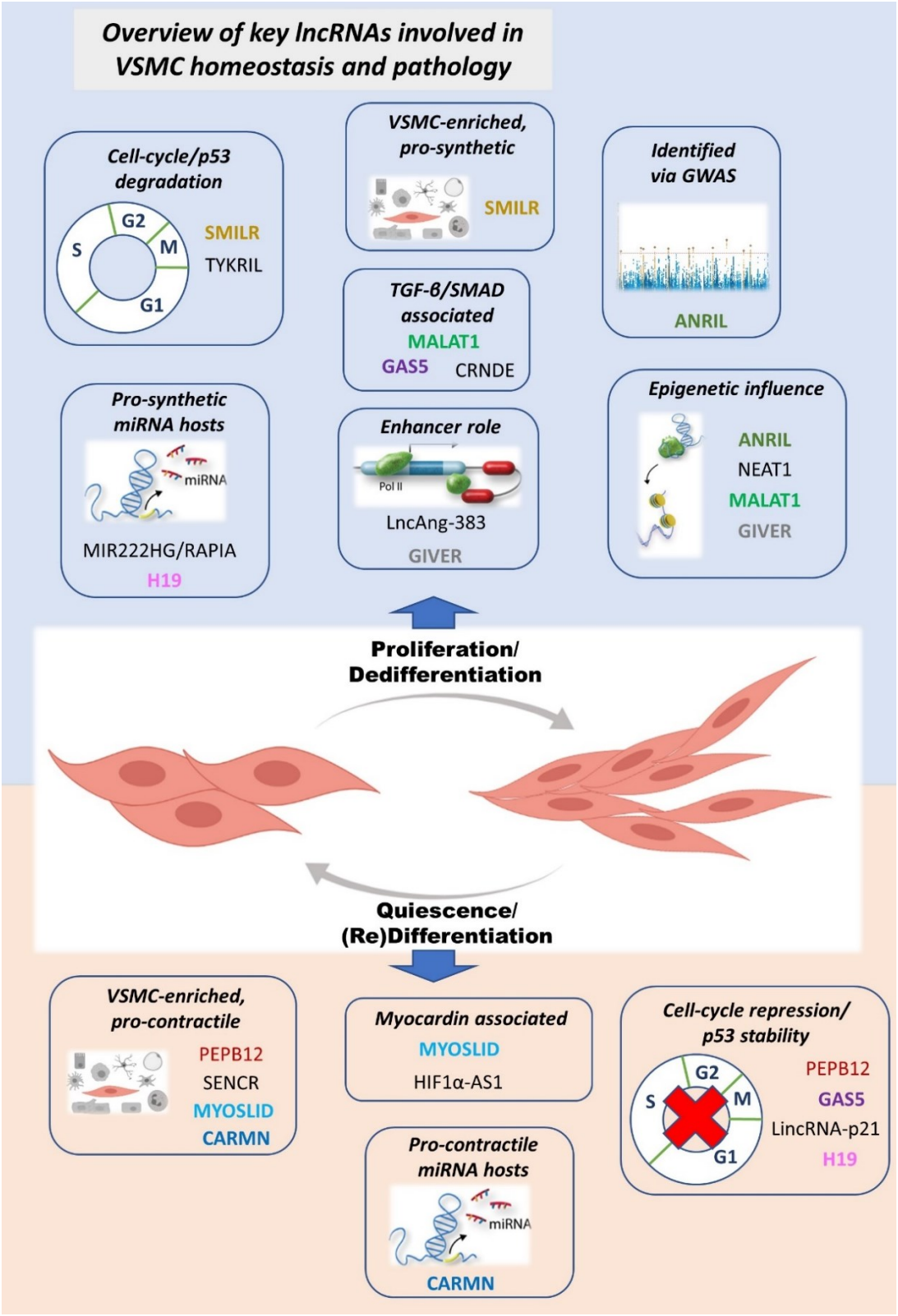


Figure 1.13 Overview of key lncRNAs involved in VSMC homeostasis and pathology

LncRNA	Type	Identification in VSMCs	Selection Criteria	Phenotype	Proposed mechanism in VSMCs	Reference
SENCR	Antisense	RNAseq, CASMCs in standard growth culture, non-reference lncRNAs included	Identified as vascular enriched via PCR cell panel	Pro-differentiation, anti-migratory	Not explored	*Bell et al., 2014*[161]
PEPB1B2	LincRNA	As above	As above	Pro-differentiation, anti-migratory, anti-proliferative	Interaction with CDK4, loss of CDK4 expression	*He et al., 2020*[162]
CARMN	LincRNA, miRNA host	RNAseq and ChIPseq for VSMC subtypes and non-VSMCs	VSMC enriched expression and chromatin marks in selected datasets, prior functionality in cardiomyocytes and production of key miRNAs	Pro-differentiation, anti-proliferative, anti-migratory	Hosts miRNAs involved in VSMC identity, miR independent role is established but mechanism unexplored	*Vacante et al., 2021*[165]
MYOSLID	Antisense	RNAseq, CASMCs with overexpression of Myocardin, non-reference lncRNAs included	Highly upregulated by Myocardin, identified as vascular enriched	Pro-differentiation, anti-proliferative, anti-migratory, stress fibre assembly	Shuttling of MLK1 from nucleus to cytoplasm, promotes pSMAD2	*Zhao et al., 2016*[166]
HIF1α-as1	Antisense	Microarray, VSMCs with overexpression or knockdown of BRG1	Highly associated with BRG1 expression	Pro-apoptotic, anti-proliferative	Not explored	*Wang et al., 2015*[167]
ANRIL	Antisense	Locus inked to CAD via GWAS, transcript in locus expressed in VSMCs	N/A	Linear ANRIL: Pro-proliferative, pro-migratory, pro-oxidative stress, anti-apoptotic Circular ANRIL: opposite	Linear: recruitment of epigenetic modifiers via Alu motifs Circular: inhibition of ribosome maturation	*Congrains et al., 2012*[145] *Holdt et al., 2013*[171] *Holdt et al., 2016*[173]
SMILR	LincRNA	RNAseq, SVSMCs exposed to PDGF-BB and/or Il1α	Highly upregulated, identified as vascular enriched via PCR cell panel	Pro-proliferative, pro-migration	Stabilising interaction with CENPF mRNA	*Ballantyne et al., 2016*[174] *Mahmoud et al., 2019*[175]
LincRNA-p21	LincRNA	Predefined target, known functional lncRNA	N/A	Anti-proliferative, pro-apoptotic	Interaction with MDM2, prevents p53 degradation	*Wu et al., 2014*[176]

Table 1.1 LncRNAs explored in VSMC identity or VSMC pathological states contributing to vasculoproliferative pathology (continued next page)

LncRNA	Type	Identification in VSMCs	Selection Criteria	Phenotype	Proposed mechanism in VSMCs	Reference
GAS5	LincRNA	RNA immunoprecipitation, SMAD factors in 10T1/2 cells, Wang et al.: predefined target	Enriched binding to SMAD3/4	Pro-differentiation (VSMCs), anti-differentiation (10T1/2s) anti-proliferative	Stabilising interaction with p53-p300, prevents nuclear translocation of β-catenin and (in 10T1/2 cells) SMAD3	*Tang et al., 2019*[177] *Wang et al,2016*[178] *Tang et al.,2017*[180]
TYKRIL	Unspecified	RNAseq, IPAH samples, PASMCs, Pericytes cultured with hypoxia	Consistent upregulation across hypertension models	Pro-proliferative, anti-apoptotic	Interaction with p53 prevents stabilising interaction with p300	*Zehender et al., 2020*[179]
CRNDE	LincRNA	Microarray, rat carotid artery balloon injury vs control	Highest upregulation	Pro-proliferative, anti-apoptotic, anti-differentiation	Stabilising interaction with SMAD3	*Li et al., 2020*[181]
MALAT1	LincRNA	CLIPseq: part of HDAC9-BRG1 complex upregulated by defects in TGFRB2 and ACTA2 RNAseq: AOSMCs, CASMCs cultured on stiff hydrogel matrices	CLIPseq only lncRNA in complex RNAseq: Known functional target	Pro-proliferative, pro-migratory, elongated morphology, anti-differentiation	Complex with HDAC9-BRG1 deposits repressive H3K27me3 chromatin marks at contractile genes	*Lino Cardenas et al., 2018*[183] *Yu et al., 2018*[184] *Brock et al, 2017*[185]
NEAT1	LincRNA	Predefined target, known functional lncRNA	N/A	Pro-proliferative, pro-migratory, anti-differentiation	Sequesters WDR5, prevents H3K4 methylation at VSMC marker promoters	*Ahmed et al., 2018*[187]
Lnc-Ang383	LincRNA (enhancer loci)	RNAseq and ChIPseq, rat AOSMCs exposed to AngII, non-reference lncRNAs included	Loci with AngII-induced transcription and H3K27ac (enhancer mark) deposition	Pro-inflammatory (whether enhancer or lncRNA effect unexplored)	*Cis* activation of neighbouring RAMP3	*Leung et al., 2013*[188] *Das et al., 2017*[189]
GIVER	LincRNA (enhancer loci)	As above	As above	Pro-proliferative, pro-inflammatory	Repression of neighbouring NR4A3, interacts with NONO to recruit HDACs and remove repressive H3K27me3	*Leung et al., 2013*[188] *Das et al., 2017*[189] *Das et al., 2018*[190]
MIR222HG/ Lnc-Ang362/ RAPIA	LincRNA, miRNA host (enhancer loci)	As above (RNAseq only)	Proximity to key miRNAs miR-221/222	Pro-proliferative	Hosts miRNAs targeting cell cycle inhibitors, sponging anti-proliferative miR-183-5p	*Leung et al., 2013*[188] *Sun et al., 2020*[193]
H19	LincRNA	Predefined target, known functional lncRNA	N/A	Pro-proliferative, anti-proliferative/pro-apoptotic	Host PTEN-targeting miR-675, miR-independent role in preventing p53 degradation	*Voellenkle et al., 2016*[195] *Lv et al., 2018*[194] *Li et al., 2018*[196]

Whilst these examples show lncRNA roles within TGF-β signalling through direct impact on SMAD factors, the ubiquitous lncRNA *MALAT1*, relatively well-described in many biological contexts, has been shown to achieve this through being an essential component of a complex that deposits repressive epigenetic marks at VSMC marker genes. *MALAT1* acts as a scaffold between HDAC9 and BRG1, together forming a complex which is upregulated in VSMCs with defective mutations in ACTA2 or the TGF-β receptor TFR2[182]. Both mutations are identified risk factors for aortic aneurysm formation, a tissue remodelling pathology directed by loss of VSMC contractility, density and homeostasis. The effects of the HDAC9-BRG1-*MALAT1* complex in VSMCs are consistent with these phenotypes, recruiting the epigenetic regulator EZH2 – part of polycomb repressive complex 2 - to deposit H3K27 methylation marks and reduce VSMC marker gene expression with a concomitant upregulation of matrix-degrading MMPs. *MALAT1* knockdown could also restore vessel wall function in a mouse aneurysm model. Though not yet examined, the role of *MALAT1* in this complex in a vasculoproliferative setting is supported by evidence that it can drive the proliferation of VSMCs in response to mechanical stress[183] and hypoxia[184], as well as HDAC9 mutations that are linked to CAD incidence[185].

Another ubiquitous, relatively well-described lncRNA - *NEAT1* - is also implicated in epigenetic regulation of VSMC phenotypic switching[186]. *NEAT1* is upregulated with dedifferentiation stimuli such as PDGF-BB, passaging in culture and balloon injury, whilst supporting a migroproliferative phenotype in VSMCs. Mechanistically, *NEAT1* was found to bind the epigenetic modifier WDR5, sequestering it from being used to deposit activating marks at Myocardin and VSMC marker gene promoters. This epigenetic role is distinct from a previously established role for *NEAT1* in the formation of paraspeckles, thought to be areas of transcript processing and splicing within the nucleus, again highlighting the possibility of multiple distinct roles for individual lncRNAs.

Other epigenetic-acting lncRNAs influencing VSMC states have been identified through their association with enhancer sites. The lncRNAs *lnc-Ang383* and *GIVER* were highlighted as they are transcribed from enhancer sites which show increased enrichment of H3K27ac - indicating accessible

chromatin - during treatment of rat AOSMCs with Angiotensin II (AngII)[187,188]. This indicate these lncRNAs may play supportive roles in any AngII-induced changes in chromatin conformation which initiate at such sites. The knockdown of *lnc-Ang383* in AOSMCs *in vitro* and rat aortas *ex vivo* showed it supported the AngII-stimulated induction of Ramp3, a neighbouring PCG involved in VSMC contraction[188]. *Lnc-Ang383* therefore may support the local effects on transcription mediated by the Ramp3 enhancer. *GIVER* is characterised in greater depth with an identified human ortholog as well as an expression profile of induction specifically within growth factor-treated vascular cells[189]. As with *lnc-Ang383*, *GIVER* was shown to support AngII-induced pro-inflammatory activity in VSMCs. This was hypothesised to be related to an interaction with the protein NONO, also a known binding partner of *NEAT1* and component of nuclear paraspeckles. Knockdown of NONO phenocopied knockdown of *GIVER* and this was hypothesised to be related to recruitment of HDAC-based remodelling complexes, similar to that shown for both *MALAT1* and *NEAT1*. This may be an example of an enhancer lncRNA supporting wider epigenetic effects beyond the local chromatin neighbourhood. Both the *lnc-Ang383* and *GIVER* enhancer-associated lncRNAs were rapidly induced and peaked within 3 hours of AngII treatment. Intriguingly, transcription at enhancer sites has been observed to generally preceed transcription at other sites (e.g. promoters of transcription factors) within multiple cell-stimuli timecourse studies, suggesting this could be a generic biological phenomena[190]. As enhancers are thought to be crucial genomic sites for facilitating chromatin accessibility through allowing looping and conformational changes, chromatin remodelling may be a key initial element of a cellular response to an external stimuli such as a pro-inflammatory mediator like AngII. The characterisation of *GIVER* is therefore intriguing as it shows the potential of enhancer-transcribed lncRNAs to support enhancer function and early chromatin remodelling during the initial response of VSMCs to a potentially pathogenic signal.

Of the remaining lncRNAs characterised in VSMCs and vasculoproliferative pathology, a noteworthy group are those which influence miRNA expression – either through acting as a host transcript required for miRNA biogenesis (such as *CARMN*), or through acting as a decoy to prevent miRNAs from degrading

other targets. An initial example within VSMC biology is another enhancer-transcribed lncRNA identified via AngII stimulation of rat AOSMCs, *lnc-Ang362*[187]. This lncRNA was identified as a host gene of miR-221 and miR-222, already shown to support FBS-stimulated VSMC proliferation and neointimal formation post-injury through shared targeting of p27[191]. Two relatively conserved areas of the human ortholog *MIR222HG* have since been found, via RNA immunoprecipitation in macrophages, to interact with miR-183-5p – a repressor of integrin B1[192]. As integrin B1 is known to play a role in supporting protein levels of cyclin A and cyclin D1, *MIR222HG* may also promote proliferation through acting as a miRNA decoy, sequestering the anti-proliferative miR-183-5p to derepress integrin B1. Another example is *H19*, mentioned previously as a chromosomal imprinting lncRNA. *H19* contains the miRNA miR-675 within a first exon and overexpression of this transcript was shown to support a proliferative phenotype in FBS-stimulated VSMCs[193]. This was explained by the targeting of PTEN by miR-675. The resulting loss of PTEN expression could reduce SRF-PTEN interaction, potentially aiding the redirection of SRF away from VSMC marker genes to proliferative genes during VSMC dedifferentiation. Both *H19* and miR-675 expression correlated with the ratio of intima to media observed after carotid artery balloon injury. However, study of *H19* in a distinct context demonstrates it also has a pro-apoptotic, anti-proliferative phenotype in AOSMCs stimulated by AngII[194]. This role of *H19* was shown to be miR-675-independent - through overexpression of a miR-675-null variant - and was closely tied to the cardiovascular development transcription factor HIF1α[195]. *H19* was found to recruit transcription factors to stimulate the HIF1α promoter whilst also interacting with HIF1α in the cytoplasm to prevent degradation of p53. Along with *CARMN* and *MIR222HG*, this provides another example of a role independent of miRNA production for a miRNA host gene, a role which, in the case of *H19,* appears to act in direct opposition to the miRNA. The possibility of these loci to switch or emphasise different phenotypic effects through different regulation of host and miRNA in different contexts is intriguing and so far a relatively unexplored area of RNA regulation.

These examples highlight the variety of mechanisms of functional lncRNAs found in VSMCs and show the capacity of lncRNAs to effect positive changes

in tissue remodelling. However, they likely constitute only a small selection of the lncRNAs with VSMC relevance and importance. The wider lncRNA field contains many lncRNAs highlighted as important through high-throughput experiments in VSMCs but with no functional characterisation. Further, lncRNAs have not been explored in areas of VSMC biology such as senescence, heterogeneity, embryonic origins or in connection to the recent advancements in defining VSMC progenitors involved in intimal growth. The overall contribution of lncRNAs to VSMC-driven disease is still to be determined with potential targets of therapeutic interest likely yet to be obtained.

1.4 Approaches to identify VSMC lncRNAs

1.4.1 Challenges to identify and characterise VSMC lncRNAs

The current picture of lncRNAs driving VSMC pathological phenotypes has been obtained through a set of ground-breaking studies laying out a large range of possibilities for lncRNA targets, function and physiological relevance. This will inform future work on the remaining portion of uncharacterised lncRNAs that could be VSMC-relevant. Such studies are still necessary to help define the extent to which lncRNA constitute therapeutic targets for modifying gene regulation networks and ultimately - to select those with the most potential to counter VSMCs directing tissue wall remodelling. However, many of the same barriers faced in these studies, and the lncRNA field in general, are yet to be broken down.

Firstly, definitively proving a lncRNA controls a phenotype *in vitro* is complex, often requiring an experimental design that can assess the influence of both the lncRNA transcript as well as the genomic locus in which it is found. An often-cited example is the lncRNA *AIRN*, a gene which silences the neighbouring, overlapping IGFR2 gene simply by being transcribed[196]. The lncRNA product is completely dispensable for this silencing; introduction of premature polyadenylation sites into the *AIRN* locus could alter transcript structure yet had no discernible effect on IGFR2 repression. For loci such as these, untangling whether the lncRNA transcript, the lncRNA promoter or other features are the driving force behind a phenotype requires awareness of the genomic context alongside careful design and interpretation of studies.

Targeting a lncRNA transcript for degradation can be achieved through use of RNA interference which directs the miRNA machinery towards the target of interest, primarily in the cytoplasm[197]. A nuclear-biased lncRNA may instead be targeted through use of antisense oligonucleotides (ASOs) which direct the RNase H machinery[146]. Both methods suffer from non-specific effects arising from either the process of transfecting nucleic acid into the cell, interference with the RNA degradation machinery or non-specific binding to other RNAs. This can be controlled somewhat through use of controls consisting of scrambled non-biological sequences but ideally requires use of multiple distinct targeting molecules targeting the same lncRNA transcript alongside[198].

Overexpression techniques using viral vectors are also often used - though unlike with coding loci, ORFs cannot be used to inform isoform selection and the selection of a single dominant isoform may be unfeasible. To avoid this issue, CRISPR-based methods can be used to activate transcription at a lncRNA promoter and activate the locus as a whole. CRISPR-based genome-editing methods can also be used for deletion or reconfiguration (e.g. inversion or addition of termination of sites) of the lncRNA locus or promoter and have been utilised if requiring to analyse a link between a particular feature of a lncRNA locus and a suspected phenotype[147,197]. The caveat with such modifications of gene systems is that the potential for disruption not directly tied to the lncRNA must be considered, for instance if the lncRNA contains an annotated promoter or enhancer for another gene within the locus. An intersection of these approaches is often required to define the source of a phenotypic effect at a lncRNA locus. Further, a high proportion of functional lncRNAs demonstrate phenotypes in a cell-specific manner[147], and so many will have to be redefined within different biological contexts of interest to ensure the relevant conclusions are being made. In VSMCs this has revealed lncRNA loci like *H19* and miR-675, where host and miRNA show opposing phenotypes[193,195], or *NEAT1*, initially described as a regulator of nuclear homeostasis but now with an additional described role in VSMC dedifferentiation[186].

With correct interpretation, strong evidence for *in vitro* phenotypes for lncRNAs can often be obtained. Showing the relevance and context of these findings *in vivo* is currently a greater challenge, with *in vitro*-described phenotypes often not reproduced or detectable when lncRNAs, even well characterised examples such as *MALAT1*, are knocked out in animal models[197,198]. This may in part be explained by non-specific effects of dsiRNA and ASOs obtained *in vitro* being mistaken for phenotypes[198]. Additionally in a physiological context, many lncRNA phenotypes may be too subtle to cause a noticeable effect or could have a redundant function with other gene regulators that can cover for any lost lncRNA function. Detection of a lncRNA phenotype may also require specific stimuli in specific tissues or cells, highlighting the necessity to base experiments upon a comprehensive expression profile for the lncRNA to define the tissues, organs and stimuli of interest. In the search

for lncRNAs controlling VSMCs, the search for a phenotype is narrowed, at least initially, to a focus upon the vasculature and well-established arterial injury models. This means genes of interest can be ruled in or out relatively simply by characterising their effect on vessel wall remodelling post-injury as has been shown with many lncRNAs described in the previous section. The lncRNA function within non-VSMC cell types in such models may also be relevant and so should also be explored to gain a wide understanding of lncRNA effect on remodelling dynamics. *SENCR* for instance, was initially described with a role in VSMC identity[160] yet is only mechanistically characterised within endothelial cell barrier integrity so far[199] with both roles likely relevant to the vasculature. Identification of *in vivo* lncRNA phenotypes is also hampered by the lack of available animal models for non-conserved lncRNAs to be characterised in. *SMILR*[174] and *TYKRIL*[178] are examples where *ex vivo* models of vessel wall injury were developed to get around this and provide evidence of physiological relevance. The inherent artificial nature of these models when compared to animal arterial injury models is paid for somewhat by their direct relevance to human patients.

Another significant barrier during lncRNA characterisation is the huge variety of mechanistic possibilities which have been uncovered for lncRNAs so far. This provides a wide range of potential investigative avenues, though the field now has an improved understanding of how to navigate these[197,198]. Determining lncRNA mechanism can be initially aided by the study of the lncRNA subcellular localisation. For instance, a functional lncRNA with a bias towards the chromatin-bound subcellular fraction may be involved in epigenetic regulation, and further may have *cis* functionality, acting to regulate the genomic locus surrounding itself. These lncRNAs would then likely show some support or inhibition of the expression of surrounding genes. LncRNAs transcribed from enhancers may also be *cis*-acting through supporting the enhancer to implement local changes in chromatin state and enhancer-promoter contacts[200]. Overlap to enhancers can thereby provide an initial hypothesis for a mechanism of action for such lncRNAs, though *GIVER* shows enhancer-transcribed lncRNAs are not necessarily limited to *cis* functionality only[189]. Conversely, lncRNAs under-represented in the chromatin-bound subcellular fraction are less likely to have *cis* functionality and so could

implement wider changes, for instance lncRNAs regulating p53 stability in the nucleus or lncRNAs sequestering miRNAs in the cytoplasm. Further clues to mechanism can be provided from the genomic locus including any overlap with small RNAs that may be hosted by the lncRNA, as is the case with *MIR222HG, H19* and *CARMN*[164,193,201]. Definitive proof of a lncRNA mechanism then rests with identifying interacting protein, DNA or RNA partners via pulldown techniques. Such interactions may be tough to prove considering the inherent tendency of RNA towards non-specific binding as well as the often low expression values of lncRNAs[197].

These challenges during the process of lncRNA characterisation, are preceded by a separate limiting factor. The specific expression profiles and low abundance of lncRNAs means their annotation, even within the extensively studied human genome, is incomplete[202]. Annotations are mostly based on the output of high-throughput methods such as RNAseq which provide a snapshot of all transcribed RNA species within a cell at once[136,137,203]. Detecting lncRNA with low and/or restricted expression levels are a problem for such methods, particularly in heterogenous samples, such as tissue or even an *in vitro* monoculture of mixed cellular subtypes. In these samples a larger variety of cell type or subtype-specific RNAs will be sampled at an overall lower rate. Unless increasing this sampling rate, i.e sequencing more RNA - often prohibitively expensive, this reduces the chance of detecting molecules with a cell-specific bias. As many lncRNAs are already low in abundance, this could reduce the chance of detection for those with some cell-specific bias further still. This is backed up by initial observations of changes in lncRNA detection rate and expression measurements between “bulk“ RNAseq and scRNAseq data which show genes of low expression in the former may have high expression in individual cells in the latter[204]. As lncRNAs have been shown to provide an *in vitro* function even at one copy per cell[205], it is important to consider the functional potential of such low-level transcripts. Another example are miRNA hosts several of which have been shown to be rapidly degraded after miRNA biogenesis[206] – several miR-hosting lncRNA transcripts are therefore likely under-represented in most high-throughput methods. Most published RNAseq datasets also enrich for RNA molecules with polyadenylated tails, which removes the chance of identifying all expressed

lncRNAs[207]. LncRNAs have therefore been hard to map out accurately and completely due to both their inherent traits as well as technical factors.

Obtaining transcriptome profiles across a range of individual cell and tissue types is the approach taken by consortiums seeking to provide a reference annotation – one which provides an approximate summary transcriptome that can be used as a basis for many studies[136,137,203]. The representation of VSMCs within the reference annotation GENCODE, considered gold standard due to the accuracy of the manual curation of this database, consists of 4 arterial subtypes of 2 biological replicates each[137,203]. Other reference sets aim to be more extensive so dispense with the slower rate of manually curated annotation. FANTOM incorporates GENCODE with 4 other reference sets, mostly produced using automated computational methods, and uses a method of 5' sequencing, CAGEseq, to more accurately delineate transcription start sites (TSS) than can be achieved using RNAseq[136]. FANTOM incorporates an increased 9 VSMC subtypes with an average of 3 biological replicates for each. However, a truly robust annotation of lncRNAs may require providing a representative profile for every cell type in the human body as well as their response to specific stimuli at different timepoints, with enough biological replicates to cover human variation - a huge undertaking. Over-reliance on reference annotation to profile lncRNA activity may therefore invite false negatives as key lncRNA genes driving the system in question are left undetected.

Another issue with lncRNAs in reference annotations such as GENCODE and FANTOM is that they overwhelmingly use "short" RNAseq data - e.g. Illumina sequencing includes a fragmentation step leading to a library with insert size range of 200-800bp – which in conjunction with computational tools to predict transcript models together have well-described deficiencies in accuracy[208,209]. Such inaccurately-assembled transcript models are often shown to have non-representative splice patterns or incomplete 5' or 3' boundaries. Importantly, these inaccuracies may lead study design to be based on incorrect or absent isoform structures, another factor which could influence reproducibility of lncRNA phenotypes *in vitro* and *in vivo*. Confusion in obtaining accurate isoform structures using high-throughput techniques may only be possible to address definitively with recent developments in long-read

RNAseq techniques allowing sequencing of longer transcript segments[162]. Regardless, as the field moves from initial characterisation of the most detectable lncRNAs towards widespread characterisation of the remaining lncRNAs, it will become more important to obtain representative, bespoke transcriptomes within model systems to ensure the full range of candidate genes are defined accurately.

1.4.2 Methods to annotate lncRNAs and use in VSMCs

This section contains material adapted from the review "Endothelial function and dysfunction in the cardiovascular system: the long non-coding road" by Bennett et al. 2019 published in Cardiovascular Research in 2019 - of which I am co-first author. Section 3 was drafted by me then edited by me in conjunction with the other authors**.**

Many of the lncRNAs actively contributing to VSMC pathological phenotypes have been initially selected for study through their known functionality in other contexts (Table 1.1). This circumvents the need to use a transcriptome-wide approach and accelerates identification of relevant, functional lncRNAs. However, it does not provide the unbiased view of lncRNA activity in the transcriptome offered by high-throughput techniques such as microarrays or the gold standard method - RNAseq. Sequencing-based techniques and analysis methods are now well-developed with computational tools providing ever-improving accuracy in building transcript models, defining their coding status and quantifying their fluctuations within systems of interest[208,210]. Use of these techniques and protocols is essential to avoid bias in the field towards well-characterised candidates and could be enlightening within areas of VSMC biology which are as yet poorly explored in terms of lncRNA activity. A flowchart showing experimental design for a lncRNA identification study is shown in Figure 1.14.

Microarrays provide a relatively simple method to obtain transcriptome level data which has highlighted several lncRNAs now characterised in VSMCs (Table 1). The arrays consist of oligonucleotide probes which produce a signal when bound specifically by the complementary RNAs in a given transcriptome sample. This constitutes a key limitation of the technology as the probe-set

must be pre-defined and therefore cannot be used to detect lncRNAs outside of reference or pre-defined annotations. Further, non-specific binding of probes means that weaker probe signals from low-expressed RNAs are hard to detect. Microarrays can still provide use in a lncRNA context however as they are an economical way to profile a high number of samples at once for pre-defined lncRNAs. For example, this could be used to provide a broad overview of changes in lncRNA activity across a large number of different VSMC subtypes as it has been beneficial to probe the relationship of lncRNA to heterogeneity in endothelial cells[211]. Microarrays could also be used to obtain the large amount of data necessary to reliably perform a co-expression analysis and determine likely areas of functionality for lncRNAs through looking at similarly regulated PCGs with known function[212].

The gold standard technique for lncRNA identification however remains RNAseq. It is sensitive, increasingly cost-effective and presents a largely unbiased, whole-transcriptome view of shifts in gene expression. RNAseq has been used to identify lncRNAs active in the VSMC response to hypoxia[178], pathological mechanical stress[213], calcification[214], growth factors and cytokines[173,215,216] as well as those enriched within VSMCs[160,161,163,165]. Overall though, the potential of this technique to uncover lncRNAs of interest within VSMC models has not yet been fully realised. Key stimuli in models of VSMC dysfunction such as oxLDL or TGF-β have not been explored via RNAseq for lncRNA activity, whilst others such as AngII and inorganic phosphate (to stimulate pro-inflammatory calcification) are explored only in rat models. Highthroughput profiling of lncRNAs in VSMCs within patient samples is also sparse, though inclusion of iPAH lung biopsy RNAseq during identification of *TYKRIL* is an initial example. Further, of the 7 studies screening human VSMC-based RNAseq data for lncRNA activity, only 2 capitalise on the ability to detect expression of completely novel transcripts outside of reference annotation, with their above-described limitations. Such studies led to identification of SENCR[160] and MYOSLID[165]. Notably, these studies were based around profiling VSMCs under homeostatic or pro-contractile conditions. Data for non-reference lncRNA annotation in VSMCs exposed to pathological stimuli or extracted from patient remodelled vessel wall is currently lacking.

A possible reason for slow uptake of RNAseq to detect non-reference VSMC lncRNAs is that the use of RNAseq data for prediction of new lncRNAs requires a specific skillset to perform algorithmic reconstruction of transcript models from short RNAseq reads (these algorithms can be 'seeded' with existing reference annotations to improve precision). Tools such as Cufflinks[217] and the more recent StringTie reporting improved completeness of transcript assembly[218] are both widely used to implement this. In addition, identified transcripts must then be filtered to determine which transcripts are non-coding with high confidence[219]. Several computational pipelines are now available aiming to aid researchers to implement this with greater ease[139,220–222]. Typical filtering steps include the use of several algorithms to evaluate coding potential and combination of their scores, exclusion of short or very lowly expressed transcripts, and exclusion of transcripts found in close proximity to annotated genes (which often correspond to unannotated extensions of untranslated regions in mRNAs). For example, transcripts containing long ORFs (i.e. possessing long stretches of uninterrupted sequence between in-frame start and stop codons) with characteristic codon frequencies and/or with high homology to existing proteins can be identified with the widely-used coding prediction calculator (CPC) scoring method[223]. Others exclude candidates with ORFs predicted to produce proteins with structural homology to known Pfam protein domains (HMMR[224]), an enrichment of short sequences characteristic of coding sequences(CPAT[220] and FEELnc[225]) or containing codons that are maintained (with the same or similar codons and without frameshifting mutations) over large evolutionary distances (RNAcode[221]). Codon-conservation tools are particularly powerful for detecting short conserved ORFs but also can have some false positives in regions of extremely high conservation and limited variation between species. Our understanding of the best use of such tools has improved since the last studies aiming to expand the VSMC lncRNA repertoire over 5 years ago[160,165]. There remains a demand for such tools to be used to update our understanding of available candidates.

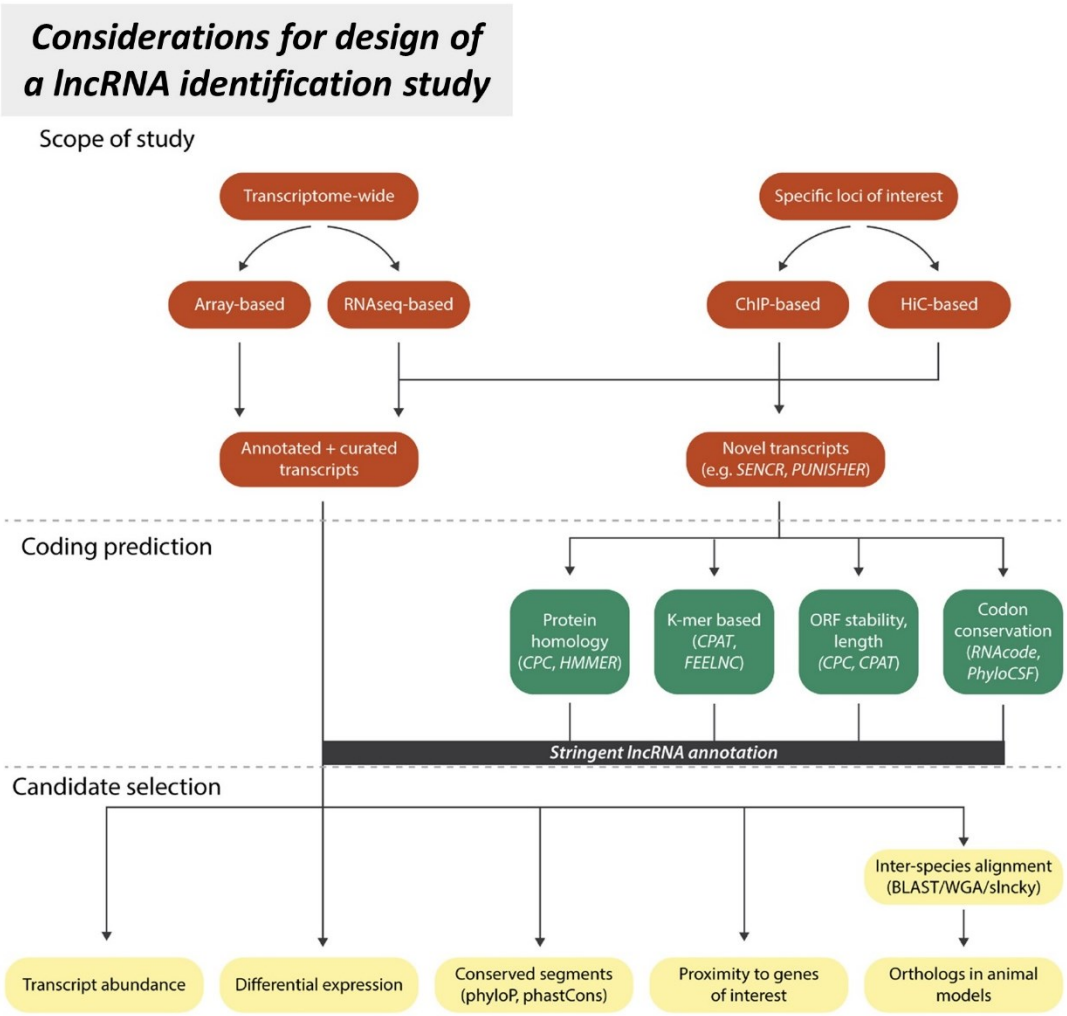


Figure 1.14 Considerations for design of a lncRNA identification study. An initial starting point is an unbiased approach using microarray or RNAseq profiling of lncRNAs in a given transcriptome. Alternatively, more targeted approaches may attempt to identify lncRNAs associated with a given region of interest for instance through identifying those transcribed from regions bound by key transcription factors (ChIPseq) or which form chromosomal contact with genes of interest (HiC). Coding prediction of identified lncRNAs may not be required if relying solely on reference annotation. Notably the criteria here listed for candidate selection is not exhaustive and may vary by context. Adapted from Bennett et al., 2019[149]

Another issue is the high amount of sequencing (read depth) required to be able to comprehensively detect and annotate lowly expressed lncRNAs. This can be costly, with an estimated depth of over 300 million paired reads per sample required to provide the highest level of lncRNA coverage[226]. However, studies sequencing at far lower depth have been fruitful in identifying substantial amounts of non-reference transcripts and interesting trends in lncRNA dynamics within specific biological contexts. Aside from the described

SENCR and *MYOSLID* studies, these include studies profiling lncRNAs over developmental timepoints in erythropoiesis[227] and endothelial cell differentiation[156], during exposure to pro-inflammatory stimuli[228,229] or from cancerous[230], chronically inflamed[231] or ischemic[232] pathological tissues. Typical findings are that non-reference lncRNAs make up a particularly large proportion of those which change in expression between experimental conditions[227,228,231], appear enriched within particular cell types[231,232] and transcribed from enhancer regions[230–232]. The latter two traits may be related as enhancer-transcribed lncRNAs generally show greater propensity towards cell-specificity[136]. Therefore, non-reference lncRNAs are not only required to get a greater definition of lncRNA activity within a biological system but also seem to possess desirable traits for finding candidates with pathological association, potential for enhancer association (with possible *cis* functionality) and potential for targeting of specific cell populations. Notably, most of these mentioned studies were based in human showing that these approaches can be fruitful even in a relatively well-annotated species.

This introductory chapter summarises the unrealised potential of lncRNAs to provide urgently needed targets that could eventually allow the long-held goal of therapeutic manipulation of VSMC phenotypes. An established set of example lncRNAs are known to control a range of such pathological VSMC states. However, to find lncRNAs with high therapeutical potential, we need an unbiased way to identify and characterise lncRNAs, without relying solely on annotated ones or lncRNAs previously studied in other contexts. This is now possible as the techniques to expand lncRNA annotations are much improved in terms of implementation and accuracy. Despite this they have so far been underused in many fields including vascular and VSMC biology. Identification of non-reference lncRNAs in other specific biological contexts has provided strong data supporting their inclusion for consideration as key functional targets. This provides a niche for a study that improves the annotation of lncRNA activity within VSMCs exposed to pathological stimuli or derived from pathological tissue.

1.5 Hypothesis and aims

1.5.1 Hypothesis

Expanding the transcriptional annotation of VSMCs responding to proliferative stimuli can provide better definition of the lncRNAs that drive vasculoproliferative diseases

1.5.2 Aims

1. Reveal the extent to which non-reference lncRNAs may contribute to VSMC proliferation and associated pathological phenotypes. Utilise the transcript assembly tool StringTie and "Pipeline for Annotation of LncRNA" (PLAR) to build bespoke expanded transcriptomes containing high-confidence GENCODE lncRNAs supplemented with any other missing lncRNAs within RNAseq data for SVSMCs exposed to IL-1α and PDGF-BB *in vitro* alongside any other appropriate RNAseq data for pathological VSMCs.

2. Match the expanded lncRNA annotations to other publicly available annotations of transcripts, TSSs, enhancers as well as expression data to provide further evidence of non-reference transcripts existence and *in vivo* relevance. Use this to probe hypotheses on VSMC-specificity and *cis*-regulation for individual lncRNAs.

3. Using the proliferation model of SVSMCs exposed to IL-1α and PDGF-BB, select uncharacterised lncRNAs with highest functional potential via identifying those which are outliers in terms of expression dynamics and abundance and screen them for a functional effect on the VSMC proliferation phenotype.

4. Examine lncRNA dynamics in a temporal context, identifying those which – as with *SMILR* and AngII-induced enhancer lncRNAs – may be centrally important to VSMC pathology through upstream initiation of pathways co-ordinating proliferation and pathological phenotypes.

Chapter 2: Materials and Methods

2.1 LncRNA discovery pipeline

2.1.1 Transcriptome Assembly

RNA sequencing (RNAseq) files were obtained from gene expression omnibus (GEO) using accession numbers GSE69637 and GSE100081. We received files for a third dataset of plaque VSMCs produced by Alloza et al.[233] by direct transfer from the authors. Data quality was checked via FastQC(0.11.9)[234]. Trimming of adaptor sequences was required for GSE69637 and done using TrimGalore(0.5.0)[235]. Custom transcriptomes consisting of GENCODE transcripts supplemented with newly-assembled transcripts were generated as follows for each dataset. STAR(2.5.1b)[236] was used to map reads to the human genome (GRCh38) indexed with GENCODEv26 (--sjdbOverhang 100). StringTie(1.3.1c)[218] was then used on these alignments to assemble transcripts (minimum length: 300bp). The StringTie assemblies for each sample were merged (using StringTie --merge) together with a filtered reference set to create an expanded transcriptome for each dataset. This filtered reference set was obtained for each dataset by removing transcripts with low expression (<0.5 FPKM for spliced transcripts and <1 FPKM for unspliced transcripts) and short transcripts (<300bp). Transcript quantification was based on RSEM(1.3.0)[237] (--bowtie2 for all uses of RSEM in this thesis). Newly-assembled transcripts and genes were provided numeric identifiers with the prefix “MSTRG.” by StringTie.

2.1.2 Pipeline for Annotation of LncRNA (PLAR)

RSEM(1.3.0)[237] was used to quantify transcripts in each dataset using their corresponding expanded transcriptome. To annotate lncRNAs, we used the published pipeline PLAR[139,238]. For downstream analysis we considered only expressed transcripts, defined as those with an average FPKM > 1 in 1 or more experimental conditions. We also discarded minor newly-assembled isoforms for each lncRNA gene; removing all newly-assembled isoforms with an average FPKM <10% of the sum of all isoform FPKMs for that gene, as used in[136]. Genes were classified based on PLAR transcript classification; genes producing any number of coding transcripts were labelled “coding”, remaining genes producing putative lncRNAs were labelled “putative lncRNAs” and

remaining genes producing high confidence lncRNA were labelled simply as “lncRNAs”. Genes were considered robustly expressed if they had an average FPKM >1 in 1 or more conditions in a dataset.

2.2 Validation of lncRNA transcript assemblies

2.2.1 Comparison of lncRNA transcripts to FANTOM CAT transcripts

The FANTOM CAT meta-annotation* co-ordinates were converted from hg19 to hg38 using liftOver from UCSC[239]. Transcripts from robustly expressed lncRNA genes were cross-referenced to the converted assembly using GFFcompare(version 0.11.2)[240]. As transcripts matching to FANTOM CAT with GFF codes “=” or “k” indicated these transcripts contained a complete chain of FANTOM CAT introns, these pairs were classed as “Full Intron Chain” matches. As transcripts matching to FANTOM CAT with codes “c”, “m”, “n” or “j” were transcripts which contained less than a complete chain of FANTOM CAT introns but at least one identical splice junction, these transcripts were classed as “Splice Junction” matches. Transcripts matching with “e” or “o” were classed “Exonic Overlap” matches whilst transcripts with no overlap (any other code or no code) were classed “No Exonic Overlap”.

*from:

https://fantom.gsc.riken.jp/5/suppl/Hon_et_al_2016/data/ assembly/lv1_raw/

2.2.2 Comparison of lncRNA transcripts to FANTOM CAT CAGEseq

Validation of lncRNA transcription start sites (TSSs) was done by comparing to the FANTOM CAT CAGE sites, again with co-ordinates converted to hg38 using liftOver from UCSC[239]. CAGE sites within 1kbps of the 5’ end of each transcript were given a transcriptional initiation evidence score (TIEScore) according to FANTOM methodology incorporating genomic distance between 5’ and CAGE site, CAGE site expression level and the transcript size[13]. We selected the highest scoring CAGE site to represent each transcript then selected which of these valid CAGE matches by applying a cut-off used by

FANTOM for defining “stringent” TSSs (TIEScore > 60, FDR <0.026) and excluded any remaining CAGE matches which implied a reduction of exon 1 to <10% of original size. LncRNAs with valid CAGE matches are referred to as CAGE-matched lncRNAs.

2.3 Use of merged VSMC transcriptomes

2.3.1 Detection of newly-assembled lncRNAs in whole plaque RNAseq

We prepared a reference transcriptome consisting of the three expanded VSMC transcriptomes merged into one (using StringTie on --merge setting). To validate newly-assembled lncRNAs expression *in vivo,* we obtained sequencing files for human carotid atherosclerotic plaques from GEO (GSE120521). As the merging process can lead to different gene structures, transcripts classed as lncRNAs in individual VSMC annotation runs were reclassified as putative lncRNAs if expressed <1 FPKM in all VSMC datasets, if showing any change in intron chain sizes, or if 5’ and 3’ extensions post-merging were >100bp. Such genes were not considered in this analysis. Carotid plaque reads were mapped and quantified to the merged annotation with RSEM(1.3.0)[237] to identify expressed high confidence lncRNA.

2.3.2 Repeatability of newly-assembled lncRNA transcript assemblies between pipeline runs

To examine the repeatability of newly-assembled lncRNA transcript assemblies generated by StringTie for those found across several of the three VSMC RNAseq datasets we used GFFcompare to compare each individual VSMC transcriptome to the merged VSMC transcriptome (created in 2.3.1). Newly-assembled transcripts from individual VSMC transcriptomes which both matched to a same transcript in the merged VSMC transcriptome as “Full intron chain” or “Splice junction” matches (see 2.2.1) were identified. These marked out transcript assemblies which were potentially built from the same transcript in each individual VSMC dataset. Transcripts in the merged VSMC

transcriptome which had multiple “Full intron chain” matches to individual VSMC transcriptomes were considered to be reproduced by StringTie in a complete form across multiple datasets.

2.3.3 Detection of VSMC lncRNAs in SVSMC timecourse analysis

To take advantage of our merged VSMC annotation (created in 2.3.1) to detect lncRNA activity within the first 24 hours of SVSMC transition to a proliferative state, SVSMCs from patient donor tissue were extracted, cultured to 3-5 passages then plated and quiesced for 48 hours as previously described[8] then stimulated with IL-1α (10ng/ml) and PDGF-BB (20ng/ml). Total RNA was then collected for SVSMCs harvested at four timepoints post-stimulation (0hrs, 4hrs, 8hrs, 24hrs, n=4) using an miRNeasy purification kit (Qiagen, Hilden, Germany) by Drs A.Mahmoud and M.Ballantyne. These were sent to Genewiz (Genewiz, Leipzig, Germany) using the “standard” RNA sequencing service for library preparation, rRNA depletion and sequencing in 150bp paired-end read format. Data quality was checked via FastQC(0.11.9)[234] and adaptor sequences trimmed using TrimGalore(0.5.0)[235]. Trimmed reads were mapped to the merged VSMC annotation to quantify transcripts expressed by SVSMCs at each timepoint using RSEM(1.3.0)[237]. As the merge process may affect transcript structure, all transcripts expressed (<0.5 FPKM for spliced transcripts and <1FPKM for unspliced transcripts) and of sufficient length (<300bp) within the merged VSMC transcriptome were re-processed using PLAR as described in section 1.1.2 to obtain a full representation of lncRNAs.

2.4 RNAseq exploratory analysis

2.4.1 Differential expression in single timepoint datasets

Differential expression was calculated on the gene level using DESeq2(version 1.28.1)[241]. We considered a gene differentially expressed if it had a fold change (FC) >1.5 or < -1.5 (i.e. Log2FC > 0.585 or < -0.585) with an adjusted p value of < 0.05 from one condition to another.

2.4.2 Differential expression in timecourse analysis (early-response genes)

Differential expression within the SVSMC timecourse RNAseq dataset again used DESeq2(version 1.28.1)[241] and required genes to have an FC >1.5 or < -1.5 (i.e. Log2FC > 0.585 or < -0.585) with an adjusted *p* value of < 0.05 from one timepoint to another to be considered differentially expressed. As genes could be defined as differentially expressed by any one of the eight available pairwise comparisons between the four timepoints, *p* values were BH-adjusted across all pairwise comparisons together. This is more stringent than the default adjustment per comparison provided by DESeq2 which may not fully correct for multiple hypothesis testing in this context as multiple pairwise comparisons per gene are considered simultaneously rather than just one. Additionally, we required genes to have a *p* value of < 0.05 when using the likelihood ratio test within DESeq2 to compare all timepoints simultaneously. Genes matching these criteria are referred to as early-response genes.

2.4.3 Gene Ontology Analysis

Over-represented gene ontologies were obtained using goseq(version 1.4)[242] on prominent clusters of differentially expressed genes. We set an FDR threshold of 0.05 on GO term significance using the BH method.

2.4.4 Enrichment or depletion of gene traits within gene groupings

Tests for over-represented traits within gene groups were done using Fisher's exact test for over-representation via the phyper(A-1, B, C, D, lower.tail =F) function in R. A is the number of genes showing that trait within a sample, B is the number of genes with that trait in a background set of genes, C is the number of remaining background genes and D is the number of genes sampled. Tests for under-represented traits within gene groups were done using Fisher's exact test for under-representation via the phyper(A, B, C, D, lower.tail = T) function in R.

2.5 Defining VSMC-enrichment or enhancer-transcription amongst lncRNAs

2.5.1 Cell type-specific lncRNAs

To assess cell-specific expression, we extracted expression data for CAGE-matched lncRNAs from the FANTOM CAGEseq expression atlas[136]. Enrichment values for each lncRNA-matched CAGE site in each primary cell type category were obtained by calculating the FC difference between the mean expression of all samples in a primary cell category against the mean expression of all other primary cell samples.

2.5.2 VSMC-enriched lncRNAs

To identify VSMC-enriched genes, we determined which CAGE-matched lncRNAs are linked via their CAGE site to genes annotated as VSMC-enriched in FANTOM CAT (defined as genes which are a) FC>5 between mean CAGEseq expression across a VSMC category vs mean in all other samples b) have a Mann-Whitney + BH adjusted p value <0.05 for this comparison and c) are expressed >0 in 50% samples in that VSMC category[136]). Genes enriched in any of the 9 VSMC-subtype categories (average of 3 samples per category) or in all combined VSMC categories together were considered VSMC-enriched. VSMC-enriched, differentially expressed lncRNAs contained newly-assembled genes which may have been identified within more than one dataset so have multiple “MSTRG.” identifiers across the individual VSMC transcriptomes. To allow easier reference to these lncRNAs, we assigned these a single unique identifier as follows. We identified which of these transcripts matched to either the same CAGEseq site or the same transcript in the merged VSMC transcriptome with GFFcompare code “=” (full intron chain). These 6 lncRNAs were labelled *VSMClnc1-6*.

2.5.3 *Identification of enhancer-transcribed lncRNA and associated PCGs*

LncRNAs were defined as enhancer-transcribed (elncRNAs) if their promoter region (-2000bps to +200bps of their 5' end - defined as such via CAGE site peak if available) contained a "double-elite" Genehancer[243] annotation of type "enhancer" (annotated through collating chromatin signature, transcriptomics and transcription factor binding data from 5 sources) available from the UCSC table browser[239]. As enhancer annotations were obtained across a range of cell and tissue types, we removed weaker enhancer annotations by setting a threshold for GeneHancer confidence scores. We set this by identifying the 1st quartile within all confidence scores for all Genehancer annotations overlapping any expressed lncRNA promoter region across the three individual VSMC transcriptomes (score of 237). In addition, we also included lncRNAs matched to a CAGE site classed as "elncRNA" in FANTOM CAT as elncRNAs[136].

2.6 Identifying candidate lncRNAs involved in *cis*-regulation of PCGs

2.6.1 *Identifying evidence for elncRNA* cis*-regulation of PCGs*

To identify PCGs that may be *cis*-regulated by differentially expressed elncRNAs in the manuscript in chapter 3, we used available interaction data. We first defined candidate regulatory pairs of PCGs and elncRNAs as those with TSSs within 250kbp of each other which were together co-induced or co-repressed in the same VSMC type/treatment. To confirm the association of elncRNA-PCG pairs, we searched enhancer-PCG interactions in Genehancer (for Genehancer sites contained in the elncRNA promoter or gene) and elncRNA-PCG interactions made using eQTL analysis in FANTOM CAT (for CAGE-matched lncRNAs).

To compare potential for *cis* regulation at elncRNA and other lncRNA loci, we correlated FCs of all differentially expressed lncRNAs in the two *in vitro* VSMC datasets in chapter 3 with any differentially expressed PCGs within 250kbps. In this correlation analysis we also included lncRNA-neighbouring

PCGs of any fold change value to allow those which may not be *cis*-regulated by lncRNAs to be included in the overall correlations.

2.6.2 Identifying evidence for cis-regulation of PCGs by lncRNAs in the SVSMC timecourse RNAseq

To identify PCGs that may be *cis*-regulated by early-response lncRNAs within the SVSMC timecourse dataset, we identified all early-response PCGs with a TSS within 250kbp of an early-response lncRNA TSS. With the increased number of datapoints in this dataset (16 samples compared to 8 for datasets in chapter 3) we decided to use Spearman's rank correlation to build further associations between candidate lncRNA-PCG pairs. *P* values were BH-corrected across all candidate lncRNA-PCG pairs.

2.7 LncRNA qRT-PCR Validation and Phenotypic Screening

2.7.1 Expression profiling candidate lncRNAs via qRT-PCR

SVSMCs from patient donor tissue were extracted, cultured to 3-5 passages then plated and quiesced for 48 hours as previously described[8] then stimulated with IL-1α (10ng/ml) and PDGF-BB (20ng/ml) or kept in 0.2% serum for a further 72h. CASMCs (Lonza, Basel, Switzerland) were also substituted for SVSMCs in the same model. Total RNA was obtained using an miRNeasy kit (Qiagen, Hilden, Germany) and quantified using NanoDrop 1000 Spectrophotometer (Thermo Fisher, Paisley, UK). cDNA was synthesised using a Multiscribe reverse transcriptase kit (Life Technologies, Paisley, UK). RNA for subcellular fraction testing was obtained using a Paris fractionation kit (Ambion). Custom-made primers (Eurofins MWG, Ebersberg, Germany) (Table S2, section 1.3.6) were used with Power SYBR Green Master Mix (Life Technologies, Paisley, UK) to perform qPCR on a QuantStudio5 thermocycler (Thermo Fisher, Paisley, UK). Ct values of candidate lncRNAs were normalised to those of the UBC housekeeping gene and fold change relative to 0.2% serum using 2-ΔΔCt calculations.

2.7.2 Phenotypic screening via EdU

DsiRNA for *AC018647.3* (IDT, Coralville, USA) and Gapmers for *MSTRG.10933* and *LINC02015* were designed and ordered using online design tools from the manufacturers (Qiagen, Hilden, Germany). *AC018647.3* and *MSTRG.10933* were knocked down after 48 hours of quiescence as described above by transfecting dsiRNA or Gapmer with RNAimax Lipofectamine (Invitrogen, Waltham, USA) in varying concentrations. After five hours in transfection media, cells were washed before continuing quiescence in 0.2% FBS for another 72 hours. *LINC02015* was knocked down by quiescing VSMCs in t75 flasks prior to detaching with Trypsin and plating at 100,000 cells per 34.8mm diameter well with Lipofectamine transfection mix. After 5 hours cells were washed before stimulation with IL-1α(10ng/ml)/PDGF-BB(20ng/ml) in 0.2% FBS for 72 hours. For Gapmer knockdowns of *MSTRG.10933* in growth conditions we followed the same protocol as for *LINC02015* though substituting IL-1α/PDGF-BB in 0.2% FBS for Smooth Muscle Cell Growth Medium 2 (PromoCell, Heidelberg, Germany) with 10% FBS as a proliferative stimuli. For EdU incorporation assays, we used a Click-iT flow cytometry cell kit (Thermo Fisher, Paisley, UK). EdU was added at a concentration of 10ng/ml after cells were washed post-transfection. Quantification of EdU incorporation was measured using an Attune Nxt system (Invitrogen, Waltham, USA).

2.8 Other

2.8.1 Normality testing

Data normality was assessed where appropriate by the shapiro.test() function within R. Datasets with low replicates (n<5) per experimental condition were considered nonparametric. Information on statistical tests in addition to that reported in this chapter is reported alongside their usage in text or figures.

2.8.2 Additional resources

Venn diagrams: http://bioinformatics.psb.ugent.be/webtools/Venn/

Graphical abstract and manuscript Fig.1 images: BioRender.com

Fisher's z in 3.4.3: comparingcorrelations.org (cocor[244] R package)

Chapter 3: Novel Transcript Discovery Expands the Repertoire of Pathologically-Associated, Long Non-Coding RNAs in Vascular Smooth Muscle Cells

3.1 Chapter 3 Introduction

A significant gap exists in the VSMC literature for a study which comprehensively identifies all lncRNAs relevant to pathological VSMC phenotypes instead of relying solely on pre-existing reference annotation data. As the process would likely reveal novel lncRNA regulators of such phenotypes, we aimed to undertake such a study and use recent developments in computational pipelines and tools to identify and annotate non-reference lncRNAs with improved accuracy and detection power. The ever-expanding range of functional genomic annotation available to all researchers now means detailed information could also be easily obtained to attempt to form hypotheses on the expression profiles, function and regulation of identified lncRNAs. We decided our study would benefit from examining multiple suitable RNAseq datasets in parallel to show the broad relevance of any conclusions across multiple VSMC types and stimuli models. This chapter describes our efforts to achieve this, starting from an initial search of publicly available VSMC RNAseq datasets, and subsequent computational processing of these datasets to build expanded transcriptomes with a stringent annotation of lncRNAs. These transcriptomes are then cross-referenced to external databases to provide contextual data on expression and enhancer-association so that initial hypotheses on lncRNA functionality could be built. In the last section, the selection and initial phenotypic screening of 3 lncRNA genes within an *in vitro* model of VSMC proliferation is described.

3.2 Chapter 3 Aims

The following were the aims to be tackled in this chapter:

1. Search for VSMC RNAseq datasets which are most appropriate for novel lncRNA discovery – those which consist of rRNA-depleted, strand-specific libraries of depth equivalent to or surpassing that of the SVSMC dataset provided by Ballantyne et al.[173]
2. Process these VSMC datasets to supplement GENCODEv26-annotated transcripts with non-GENCODE transcripts - using the transcript assembly tool StringTie in tandem with the previously published "pipeline for identification of lncRNAs" PLAR[139,238] to identify a complete set of high-confidence lncRNAs.
3. Validate the accuracy of the non-GENCODE lncRNA transcripts assembled by StringTie (hereon termed newly-assembled lncRNA transcripts) through comparing exonic structures and 5' ends to transcript models and TSS catalogues respectively within the FANTOM CAT annotation.
4. Explore the *in vivo* relevance of newly-assembled transcripts by assessing their expression in whole plaque tissue RNA-seq.
5. Explore the expression dynamics of the lncRNAs and assess newly-assembled lncRNAs for their association to pathological states and/or response to pathological stimuli.
6. Explore the cell-type specificity of the VSMC lncRNAs using the FANTOM expression atlas and identify VSMC-enriched lncRNAs.
7. Identify lncRNAs that appear to be transcribed from enhancer regions and their potential regulation of nearby PCGs via *cis*-regulation, using FANTOM enhancer lncRNA annotation as well as the Genehancer database.
8. Select candidates with strong functional potential based on expression dynamics and high abundance for phenotypic screening within the SVSMC proliferation model using siRNA or GapmeR knockdown prior to EdU incorporation assays

The work done towards these aims are outlined in the following peer-reviewed publication as well as subsequently in additional work.

3.3 Manuscript

Here follows the publication “Novel transcript discovery expands the repertoire of pathologically-associated long non-coding RNAs in vascular smooth muscle cells” of which I am first author. The computational analysis and lab experiments were performed by me. The manuscript was written and edited by me in conjunction with input from Dr Amira Mahmoud as well as joint last authors Prof. Andrew H. Baker and Dr. Julie Rodor. Other specific contributions from other authors are listed at the end of the publication. Supplementary figures and summary tables (S1-4) are all included whilst raw data supplementary tables (S5-S19) are omitted (though available online) due to their extensive sizes. All methods can be found in Chapter 2 of this thesis. As part of the publication, we also provide a web-based application where the expanded VSMC lncRNA annotations can be viewed interactively, the link is provided again here for ease:

https://bakergroup.shinyapps.io/VSMClncRNAannotation/

3.3.1 Preliminary Work: RNAseq datasets suitable for lncRNA discovery

During initial planning of the below publication we decided that applying the computational pipeline to multiple VSMC RNAseq datasets in parallel would provide a basis for a stronger and broader study applicable to various VSMC subtypes and pathological states. We therefore collated information on all available VSMC RNAseq datasets within our research group or online at the Gene Expression Omnibus (GEO) repository to determine whether they would be appropriate to mine for non-reference lncRNA activity. The term "vascular smooth muscle cell" was used to search for RNAseq data in GEO. Any type of human VSMC exposed to any culture conditions was acceptable for consideration. We also obtained the Alloza et al. dataset[233] for consideration through approaching the authors directly. As we had already processed the SVSMC-based dataset from Ballantyne et al., using the lncRNA discovery pipeline, this was used as a benchmark in terms of sequencing depth. Ideally, datasets would also match the SVSMC RNAseq in terms of number of replicates as well as library preparation - using a strand specific-protocol to create paired-end reads from rRNA-depleted samples.

Table 3.1 shows the 7 human VSMC datasets we found in this search. The Zhao et al.[165] and Bell et al.[160] datasets were excluded through being previously explored for non-reference lncRNA activity as well as for a lack of any pathological stimulus and rRNA-depletion. The Vacante[164] dataset was also excluded for containing libraries from cells treated with vehicle or scrambled GapmeR treated VSMCs which could cause non-physiological off-target effects that could potentially obscure conclusions on relevance of novel lncRNA annotations. Altogether, only 3 were rRNA-depleted and of sufficient depth for use. Despite the importance of VSMCs to CVD, few publicly available VSMC RNAseq datasets appropriate for novel lncRNA discovery were present in the literature. This highlights the work remaining to be done generally in exploring the human transcriptomes of pathological VSMCs but also the current gap in the field for a focused lncRNA annotation effort such as ours

Study	Annotation expansion	VSMC type	Stimulants/Conditions	Purification	Replicates	Depth (Average No. Million Reads/Sample)
Ballantyne et al. 2016	None	SVSMC	IL-1α + PDGF-BB	rRNA depletion	n=4	68 paired-end
Zhao et al. 2016	**De novo**	CASMC	**Vehicle**/Ad-MYOCD	**PolyA-enrichment**	**n=2**	67 single-end
Bell et al. 2014	**De novo**	CASMC	Standard culture only	**PolyA-enrichment**	**n=3**	**28 single-end**
Yu et al. 2018	None	AOSMC/CASMC	Soft/stiff hydrogel	rRNA depletion	n=4	118 paired-end
Vacante et al. 2021	None	CASMC	**GapCtrl**/GapCARMN + Chol/Migration/PDGF-BB	**PolyA-enrichment**	n=3	**24 paired-end**
Alloza et al. 2018	None	Plaque-derived SMCs	Asymptomatic/Symptomatic plaques	rRNA depletion	n=7	72 paired-end
Pan et al. 2018	None	AOSMCs/AOECs	TNFα, IL-1β	rRNA depletion	n=3	**12 single-end**
Miller et al. 2016	None	CASMCs	PDGF-BB, TGF-β	rRNA depletion	**n=2**	**27 paired-end**

Table 3.1 Available datasets for consideration to improve lncRNA annotation in VSMCs in pathological states (as of 2019). The Vacante et al. dataset was available in-house. Depth estimates based on read length and total base number recorded for each sample in each dataset in GEO. Text in bold are factors which mean methods to expand lncRNA annotation may not be fruitful in this dataset. Rows highlighted in gray are those taken forward for transcriptome expansion

3.3.2 Manuscript Introduction

International Journal of Molecular Sciences

MDPI

Article

Novel Transcript Discovery Expands the Repertoire of Pathologically-Associated, Long Non-Coding RNAs in Vascular Smooth Muscle Cells

Matthew Bennett [1], Igor Ulitsky [2], Iraide Alloza [3], Koen Vandenbroeck [3,4], Vladislav Miscianinov [1], Amira Dia Mahmoud [1], Margaret Ballantyne [1], Julie Rodor [1] and Andrew H. Baker [1,*]

1 Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh, 47 Little France Crescent, Edinburgh EH16 4TJ, UK; s1795508@ed.ac.uk (M.B.); vlad.miscianinov@biohabit.co.uk (V.M.); amahmoud@exseed.ed.ac.uk (A.D.M.); magz216@hotmail.com (M.B.); julie.rodor@ed.ac.uk (J.R.)
2 Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel; igor.ulitsky@weizmann.ac.il
3 Inflammation & Biomarkers Group, Biocruces Bizkaia Health Research Institute, Cruces Plaza, 48903 Barakaldo, Spain; iraide.alloza@ehu.eus (I.A.); k.vandenbroeck@ikerbasque.org (K.V.)
4 Ikerbasque, Basque Foundation for Science, 3 María Díaz Haroko Kalea, 48013 Bilbao, Spain
* Correspondence: Andy.Baker@ed.ac.uk

Citation: Bennett, M.; Ulitsky, I.; Alloza, I.; Vandenbroeck, K.; Miscianinov, V.; Mahmoud, A.D.; Ballantyne, M.; Rodor, J.; Baker, A.H. Novel Transcript Discovery Expands the Repertoire of Pathologically-Associated, Long Non-Coding RNAs in Vascular Smooth Muscle Cells. *Int. J. Mol. Sci.* **2021**, *22*, 1484. https://doi.org/10.3390/ijms22031484

Academic Editors: Miguel Hueso and Estanislao Navarro
Received: 13 January 2021
Accepted: 29 January 2021
Published: 2 February 2021

Abstract: Vascular smooth muscle cells (VSMCs) provide vital contractile force within blood vessel walls, yet can also propagate cardiovascular pathologies through proliferative and pro-inflammatory activities. Such phenotypes are driven, in part, by the diverse effects of long non-coding RNAs (lncRNAs) on gene expression. However, lncRNA characterisation in VSMCs in pathological states is hampered by incomplete lncRNA representation in reference annotation. We aimed to improve lncRNA representation in such contexts by assembling non-reference transcripts in RNA sequencing datasets describing VSMCs stimulated in vitro with cytokines, growth factors, or mechanical stress, as well as those isolated from atherosclerotic plaques. All transcripts were then subjected to a rigorous lncRNA prediction pipeline. We substantially improved coverage of lncRNAs responding to pro-mitogenic stimuli, with non-reference lncRNAs contributing 21–32% for each dataset. We also demonstrate non-reference lncRNAs were biased towards enriched expression within VSMCs, and transcription from enhancer sites, suggesting particular relevance to VSMC processes, and the regulation of neighbouring protein-coding genes. Both VSMC-enriched and enhancer-transcribed lncRNAs were large components of lncRNAs responding to pathological stimuli, yet without novel transcript discovery 33–46% of these lncRNAs would remain hidden. Our comprehensive VSMC lncRNA repertoire allows proper prioritisation of candidates for characterisation and exemplifies a strategy to broaden our knowledge of lncRNA across a range of disease states.

Keywords: vascular smooth muscle cells; long non-coding RNAs; RNA sequencing; enhancers

1. Introduction

The principal role of vascular smooth muscle cells (VSMCs) in their differentiated state is to provide contractile force in the vessel wall to ensure proper circulation. However, a high level of plasticity relative to other cell types is well established [1–3], with phenotypes, such as proliferation, migration, and extracellular matrix production, often displayed at the expense of contractility. This adaptability can aid vessel growth and repair in response to a wide range of biochemical signals or mechanical stresses [1]. Conversely, it also contributes to vessel wall remodelling during some of the most prevalent and life-threatening cardiovascular diseases, such as atherosclerosis, pulmonary hypertension, and restenosis [2]. The molecular mechanisms controlling the wide variety of phenotypic changes involved in such diseases are controlled by both coding and non-coding genes

at both the genetic and epigenetic level [3,4]. However, they are not yet understood sufficiently to effectively target them therapeutically.

Long non-coding RNAs (lncRNAs), RNA transcripts > 200 bp in length that do not produce proteins, are mostly uncharacterised. Yet several have been found to be key to a range of vital cellular processes via contributing to epigenetic, transcriptional, or translational regulation [5,6]. However, only a few hundred are experimentally characterised out of tens of thousands annotated through genome-wide sequencing efforts [7]. Therefore, lncRNAs represent a potential cache of novel mechanistic processes crucial for cell function. In addition, many lncRNAs show a high level of specificity in terms of their expression across tissues, cells, and timepoints during development or stimuli responses [8,9], raising the prospect of using them as markers or therapeutic targets to tackle aberrant cell behaviour. For example, recent studies have described *SMILR* [10], *MYOSLID* [11], and *SENCR* [12] with biased expression to VSMCs or the vasculature and as key regulators of cell cycle, migration and differentiation state. However, lncRNA discovery in VSMCs remains limited and relatively unexplored. Therefore, the overall scale of lncRNA contribution to aberrant VSMC behaviour during tissue remodelling is still an open question that crucially needs to be addressed.

An obstructing factor in determining lncRNA regulation of VSMCs is that the annotation of such genes is incomplete, even across the extensively annotated human genome [13]. Even reference annotations, such as GENCODE [14], considered gold standard [13], and the more extensive FANTOM CAT [15], which integrates 5 smaller reference annotations, are missing lncRNAs. The datasets used to create such annotations cannot represent all possible biological and pathological settings so reference annotations are inherently incomplete. In addition, low abundance and lack of poly-adenylated tails for some lncRNAs means many are difficult to detect in transcriptomic datasets, which are polyA-enriched or of insufficient sequencing depth. A tendency for cell-specific expression also hinders their detection in samples containing heterogenous mixtures of cell types, such as tissues. Accordingly, several efforts have aimed to expand human lncRNA annotations beyond reference annotations. For example, building transcripts de novo from sequencing data was found to yield a substantially greater number of lncRNAs of interest in contexts, such as erythropoiesis [16], formation of CD8+ memory T cells [17], and psoriatic skin tissue [18]. Novel lncRNAs obtained from such approaches have demonstrated particularly high expression specificity, in particular cell type specificity and/or high likelihood of differential expression between conditions or stimuli. They have also shown a particular association with enhancer sites, activators of localised transcription often involved in epigenetic control of cellular states [19]. Enhancers that produce lncRNAs show increased signs of influence on neighbouring transcription relative to other enhancers. These lncRNAs are likely to aid formation of regulatory complexes at many of these enhancer sites [20]. Efforts to improve human VSMC lncRNA annotation include studies on coronary artery VSMCs (caSMCs) maintained in standard growth conditions [12] or pushed toward a differentiated phenotype via overexpression of *MYOCD* [11], which together highlighted the pro-contractile lncRNAs *SENCR* and *MYOSLID*. However, no novel transcript discovery methods have yet been applied to VSMCs responding to stimuli that induce proliferative or pro-inflammatory phenotypes. This means a subset of lncRNAs with potentially crucial roles in VSMC-directed tissue remodelling could remain hidden.

To obtain a more complete representation of the lncRNAs expressed in pathologically active VSMCs, we obtained three published RNA sequencing (RNAseq) datasets describing in vitro stimulation of VSMCs into proliferative, migratory, or pro-inflammatory phenotypes or isolated from diseased tissue in vivo. We then combined a novel transcript discovery approach with a stringent lncRNA annotation pipeline, thereby markedly improving the coverage of stimuli-responsive, VSMC-enriched and enhancer-transcribed lncRNAs within VSMC pathology. Our study highlights lncRNAs with high potential to control VSMC pathological states.

View our data interactively at: https://bakergroup.shinyapps.io/VSMClncRNAannotation/.

3.3.3 Manuscript Results

2. Results

2.1. A Bioinformatic Approach to Provide a More Complete Annotation of lncRNAs Expressed in VSMCs in Basal and Pathological Conditions

Several LncRNAs are known to be involved in VSMC phenotypic transitions occurring in vessel wall remodelling [10–12]. However, a full accounting of lncRNAs expressed in these transitions does not yet exist. Accordingly, to gain in-depth representation of the lncRNAs expressed in pathological VSMCs, we applied a transcript discovery pipeline to published VSMC RNAseq datasets selected based on specific criteria (Figure 1a). We focused on high depth, paired-end total RNA sequencing datasets to identify all lncRNAs and their gene structures (including non-polyA tailed and/or lowly expressed lncRNAs) and we selected two in vitro datasets fulfilling these criteria. The first dataset describes primary human saphenous vein SMCs (svSMCs) either quiesced in 0.2% FBS or treated with interleukin-1α (Il-1α) and/or platelet-derived growth factor-BB (PDGF-BB) [10]. The second dataset describes primary aortic (aoSMCs) or coronary artery (caSMCs) VSMCs, plated in 5% FBS media onto soft or stiff culture matrices [21]. These conditions model a convergence of pro-inflammatory and pro-mitogenic signals, or mechanical stretch in the vessel wall, both of which promote proliferation and disruption of contractility. We also selected an in vivo dataset describing VSMCs isolated and sequenced directly from enzymatically digested carotid plaques derived from symptomatic or asymptomatic patients, defined as such based on lumen size and occurrence of cardiovascular events prior to surgery [22]. Together, these three datasets document a broad span of VSMC types and phenotypes contributing to vessel wall remodelling.

The transcriptome analysis used GENCODE annotation as a reference and included a transcript assembly step to further identify transcripts not previously described in GENCODE (newly assembled transcripts). This approach allows the identification of novel isoforms for GENCODE genes, but also allows the identification of novel genes (newly assembled genes). This analysis was carried out independently for the three datasets. These expanded transcriptomes consisted of ~80,000–90,000 transcripts in total with 0.6–1.5% transcribed from newly assembled genes (Table S1). To identify high confidence lncRNAs from these complete transcriptomes, we used "Pipeline for Annotation of LncRNAs" (PLAR) [23], which filters lowly-expressed or artefactual transcripts and assesses coding potential based on three distinct tools. As expected, the bulk of expressed transcripts were annotated as protein coding (Figure S1A). However, a high confidence set of ~2500–3000 lncRNA annotations were predicted within each transcriptome, with 6–7% deriving from newly assembled gene loci (Table S1). Our analysis of robustly expressed genes showed that newly assembled lncRNA genes produced transcripts with comparable lengths to transcripts from GENCODE protein-coding genes (PCGs) or GENCODE lncRNA genes (Figure S1B–E). Of these genes, newly assembled and GENCODE lncRNAs have a lower expression compared to PCGs, as expected. Further, the GENCODE lncRNA genes were also more abundant than the newly assembled lncRNA genes but only by a median difference of ~1 FPKM (Figure S1F–I).

To show the validity of the transcript discovery pipeline for lncRNA identification, we assessed if the newly assembled lncRNA transcripts identified in the three datasets were observed in other reference databases. Using GFFcompare [24], we cross-referenced expressed GENCODE and newly assembled lncRNA transcript structures to transcripts annotated in FANTOM CAT [15], a particularly extensive reference annotation. We observed 72% of GENCODE and 40% of newly assembled lncRNAs matched to a FANTOM transcript containing the exact same chain of introns whilst another 25% of GENCODE and 40% of newly assembled lncRNAs contained at least 1 matching splice junction site (Figure 1b). The validation of the complete or partial gene structures for a large proportion of the newly assembled lncRNAs in other annotation sets (derived from other contexts) provide confidence in the identified transcripts and evidence of the lncRNA expression in different datasets. To find further corroborating evidence of transcription for our lncRNAs, we used FANTOM CAT CAGEseq data which accurately defines transcription start sites

(TSSs) in ~1800 distinct human samples through sequencing the site of RNA 5′ capping [15]. We identified 74.3% of newly assembled and 86.9% of GENCODE lncRNA genes across all datasets matched to experimentally validated TSSs in FANTOM CAGEseq data (hereafter referred as CAGE-matched lncRNAs) (Figure 1c). The position of these CAGEseq matches indicates the first exons of newly assembled lncRNAs from our analysed datasets were largely complete at their 5′ ends (incomplete by median of 8% of their initial size) (Figure S2).

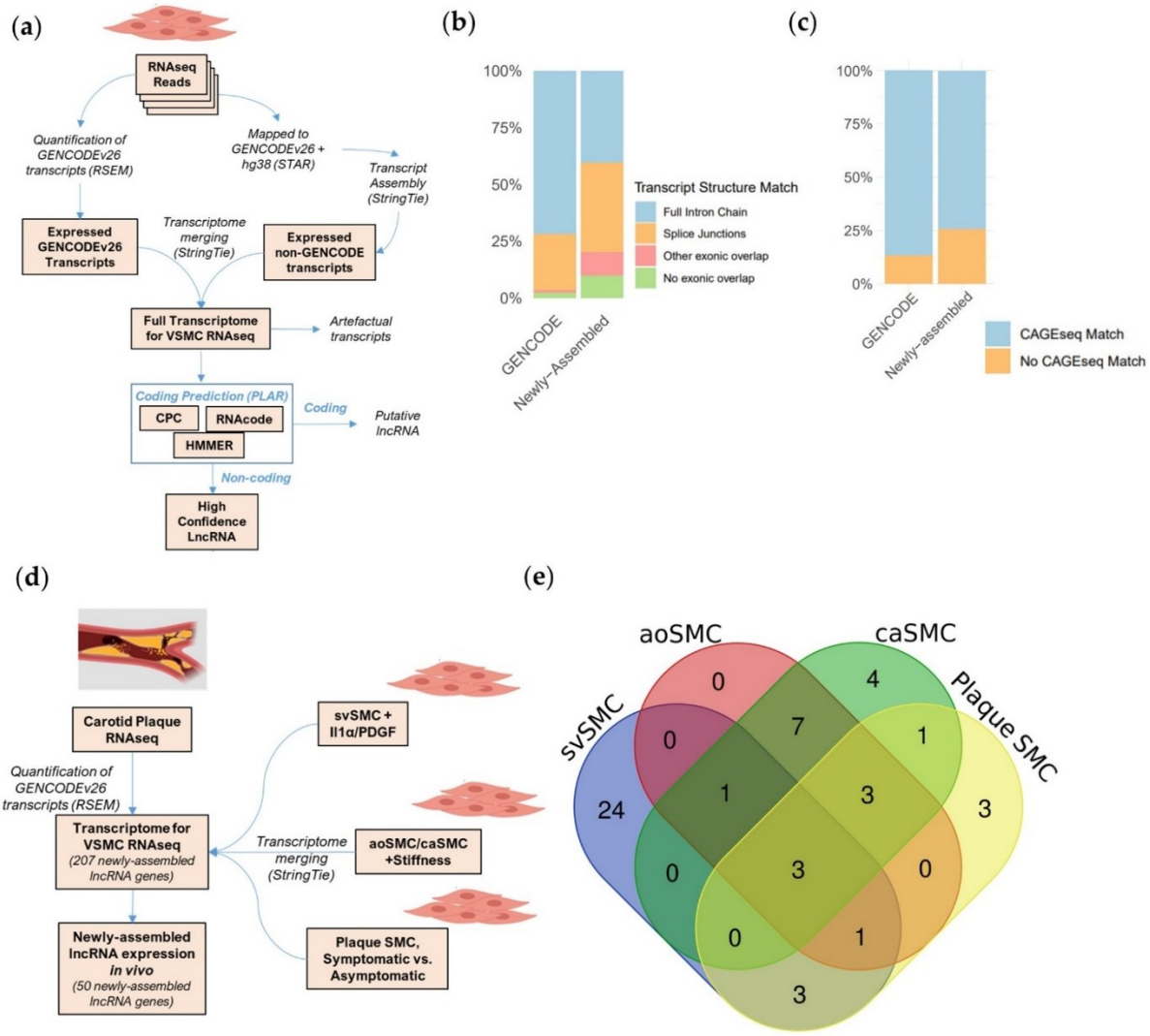


Figure 1. Identification of a high-confidence lncRNA repertoire for vascular smooth muscle cells (VSMCs) in physiological and pathological states. (**a**) Strategy to supplement GENCODE annotation with newly assembled transcripts and annotate lncRNAs. (**b**) Proportion of GENCODE and newly assembled lncRNA transcripts with structures matching FANTOM CAT annotation. (**c**) Proportion of GENCODE and newly assembled lncRNAs with a matching CAGE sequencing (CAGEseq) site in FANTOM CAT database. (**d**) Method to detect newly assembled lncRNAs in carotid plaque RNAseq. (**e**) Newly assembled lncRNAs derived from VSMC datasets detected in whole plaque tissue.

To gain perspective on the in vivo relevance of newly assembled lncRNAs, we assayed their expression in an RNAseq dataset of carotid plaque tissue. VSMCs are major components of atherosclerosis plaque, with many demonstrating phenotypic modulation [25]. To assess all newly assembled lncRNAs simultaneously, we merged the three expanded

annotations into a non-redundant transcriptome containing 255 newly assembled lncRNA transcripts from 207 lncRNA genes (Figure 1d). Analysis of the plaque-derived RNAseq with this non-redundant merged transcriptome demonstrated that 50 (24%) of the newly assembled lncRNA genes were detectable in whole plaque tissue. This is a substantial detection rate, considering that plaques are heterogenous and contain non-VSMC cells contributing to the RNAseq. In addition, the newly assembled lncRNAs were identified in VSMCs from different vessel types grown in distinct conditions and so might not be expressed in plaques. Interestingly, these 50 lncRNAs come from all four independent transcriptomes (Figure 1e), showing each new annotation provides in vivo relevant transcripts. Notably, 24 of these lncRNAs were identified exclusively using the svSMC dataset showing this annotation particularly improved coverage of plaque-expressed lncRNAs.

Together, these analyses expand the representation of lncRNAs expressed in basal and pathological VSMCs in vitro and in vivo and provide confidence in the newly assembled gene structures.

2.2. Newly Assembled Genes Substantially Increase the Number of VSMC lncRNAs Detected in Response to Pathological Stimuli

With confidence in our expanded lncRNA annotation, we next sought to comprehensively identify all lncRNAs differentially expressed between conditions in the RNAseq datasets using DESeq2 [26]. For the svSMC dataset, we identified 162 differentially expressed lncRNAs between control and IL-1α/PDGF-BB stimulation (absolute fold change > 1.5, $p < 0.05$ *) out of the 598 robustly expressed lncRNA genes. Notably, newly assembled genes represented 32% of differentially expressed lncRNAs, more than would be expected by chance considering their proportion of total expressed lncRNAs (18%, $p < 0.0001$ ****, Fisher's exact test) (Figure 2a). Differential expression dynamics were validated via qRT-PCR in the same svSMC proliferation model for 7 GENCODE and three newly assembled lncRNAs chosen from the top 10% lncRNAs with the highest fold changes (Table S2 and S3). We obtained a high and significant correlation between the qRT-PCR and RNA-seq fold changes ($R = 0.97$, $p < 0.0001$ ****) (Figure 2b). In the VSMC response to stiff culturing dataset, we identified 143 out of 551 and 168 out of 539 differentially expressed lncRNAs for aoSMC and caSMC, respectively. Again newly assembled lncRNAs made up a larger proportion of the differentially expressed.

lncRNAs then would be expected by chance, considering their proportion of total expressed lncRNAs (Figure 2c,d) (25% vs. 15%, $p < 0.001$ *** and 21% vs. 15%, $p < 0.01$ **, Fisher's exact test for aoSMC and caSMC, respectively). This indicates a particular tendency for newly assembled lncRNA genes to respond to IL-1α/PDGF-BB or increased vascular stiffness, both pathologically associated as pro-mitogenic stimuli.

For the plaque-isolated VSMC dataset, though a comparable portion of expressed lncRNAs were newly assembled (16%), only 4 GENCODE lncRNAs and no newly assembled lncRNA genes were found differentially expressed between VSMCs from symptomatic and asymptomatic plaques. This small number of lncRNAs reflects the smaller number of transcriptional changes in between this dataset, even for protein coding genes (PCGs) (<1% of expressed genes) (Table S4), likely explained by the higher heterogeneity of plaque-derived VSMCs compared to cultured VSMCs.

To gain perspective on the role the differentially expressed lncRNAs may play in determining VSMC phenotypic state, we hierarchically clustered all differentially expressed genes based on their expression profile across all four conditions in either the saphenous vein-based model (Figure 2f) or the ao/caSMC stiff culture-based model (Figure 2g). We identified six clusters of gene expression changes in each dataset and used gene ontology analysis (using goseq [27]) to associate them with biological processes, cellular components and molecular functions (Figure 2h,i) (Tables S5–S16). GENCODE and newly assembled lncRNAs were found in most clusters, suggesting the contribution of distinct lncRNAs across the different identified processes. We noted a large proportion of newly assembled lncRNAs in cluster 3 and 4 of the svSMC dataset, involved in cytokine and immune response, and in cluster 6 of the ao/caSMC dataset, involved in ribosome, Cajal body and

mitochondrial activity. This indicates that newly assembled lncRNAs may be particularly relevant to these specific processes or compartments.

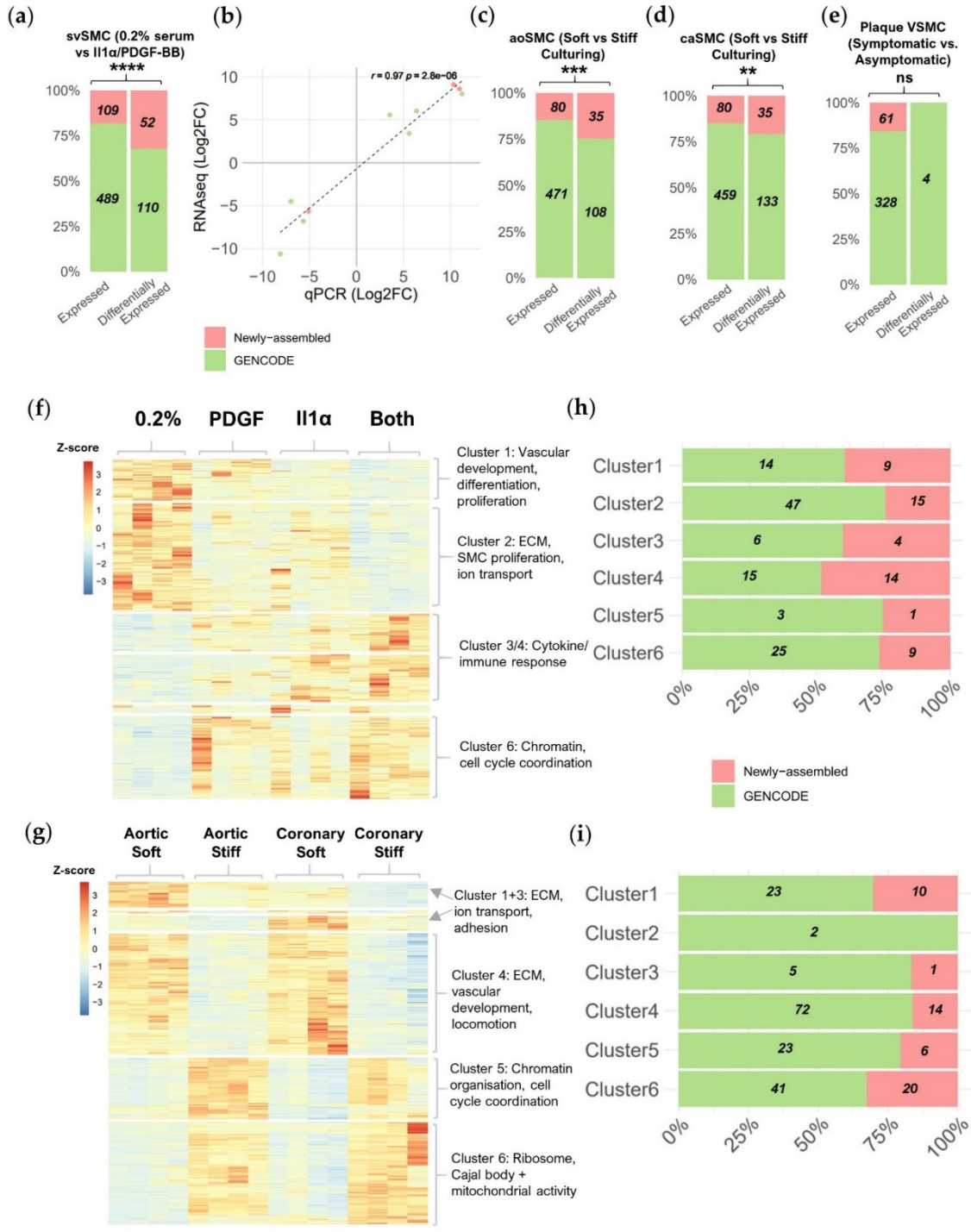


Figure 2. Newly assembled lncRNAs show a tendency to respond to pro-mitogenic stimuli. (**a**) Proportion of GENCODE and newly assembled lncRNAs expressed or differentially expressed in the saphenous vein VSMC (svSMC) dataset (Fisher's exact test, background of all expressed genes). (**b**) Validation of expression dynamics in the svSMC RNAseq dataset by qRT-PCR for six lncRNAs (Spearman's rank, $p = 6.7 \times 10^{-4}$). (**c–e**) Same as (**a**) but for remaining VSMC datasets. Expression heatmap of differentially expressed genes expression responding to (**f**) interleukin-1α (IL-1α)/platelet-derived growth factor-BB (PDGF-BB) or (**g**) stiff-culturing clustered hierarchically based on their variation across all samples with selected over-represented gene ontologies derived for each cluster. Proportion of GENCODE and newly assembled lncRNAs in each cluster of differentially expressed genes responding to (**h**) IL-1α/PDGF or (**i**) stiff-culturing. (**a**,**c–e**) $p < 0.0001$ ****, $p < 0.001$ ***, $p < 0.01$ **, ns not significant.

Together our differential expression analysis showed that newly assembled lncRNAs were more likely to be stimuli-responsive than reference lncRNAs, highlighting the importance of novel transcript discovery in pathological contexts.

2.3. Novel Transcript Discovery Increases the Representation of VSMC-Enriched lncRNAs with Pathological Association

Genes with cell type and/or state specific expression can hold particular functional relevance in the condition to which their expression is biased [28], and could be targeted by gene therapy approaches with minimal effects on neighbouring cells. Cell type-enriched lncRNAs could be less likely to be annotated in GENCODE than ubiquitously expressed lncRNAs. We therefore sought to examine whether by identifying genes absent from GENCODE, we concordantly increase the coverage of VSMC-enriched lncRNAs.

To assess the tendency for cell-type enriched expression of GENCODE and newly assembled lncRNAs, we again used the FANTOM CAT CAGEseq library, which contains expression data for 69 primary cell categories (including 9 VSMC subtypes) and 174 tissue categories consisting of 744 samples in total. The expression of 801 of the CAGE-matched lncRNAs (defined in Figure 1d) was accessible in FANTOM data (Table S17). The cell-type specificity of newly assembled and GENCODE lncRNAs was assessed by obtaining enrichment values for each lncRNA within each primary cell category compared to all other primary cell categories. We noted higher enrichment values for newly assembled lncRNAs in a select group of mesenchymal cell type categories including VSMCs. Enrichment values were much lower in several other primary cell categories, including leukocytes, endothelial, and epithelial cells (Figure S3A). In contrast, GENCODE lncRNAs enrichment was less variable across cell types (Figure S3B), suggesting the higher specificity of expression of newly assembled lncRNAs compared to GENCODE lncRNAs, which were more ubiquitous.

We next aimed to identify VSMC-enriched lncRNAs. FANTOM CAT has previously defined genes with cell-type enriched expression as those with a 5-fold or greater enriched expression in a given category when compared to all other categories [15]. Therefore, by selecting CAGE-matched lncRNAs with enriched expression in at least one VSMC type in the FANTOM CAT library, we were able to identify 72 VSMC-enriched lncRNAs across all datasets. Amongst the CAGE-matched lncRNAs, significantly more newly assembled lncRNAs were found to be VSMC-enriched than would be expected by chance considering their proportion of total expressed lncRNAs in svSMC (31% of VSMC-enriched vs. 12% of expressed lncRNAs, $p < 0.001$ ***, Fisher's exact test), caSMC (31% of VSMC-enriched vs. 10% of expressed lncRNAs, $p < 0.001$ ***, Fisher's exact test) or plaque SMCs (20% of VSMC-enriched vs. 10% of expressed lncRNAs, $p = 0.03$ *, Fisher's exact test) (Figure 3a,c,d). This trend was maintained albeit with borderline significance for aoSMCs (20% vs. 10%, $p = 0.07$, Fisher's exact test) (Figure 3b). Overall, this demonstrates that the newly assembled lncRNAs have a greater tendency for VSMC-enriched expression when compared to GENCODE lncRNAs.

To reveal the contribution of VSMC-enriched lncRNAs to VSMC phenotypic modulation, we evaluated their likelihood of differential expression in response to pathological stimuli. In the svSMC dataset, we found an increased tendency for VSMC-enriched lncRNAs to be responsive to IL-1α/PDGF-BB as compared to other lncRNAs (18% of differentially expressed vs. 8% of expressed lncRNAs, $p < 0.001$ ***, Fisher's exact test) (Figure 3e). This effect was also observed if considering lncRNAs responsive to stiffness in aoSMC (10% of differentially expressed vs. 6% of expressed lncRNAs, $p = 0.03$, Fisher's exact test) and caSMC (13% of differentially expressed vs. 8% of expressed lncRNAs, $p = 0.02$, Fisher's exact test) (Figure 3e–g), but not in the plaque VSMC dataset (as expected due to the low number of differentially expressed genes). VSMC-enriched lncRNAs are therefore particularly likely to be involved in regulating VSMC response to physiological stimuli.

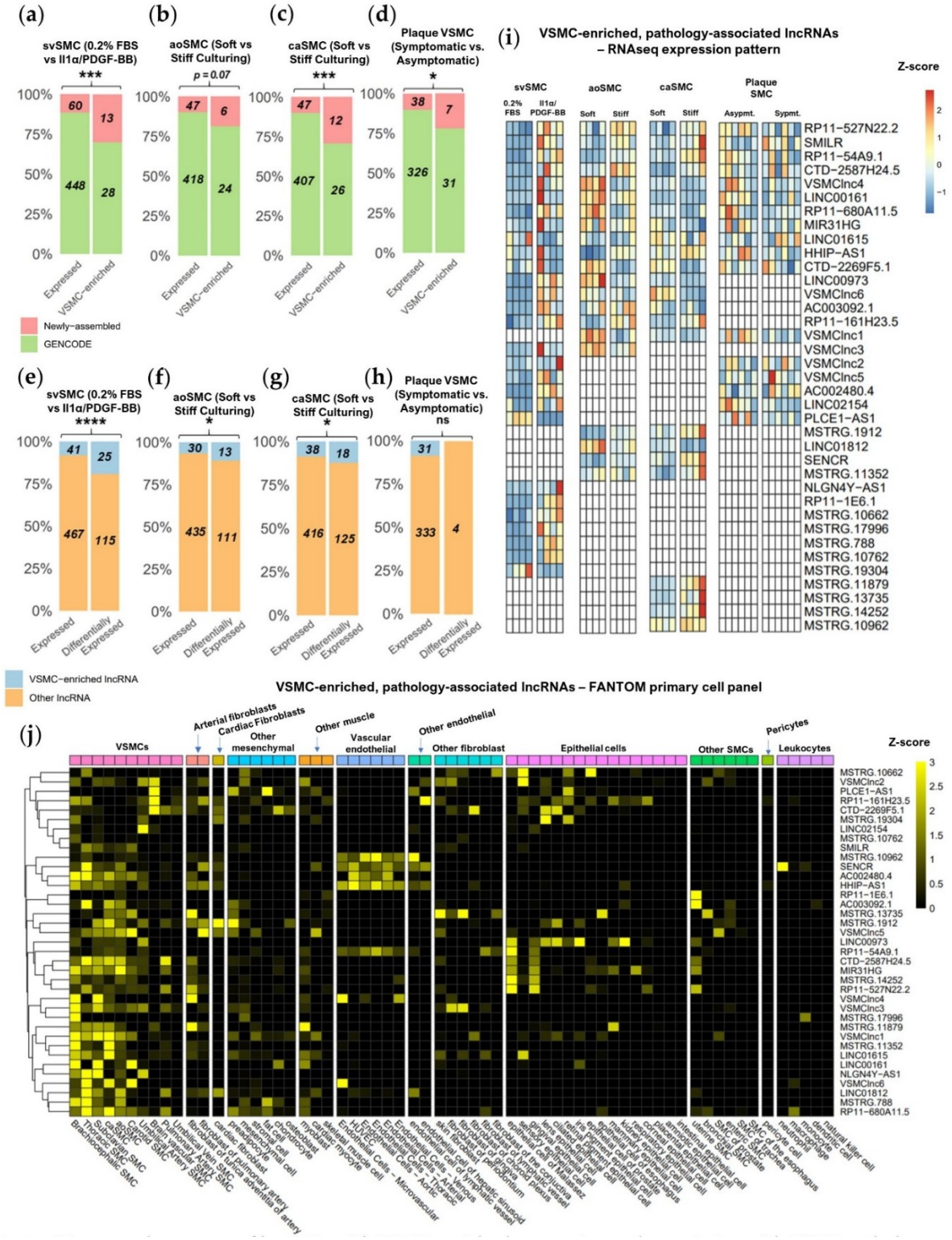


Figure 3. Increased coverage of lncRNAs with VSMC-enriched expression and association with VSMC pathology. (**a**–**d**) Proportion of newly assembled vs. GENCODE lncRNAs within CAGE-matched lncRNAs and CAGE-matched, VSMC-enriched lncRNAs. (**e**–**h**) Proportion of VSMC-enriched or non VSMC-enriched lncRNAs that are expressed or differentially expressed (Fisher's exact test for (**a**–**h**), background of CAGE-matched lncRNAs). (**i**) Expression heatmap of all 37 VSMC-enriched and differentially expressed lncRNAs responding to IL-1α/PDGF-BB or stiff culturing as well as within plaque VSMCs (Z score generated for each VSMC type individually, white cells indicate no robust expression). (**j**) Expression heatmap of all 37 VSMC-enriched and differentially expressed lncRNAs across vascular cell types and other relevant cell types in the FANTOM CAGE expression atlas. (**a**–**h**) $p < 0.0001$ ****, $p < 0.001$ ***, $p < 0.05$ *, ns not significant.

In total, we identified 37 VSMC-enriched lncRNAs, including 17 newly enriched lncRNAs, responding to either IL-1α/PDGF-BB or stiff-culturing (Figure 3i). The differential expression of these lncRNAs appeared mostly exclusive to a specific dataset with only four lncRNAs upregulated by both IL-1α/PDGF-BB in svSMC and stiff culturing in aoSMC or caSMC. This suggests that the VSMC-enriched lncRNAs respond to different stimuli or are expressed and regulated in specific VSMC subtypes. We also noted that 13 of the VSMC-enriched lncRNAs induced with IL-1α/PDGF-BB in svSMCs were mostly absent in quiescent svSMCs and robustly expressed in vivo in plaque-isolated VSMCs. Together this expression pattern strongly implicates the involvement of these lncRNAs in VSMC transitions to pathological states. In support of this, one lncRNA showing this pattern, SMILR, has known VSMC enrichment along with roles in both VSMC proliferation and vessel wall remodelling [10].

Interestingly, of all 37 VSMC-enriched, differentially-expressed lncRNAs, many showed a greater VSMC enrichment value than *SMILR* [10] and *SENCR* [12] the 2 lncRNAs in the list already characterised as functional and VSMC-enriched. This indicates existence of several lncRNAs with particularly high VSMC expression bias. For example, newly assembled lncRNA *VSMClnc6* and GENCODE lncRNAs *NLGN4Y-AS1* and *AC002480.4*, all induced by IL-1α/PDGF-BB, show a VSMC enrichment greater than 20-fold. In addition to their high VSMC enriched expression, some lncRNAs were also expressed to a lesser extent in arterial or cardiac fibroblasts, other mesenchymal types or other muscle types (Figure 3j and Figure S3C), suggesting related function in these cell types. Many also show expression bias to certain VSMC subtypes suggesting they may have some specialised function in particular vascular beds. Of note, VSMC-enrichment does not preclude lncRNAs from enrichment in other cell types including those involved in vessel wall remodelling. However, though some expression in vascular endothelial cells was observed this was largely confined to seven lncRNAs. Very little expression was seen across epithelial cells, non-vascular SMCs, leukocytes or pericytes.

Taken together, we show that newly assembled lncRNAs have a greater tendency for VSMC-enriched expression. In turn, VSMC-enriched lncRNAs are substantially associated with VSMC response to IL-1α/PDGF-BB or stiff-culturing. These pathological-associated and VSMC-enriched lncRNAs, therefore, represent potential candidate for the therapeutic targeting of VSMC phenotypic changes.

2.4. Novel Transcript Discovery in VSMCs Increases Evidence of Enhancer-Transcribed lncRNAs

As non-reference lncRNAs have previously shown a particular association with enhancers [18] and lncRNA-producing enhancers have shown greater signs of activity influencing neighbouring PCG expression [20], we aimed to identify elncRNAs and their potentially regulated PCGs.

To identify elncRNAs, we selected lncRNAs with a 5′ region overlapping a GeneHancer [29] enhancer site or matched to a CAGE site previously classed as an "elncRNA" in FANTOM CAT (Table S18). We found 110, 90, and 87 expressed elncRNAs in the svSMC, aoSMC or caSMC datasets, respectively. Newly assembled lncRNAs made up a higher proportion of the elncRNAs than would be expected by chance considering their proportion of all expressed lncRNAs (26% elncRNAs vs. 18% expressed lncRNAs $p = 0.01$, 23% elncRNAs vs. 15% expressed lncRNAs $p = 0.01$, 23% elncRNAs vs. 15% expressed lncRNAs $p = 0.02$, for svSMC, aoSMC, and caSMC respectively, Fisher's exact test) (Figure 4a–c). Hence, the newly assembled lncRNAs were particularly likely to be enhancer-transcribed.

We also observed a higher proportion of elncRNAs were differentially expressed in response to IL-1α/PDGF-BB than would be expected by chance compared to their proportion of all expressed lncRNAs (Figure 4d; 31% differentially expressed lncRNAs vs. 18% expressed lncRNAs $p < 0.0001$ ****, Fisher's exact test). This was also true for lncRNAs differentially expressed in response to stiff culturing in aoSMC (Figure 4e; 24% vs. 16% $p < 0.01$ **) and caSMC (Figure 4f; 21% vs. 16% $p = 0.02$ **). ElncRNAs are therefore more likely than other lncRNAs to be differentially expressed in VSMCs responding to these

pathological stimuli, with IL-1α/PDGF-BB in particular eliciting a large elncRNA response. Around 34% of these regulated elncRNAs were newly assembled (Figure 4g).

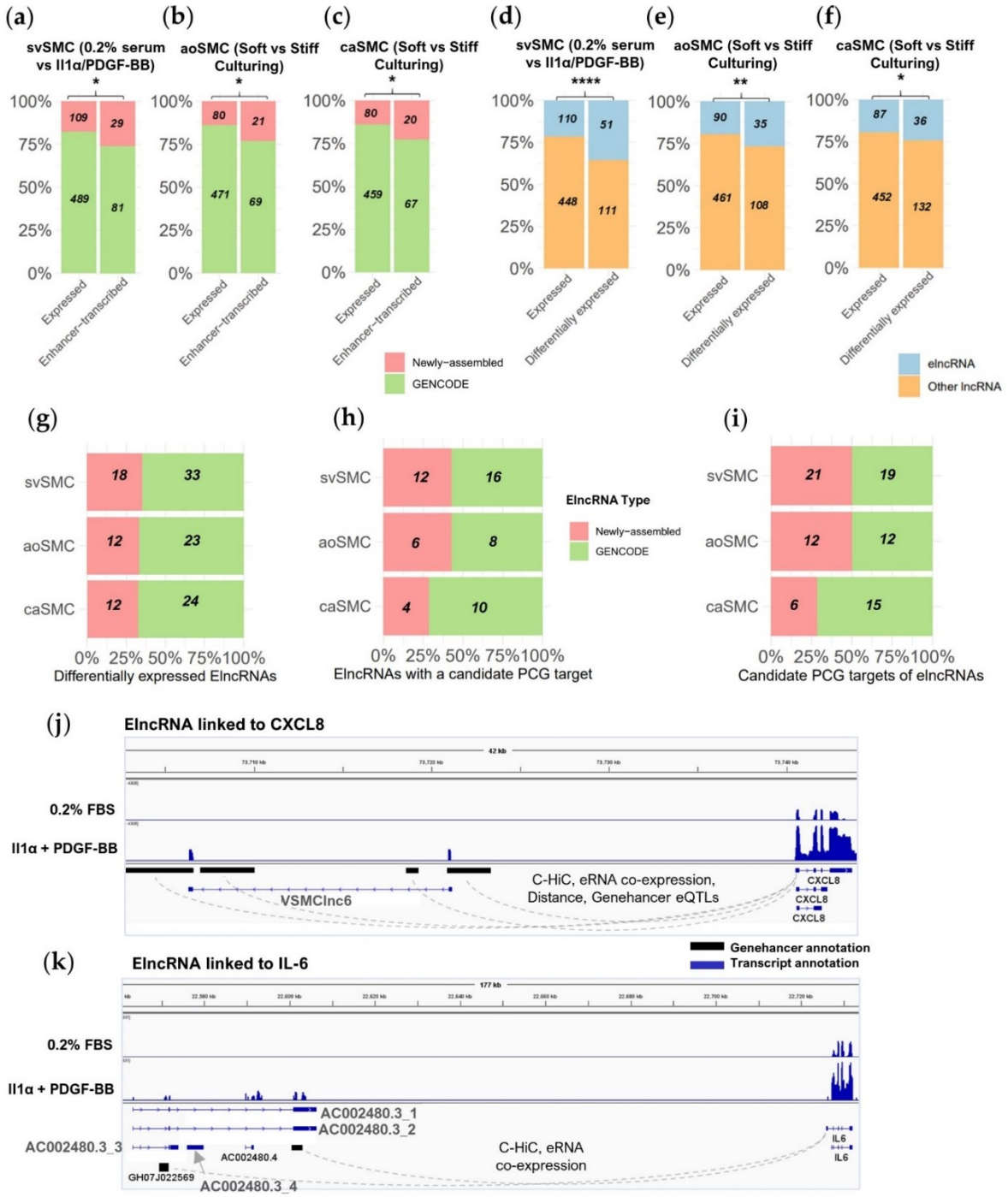


Figure 4. Increased coverage of elncRNAs and their association with VSMC pathology. (**a–c**) Proportion of newly assembled vs. GENCODE lncRNAs for expressed lncRNAs or elncRNAs in the three in vitro VSMC datasets (Fisher's exact test, background of expressed lncRNAs). (**d–f**) Proportion of elncRNA vs. other lncRNAs amongst expressed or differentially expressed lncRNAs (Fisher's exact test, background of expressed lncRNAs). (**g**) Proportions of newly assembled vs. GENCODE differentially expressed elncRNAs (**h**) Proportions of newly assembled vs. GENCODE elncRNAs with a candidate PCG target. (**i**) Proportion of candidate PCG targets identified for newly assembled and GENCODE elncRNA (any proximal to both types are considered as GENCODE). (**j**,**k**) Representative RNAseq coverage at two genomic regions with elncRNAs (AC002480.3 and VSMClnc6) linked to cytokine (*IL-6*) and chemokine (*CXCL8*) genes by GeneHancer interaction data. (**a–f**) $p < 0.0001$ ****, $p < 0.01$ **, $p < 0.05$ *.

ElncRNAs may increase the activity of their associated enhancer thereby promoting expression of a proximal PCG [20]. To identify candidate PCGs that may be regulated in this manner, we identified all PCGs located within 250 kbp of a differentially expressed elncRNA and co-induced or co-repressed with the elncRNAs (Figure S4). We found candidate PCG targets for 55% of IL-1α/PDGF-BB-responsive elncRNAs and 39% of stiffness-responsive elncRNAs in aoSMC or caSMC (Figure 4h and Figure S4). For the svSMCs and aoSMCs, the number of candidate PCG targets were doubled as a result of the inclusion of newly assembled elncRNAs (Figure 4i).

To find external evidence supporting the regulation of these PCGs by elncRNAs, we assessed if any were linked to elncRNAs in GeneHancer or FANTOM CAT interaction data. GeneHancer interaction annotations are based on (1) physical association with a PCG promoter (using Capture Hi-C data); (2) presence of SNPs linked to changes in PCG expression (expression quantitative trait loci—eQTLs); (3) presence of motifs shared with a nearby PCG promoter for use by a co-expressing transcription factor; or (4) production of enhancer RNAs that co-express with nearby PCGs. In FANTOM CAT, lncRNA/PCG pairs are linked by eQTL-associated SNPs [15]. We found 13 differentially expressed elncRNAs (6 newly assembled) with evidence linking them to candidate PCG targets (Table S19). Interestingly, 6 out of these 13 candidate PCG targets have been previously shown to contribute to SMC pathology or to be involved in pathological proliferative phenotypes (Table 1). For example, elncRNAs *VSMClnc6* and *AC002480.3* are both induced by IL-1α and PDGF-BB in svSMC and linked to expression of the cytokine IL-6 and the chemokine CXCL8 via GeneHancer interaction data (Figure 4j,k). IL-6 and CXCL8 are key pro-inflammatory mediators known to be induced by IL-1α and characterised as promoting VSMC proliferation, migration, pro-inflammatory activity, and vascular remodelling [30–34]. As mentioned above, VSMClnc6 has a high VSMC enrichment value, raising the prospect that this elncRNA may be part of a VSMC-enriched mechanism regulating the expression of *CXCL8*.

Table 1. Differentially expressed elncRNAs with predicted protein-coding gene (PCG) targets and links to VSMCs, proliferation and/or migration.

ElncRNA Name [1]	Interaction Evidence	PCG	VSMC Type + Stimulus	Stimulus Response	Max. FANTOM VSMC Enrichment	PCG VSMC Characterisation
VSMClnc6 (svSMC + stiff-culturing)	CHi-C, eRNA co-expression, Distance, GeneHancer eQTLs	*CXCL8*	svSMC + IL-1α/PDGF-BB	Co-induced	LncRNA: 41-fold PCG: None	Activates VSMC proliferation + migration [30,31]
AC002480.3	CHi-C, eRNA co-expression	*IL-6*	svSMC + IL-1α/PDGF-BB	Co-induced	LncRNA: None (overlaps AC002480.4: 20-fold) PCG: 19-fold	Activates VSMC proliferation/migration + osteoblast phenotype [32–34]
LINC00973	FANTOM eQTL analysis	*DCBLD2*	All datasets	Co-induced with IL-1α/PDGF-BB Co-repressed with stiffness	LncRNA: 7-fold PCG: 5-fold	Regulates PDGFR + Il-8 expression, Marker of vascular remodelling [35]
AC009229.5	FANTOM eQTL analysis	*CYP1B*	aoSMC + stiffness	Co-repressed	LncRNA: None PCG: None	Mediates angiotensin II-induced VSMC proliferation + migration [36]
MSTRG.10933 (svSMC)	Distance, GeneHancer eQTLs	*GLS*	svSMC + IL-1α/PDGF-BB	Co-repressed	LncRNA: None PCG: None	Required for TGFβ-induced myofibroblast differentiation + pro-fibrotic marker expression [37]
NR2F2-AS1	GeneHancer eQTLs, CHi-C	*NR2F2 (COUP-TFII)*	svSMC + IL-1α/PDGF-BB	Co-repressed	LncRNA: None PCG: None	Ablation leads to osteoblast-like phenotype in mesenchymal precursors [38]

[1] Newly assembled lncRNAs given name (see methods) along with their respective RNAseq dataset in brackets.

3.3.4 Manuscript Discussion

3. Discussion

In this study, we used novel transcript discovery and a rigorous lncRNA prediction pipeline on two in vitro VSMC and one in vivo VSMC RNAseq datasets to expand the lncRNA annotation of VSMCs in pathological states. We identified 61–109 newly assembled lncRNAs expressed in each of these datasets. Interestingly, these newly assembled lncRNAs were more likely to be differentially expressed in response to pathological stimuli, to

be VSMC-enriched, and/or transcribed from enhancer regions (elncRNAs), compared to previously annotated lncRNAs. As enhancers regulate local transcription, we also predicted neighbouring PCGs regulated by the elncRNAs and show many of these PCGs are involved in VSMC pathology. Taken together, we demonstrate that inclusion of non-reference transcripts is crucial to get a complete representation of lncRNAs and their contribution to the regulation of VSMC pathology. VSMCs in such states are rarely used to annotate non-reference transcripts and so our expanded repertoire is a unique source of potential regulatory mechanisms that could contribute to many vascular pathologies.

Both in vitro datasets analysed in this study model increased VSMC proliferation [10,21] and this was confirmed in our analysis through identification of clusters of upregulated genes associated with cell-cycle processes. The presence of lncRNAs in these clusters shows their potential role in regulating proliferation. We also predict that two elncRNAs within the IL-1α/PDGF-BB induced cell-cycle cluster, *AC002480.3* and *VSMClnc6*, regulate the cytokine *IL-6* and chemokine *CXCL8* genes respectively. These two factors are known to promote VSMC proliferation [31,33], yet are also indicative of the pro-inflammatory activity induced by IL-1α in the svSMC dataset. Activation of pro-inflammatory cascades likely aids the proliferation of VSMCs, in part through a synergistic potentiation of their response to PDGF-BB [39]. Additionally, VSMC pro-inflammatory activity is at the core of vessel wall remodelling contexts where a sustained inflammatory response, often from senescent VSMCs, promotes the influx of myeloid cells [3]. We see clusters of genes associated with cytokine response induced with IL-1α alone or combined IL-1α/PDGF-BB stimulation and a particularly large number of newly assembled lncRNAs were found in these clusters. This suggests we also capture lncRNA activity involved in IL-1α-driven pro-inflammatory VSMC phenotypes, as well as cell-cycle processes.

In addition to the *IL-6*- and *CXCL8*-associated elncRNAs, we were able to identify several other elncRNAs likely to regulate PCGs with established roles in VSMC pathology. The elncRNA *LINC00973* for instance is co-induced during IL-1α/PDGF-BB stimulation with *DCBLD2*, a PCG known to regulate PDGFR surface levels to promote VSMC proliferation [35]. *GLS* and *NR2F2* (aka *COUPTFII*) are two particularly notable elncRNA-associated PCGs that are co-repressed with IL-1α/PDGF-BB and are involved in promoting pro-fibrotic activity in myofibroblasts (*GLS* [37]) or defining mesenchymal lineage in the vasculature (*NR2F2* [38]). ElncRNAs are only one of many types of components involved in changing the chromatin accessibility at sites associated with VSMC pathology [4]. Therefore, any further characterisation of the elncRNAs highlighted here must put their potential mechanism in context with any recruitment of transcription factors or histone-modifying enzymes.

Our approach to fully define lncRNA contribution could be used in transcriptomics analysis of other contexts of VSMC pathological activation that remain to be explored. For instance subsets of VSMCs in the vessel wall appear particularly prone to phenotypic modulation including those derived from adventitial stem cells and responsible for neointimal VSMC proliferation in mouse injury models [40,41]. The pipeline could also improve lncRNA coverage in poorly annotated animal models of cardiovascular disease. High turnover of lncRNA sequences during evolution means conserved lncRNAs are rare despite their high potential for function [15]. Expanding the lncRNA annotation of human and animal models is key to maximise discovery of such relationships. For example, lncRNA annotation in rat VSMCs stimulated with pro-inflammatory Angiotensin-II has extended coverage of Angiotensin-II-responsive lncRNAs in rats [42]. Using a matching approach in human VSMCs would allow comprehensive detection of rat-human conserved lncRNAs, which could be characterised in vivo in rats to provide relevant data for clinical relevance in human. For the same reason, it would be beneficial to match the lncRNAs highlighted in our study to orthologous lncRNAs in analogous animal models of VSMC proliferation, with our study providing a template methodology to achieve this.

Our use of an expression atlas to define cell-type enrichment in the VSMC for the expanded lncRNA repertoire aids predictions of lncRNA location and function in the

vessel wall. The repertoire showed highest enrichment values within mesenchymal cells, including VSMCs. We identify 37 VSMC-enriched lncRNA which responded to either IL-1α/PDGF-BB or stiff-culturing (with 17 of these obtained through novel transcript discovery) and show their expression appears largely limited to VSMCs with a secondary tendency for expression in arterial adventitial fibroblasts. This could reflect the inherent similarity between VSMCs and fibroblasts or also be indicative of the fibroblast-like transcriptional profile that phenotypically modulated VSMCs have been observed to take on in vivo [25]. We show that a majority of these lncRNAs have minimal expression in other cell types involved in vessel wall remodelling, such as endothelial cells or leukocytes, which could be of value therapeutically. For example, targeting VSMC proliferation whilst preserving the endothelial barrier as a protective layer could be an effective strategy to improve the clinical outcome of late vein graft failure [43].

Similarly to cell-type specific lncRNAs, those with stimuli-specific expression are likely missing from reference annotation which cannot cover all biological conditions [13]. We see a particularly high proportion of newly assembled lncRNAs in induced clusters of genes associated with cytokine/chemokine response (46% newly assembled) and ribosomal/mitochondrial/Cajal activity suggestive of increased biosynthesis possibly related to proliferation (33% newly assembled). This could be explained by a low representation of these pathways in GENCODE. If so, many newly assembled lncRNAs may represent stimulus-specific lncRNAs, a trait that could in turn explain why newly assembled lncRNAs generally showed a high tendency to be differentially expressed. Further studies are required to study these stimuli-induced lncRNAs in non-VSMCs to determine if their expression is activated by the same stimulus on other cell types.

FANTOM and GeneHancer databases were used in this study to validate the structure of the identified lncRNAs and provide further characterisation in terms of expression or location relative to enhancer regions. Further analyses would benefit from using datasets matching to the VSMC type/stimuli in the datasets rather than these generic databases. For instance, no CAGE sites were found in FANTOM data for 26% of newly assembled lncRNAs. However, some of these missing TSSs could be identified using CAGEseq from VSMCs in the same context and these may represent particularly specifically expressed lncRNAs. Similarly, the identification of elncRNAs and their paired PCGs could be improved by studying chromatin marks in the same VSMC context as the RNAseq and by applying recently-developed, to link promoters and enhancers [44].

The tendencies identified amongst newly assembled genes for differential expression, cell-specificity and enhancer association are in broad agreement with other human lncRNA annotation efforts describing skin psoriasis [18] and cell state transitions [16,17]. We underscore the value of using such pipelines to highlight areas not yet covered by reference annotation, even in the extensively annotated human genome. RNAseq is an unbiased approach in comparison to microarray technology yet is under-utilised if relying solely on predefined reference annotation. An ever-increasing amount of high-quality sequencing data is available to profile lncRNA expression in a similar manner.

Reference annotation cannot capture the full transcriptional variety of all cellular states at all times. Therefore, focused annotation efforts such as in this study allow capturing missing details. We reinforce that these missing details provide important definition by identifying key components of the lncRNAs associated with VSMC pathology. We demonstrate an easily implemented approach to achieve this and provide a resource to help identify key candidate regulators of VSMC pathological states for investigation. We recommend similar strategies to comprehensively map lncRNA and maximise knowledge of their function in homeostasis and disease more broadly, as well as their potential for therapeutic manipulation.

3.3.5 Manuscript References

References

1. Alexander, M.R.; Owens, G.K. Epigenetic Control of Smooth Muscle Cell Differentiation and Phenotypic Switching in Vascular Development and Disease. *Annu. Rev. Physiol.* **2012**, *74*, 13–40. [CrossRef]
2. Frismantiene, A.; Philippova, M.; Erne, P.; Resink, T.J. Smooth Muscle Cell-Driven Vascular Diseases and Molecular Mechanisms of VSMC Plasticity. *Cell. Signal.* **2018**, *52*, 48–64. [CrossRef]
3. Sorokin, V.; Vickneson, K.; Kofidis, T.; Woo, C.C.; Lin, X.Y.; Foo, R.; Shanahan, C.M. Role of Vascular Smooth Muscle Cell Plasticity and Interactions in Vessel Wall Inflammation. *Front. Immunol.* **2020**, *11*, 3053. [CrossRef]
4. Gurung, R.; Choong, A.M.; Woo, C.C.; Foo, R.; Sorokin, V. Genetic and Epigenetic Mechanisms Underlying Vascular Smooth Muscle Cell Phenotypic Modulation in Abdominal Aortic Aneurysm. *Int. J. Mol. Sci.* **2020**, *21*, 6334. [CrossRef]
5. Dykes, I.M.; Emanueli, C. Transcriptional and Post-Transcriptional Gene Regulation by Long Non-Coding RNA. *Genom. Proteom. Bioinform.* **2017**, *15*, 177–186. [CrossRef]
6. Haemmig, S.; Simion, V.; Yang, D.F.; Deng, Y.H.; Feinberg, M.W. Long Noncoding RNAs in Cardiovascular Disease, Diagnosis, and Therapy. *Curr. Opin. Cardiol.* **2017**, *32*, 776–783. [CrossRef]
7. Volders, P.J.; Anckaert, J.; Verheggen, K.; Nuytens, J.; Martens, L.; Mestdagh, P.; Vandesompele, J. Lncipedia 5: Towards a Reference Set of Human Long Non-Coding Rnas. *Nucleic Acids Res.* **2019**, *47*, D135–D139. [CrossRef]
8. Washietl, S.; Kellis, M.; Garber, M. Evolutionary Dynamics and Tissue Specificity of Human Long Noncoding RNAs in Six Mammals. *Genome Res.* **2014**, *24*, 616–628. [CrossRef]
9. Gloss, B.S.; Dinger, M.E. The Specificity of Long Noncoding RNA Expression. *Biochim. Biophys. Acta Gene Regul. Mech.* **2016**, *1859*, 16–22. [CrossRef]
10. Ballantyne, M.D.; Pinel, K.; Dakin, R.; Vesey, A.T.; Diver, L.; Mackenzie, R.; Garcia, R.; Welsh, P.; Sattar, N.; Hamilton, G.; et al. Smooth Muscle Enriched Long Noncoding RNA (SMILR) Regulates Cell Proliferation. *Circulation* **2016**, *133*, 2050–2065. [CrossRef]
11. Zhao, J.J.; Zhang, W.; Lin, M.Y.; Wu, W.; Jiang, P.T.; Tou, E.; Xue, M.; Richards, A.; Jourd'heuil, D.; Asif, A.; et al. MYOSLID Is a Novel Serum Response Factor-Dependent Long Noncoding RNA That Amplifies the Vascular Smooth Muscle Differentiation Program. *Arterioscler. Thromb. Vasc. Biol.* **2016**, *36*, 2088–2099. [CrossRef]
12. Bell, R.D.; Long, X.C.; Lin, M.Y.; Bergmann, J.H.; Nanda, V.; Cowan, S.L.; Zhou, Q.; Han, Y.; Spector, D.L.; Zheng, D.Y.; et al. Identification and Initial Functional Characterization of a Human Vascular Cell-Enriched Long Noncoding RNA. *Arterioscler. Thromb. Vasc. Biol.* **2014**, *34*, 1249–1259. [CrossRef]
13. Uszczynska-Ratajczak, B.; Lagarde, J.; Frankish, A.; Guigó, R.; Johnson, R. Towards a Complete Map of the Human Long Non-Coding RNA Transcriptome. *Nat. Rev. Genet.* **2018**, *19*, 535–548. [CrossRef]
14. ENCODE Consortium. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **2012**, *489*, 57–74. [CrossRef]
15. Hon, C.C.; Ramilowski, J.A.; Harshbarger, J.; Bertin, N.; Rackham, O.J.L.; Gough, J.; Denisenko, E.; Schmeier, S.; Poulsen, T.M.; Severin, J.; et al. An Atlas of Human Long Non-Coding RNAs with Accurate 5' Ends. *Nature* **2017**, *543*, 199–204. [CrossRef]
16. Alvarez-Dominguez, J.R.; Hu, W.; Yuan, B.; Shi, J.; Park, S.S.; Gromatzky, A.A.; Van Oudenaarden, A.; Lodish, H.F. Global Discovery of Erythroid Long Noncoding RNAs Reveals Novel Regulators of Red Cell Maturation. *Blood* **2014**, *123*, 570–581. [CrossRef]
17. Hudson, W.H.; Prokhnevska, N.; Gensheimer, J.; Akondy, R.; McGuire, D.J.; Ahmed, R.; Kissick, H.T. Expression of Novel Long Noncoding RNAs Defines Virus-Specific Effector and Memory CD8 + T Cells. *Nat. Commun.* **2019**, *10*. [CrossRef]
18. Tsoi, L.C.; Iyer, M.K.; Stuart, P.E.; Swindell, W.R.; Gudjonsson, J.E.; Tejasvi, T.; Sarkar, M.K.; Li, B.; Ding, J.; Voorhees, J.J.; et al. Analysis of Long Non-Coding RNAs Highlights Tissue-Specific Expression Patterns and Epigenetic Profiles in Normal and Psoriatic Skin. *Genome Biol.* **2015**, *16*, 24. [CrossRef]
19. Andersson, R.; Gebhard, C.; Miguel-Escalada, I.; Hoof, I.; Bornholdt, J.; Boyd, M.; Chen, Y.; Zhao, X.; Schmidl, C.; Suzuki, T.; et al. An Atlas of Active Enhancers across Human Cell Types and Tissues. *Nature* **2014**, *507*, 455–461. [CrossRef]
20. Gil, N.; Ulitsky, I. Production of Spliced Long Noncoding RNAs Specifies Regions with Increased Enhancer Activity. *Cell Syst.* **2018**, *7*, 537–547. [CrossRef]
21. Yu, C.K.; Xu, T.; Assoian, R.K.; Rader, D.J. Mining the Stiffness-Sensitive Transcriptome in Human Vascular Smooth Muscle Cells Identifies Long Noncoding RNA Stiffness Regulators. *Arterioscler. Thromb. Vasc. Biol.* **2018**, *38*, 164–173. [CrossRef]
22. Alloza, I.; Goikuria, H.; Idro, J.L.; Triviño, J.C.; Fernández Velasco, J.M.; Elizagaray, E.; García-Barcina, M.; Montoya-Murillo, G.; Sarasola, E.; Vega Manrique, R.; et al. RNAseq Based Transcriptomics Study of SMCs from Carotid Atherosclerotic Plaque: BMP2 and IDs Proteins Are Crucial Regulators of Plaque Stability. *Sci. Rep.* **2017**, *7*, 1–12. [CrossRef]
23. Hezroni, H.; Koppstein, D.; Schwartz, M.G.; Avrutin, A.; Bartel, D.P.; Ulitsky, I. Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Rep.* **2015**, *11*, 1110–1122. [CrossRef]
24. Pertea, M.; Pertea, G. GFF Utilities: GffRead and GffCompare. *F1000Research* **2020**, *9*, 304. [CrossRef]
25. Wirka, R.C.; Wagh, D.; Paik, D.T.; Pjanic, M.; Nguyen, T.; Miller, C.L.; Kundu, R.; Nagao, M.; Coller, J.; Koyano, T.K.; et al. Atheroprotective Roles of Smooth Muscle Cell Phenotypic Modulation and the TCF21 Disease Gene as Revealed by Single-Cell Analysis. *Nat. Med.* **2019**, *25*, 1280–1289. [CrossRef]
26. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*. [CrossRef]
27. Young, M.D.; Wakefield, M.J.; Smyth, G.K.; Oshlack, A. Gene Ontology Analysis for RNA-Seq: Accounting for Selection Bias. *Genome Biol.* **2010**, *11*. [CrossRef]

28. Breschi, A.; Muñoz-Aguirre, M.; Wucher, V.; Davis, C.A.; Garrido-Martín, D.; Djebali, S.; Gillis, J.; Pervouchine, D.D.; Vlasova, A.; Dobin, A.; et al. A Limited Set of Transcriptional Programs Define Major Cell Types. *Genome Res.* **2020**, *30*, 1047–1059. [CrossRef]
29. Fishilevich, S.; Nudel, R.; Rappaport, N.; Hadar, R.; Plaschkes, I.; Iny Stein, T.; Rosen, N.; Kohn, A.; Twik, M.; Safran, M.; et al. GeneHancer: Genome-Wide Integration of Enhancers and Target Genes in GeneCards. *Database* **2017**. [CrossRef]
30. Kim, H.Y.; Kang, Y.J.; Song, I.H.; Choi, H.C.; Kim, H.S. Upregulation of Interleukin-8/CXCL8 in Vascular Smooth Muscle Cells from Spontaneously Hypertensive Rats. *Hypertens. Res.* **2008**, *31*, 515–523. [CrossRef]
31. Qin, Y.; Fan, F.; Zhao, Y.; Cui, Y.; Wei, X.; Kohama, K.; Gordon, J.R.; Li, F.; Gao, Y. Recombinant Human CXCL8(3-72)K11R/G31P Regulates Smooth Muscle Cell Proliferation and Migration through Blockage of Interleukin-8 Receptor. *IUBMB Life* **2013**, *65*, 67–75. [CrossRef]
32. Wang, Z.; Newman, W.H. Smooth Muscle Cell Migration Stimulated by Interleukin 6 Is Associated with Cytoskeletal Reorganization. *J. Surg. Res.* **2003**, *111*, 261–266. [CrossRef]
33. Ikeda, U.; Ikeda, M.; Oohara, T.; Oguchi, A.; Kamitani, T.; Tsuruya, Y.; Kano, S. Interleukin 6 Stimulates Growth of Vascular Smooth Muscle Cells in a PDGF-Dependent Manner. *Am. J. Physiol. Heart Circ. Physiol.* **1991**, *260*, H1713–H1717. [CrossRef]
34. Kurozumi, A.; Nakano, K.; Yamagata, K.; Okada, Y.; Nakayamada, S.; Tanaka, Y. IL-6 and SIL-6R Induces STAT3-Dependent Differentiation of Human VSMCs into Osteoblast-like Cells through JMJD2B-Mediated Histone Demethylation of RUNX2. *Bone* **2019**, *124*, 53–61. [CrossRef]
35. Sadeghi, M.M.; Esmailzadeh, L.; Zhang, J.; Guo, X.; Asadi, A.; Krassilnikova, S.; Fassaei, H.R.; Luo, G.; Al-Lamki, R.S.M.; Takahashi, T.; et al. ESDN Is a Marker of Vascular Remodeling and Regulator of Cell Proliferation in Graft Arteriosclerosis. *Am. J. Transplant.* **2007**, *7*, 2098–2105. [CrossRef]
36. Yaghini, F.A.; Song, C.Y.; Lavrentyev, E.N.; Ghafoor, H.U.B.; Fang, X.R.; Estes, A.M.; Campbell, W.B.; Malik, K.U. Angiotensin II-Induced Vascular Smooth Muscle Cell Migration and Growth Are Mediated by Cytochrome P450 1b1-Dependent Superoxide Generation. *Hypertension* **2010**, *55*, 1461–1467. [CrossRef]
37. Bernard, K.; Logsdon, N.J.; Benavides, G.A.; Sanders, Y.; Zhang, J.; Darley-Usmar, V.M.; Thannickal, V.J. Glutaminolysis Is Required for Transforming Growth Factor-B1–Induced Myofibroblast Differentiation and Activation. *J. Biol. Chem.* **2018**, *293*, 1218–1228. [CrossRef]
38. Pereira, F.A.; Yuhong, Q.; Zhou, G.; Tsai, M.J.; Tsai, S.Y. The Orphan Nuclear Receptor COUP-TFII Is Required for Angiogenesis and Heart Development. *Genes Dev.* **1999**, *13*, 1037–1049. [CrossRef]
39. Chen, C.N.; Li, Y.S.J.; Yeh, Y.T.; Lee, P.L.; Usami, S.; Chien, S.; Chiu, J.J. Synergistic Roles of Platelet-Derived Growth Factor-BB and Interleukin-1β in Phenotypic Modulation of Human Aortic Smooth Muscle Cells. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 2665–2670. [CrossRef]
40. Chappell, J.; Harman, J.L.; Narasimhan, V.M.; Yu, H.; Foote, K.; Simons, B.D.; Bennett, M.R.; Jørgensen, H.F. Extensive Proliferation of a Subset of Differentiated, yet Plastic, Medial Vascular Smooth Muscle Cells Contributes to Neointimal Formation in Mouse Injury and Atherosclerosis Models. *Circ. Res.* **2016**, *119*, 1313–1323. [CrossRef]
41. Tang, J.; Wang, H.; Huang, X.; Li, F.; Zhu, H.; Li, Y.; He, L.; Zhang, H.; Pu, W.; Liu, K.; et al. Arterial Sca1+ Vascular Stem Cells Generate De Novo Smooth Muscle for Artery Repair and Regeneration. *Cell Stem Cell* **2020**, *26*, 81–96. [CrossRef]
42. Leung, A.; Trac, C.; Jin, W.; Lanting, L.; Akbany, A.; Saetrom, P.; Schones, D.E.; Natarajan, R. Novel Long Noncoding RNAs Are Regulated by Angiotensin II in Vascular Smooth Muscle Cells. *Circ. Res.* **2013**, *113*, 266–278. [CrossRef]
43. De Vries, M.R.; Simons, K.H.; Jukema, J.W.; Braun, J.; Quax, P.H.A. Vein Graft Failure: From Pathophysiology to Clinical Outcomes. *Nat. Rev. Cardiol.* **2016**, *13*, 451–470. [CrossRef]
44. Vangala, P.; Murphy, R.; Quinodoz, S.A.; Gellatly, K.; McDonel, P.; Guttman, M.; Garber, M. High-Resolution Mapping of Multiway Enhancer-Promoter Interactions Regulating Pathogen Detection. *Mol. Cell* **2020**, *80*, 359–373. [CrossRef]
45. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed on 1 February 2021).
46. Krueger, F. Trim Galore! 2012. Available online: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed on 1 February 2021).
47. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]
48. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [CrossRef]
49. Li, B.; Dewey, C.N. RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinform.* **2011**, *12*. [CrossRef]
50. Ulitsky, I. Pipeline for LncRNA Annotation from RNA-Seq Data (PLAR). 2018. Available online: https://www.weizmann.ac.il/Biological_Regulation/IgorUlitsky/PLAR (accessed on 1 February 2021).

3.3.6 Manuscript Supplement

Novel transcript discovery expands the repertoire of pathologically-associated, long non-coding RNAs in vascular smooth muscle cells

Matthew Bennett[1], Igor Ulitsky[2], Iraide Alloza[3], Koen Vandenbroeck[3,4], Vladislav Miscianinov[1], Amira Dia Mahmoud[1], Margaret Ballantyne[1], Julie Rodor[1], Andrew H Baker[1]*†*

1. *Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh, 47 Little France Crescent, Edinburgh, UK*
2. *Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel*
3. *Inflammation & Biomarkers Group, Biocruces Bizkaia Health Research Institute, Cruces Plaza, 48903 Barakaldo, Spain*
4. *Ikerbasque, Basque Foundation for Science, 3 María Díaz Haroko Kalea, 48013 Bilbao, Spain*

**Equal contributions † Corresponding author* Andy.Baker@ed.ac.uk

Supplementary Info

See shinyApp for interactive data presentation and to download lncRNA annotations in .bed file format: https://bakergroup.shinyapps.io/VSMClncRNAannotation/

Table S1 Transcriptome assembly and lncRNA prediction statistics

Table S2 Primer sequences for selected lncRNAs quantified by qRT-PCR in the svSMC proliferation model

Table S3 Fold changes for selected lncRNAs in the svSMC model as determined by qRT-PCR or RNAseq

Table S4 Number of differentially expressed PCGs activity in the 3 VSMC RNAseq datasets

Tables S5-10 Significant GO terms for clusters 1-6 of differentially expressed genes in the svSMC RNAseq dataset

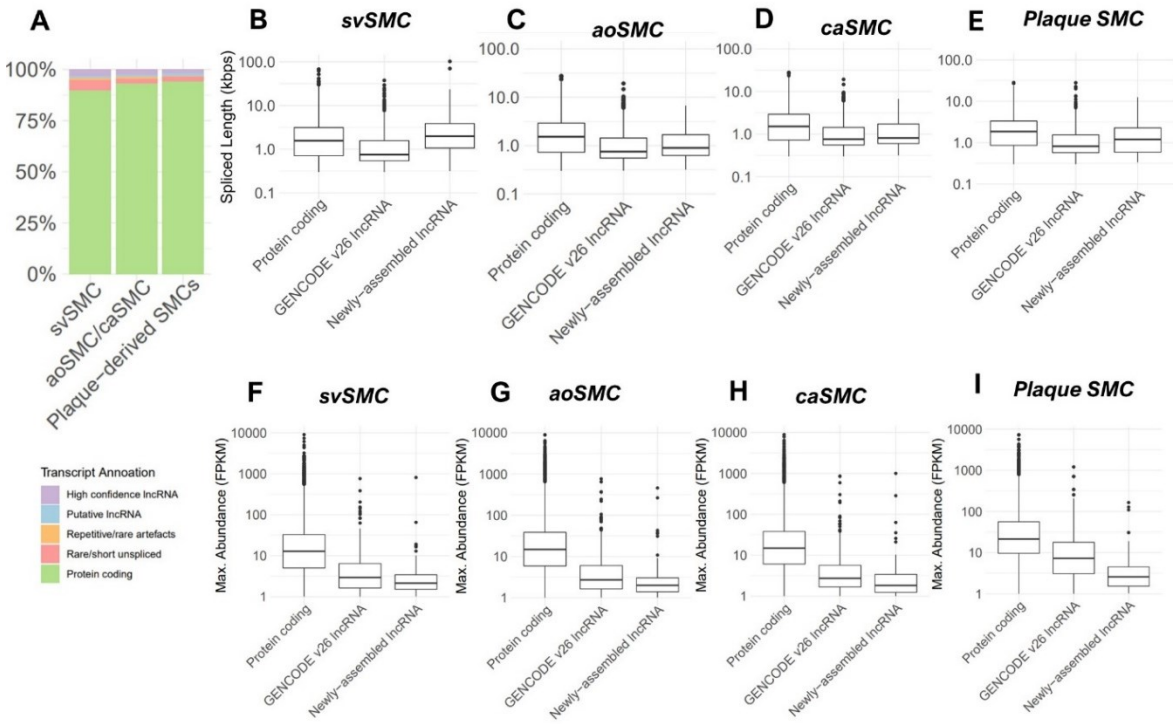
Tables S11-16 Significant GO terms for clusters 1-6 of differentially expressed genes in the stiff-culturing RNAseq dataset

Table S17 VSMC enrichment values for all lncRNAs detected as expressed in the 3 VSMC RNAseq datasets

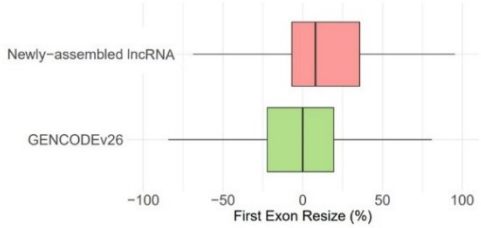
Table S18 All lncRNAs detected as expressed and enhancer-transcribed in the 3 VSMC RNAseq datasets

Table S19 All enhancer-transcribed lncRNAs with evidence linking them to expression of a protein coding genes in the 3 VSMC RNAseq datasets

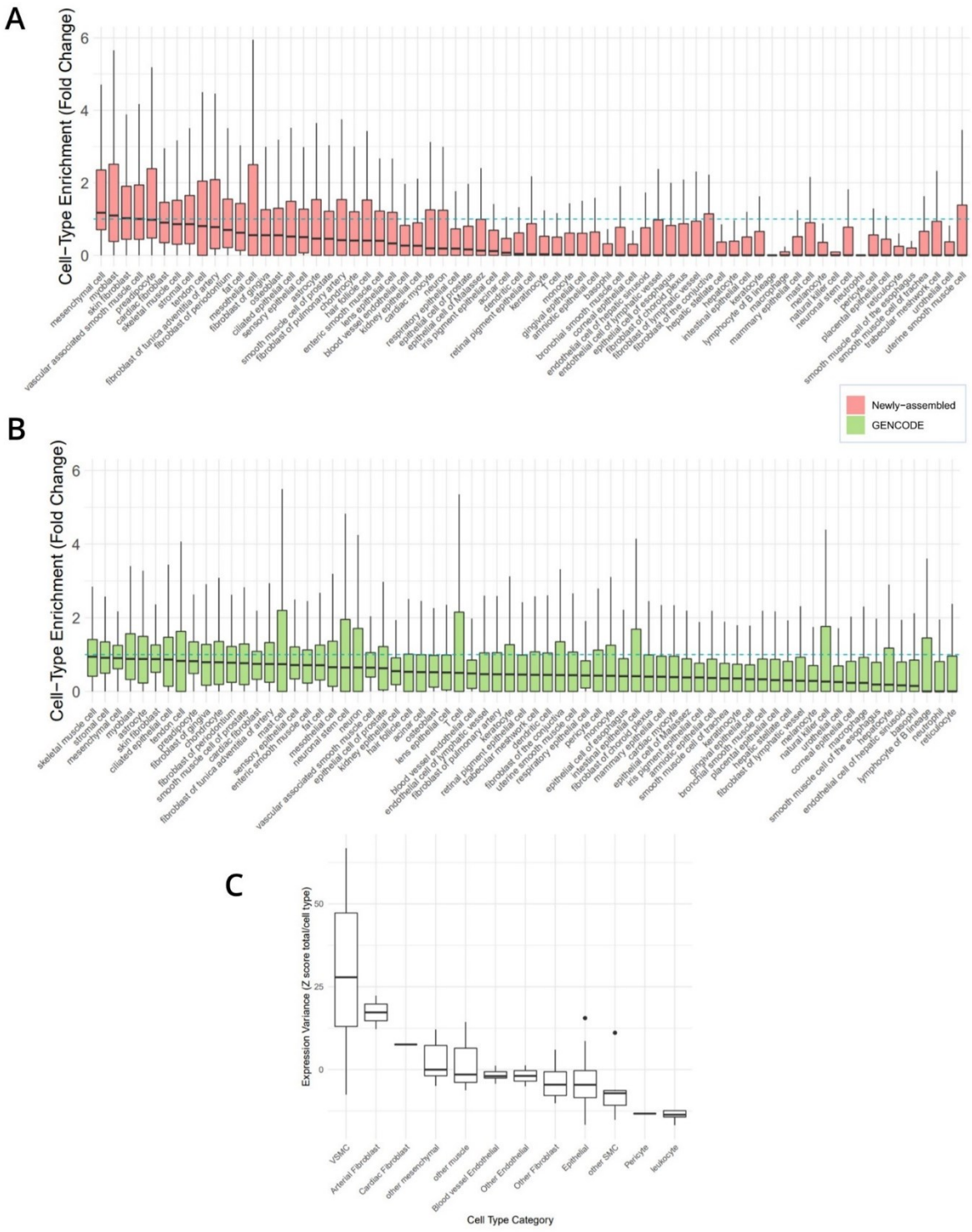
Supplementary Figures



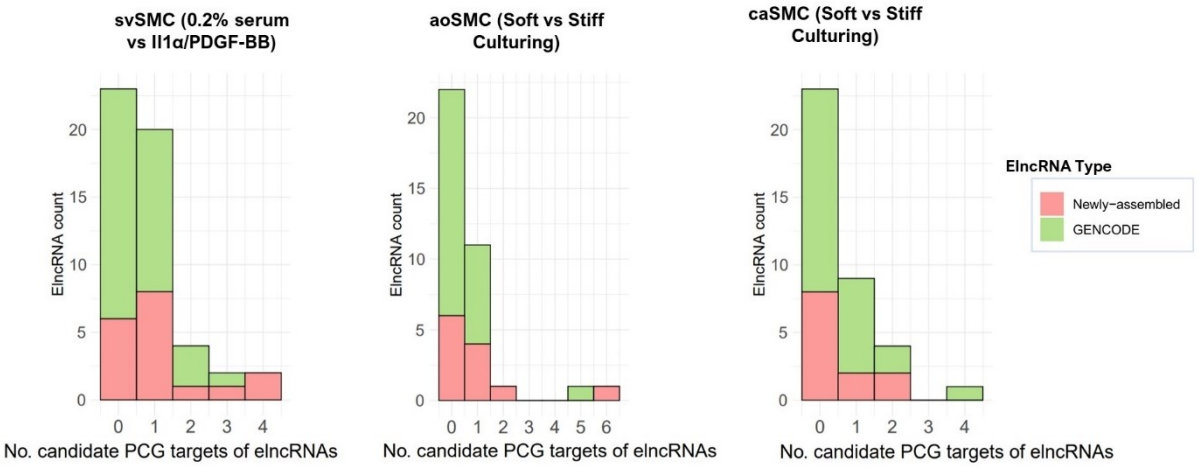
Supplementary Figure 1. Transcript annotation and characterisation of the 3 VSMC RNA-seq datasets. (A) Proportion of transcript types. (B-E) Spliced length and (F-I) abundance of split into GENCODE lncRNA, newly-assembled lncRNA or GENCODE PCG transcripts. Only transcripts with an expression >1FPKM were considered.



Supplementary Figure 2. Accuracy of the 5’ end transcript structure in comparison to its matched CAGE site in FANTOM databases (expressed as percentage change in the size of the first exon).



Supplementary Figure 3. Cell-type enrichment values for (A) newly-assembled and (B) GENCODE lncRNAs in 69 primary cell categories ranked by median. (C) Total Z score across all VSMC-enriched lncRNAs for each cell category in figure 3J



Supplementary Figure 4. Number of newly-assembled and GENCODE differentially expressed elncRNAs in the svSMC, aoSMC and caSMC datasets vs. number of candidate elncRNA PCG targets

Table S1

Numbers of GENCODE and newly-assembled transcripts annotated as lncRNAs in 3 selected VSMC RNAseq datasets

VSMC type	*Conditions*	*No. transcripts in full transcriptome (GENCODE + Newly-assembled)*	*No. transcripts from newly-assembled genes*	*High confidence lncRNA transcripts*	*High confidence lncRNA transcripts from newly-assembled genes*
Saphenous vein	0.2% FBS, IL-1α, PDGF-BB, IL-1α + PDGF-BB	66500 (82%) + 14151 (18%)	1223 (1.5%)	3005 (3.7%)	213 (7.1% of lncRNAs)
Aortic, Coronary	Soft culture (2-4kPa), Stiff culture (25kPa)	74891 (82%) + 16037 (18%)	701 (0.8%)	2956 (3.3%)	195 (6.6% of lncRNAs)
Plaque-isolated	Symptomatic Plaques, Asymptomatic Plaques	69113 (87%) + 10303 (13%)	502 (0.6%)	2508 (3.2%)	143 (5.7% of lncRNAs)

Table S2

Primer sequences for selected lncRNAs quantified by qRT-PCR in the svSMC proliferation model

MSTRG.743_F	tccaagactaccaaggtgttac
MSTRG.743_R	aggccaatgaaaattgagtgtc
MSTRG.13896_F	GTCACAGTAATCATGGCACTTG
MSTRG.13896_R	agatatgacttgctcctccttg
AC002480_F	gcaggggacacatttattgatg
AC002480_R	TTCATCCACCCATCCAGTAAtc
LINC02015a_F	gcattcaagggagtaagaagaac
LINC02015a_R	ctgctgaagggagtctatcttg
RP11-1E6_F	TCATCACCCTTGTTATCCCATG
RP11-1E6_R	TGGCACTTATAGGAAAGGACAG
MSTRG.10933ii_F	tgatgtgtttgagtgtggattc
MSTRG.10933ii_R	cctgaagaaataatggcccaac
LINC01013_hg38_4if	CTCCAATCTGCTGCCCCTAA
LINC01013_hg38_4ir	AGATTTCCCTGGGGCTTTTGG
AC003092_hg38_6if	TGAGATTTCAGCTCTCCTGATCC
AC003092_hg38_6ir	GCCTCCACTGTGAGTAAAGACA
AC018647_hg38_3f	GGACTCTGCCACCTTTCGTA
AC018647_hg38_3r	TAGGCACTGCCACAGAGTAAC
PLCE-AS1_hg38_4f	CAGTGGAATGGTCCGAGTGT
PLCE-AS1_hg38_4r	AGGCTGGCATGTCCTTCATT

Table S3

Fold changes for selected lncRNAs in the svSMC model as determined by qRT-PCR or RNAseq

Name	Log2FC.qPCR.	LncRNA_Annotation	LogFC_BO
MSTRG.10933	-5.10	Newly-assembled	-5.65
MSTRG.13896	10.96	Newly-assembled	8.60
MSTRG.743	10.34	Newly-assembled	9.09
AC002480.4	3.55	GENCODE	5.56
AC003092.1	6.39	GENCODE	6.02
AC018647.3	-8.10	GENCODE	-10.60
LINC01013	-6.96	GENCODE	-4.49
LINC02015	5.59	GENCODE	3.41
PLCE1-AS1	-5.65	GENCODE	-6.80
RP11-1E6.1	11.22	GENCODE	8.01

Table S4

Differentially expressed PCGs activity in the 3 VSMC RNAseq datasets

Dataset	*VSMC type*	*Comparison*	*Expressed PCGs*	*Differentially expressed PCGs (% expressed PCGs)*
Ballantyne et al. 2016	Saphenous vein	0.2% FBS vs. IL-1α + PDGF-BB	10856	2292 (21.1%)
Yu et al. 2018	Aortic	Soft culture (2-4kPa) vs. Stiff culture (25kPa)	11067	2952 (26.7%)
Yu et al. 2018	Coronary	Soft culture (2-4kPa) vs. Stiff culture (25kPa)	11059	3796 (34.3%)
Alloza et al. 2017	Plaque-isolated	Symptomatic Plaques vs. Asymptomatic Plaques	10543	46 (0.4%)

3.4 Additional Analysis

3.4.1 Newly-assembled transcripts found in multiple VSMC datasets are typically assembled with the same intron chains

In the published manuscript, we probed newly-assembled lncRNA transcripts through comparing their exonic structures and TSSs to those present in the FANTOM CAT library. This allows us to judge the completeness of the transcript models obtained from use of our pipeline in VSMCs by assessing if similar intron chains or 5' sites have been captured in previous analysis of the much larger and heterogenous sample set within FANTOM.

Another way to gauge the completeness of our newly-assembled transcripts is to compare the structures of those which are found in multiple of the three VSMC datasets. If such transcripts are reproduced in their entirety in each parallel run of the pipeline then this indicates that the pipeline is robustly building complete exonic structures. Alternatively, if they are generally found to be reproduced with missing exons or incomplete intron chains, this could be a technical issue – i.e. the pipeline is not consistent in transcript assembly - and/or due to biological differences – i.e. the transcript expression is too low or another isoform is expressed between datasets.

We assessed this by comparing each of the three VSMC transcriptomes to the transcriptome obtained through merging all 3 together (publication Figure 1D). Newly-assembled transcripts potentially found in multiple annotation runs were identified as those which matched to the same transcript in the merged transcriptome via at least 1 splice junction (Figure 3.1a). This resulted in 31 newly-assembled lncRNA transcripts which were potentially expressed in more than one of the VSMC datasets. Of these 24 (77%) matched to each other via a full intron chain (Figure 3.1b). We therefore estimate that 77% of the newly-assembled transcripts originating from the same gene in multiple VSMC datasets were independently reconstructed by the pipeline with an overlapping set of introns. The remaining 23% may not match due to differential transcript abundance, varying isoforms across the datasets or inconsistencies in the transcript assembly process. Overall, this suggests our pipeline was effective at providing consistent exonic structures for the same transcript in multiple runs.

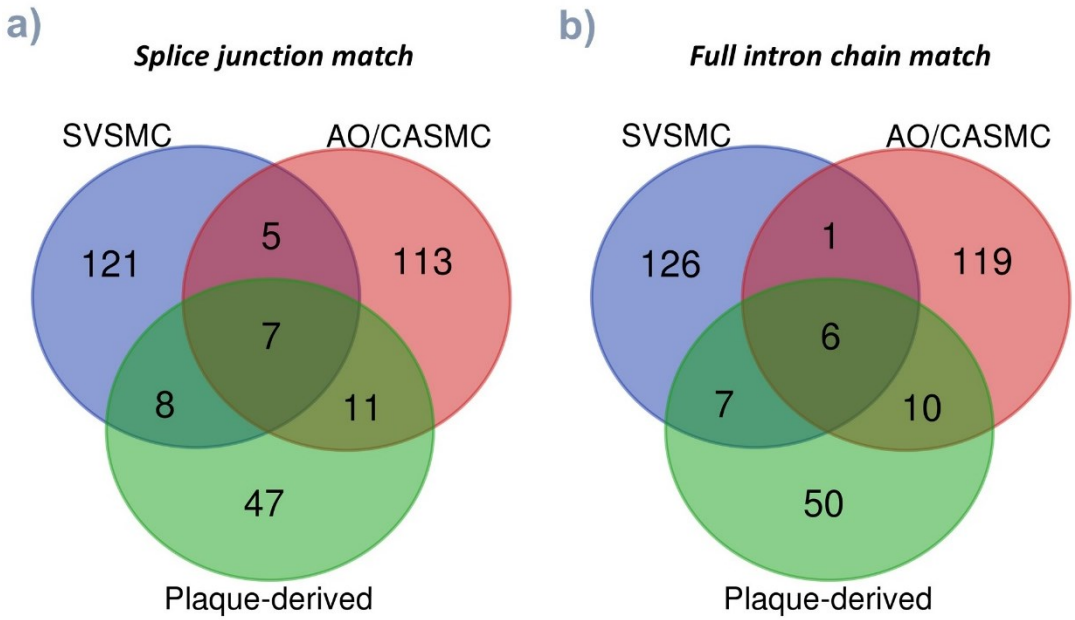


Figure 3.1 Comparing newly-assembled lncRNA transcripts between VSMC datasets. a) all matches by at least 1 splice junction b) all matches by an exact intron chain

3.4.2 Towards an unbiased view of VSMC enrichment amongst pathology-associated lncRNAs

Using the FANTOM expression atlas gives a comprehensive view of cell-type specificity due to the extensive sample set they use in their analysis. This expression atlas is therefore ideal to identify VSMC-enriched lncRNAs which could be ideal candidates for therapeutic targeting. In the publication, we identified 37 such lncRNAs, of which 17 – or 46% - were newly-assembled. Hence, including the cohort of extra lncRNAs found using our pipeline nearly doubled the pool of VSMC-enriched, stimuli-responsive lncRNAs. This is likely related to the fact that they are enriched within both VSMC-enriched lncRNAs (publication Figure 3a-d) and stimuli-responsive lncRNAs (publication Figure 2a,c-e).

Here we describe these 37 lncRNAs in terms of their VSMC enrichment strength and the number of VSMC subtypes/categories in which they are enriched (Figure 3.2), aiming to provide a holistic view on VSMC-enriched lncRNAs in pathology. *SENCR* and *SMILR*, previously identified with vascular or VSMC enrichment respectively[160,173], are present within the cohort of 37 stimuli-responsive and VSMC-enriched lncRNAs. Within the expression atlas,

these lncRNAs show an 8-9-fold higher expression in a VSMC category (including 9 VSMC subtype categories and 1 meta-category containing all subtypes together) relative to average expression across all other sample categories (Figure.3.2). Intriguingly, 16 uncharacterised lncRNAs surpass *SENCR* and *SMILR* in terms of their maximum VSMC enrichment, with the newly-assembled gene *MSTRG.13896* (renamed *VSMClnc6* in the publication and from hereon) showing a 40-fold enrichment in VSMCs and the GENCODE lncRNA *NLGN4Y-AS1* showing 80-fold enrichment in VSMCs. Further, *SENCR* and *SMILR* were found enriched in 1 or 3 VSMC categories respectively whilst 8 lncRNAs were enriched in 4 or more VSMC subtypes. 5 such lncRNAs were newly-assembled and include *VSMClnc6* revealed as an outlier in terms of VSMC-enrichment strength and enrichment in VSMCs of multiple vascular beds. For these lncRNAs, VSMC enrichment occurs more generally across the multiple vascular beds profiled in the FANTOM expression atlas, with the important caveat that the atlas is not exhaustive and contains certain biases. Indeed, only 1 of the 9 VSMC subtypes is venous (umbilical vein) and 4 of the 9 subtypes have likely origins from a single embryonic region (neural crest region for the aorta, carotid, coronary and subclavian arteries).

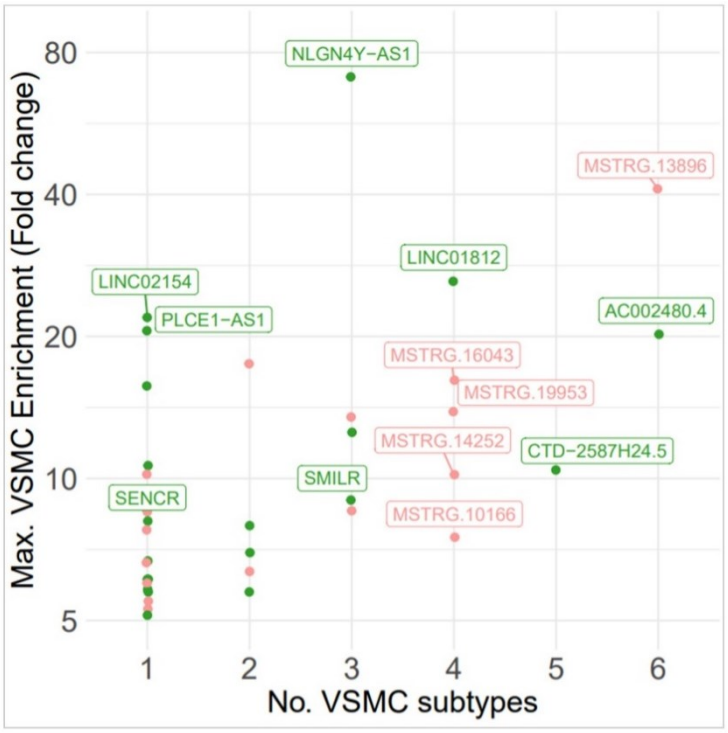


Figure 3.2 Patterns of VSMC enrichment strength across all FANTOM VSMC subtypes for stimuli-responsive, VSMC-enriched lncRNAs. Green = GENCODE lncRNA, red = newly-assembled lncRNA.

This further analysis of VSMC enriched lncRNAs underlines a central message of the publication – that previously hidden non-reference lncRNAs substantially improve the pool of high potential candidates with evidence of particular relevance to VSMC function over the functioning of other cells. The additional detail of VSMC expression patterns provided here can also be used to aid the often ambiguous or arbitrary process of lncRNA candidate selection (Table S17 in the publication). For instance, selecting lncRNAs enriched in 4 or more VSMC subtypes and/or with enrichment strength over 20 predicts that 11 of the 37 lncRNAs have particularly strong VSMC association and/or broad relevance across vascular beds.

3.4.3 Further exploration of lncRNA enhancer association and potential for cis regulation

In publication Figure 4, we present data showing the tendency of newly-assembled lncRNAs to be transcribed from enhancers, suggesting our pipeline may aid in uncovering lncRNA-driven *cis* regulatory mechanisms which would be obscured if relying solely on reference annotation. To provide a full catalogue of enhancer lncRNAs we utilised two sources of enhancer annotation – GeneHancer[243] and FANTOM CAT[136]. The methodology used to annotate enhancers differs somewhat between these two databases. GeneHancer integrates five enhancer databases and scores them according to criteria such as number of identifying databases, number of transcription factor binding sites and production of eRNAs. In contrast FANTOM CAT relies solely on the Roadmap epigenome database and does not provide a score. Both methods rely heavily on ChIPseq data of common histone marks such as H3K27ac as well as sequencing techniques to identify DNase hypersensitive regions where DNase is used to digest genomic DNA at regions of accessible chromatin. Discrepancies between these two methods are here explored to help judge the validity of our elncRNA classfication.

FANTOM CAT provided the largest number of lncRNA TSS-overlapping enhancer regions with 276 enhancer-transcribed lncRNA transcripts compared to 166 using the GeneHancer annotation (Figure 3.3). 206 (75%) lncRNA enhancer annotations obtained from FANTOM CAT were absent from GeneHancer. FANTOM CAT therefore provided the largest number of lncRNA-overlapping enhancer regions, most of which were not found in GeneHancer.

For the 206 FANTOM CAT enhancer annotations missed by GeneHancer, 129 (63%) of these were absent due to a lack of GeneHancer annotations and 28 (14%) due to low scoring of GeneHancer annotations. The remaining 49 FANTOM CAT enhancer annotations missed by GeneHancer were those classed as “promoter/enhancers” annotations in GeneHancer as they show signatures of both promoter and enhancer loci. Conversely, looking at GeneHancer enhancer annotations missed by FANTOM CAT, 72 (75%) of these are due to a non-enhancer classification (promoter or other) provided by FANTOM CAT at these TSSs. Therefore, here the discrepancy is largely due

to different classifications provided between FANTOM and GeneHancer, rather than a lack of FANTOM CAT coverage of GeneHancer annotations.

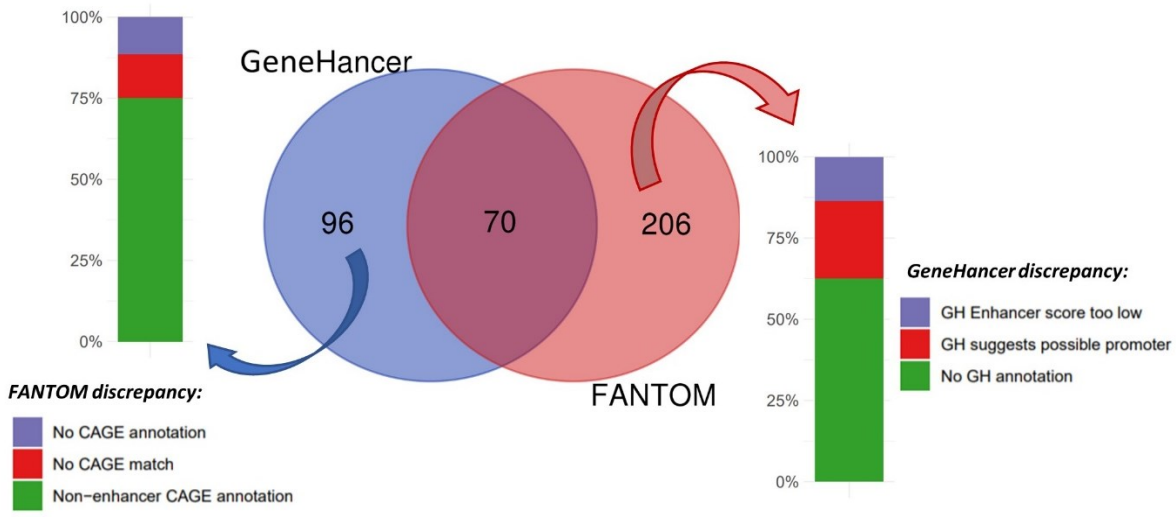


Figure 3.3 Overlap of enhancer annotation provided for lncRNA transcripts by the GeneHancer or FANTOM CAT databases along with reasons for any discrepancies between both methods.

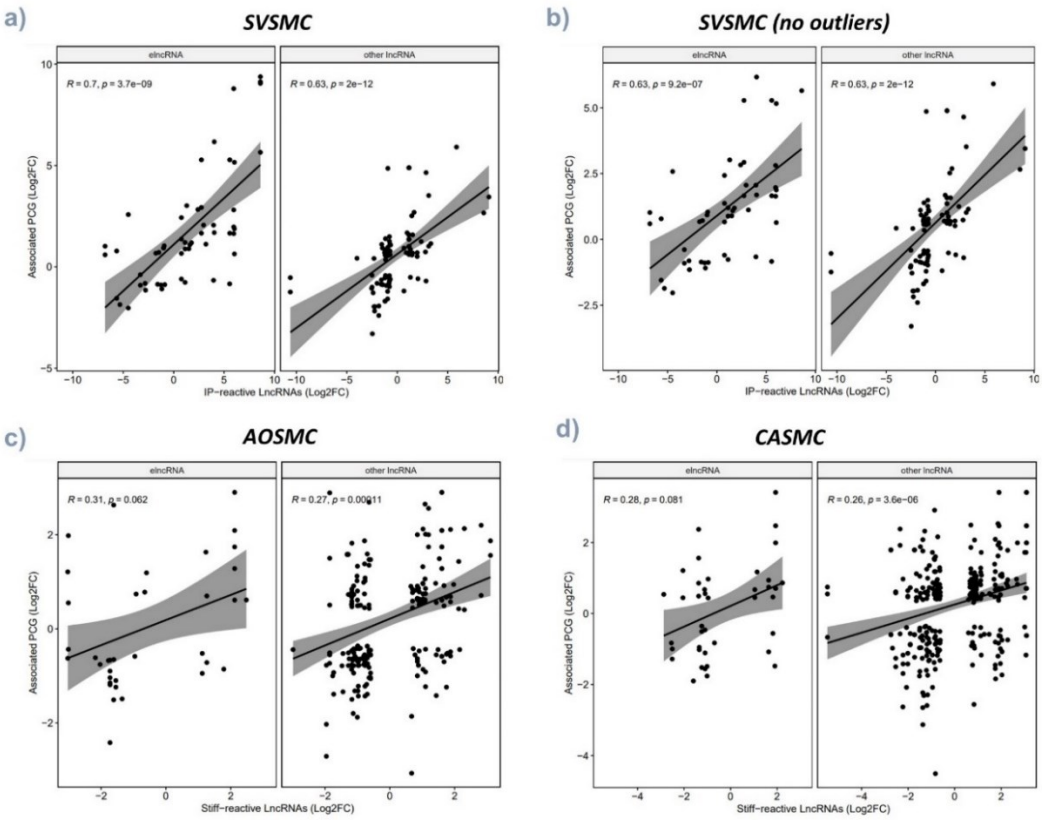
The two databases therefore complement each other if used for elncRNA classification. FANTOM CAT is more extensive and contains accurate TSSs through use of CAGEseq but can be supplemented by combining with the 5 additional enhancer databases collated by GeneHancer. Though there are discrepancies in classifications of enhancers and promoters between the two, this reflects the loosely defined classification system and porous border between these two loci types in the field at present[243]. This analysis provides context and justifies our inclusive approach for defining elncRNAs - keeping all enhancer annotations from both methods allows us to make initial conclusions on their association with our expanded lncRNA annotations.

To provide validation of our elncRNAs we searched for evidence of their capacity to effect *cis* regulation on surrounding genes. We compared the fold changes of candidate PCG targets of differentially expressed elncRNAs (as defined in the manuscript) to those within 250kbp range of any other differentially expressed lncRNA. We next assessed if there was an overall stronger correlation seen between stimuli-responsive elncRNAs and neighbouring stimuli-responsive PCGs than for other lncRNAs. Stronger correlation could indicate greater tendency towards shared regulation at

elncRNA over other lncRNA loci which could in turn indicate greater tendency for lncRNA-dependent *cis*-activation at such loci.

We saw strong correlation in the SVSMC dataset for elncRNAs and neighbouring PCGs as well as other lncRNAs and their neighbouring PCGs (rho 0.70 and rho 0.63 SVSMC) (Figure 3.4a). However, the observed increase in correlation for elncRNAs was not significant (p >0.05, Fisher's z) and seemed reliant on 3 candidate elncRNA and PCG pairings which were outliers in terms of their high fold changes (>3-fold the interquartile range from the 3rd quartile). Excluding these 3 targets reduced correlation strength (rho) at elncRNA loci closer to that seen for other lncRNAs (Figure 3.4b). Interestingly, these 3 PCGs with outlying fold changes were in the CXCL locus of chemokines, in the region 250kbp downstream of just one elncRNA - *VSMClnc6* highlighting this region as particularly highly induced during proliferation. Of these, CXCL8 already has available data supporting an elncRNA-PCG regulatory partnership (publication Table 1). In the AOSMC or CASMC datasets we saw no significant correlation of elncRNA and PCG fold changes – contrasting with the significant fold change correlations around other lncRNAs This may be due to the lower pool of differentially expressed elncRNAs found in these datasets. Overall, we could identify correlations, and so potential for *cis*-regulatory mechanisms, at both elncRNA and other lncRNA loci but this was inconsistent across the proliferation models used with no evidence for stronger correlation near elncRNAs compared to other lncRNAs in any dataset (Fisher's z, p >0.05). As *cis*-regulatory mechanisms may also be repressive, we repeated this analysis using absolute values for fold changes (negative values converted to positive) but saw no change in this conclusion (Figure 3.3e-h). We therefore do not identify an increased tendency for *cis* regulation mechanisms at elncRNA loci relative to other lncRNA loci using this approach.

Fold change values:



Absolute fold change values:

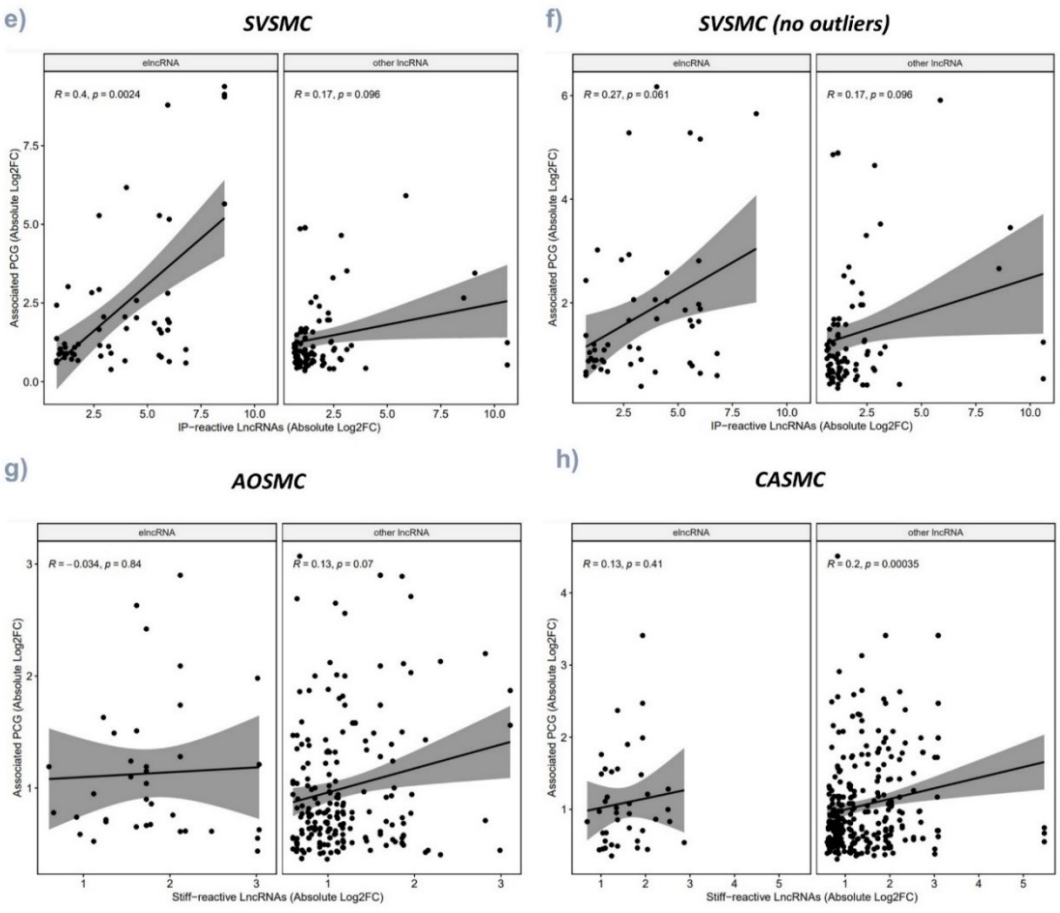


Figure 3.4 a-h) Correlation of FCs between differentially expressed elncRNAs or other lncRNAs with their neighbouring differentially-expressed PCGs. "No outliers" refers to removal of the strongest outlying PCG fold change values exceeding 1.5 times the IQR from the 3rd quartile (correlation coefficients are Spearman's rank)

3.4.4 Screening of candidate pathology-associated lncRNAs in SVSMCs (Preliminary work)

To identify lncRNA with definitive impact on VSMC proliferation, we aimed to knockdown candidate lncRNAs and assess any functional consequences in a VSMC proliferation model *in vitro*. As the SVSMC dataset was originally generated using an easily implemented proliferation model available within our research group, we began by identifying differentially expressed lncRNAs of interest within this dataset. An initial pool of lncRNAs were selected for consideration based on their expression dynamics (Figure 3.5), lack of previous characterisation (Table 3.2). This selection and characterisation were carried out prior to CAGE annotation, VSMC-specificity and full enhancer annotation so these analyses were not taken into account at the time.

From this initial pool of lncRNAs, *AC018647.3* was selected for functional screening as it displayed the largest fold change of any lncRNA gene in response to IL-1α/PDGF-BB co-stimulation as well as individual growth factor treatments (Figure 3.5 + 3.6a). It was also highly expressed relative to other lncRNAs, had a highly conserved first exon and we noted an overlap to a GeneHancer annotation with predicted associations to both neighbouring genes, one of which (*SEPT7*) is a homolog to yeast Cdc10, a structural element involved in cytokinesis[245]. In analyses performed after selection we noted that the enhancer annotation of *AC018647.3* was too low scoring in GeneHancer to be considered an elncRNA by our later definition though we do not rule out a potential *cis* effect on surrounding genes. *MSTRG.10933* was another lncRNA with relatively high IL-1α/PDGF-BB repression and showed a predicted regulatory link to glutaminase (*GLS*) located antisense upstream via a GeneHancer eQTL located in the first exon of the lncRNA. *MSTRG.10933* was defined as an elncRNA in later analyses along with the link to *GLS* (publication Table 1) with which it shows a similar pattern of regulation as the repression of both genes with PDGF-BB is potentiated more than 1.5x times upon co-stimulation with IL-1α (Figure 3.5 + 3.6b). This could be pathologically

relevant as *GLS* is implicated in supporting metabolic requirements of VSMCs proliferating in response to stiffness[105] as well as the expression of pro-fibrotic markers and TGFβ-dependent differentiation of myofibroblasts[246]. Finally, *LINC02015* was chosen as it is upregulated by IL-1α and IL-1α/PDGF-BB with a relatively high FPKM (Figure 3.5 + 3.6c). None of these selected genes were defined as VSMC-enriched in later analysis. However, the traits identified here mark them as worthy of investigation.

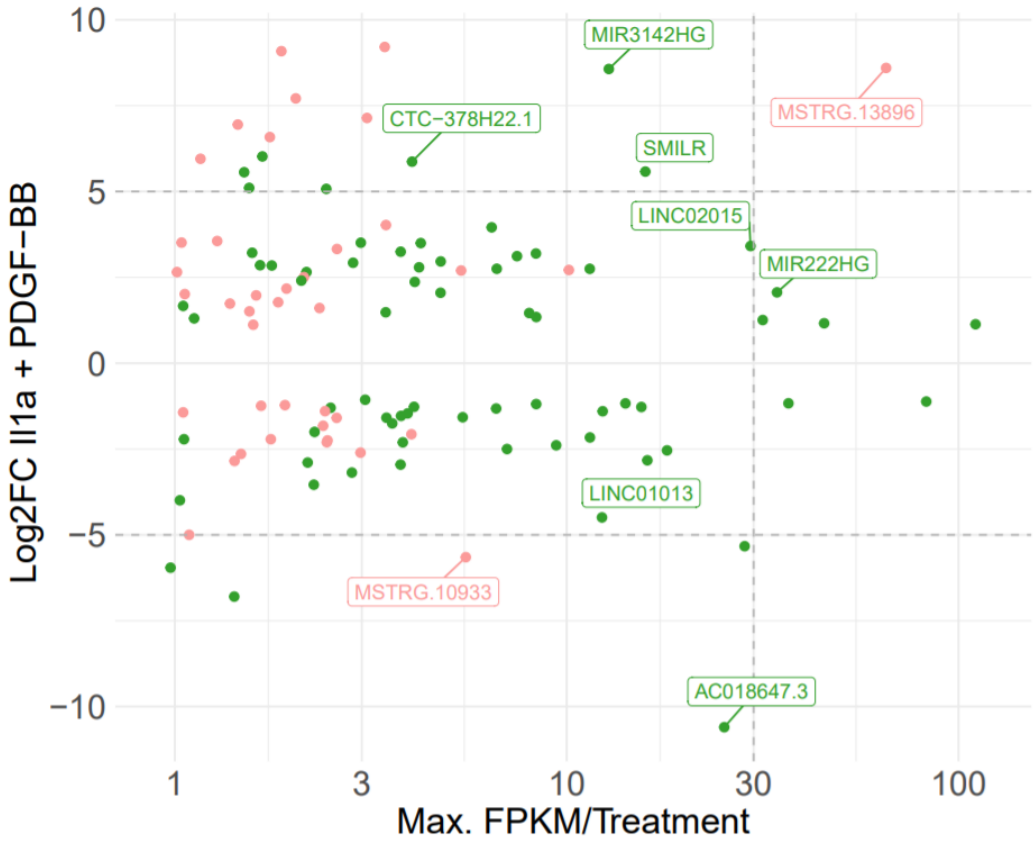


Figure 3.5 Expression dynamics of differentially expressed lncRNAs in the SVSMC dataset. Dashed grey lines indicate thresholds for notably high fold changes and/or FPKM. Uncharacterised lncRNAs considered for phenotypic screening as well as known proliferative VSMC lncRNAs (SMILR and MIR222HG) are labelled. Those initially considered for phenotypic screening are circled and further examined in Table 3.2

Gene Name	Class*	IL-1α/PDGF-BB effect	Novelty	Notable Conservation PhyloP/PhastCons	Most Distant Annotated Orthologs**	GeneHancer Interactions*	DE surrounding genes (within 1mbp)	Notable GO surrounding genes
AC018647.3	lincRNA/ eRNA	All treatments gave complete loss of expression	No characterisation	Strong in First Exon	Primate	SEPT7, HERPUD2	SEPT7, HERPUD2	SEPT7: Cell cycle, septin complex
MSTRG.13896 (VSMClnc6)	eRNA	No PDGF-BB effect	Unannotated/ **Characterised*** **	Neither exon	n/a	CXCL8	CXCL locus	Inflammatory response, chemotaxis, adhesion, proliferation
MIR3142HG	miRNA host	PDGF-BB no individual effect but potentiated IL-1a induction	**Host for well characterised miR-146a (2nd exon)[244]**	Surrounding 1st Exon, miR-3142 and miR-146a	Tetrapod	None	PTTG1	Chromosome organisation
LINC02015	lincRNA	PDGF-BB no individual effect but reduced IL-1α induction	Part of prognostic lncRNA signature in brain tumour[245]	Strong in upstream enhancers, several intronic spikes	Syntenic only	None (but two upstream)	None	n/a
LINC01013	lincRNA	All treatments gave similar repression	**Published EpiMT, fibronectin, cell invasion in anaplastic large cell lymphoma[246]**	Slight increase in 3' exon	Primate	None	CTGF, MOXD1	CTGF: Ossification, proliferation, migration, adhesion
MSTRG.10933	lincRNA/ eRNA	IL-1α no individual effect but potentiated PDGF-BB repression	Unannotated/ No characterisation	None in exons, in intron-contained enhancer	n/a	NEMP2, GLS	STAT1, GLS	Glutamate biosynthesis, transcription factor, proliferation, apoptosis, angiogenesis
CTC-378H22.1	bidirectional lncRNA/ miRNA host	PDGF-BB no individual effect but potentiated IL-1a induction	**miR-4323 (intronic) induces bladder SMC proliferation via ERK 1/2 under cyclic hydrodynamic pressure[247]**	Strong in promoter, strong in intron	Mouse	None (two upstream)	POU2F2	Transcription factor activity

Table 3.2 Collated expression dynamics, genomic annotation, literature and potential functionality for SVSMC lncRNA candidates considered for phenotypic screening. Rows highlighted in grey are the lncRNAs selected for experimentation. * Genes were classed prior to CAGE matching and full enhancer annotation analysis including use of interaction data. ** As determined using orthology data available for GENCODE genes at https://www.weizmann.ac.il/Biological_Regulation/lgorUlitsky/PLAR2 * MSTRG.13896/VSMClnc6 was later identified as a lncRNA already being characterised by a collaborator.**

Additional Analysis

We began characterisation by validating differential expression of these three candidate lncRNAs in the SVSMC proliferation model by RT-qPCR ($p < 0.01$, paired T test for *AC018647.3* and *MSTRG.10933*, $p < 0.001$, paired T test for *LINC02015*) (Figure 3.6d-f). Additionally, we examined their expression when substituting the SVSMCs in the proliferation model with CASMCs. This change has previously been demonstrated to result in much lower levels of proliferation[173,174] and indeed differential expression of the lncRNAs with IL-1α/PDGF-BB stimuli was much dampened, with no significant change occurring for *AC018647.3* or *MSTRG.10933* ($p = 0.06$, t test for paired samples: *AC018647.3,* ns, t test for unpaired samples: *MSTRG.10933* and *LINC02015*) (Figure 3.6d-f). The maximal expression of the three lncRNAs in CASMCs was also much lower than SVSMC. Together this suggests any function they may have in the proliferation model may be particularly emphasised in SVSMCs over CASMCs. Subcellular fractionation showed a strong nuclear bias for all genes (Figure 3.6g). Whilst Gapmer-mediated knockdown, thought to more efficiently target nuclear transcripts than dsiRNA[197], was selected as an approach for *LINC02015* and *MSTRG.10933*, for *AC018647.3*, a dsiRNA-mediated knockdown proved effective in initial tests and so was used subsequently.

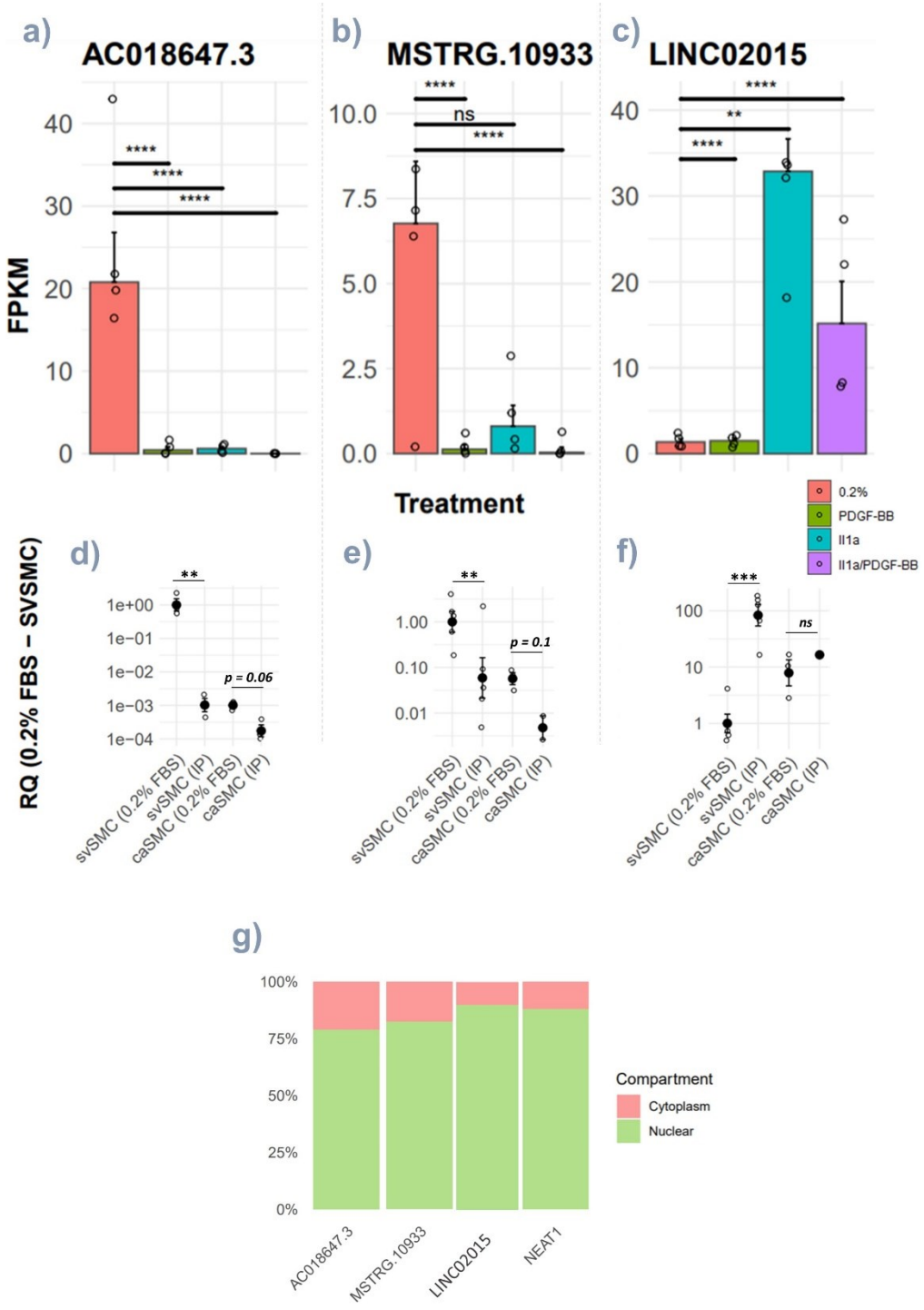


Figure 3.6 Expression of the 3 selected lncRNAs for phenotypic screening in a-c) the SVSMC RNAseq dataset (bars are median FPKM, error bars are standard error, significance obtained using DESeq2) d-f) RT-qPCR assays of the SVSMC proliferation model, an analogous CASMC model (black dots represent mean, error bars represent standard error, significance obtained using paired T tests on dCt values for all comparisons shown - other than the CASMC comparisons for MSTRG.10933 and LINC02015 for which one biological replicate contained failed PCR data) g) Subcellular localisation of lncRNAs in quiesced cells in the proliferation model or after 72 hours IL-1α/PDGF-BB treatment for LINC02015. p <0.00001 ****, p<0.01 **, p>0.05 ns

Additional Analysis

We hypothesised that *LINC02015* promotes SVSMC proliferation after being induced by IL-1α/PDGF-BB and so active degradation of *LINC02015* transcripts would therefore reduce proliferation rates in an EdU assay. The procedure for transfection of knockdown reagents was performed prior to 72 hours of co-stimulation with IL-1α/PDGF-BB. Gapmer-mediated knockdown resulted in no significant repression of *LINC02015* expression after 72 hours though high variability between patients suggests further testing is required (p >0.05, ANOVA on Iman and Conover ranked non-parametric data) (Figure 3.7).

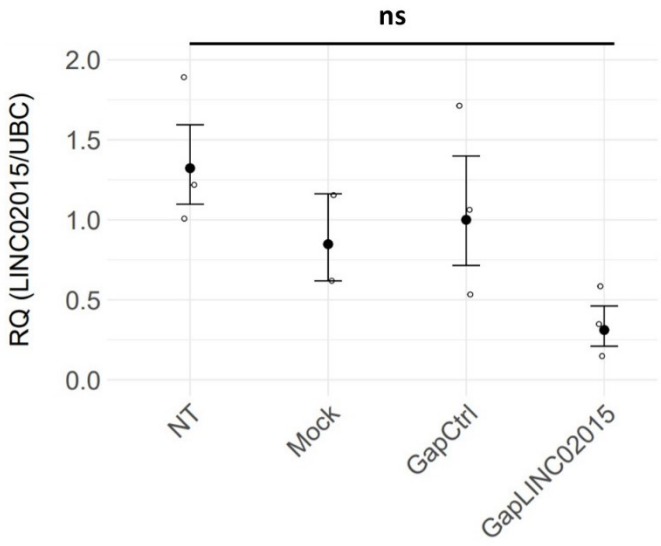


Figure 3.7 LINC02015 expression in Gapmer-mediated LINC02015 knockdown in SVSMC treated with IL-1α/PDGF-BB (black dots are mean and standard error of either dCt represented as RQ, one-way ANOVA on Iman and Conover nonparametric ranked data, ns not significant)

For the lncRNAs repressed by IL-1α/PDGF-BB we hypothesised that these lncRNAs represent mechanisms to actively inhibit VSMC proliferation and that knockdown of these lncRNAs could lead to an increase of proliferation in the absence of any stimulation. Transfection of knockdown reagents was thus performed after 48 hours of SVSMC quiescence with 0.2% FBS before using EdU assays to look for increased proliferation over the next 72 hours. *AC018647.3* could be effectively knocked down (~10-20% of DsiCtrl) in quiesced VSMCs via dsiRNA (p <0.01, ANOVA, p <0.01, Tukey's post-hoc for DsiCtrl vs. DsiLnc8, Iman and Conover ranked non-parametric data) (Figure 3.8a) but this yielded no change in proliferation via EdU incorporation (p>0.1,

ANOVA on Iman and Conover ranked non-parametric data) (Figure 3.8b) showing it is not a direct repressor of VSMC proliferation.

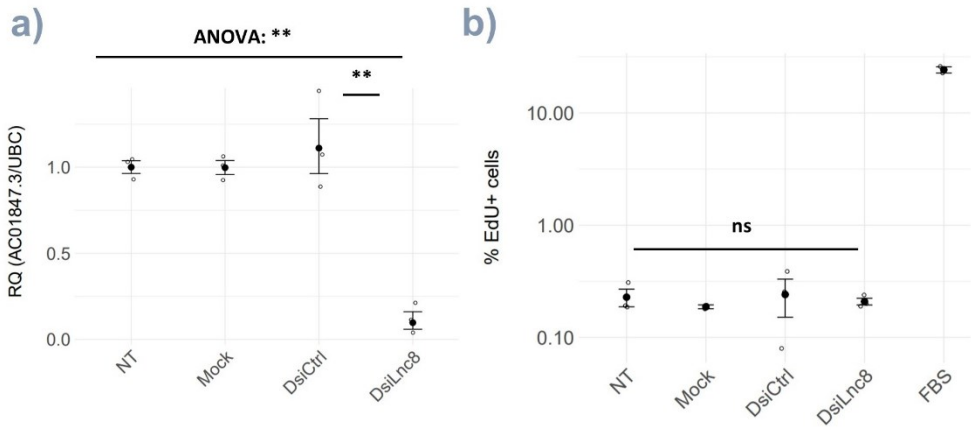


Figure 3.8 a) AC018647.3 expression in DisRNA-mediated AC018647.3 knockdown in SVSMC quiesced in 0.2% FBS and b) corresponding EdU assay (black dots show mean and standard error of either dCt represented as RQ or EdU+ %, one-way ANOVA followed by Tukey's post-hoc test, p < 0.001 *, ns not significant)**

For *MSTRG.10933*, initial titrations of Gapmer concentrations to determine conditions for efficient knockdown showed strong (~95%) inhibition of the lncRNA expression with 5nM of an initial Gapmer (Gap1) and weaker (~76%) inhibition with a second Gapmer (Gap2) at 20nM, showing marked differences in effect size between Gapmers (Figure 3.9a). In Gap1-treated SVSMCs, reduced cell elongation and cell density were visible using 5nM concentrations of Gap1 and this became more obvious at larger concentrations of 10 and 20nM (Figure 3.9b). Gap2 showed no visible morphological effect at any concentration. These initial observations could be related to non-specific effects induced by Gap1 or could suggest compromised cell viability if *MSTRG.10933* expression is knocked down to the low levels seen with use of Gap1. Subsequent repeat knockdowns confirmed a much higher potency of Gap1 over Gap2 (Figure 3.9c). Morphological effects from Gap1 treatment were more subtle in these repeats (Figure 3.9d), potentially due to the increased confluence of VSMCs and lack of higher doses of Gap1 used in these next runs. To confirm any effect on cell density or elongation linked to *MSTRG.10933* expression rather than non-specific effects from the sequence design of Gap1 would require further testing using additional Gapmers and biological replicates.

Additional Analysis

As these initial observations suggested loss of *MSTRG.10933* may reduce overall cell numbers, our initial hypothesis that *MSTRG.10933* inhibits cell proliferation could be incorrect. We therefore decided to initially explore an alternative hypothesis that *MSTRG.10933* instead promotes cell viability. To investigate this, we performed knockdown of *MSTRG.10933* in SVSMCs within complete growth media (10% FBS), as any loss of viability would be more apparent via EdU than if performed in quiescent conditions where the proportion of EdU cells is typically <1%. In initial testing, we again saw a much higher potency of Gap1 to reduce expression of the lncRNA, with Gap2 not effective at all in the proliferation model using complete growth media (Figure 3.9e). EdU incorporation appeared dampened in this model in SVSMCs treated with Gap1 at 10 (~50-80% of Gap Ctrl) and 20nM (~27-32% of GapCtrl) (Figure 3.9f). This provides further evidence that *MSTRG.10933* may indeed be functional. Considered alongside the initial observations on changes in cell morphology and loss of density, the loss of EdU signal could be interpreted as related to loss of cell survival rather than inhibition of proliferation. Testing with more repeats and additional Gapmer designs would be needed to confirm this.

Overall, these knockdown approaches on three lncRNA candidates could not confirm the role of these lncRNAs on proliferation.

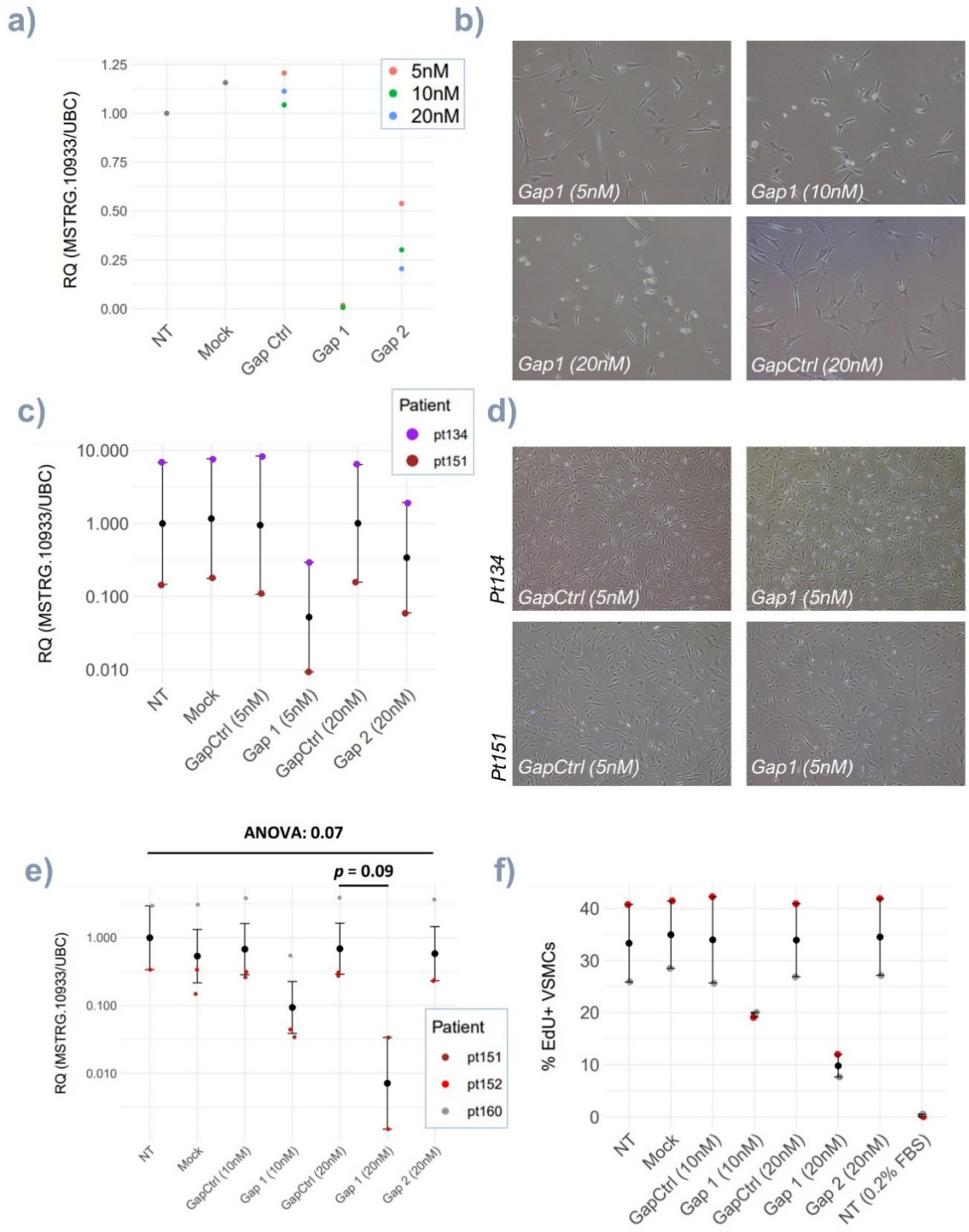


Figure 3.9 a) Optimisation of Gapmer-mediated MSTRG.10933 knockdown in SVSMC quiesced in 0.2% FBS and b) corresponding images of examples of cell morphology with Gap1 concentration gradients or GapCtrl (n=1). c) MSTRG.10933 expression during repeat knockdowns in quiesced SVSMC (n=2) and d) corresponding images of example cell morphology with Gap1 or GapCtrl. e) MSTRG.10933 expression during Gapmer-mediated MSTRG.10933 knockdown in SVSMCs grown in complete growth media (10% FBS) (n=3) and corresponding EdU assay. (Black dots show mean and standard error (for the n=3 panel) of either dCt represented as RQ or EdU+ %, one-way ANOVA followed by Tukey's post-hoc test)

3.5 Additional Analysis Discussion

3.5.1 Options to continue phenotype screening for candidate pathology-associated lncRNAs

We identified 3 lncRNAs with potential to influence proliferation in the SVSMC model and explored this in knockdown experiments coupled to EdU assays. Whilst *AC018647.3* knockdown was effective via use of dsiRNA, we saw no reactivation of SVSMC proliferation during quiescence. Notably, this lncRNA has a nuclear component which may not have been targeted effectively by the dsiRNA reagent[197]. Definitive conclusions for functionality of *LINC02015* within the model would require repeat testing either using further biological replicates or with a second, more effective Gapmer to provide an effective knockdown of the gene. For *MSTRG.10933*, data suggesting functional potential is obtained but confounded by use of a second, less effective Gapmer. Hence, for all three candidate lncRNAs a requirement for further knockdown optimisation, ideally with novel Gapmer designs, to provide an effective, consistent knockdown of the lncRNA is identified.

It is also possible that, despite the traits they possess making them worthy of functional screening as laid out in Table 3.2, their induction or repression given IL-1α/PDGF-BB stimulation may be incidental to VSMC proliferation. However, lncRNA functionality is known to be able to occur within specific temporal windows and may also be particular to a specific pathway or phenotype. Follow-up exploration of the candidates within this model could therefore vary the experimental set-up to try and optimise the best conditions in which to detect a function via knockdown. Phenotypes other than proliferation that are activated by IL-1α/PDGF-BB stimulation could also be explored such as VSMC quiescence, differentiation state, migration, invasion and release of pro-inflammatory mediators. For example, the candidate lncRNAs upregulated during quiescence could be assessed for an effect on VSMC contractility using electric cell-substrate impedance sensing as recently demonstrated[251], or an effect on VSMC viability using widely-used MTT assays.

3.5.2 Evidence for cis activation mechanisms at elncRNA loci and other lncRNA loci

Enhancer association is particularly of interest in this chapter as we find in the manuscript that non-reference lncRNAs were found enriched amongst elncRNAs, and that elncRNAs appeared to have a particularly large contribution to VSMC pathology. Determination of the potential functional impact of elncRNAs would be aided by identifying increases in PCG co-expression, activation and/or repression suggestive of *cis* function around elncRNA loci compared to other lncRNA loci. *Cis*-regulation of neighbouring genes is a trait also found in many lncRNAs which are not enhancer-transcribed[252] and has already been explored via correlation with neighbouring PCGs in various prior studies. However, we would expect this to be a stronger effect around elncRNAs than other lncRNAs considering the role enhancers play in gene regulation. Identification of such effects would provide validation for our enhancer annotation method and aid direction of future experiments aiming to identify possible mechanisms of action.

Whilst we identify correlation of fold changes for elncRNAs and neighbouring PCGs as well as other lncRNAs and neighbouring PCGs in the SVSMC dataset, this was not seen for elncRNAs in the AOSMC and CASMC datasets. This is potentially due to low numbers of differentially expressed elncRNAs in the latter two. The different stimuli used across these *in vitro* datasets appear to have varying influence on elncRNA loci, as 51 of the 162 lncRNAs affected by IL-1α/PDGF-BB are elncRNAs (31.5%) compared to only 35 of the 143 stiffness-responsive lncRNAs in AOSMCs (24.5%) and 36 of the 168 such lncRNAs in CAMSCs (21.4%). Fold changes are inherently variable between datasets and abundance levels[241], whilst they were also generally lower in the stiffness datasets which all may dampen correlation strength. As we performed this analysis by examining the aggregate correlation of all lncRNA-PCG pairs in each dataset this may obscure details in differences in capacity for cis-regulation between elncRNAs and other lncRNAs. An improvement in future could be to obtain enough replicates to perform correlation of individual lncRNA-PCG pairs in terms of expression level. As mentioned in the manuscript discussion, techniques to experimentally match lncRNAs to PCGs may also be worth consideration in future[253].

3.6 Chapter 3 Summary

This chapter provides an expanded view of lncRNAs relevant to VSMC pathology by identifying the non-reference lncRNAs associated with VSMC transitions to pathological states. In doing so we:

1. Create a more representative catalogue of the associated lncRNAs than has been previously obtained in pathological VSMCs, alongside a more powerful contextualisation of these lncRNAs that provides detail on their VSMC-enrichment and association with enhancers.

2. Provide a rigorous framework that could be applied to other cell types or model systems to improve understanding of lncRNAs in other specific biological niches. This becomes ever easier to implement with exponential increases in the amount of publicly available sequencing data and advances in the accuracy or speed of transcript assembly or coding prediction tools.

3. Tackle a well described limitation in transcript assembly approaches, specifically the accuracy of the 5' limits and exonic structures of the provided assemblies. We validate these for the majority of non-reference lncRNAs within our expanded VSMC annotation through comparing them to transcript structures and TSSs obtained in other biological contexts profiled in FANTOM CAT[136] as well as by assessing the repeatability of transcript structures assembled in one or more of the three VSMC datasets.

4. Use external expression atlas or genomic annotation data to build hypotheses on differences in lncRNA regulation between cell type and potential to interact with enhancer function.

A clear gap is proof of functionality within the as yet uncharacterised lncRNAs of the annotation. No selected candidate could be shown to influence proliferation conclusively. We do however provide a resource to aid further selection of candidates in future. This could be improved, particularly in

annotation of enhancers and their regulated PCGs which could well benefit from use of bespoke enhancer annotation data or use of an expanded number of RNAseq replicates within future datasets to obtain more robust correlations of lncRNA and PCG neighbours. Together this may provide stronger inference of candidate elncRNAs. Overall however, this chapter provides a strong foundation to further investigate lncRNAs in VSMC pathology as a whole and help understand where they fit in a wider picture of the regulatory framework that guides VSMC transitions. This will be explored in the next chapter.

Chapter 4: Identifying candidate lncRNAs involved in the initiation of VSMC transitions to pathological states

4.1 Chapter 4 Introduction

The expanded VSMC annotation described so far in this thesis comprises a valuable tool to explore the contribution of lncRNA to VSMC pathology. A constant underlying issue in studies which rely solely on reference annotation to characterise lncRNAs is whether the set of lncRNAs being analysed are fully representative of those present within the model system in question. The previous chapter goes some way to address this largely untouched issue in the context of VSMC pathological transitions and provides a solid foundation to further explore the underlying lncRNA contribution in specific relevant areas. One such area is vein graft failure pathology where an opportunity for direct *ex vivo* gene therapy treatment of grafts prior to implantation is provided during the surgical procedure. In most vasculoproliferative remodelling contexts, the aim when targeting VSMCs is to reverse their dedifferentiation and proliferation. However, during vein grafting the proliferation of VSMCs is thought to be initially triggered by the loss of endothelium during the harvesting and preparation of the vein as well as insertion into a high-pressure arterial environment[9]. A unique opportunity to modify VSMC behaviour at the point of their pathological stimulation is therefore presented. As the extent of this initial VSMC proliferation appears to determine later graft patency (see section 1.1.4), modifying the initial transcriptional response to proliferative stimuli originating from the surgical procedure could be a viable route to reduce long-term graft failure rates. Discovery and validation of genes at the forefront of stimuli responses in this setting is therefore of high interest. Further, targeting such regulators may also be advantageous in blocking upstream induction of multiple proliferative or otherwise pathological pathways or vice-versa for homeostatic pathways.

LncRNAs could offer promise to provide such candidates as they are enriched within the immediate-early genes thought to regulate initial the initial two hours of stimulus responses in VSMCs[254] and have a particular tendency to act in specific temporal windows (e.g developmental timepoints[155]) compared to PCGs. Further reason to suspect their involvement at early timepoints comes from their association with enhancers – which appear particularly dynamic in the initial transcriptional responses to stimuli[190]. As with lncRNAs, enhancer activity also occurs in a cell-specific manner[255], and so

identifying their interplay in VSMCs could lead to targeting of mechanisms of particular relevance to VSMCs over other cells or limiting off-target effects on other cell types. A promising example of an early-response, VSMC-enriched lncRNA is the cell cycle activator *SMILR*, which is induced prior to observable cell division in the SVSMC *in vitro* model[174]. Knockdown of *SMILR* destabilises mRNA for the core cell cycle gene *CENPF* at an early stage in the transition from quiescence to proliferation and could be targeted to reduce early proliferative drivers whilst minimising interference with endothelial cells. This would be particularly interesting to attempt in vein grafting procedures where *SMILR* knockdown could reduce vessel wall expansion without preventing the endothelial regeneration required for graft longevity.

In this chapter we hypothesised that early lncRNA dynamics after exposure to a mitogenic stimulus but prior to cell division may have substantial influence on the eventual fate of VSMCs and that investigating this in the SVSMC proliferation model would reveal high value therapeutic candidates like *SMILR* that could reduce or prevent early VSMC-directed vessel wall remodelling leading to vein graft failure (Figure 4.1). To address this we decided to leverage our previous work to build an expanded VSMC lncRNA annotation to allow a more comprehensive analysis of pathologically-relevant, VSMC-enriched and enhancer-transcribed lncRNAs. We used this in conjunction with RNAseq data of four timepoints within the previously determined 24 hour window between IL-1α/PDGF-BB exposure and cell division within the SVSMC proliferation model[174] and set early-response lncRNAs into context alongside other involved genes such as transcription factors (TFs), and cell cycle genes. Our transcriptomic analyses suggest an interplay of TFs, enhancers and lncRNAs within the first four hours that sets the VSMCs on a course towards later induction of cell cycle genes between 8 and 24 hours. We predict functional lncRNA-PCG *cis* pairings active within this crucial four hour period to identify candidate loci involved in early gene regulatory activity leading up to VSMC proliferation.

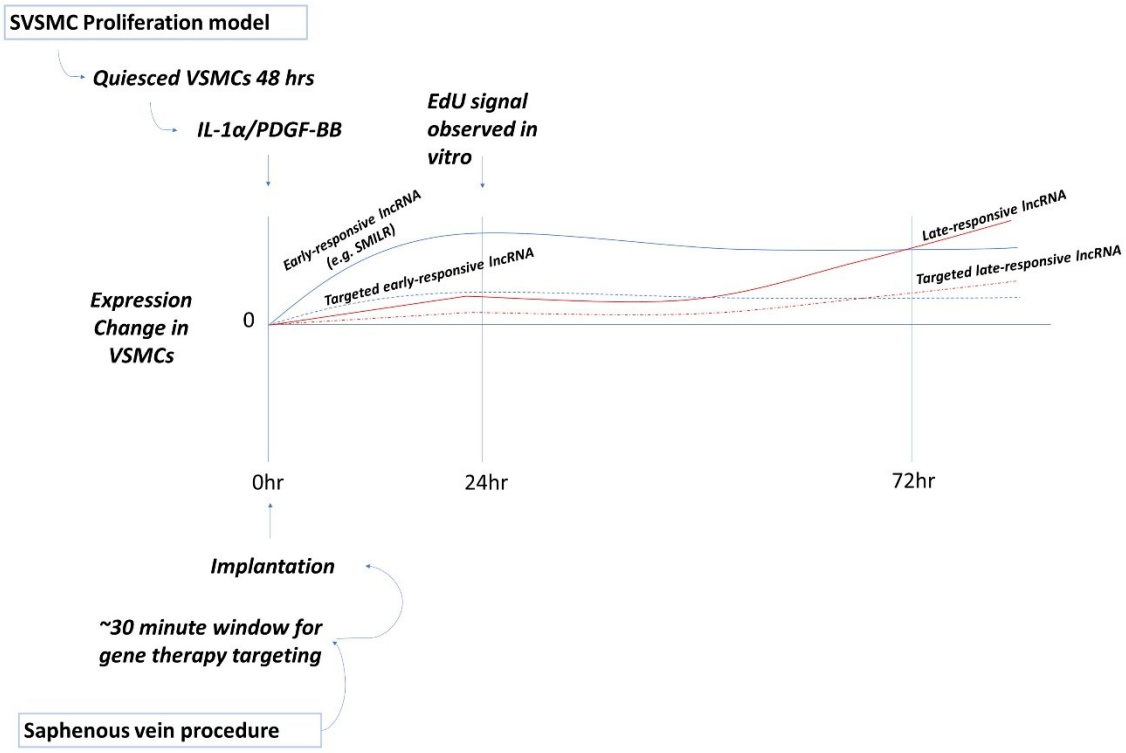


Figure 4.1 The early phases after introduction of the mitogenic stimuli within the SVSMC proliferation model can be mined to identify early-response genes that occur prior to observable proliferation. Findings in these early phases can be used to infer gene regulation activity occurring within the analogous vein graft setting where VSMC stimulation is also thought to initiate at an acute timepoint, here due to surgical processing prior to implantation.

4.2 Chapter 4 Aims

1. Use the expanded VSMC transcriptome to comprehensively characterise overall lncRNA dynamics in the model first 24 hours of IL-1α/PDGF-BB stimulated SVSMCs, set into context alongside cohorts of TFs and cell cycle genes

2. Examine the potential for lncRNAs to drive early stimuli response through *cis* regulation of their neighbouring PCGs

3. Use correlation to build further evidence of early response lncRNA *cis* regulation of PCGs and identify those with relevance to VSMC pathology

4.3 Chapter 4 Results

4.3.1 LncRNA dynamics in the early phases of SVSMCs stimulated with IL-1α/PDGF-BB

LncRNAs acting as gene regulators in the early phases of response to a mitogenic stimulus may have additional value relative to other lncRNAs as gene therapy targets to modify VSMC proliferation. To help identify such targets we first sought to gain a general overview of transcriptional dynamics across all genes involved in the early response to IL-1α/PDGF-BB within the SVSMC proliferation model described in chapter 3. Specifically, we wanted to examine expression changes in the temporal window between exposure to proliferative stimuli and first observance of cell division in this model at 24 hours[174]. RNAseq data was therefore obtained from within the SVSMC proliferation model by Drs A. Mahmoud and M. Ballantyne, this time using four timepoints in the first 24 hours (0, 4, 8 and 24 hours) post-IL-1α/PDGF-BB stimulation (unpublished). SVSMCs were sourced from four separate patient explants. Unstranded libraries were sequenced at comparable depth to the SVSMC dataset used in chapter 3 (35-45 million paired-end reads per sample).

To aim for a comprehensive analysis of all expressed lncRNAs we mapped reads from these libraries to a reference annotation consisting of the 3 expanded VSMC annotations (derived from datasets of SVSMC, AOSMC, CASMC and plaque-derived VSMCs) and GENCODEv26 merged into a non-redundant transcriptome as used in the manuscript presented in chapter 3. As the merging process can alter the structure of transcripts identified independently in each dataset, we used PLAR again to re-assess the coding status of all transcripts. This involved the removal of transcripts which were lowly-expressed or likely artefacts and processing the remaining transcripts with 3 distinct coding prediction tools. This provided a final list of 11815 expressed genes, 415 of which were genes absent from GENCODE (newly-assembled genes). A total of 10212 PCGs and 558 expressed lncRNAs genes (99 of which were newly-assembled) were present.

Identification of dynamically expressed genes over the four timepoints utilised two distinct methods within the DESeq2 tool[241]. The first used the likelihood-ratio test (LRT) to examine changes in gene expression across all timepoints simultaneously and identify those with repeatable changes between biological replicates through providing a single *p* value. The second took a more fragmented approach, using all 6 pairwise comparisons of timepoints to provide each gene with a set of 6 fold changes and *p* values (Figure 4.2a). This allowed us to select genes with substantial enough fold changes to be of interest in any of the comparisons between timepoints. A large overlap existed between the two methods (Figure 4.2b) but using both allowed us to rule out the proportions of genes with inconsistent changes between replicates over the full 24 hours (p >0.05 via LRT) or inconsequential changes in expression (p >0.05 and/or absolute fold change values of under 1.5 for all timepoint comparisons). Altogether, 4395 differentially expressed genes (including 234 lncRNAs) were identified over the timecourse experiment and are here-on referred to as early-response genes.

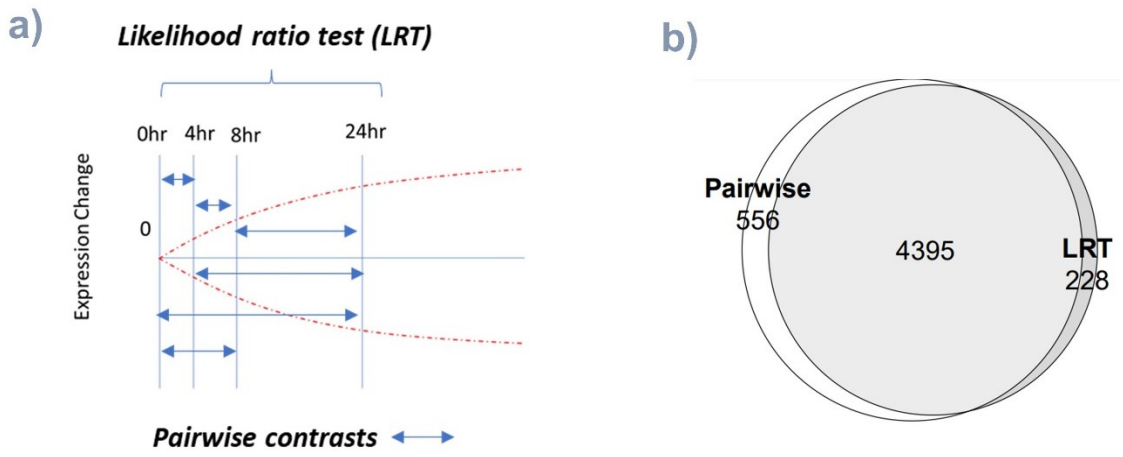


Figure 4.2 a) 2 distinct methods to identify genes with significant differential expression within the SVSMC timecourse experiment and b) their overlap in terms of number of identified early-response genes

To begin to explore early-responsive lncRNAs we examined their overall induction or repression in the 3 discrete temporal windows making up the timecourse (0-4 hours, 4-8 hours and 8-24 hours) (Figure 4.3a and b). We found lncRNAs were enriched within genes induced in the first four hours (79 lncRNAs, p <0.0001, BH-corrected Fisher's exact test, background of all early-

response genes for all *p* values in this section) and depleted amongst genes induced in the 4-8 hour window (9 lncRNAs, $p < 0.05$). LncRNA genes were neither enriched nor depleted within genes induced in the later 8-24 hour window (59 lncRNAs, $p > 0.05$) but were depleted within genes repressed in this window (38 lncRNAs, $p < 0.01$). Hence, though lncRNAs were induced in the 8-24 hour window, the class as a whole shows greatest tendency for induction prior to this in the 0-4 hour window. Interestingly, principal component analysis of all samples indicated largest gene expression variance between samples was between those collected at zero and four hours (Figure 4.3c), indicating most of the transcriptional changes in the 24 hour period of SVSMCs treated with IL-1α/PDGF-BB occur within an initial 4 hour burst of transcriptional activity - the same phase in which lncRNA induction is enriched.

VSMC-enriched and enhancer-transcribed lncRNAs (elncRNAs) are of interest for characterisation for the opportunity to target VSMC-specific and/or enhancer-related control mechanisms. Our analysis in chapter 3 also showed their tendency to be dynamically regulated during VSMC pathological transitions. We defined these groups of lncRNAs (as described in chapters 2 and 3) and saw again a tendency towards induction within four hours for both VSMC-enriched lncRNAs (12 lncRNAs, $p < 0.01$, 0-4 induced genes) and elncRNAs (22 lncRNAs, $p < 0.001$, 0-4 induced genes) (Figure 4.3a and b). Intriguingly this pattern did not hold for the remaining lncRNAs (51 lncRNAs, $p > 0.05$, 0-4 induced genes), suggesting lncRNAs with VSMC-enrichment and/or enhancer-transcription may have particular importance for regulating the initial 0-4 hour burst of transcriptional activity in response to IL-1α/PDGF-BB in this model. Use of the expanded VSMC annotation provided a substantial number of newly-assembled lncRNAs induced in 0-4 hours (15 lncRNAs or 20% of all 0-4 hour induced lncRNAs) and throughout, supporting use of the approach outlined in chapter 3 to identify a more complete cohort of early-response lncRNAs in the initial phases of SVSMC response to mitogenic stimuli.

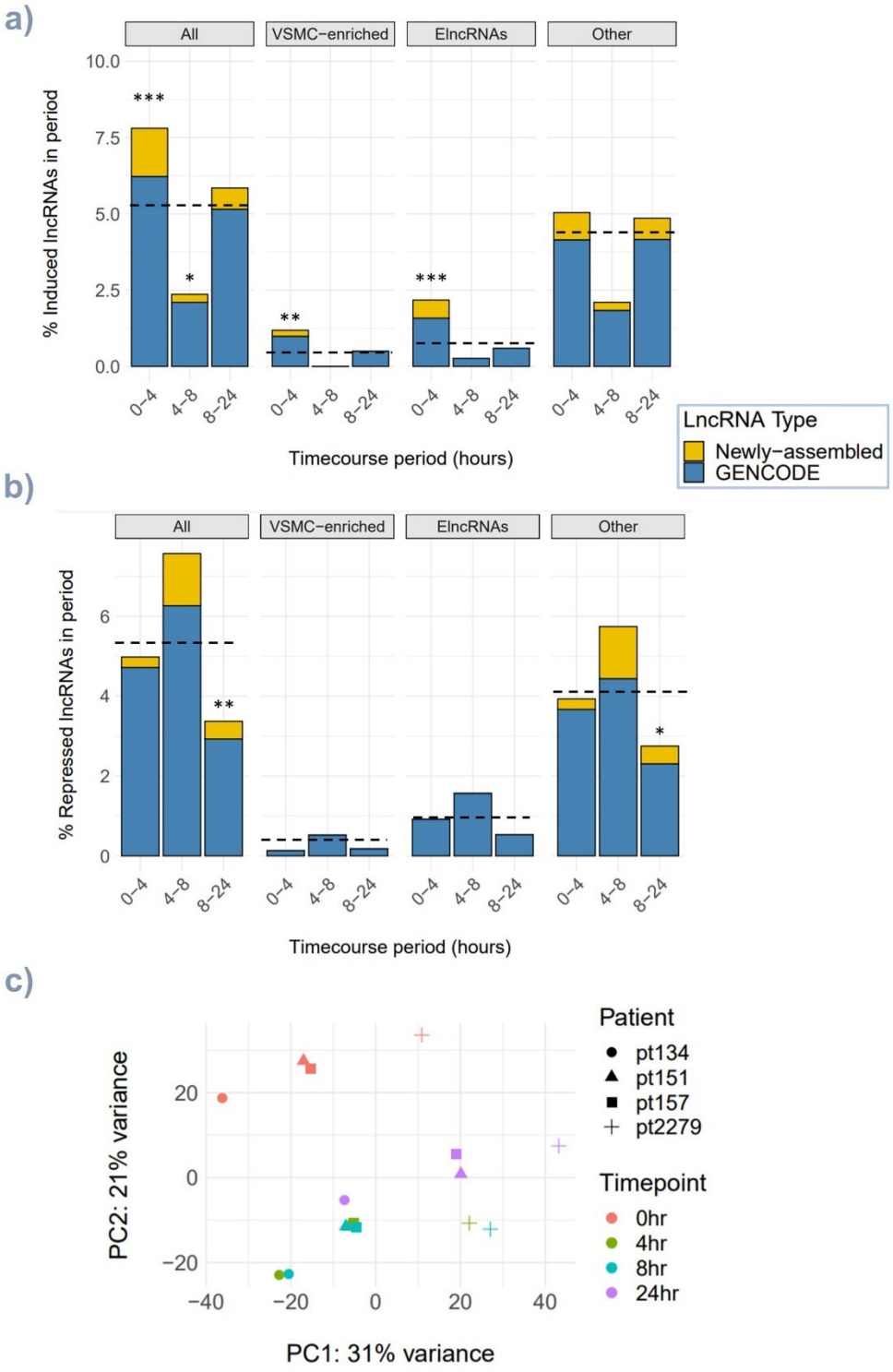


Figure 4.3 Enrichment or depletion of lncRNAs genes within either a) induced or b) repressed genes in each discrete window of the timecourse (Fisher's exact test + BH correction, background of all early-response genes indicated by dashed lines). c) PCA of all read counts per sample in the SVSMC timecourse RNAseq normalised within DESeq2 (using the rld method). (p <0.05 *, p <0.01 **, p <0.001 *, bars with no annotation indicate no significant enrichment or depletion).**

To set this lncRNA activity into context within the change in VSMC state, we began by identifying where the effector genes which implement the cell cycle were in our data, utilising a recently published cohort of the core genes implicated in control of the S, G2 and mitotic phases of the cell cycle (referred to hereon as cell cycle genes)[256]. Cell cycle genes were highly enriched within genes induced in the final 8-24 hour window (p <0.0001) (Figure 4.4a) but were depleted amongst genes induced in the initial 0-4 hour window and any window of repressed genes (p <0.001) (Figure 4.4b). Notable cell cycle genes heavily upregulated in the 8-24 hour window include proliferation markers, mitotic components and cell cycle regulators such as *MKI67*, *CDK1*, various *CENP* genes and the S/G2 checkpoint regulator *CDK2*. The lncRNA *SMILR* which controls the cell cycle in VSMCs was also present amongst such genes. This induction in the later phase of the 24 hour period mirrors the EdU assay data collected for the SVSMC model in[174] and indicates the overwhelming majority of effector genes implementing cell division are induced downstream of transcriptional changes occurring in the first four hours.

To look for gene regulators which may influence the later activation of these cell division genes, we examined TF genes using a recently published cohort[257]. In contrast to cell cycle genes, TFs showed a tendency for induction within four hours (p <0.01) and were also present in changes at 4-8 and 8-24 hours. They were therefore induced earlier than cell cycle genes but showed less extreme bias to induction in one window (Figure 4.4a). They were also particularly repressed within four hours and between 4-8 hours (p <0.0001) which further underlines the localisation of TF expression fluctuations to within early transcriptional changes (Figure 4.4b). We found several TFs of interest were transiently induced with a four hour peak in our dataset including those consistent with VSMC dedifferentiation such as *ELK1, ETS1* and NF-kB components *RelB*, *NFKB1* and *NFKB1* as well as TFs associated with early cell responses to mitogenic stimuli such as *JUN* and *FOSB*.

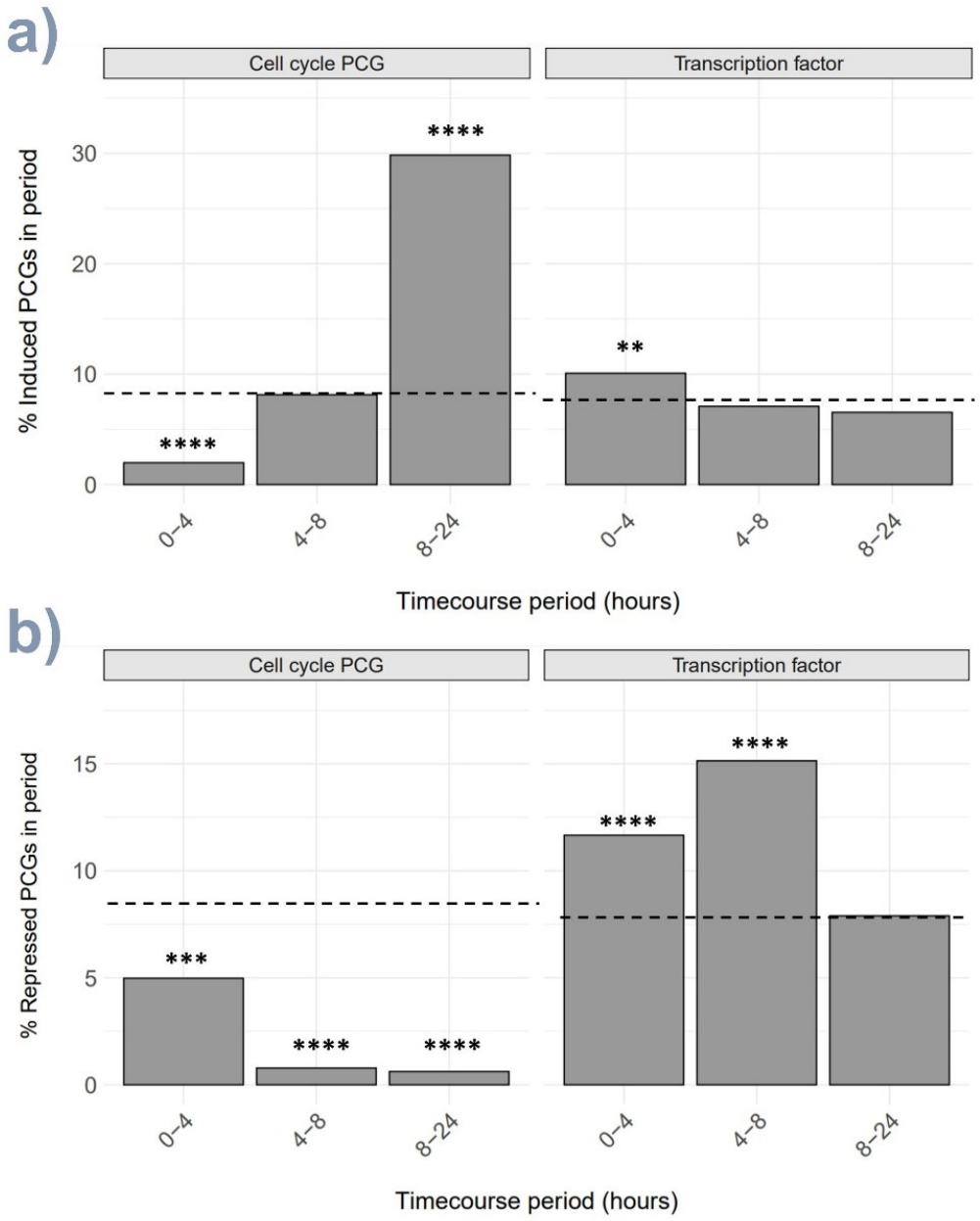


Figure 4.4 Enrichment or depletion of cell cycle or TF genes within either a) induced or b) repressed genes in each discrete window of the timecourse (Fisher's exact test + BH correction, background of all early-response genes indicated by dashed lines). (p <0.05 *, p <0.01 **, p <0.001 ***, p <0.0001 **, bars with no annotation indicate no significant enrichment or depletion).**

Together this provides an overview of early-responses to proliferative stimuli in the SVSMC model. During the initial four hours, the bulk of transcriptional changes that are to occur within 24 hours have taken place. This indicates an initial burst of transcriptional activity occurs in this period, and this appears to be enriched with lncRNA (particularly VSMC-enriched lncRNAs and elncRNAs) and TF induction as well as TF repression. This initial activity is likely to lay the foundation for the future path of the VSMCs and act upstream

of the S/G2-M phase cell cycle genes which before 8 hours are relatively unaffected. These genes are instead upregulated over the 8-24 hour window which precedes or coincides with the initial rounds of cell division observed by Mahmoud et al.[174]. Genes which are active within the early 0-4 hour timeframe may constitute particularly high value targets to dampen or prevent pathways induced by IL-1α/PDGF-BB before their effects take hold and induce the VSMC transition to a proliferative state.

4.3.2 Evidence for widespread lncRNA-dependent cis activation of PCGs and TFs prior to cell division

The enrichment of lncRNAs amongst the earliest induced genes in the first four hours suggests a role for them in early redirection of VSMCs towards proliferative phenotypes. As many lncRNAs are known to regulate neighbouring genes using *cis* regulation[252] we assessed whether the rate of PCG differential expression in the first four hours was increased for those which were neighbours (within 250kbp) of these earliest lncRNA responders. As the RNAseq libraries were sequenced using an unstranded protocol, we discounted PCGs which overlapped the lncRNA from consideration as *cis* potential pairs and focus our analysis solely on *cis* regulatory potential of intergenic lncRNAs.

Considering only genes which were differentially expressed in the first four hours, we found 59 of 1634 early-response PCGs in the flanks of 117 early-response lncRNAs. The rate of differential expression amongst these PCGs was 1.51x greater than that found genome-wide (p <0.01, background of 0-4 hour early-response PCGs) (Figure 4.5a) – an effect not observed for PCGs flanking lncRNAs which were stable (i.e. unaffected by IL-1α/PDGF-BB) in this initial four hour window. Considering all early-response genes across the 24 hours we found 223 of 4116 early-responsive PCGs were in the flanks of the 234 early-responsive lncRNAs. These PCGs however were only 1.16x more likely to be early-responsive if in the flanks of early-response lncRNAs compared to PCGs genome-wide (p <0.01, BH-corrected Fisher's exact test for p values in this section unless stated, background of all early-response PCGs) (Figure 4.5b). This suggests a stronger link between changes in

lncRNA expression and neighbouring PCG expression exists within the first four hours compared to over the full 24 hours.

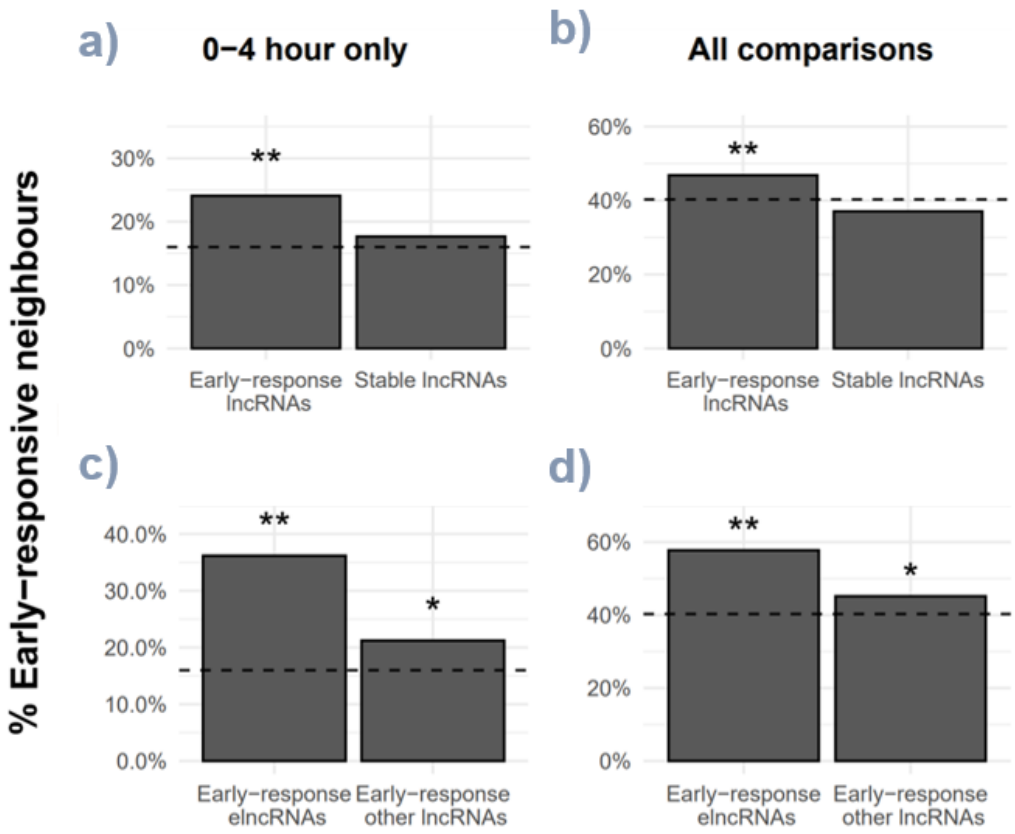


Figure 4.5 Enrichment or depletion of early-response PCGs within PCGs neighbouring early-response or stable lncRNAs considering either a) the 0-4 hour window or b) the full timecourse experiment. c-d) considers PCGs around early-response elncRNAs or other early-response lncRNAs. Fisher's exact test + BH correction by row, background of all PCGs genome-wide indicated by dashed lines. (p <0.05 *, p <0.01 ** bars with no annotation indicate no significant enrichment or depletion).

As enhancers appeared to be a key source of many of the lncRNAs induced within four hours we checked whether elncRNAs could support enhancer-based *cis* regulation in this timeframe and so have more pronounced signs of *cis* regulation than other lncRNAs. We found 17 early-response PCGs near 29 early-response elncRNAs and 42 early-response PCGs near 88 early-response other lncRNAs in the first four hours. Though early-response PCGs were enriched amongst PCGs in the flanks of both types of lncRNAs relative to PCGs genome-wide, this was particularly pronounced for those in the flanks of elncRNAs (2.26x more likely, p <0.01, background of all early-response PCGs) compared to those in flanks of other lncRNAs (1.33x more likely, p

<0.05, background of all early-response PCGs) (Figure 4.5c). The effect was again decreased when considering differential expression across the full 24 hours (41 early-responsive PCGs in proximity to 41 early-responsive elncRNAs and 186 early-responsive PCGs around 71 other early-responsive lncRNAs). In this timeframe, PCGs near early-response elncRNAs were 1.43x more likely to be early-responsive relative to PCGs genome-wide (p <0.01, background of 0-4 hour early-response PCGs) whilst for those near other lncRNAs this was only 1.12x more likely (p <0.05, background of 0-4 hour early-response PCGs) (Figure 4.5d). Hence, though both elncRNAs and other lncRNAs show signs of *cis* regulation, this is more pronounced at elncRNAs suggesting the influence of their requisite enhancer regions.

As TFs and lncRNAs were enriched within 0-4 hour early-response genes, we also looked for signs of lncRNA-dependent *cis* regulation of TFs. Considering early-response genes identifiable in the first four hours, we found 16 of 191 early-response TFs were in the flanks of 117 early-response lncRNAs. TFs were 2.26x more likely to be differentially expressed if neighbouring such lncRNAs compared to TFs genome-wide (p <0.01, background of 0-4 hour early-response TFs) (Figure 4.6a). The association of lncRNA and neighbouring gene differential expression is therefore present for TFs as well as all PCGs generally. No enrichment of differentially expressed TFs was seen around stable lncRNAs. The effect was again reduced when expanding the analysis to all timepoints; we found 26 of 339 early-response TFs were flanking the 234 early-response lncRNAs. Here, the rate of differential expression was only 1.51x greater for TFs in the flanks of early-response lncRNAs compared to TFs genome wide (p <0.01, background of all early-response TFs) (Figure 4.6b). We also noted that in the first four hours candidate cis-regulated PCGs of early-response lncRNAs were more likely to be TFs than differentially expressed PCGs genome-wide (2.32x more likely, p < 0.001) (Figure 4.6c). TFs were also enriched to a lesser extent as neighbours of unchanging lncRNAs (1.48x more likely, p < 0.01) suggesting this effect is in part due to a tendency for genomic proximity of lncRNAs and TFs. However, as the stronger association occurs for differentially expressed lncRNAs, this further underlines the likely co-regulation and interplay of TF and lncRNAs within the first four hours. Overall, we provide evidence suggesting lncRNA-

dependent *cis* regulation of TFs is particularly acute within the first four hours with an enrichment of TFs as candidate *cis*-partners of early response lncRNAs.

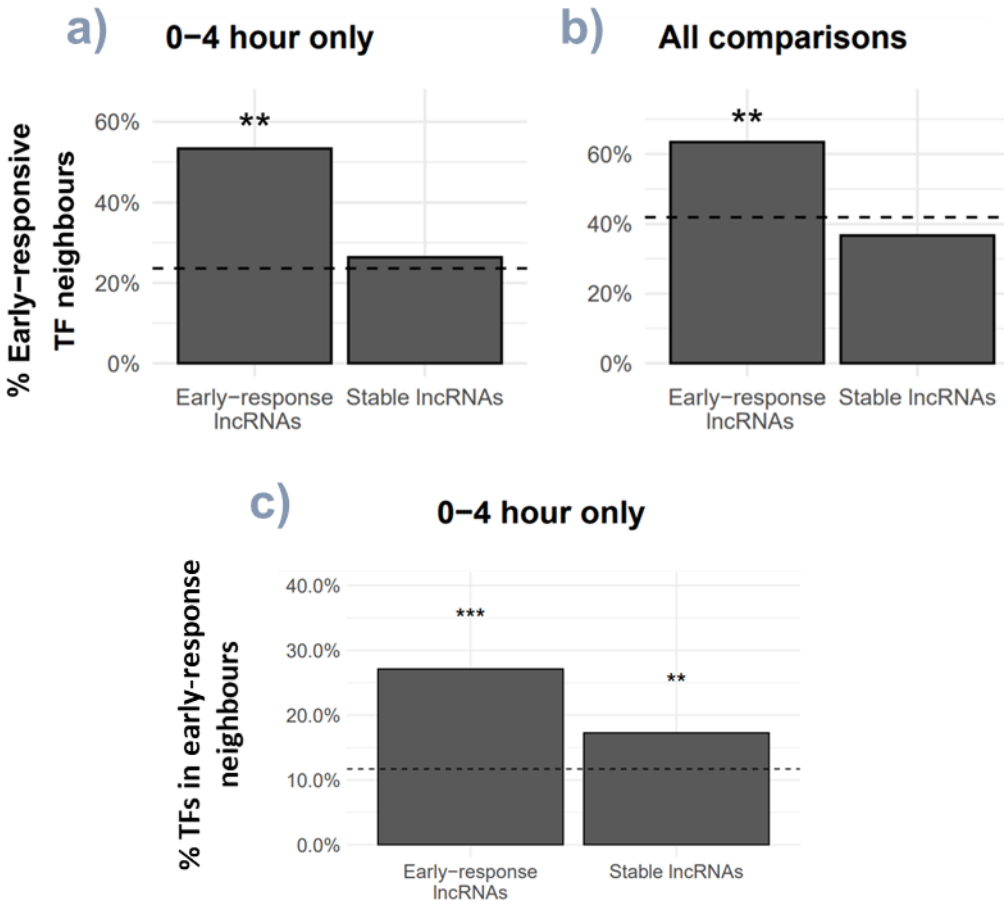


Figure 4.6 Enrichment or depletion of early-response genes within TFs neighbouring early-response or stable lncRNAs considering either a) the 0-4 hour window or b) the full timecourse experiment. c) Proportion of TFs among early-response genes around early-response lncRNAs or stable lncRNAs. Fisher's exact test + BH correction by row, background of all PCGs genome-wide indicated by dashed lines. (p <0.05 *, p <0.01 **, p < 0.001 *, bars with no annotation indicate no significant enrichment or depletion).**

These conclusions could be confounded by different numbers of PCG neighbours for different lncRNA types, for example if one lncRNA type is often found in particularly gene-dense loci this may reduce the proportion of regulated neighbouring PCGs and dilute any signal of *cis* regulation. We therefore also checked whether the lncRNA groups analysed were in more PCG-dense loci but found no significant difference between lncRNA groups (p >0.05, Kruskal-Wallis) (Figure 4.7). Differences in number of lncRNA neighbours is not likely to unduly affect our conclusions.

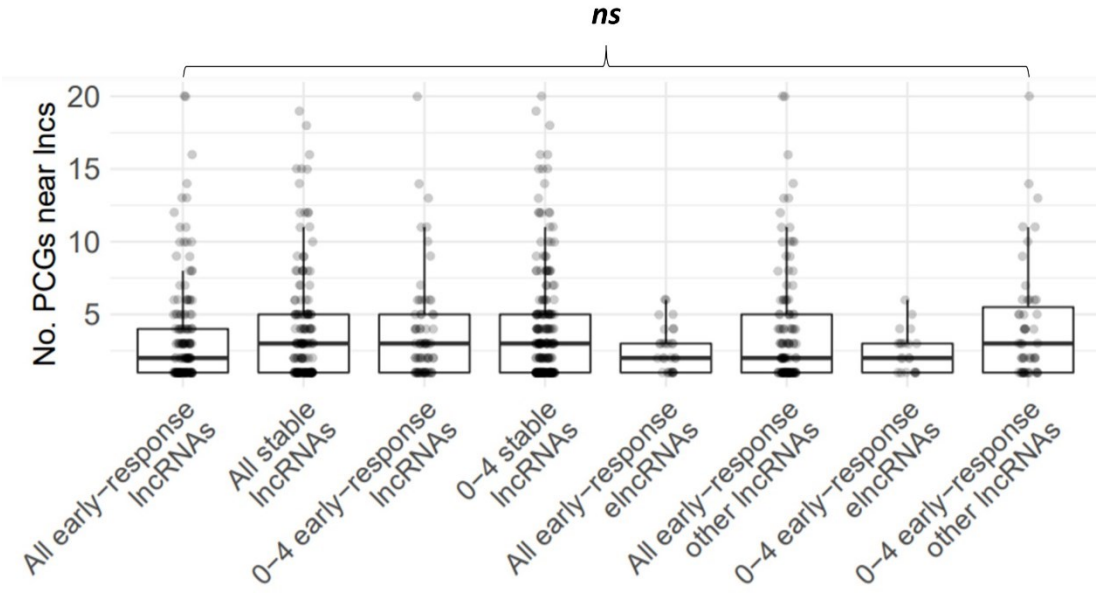


Figure 4.7 Boxplots showing number of PCG neighbours for each lncRNA type investigated in Figure 4.5a-d and Figure 4.6a-b (Kruskal-Wallis)

Overall, we find PCGs are more likely to be early-responsive if in the flanks of early-responsive lncRNAs but not stably expressed lncRNAs, suggesting widespread lncRNA-dependent *cis* regulation. This is particularly acute in the 0-4 hour timeframe, for TFs and for PCGs near elncRNAs. This suggests lncRNA-dependent *cis* regulation is particularly active in the first four hours around enhancer loci and may enable wider downstream regulation of VSMC state through regulation of PCGs and TFs.

4.3.3 Candidate lncRNAs involved in early cis regulation of VSMC state

To now focus in on lncRNAs with strongest evidence for a *cis* regulatory role during initiation of the VSMC transition towards cell division, we examined the correlation in expression between early-response lncRNAs and their neighbouring early-response PCGs across the full 24 hour timecourse experiment. We identified the *VSMClnc6-CXCL8* pairing (BH-corrected Spearman's p = 0.096) already shown to be of interest through use of GeneHancer interaction data in chapter 3 and used this as a benchmark for finding further potential pairings of interest. Based on this we set a permissive threshold for determining significant correlation (BH-corrected Spearman's p

<0.1). Using this, we found that of all 223 total early-response PCGs in the flanks of the 234 early-response lncRNAs across the timecourse, 69 (30.2%) were significantly correlated with their lncRNA neighbour. These were found across the full range of distances between lncRNA and PCG TSSs and included 33 pairings of early-response genes found within the crucial four hour timeframe (Figure 4.8a). Notably, these 0-4 hour correlated pairings including 32 PCGs, 54.2% of the 59 0-4 hour early-response PCGs flanking 0-4 hour early-response lncRNAs (Figure 4.8b). This was a significant enrichment of correlating lncRNA-PCG partnerships within the earliest 4 hour phase of VSMC response to mitogenic stimulus (p <0.0001, Fisher's exact test, background of all early-response PCGs in flanks of all early-response lncRNAs) – an effect which remained if using a more stringent cut-off of p <0.05 for identifying correlating pairings (p <0.01, Fisher's exact test, background of all early-response PCGs in flanks of all early-response lncRNAs). This again shows the greatest potential for lncRNA-dependent *cis* regulatory activity in the first four hours of the timecourse.

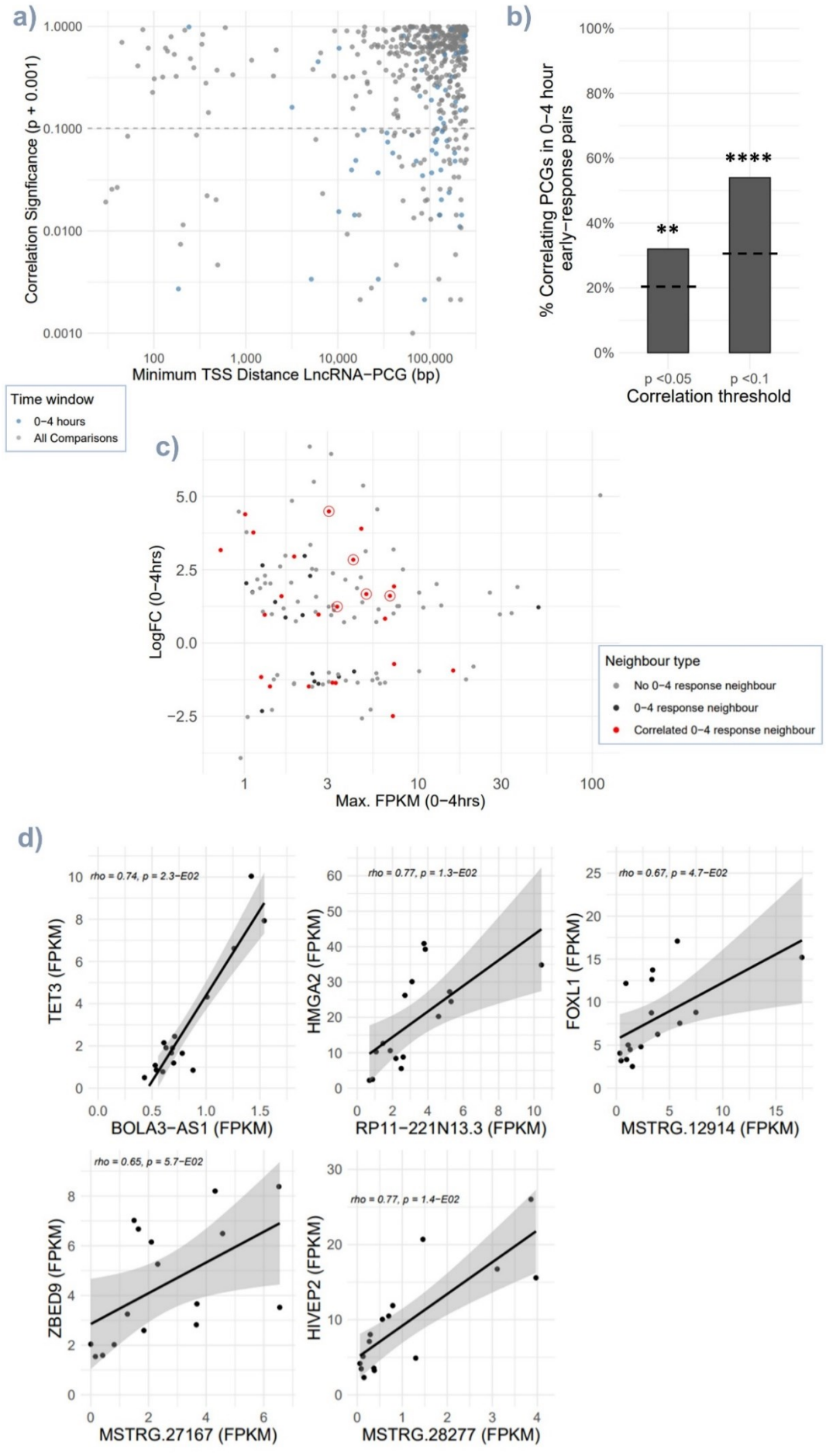

a)
Correlation Significance (p + 0.001)
1.0000
0.1000
0.0100
0.0010
100
1,000
10,000
100,000
Minimum TSS Distance LncRNA−PCG (bp)
Time window
0−4 hours
All Comparisons
b)
% Correlating PCGs in 0−4 hour early−response pairs
100%
80%
60%
40%
20%
0%
**
****
p <0.05
p <0.1
Correlation threshold
c)
LogFC (0−4hrs)
5.0
2.5
0.0
−2.5
1
3
10
30
100
Max. FPKM (0−4hrs)
Neighbour type
No 0−4 response neighbour
0−4 response neighbour
Correlated 0−4 response neighbour
d)
TET3 (FPKM)
rho = 0.74, p = 2.3−E02
BOLA3−AS1 (FPKM)
HMGA2 (FPKM)
rho = 0.77, p = 1.3−E02
RP11−221N13.3 (FPKM)
FOXL1 (FPKM)
rho = 0.67, p = 4.7−E02
MSTRG.12914 (FPKM)
ZBED9 (FPKM)
rho = 0.65, p = 5.7−E02
MSTRG.27167 (FPKM)
HIVEP2 (FPKM)
rho = 0.77, p = 1.4−E02
MSTRG.28277 (FPKM)

Figure 4.8 a) Scatterplot showing significance of early-response lncRNAs and neighbouring early-response PCG correlations over a range of distances between pairs, dashed line indicates a permissive significance threshold of (p <0.1, Spearman's rank). b) Enrichment of correlating early-response lncRNA-PCG pairs within those found in the 0-4 period using a permissive or more stringent threshold for correlation significance (Fisher's exact test, dashed line is background of correlating pairs found for early-response lncRNA-PCG pairs within all timepoints) c) Scatterplot indicating abundance and strength of induction or repression for all lncRNAs activated in the 0-4 hour timepoint as well as presence of any correlating early-response PCG neighbours in the same time frame. Red-ringed lncRNAs are those which are uncharacterised and link to a TF, correlations of these lncRNA-TF pairs are shown in d). (p <0.01 **, p <0.0001 **)**

We now defined neighbouring lncRNA-PCG pairs of interest based on their joint differential expression in the initial four hours and significant correlation over the full timecourse. Of the 117 differentially expressed lncRNAs in the four hour window, 38 had a PCG neighbour which was a four hour early-response gene and 23 correlated in expression with this PCG neighbour over the timecourse (Figure 4.8c). To identify which of these 23 could be of particular interest for further validation and characterisation, we first performed a literature search on their 32 correlated PCGs, showing several have a previously identified association with VSMC pathology, inflammation or other cardiovascular disease in the literature. We also classed the 11 TFs found within the 32 linked PCGs as of particular interest due to their potential to carry out wide-ranging changes on gene expression. We noted lncRNAs located in developmental *HOX* gene clusters accounted for 6 of these TFs whilst 5 uncharacterised lncRNAs (including 3 newly-assembled lncRNAs) were found linked to the remaining 5 (Figure 4.8d). Profiles for the 13 lncRNAs associated with TFs and/or the PCGs highlighted in the literature search are displayed in Table 4.1 and can be used as a basis to explore these candidate *cis* pairings experimentally.

In summary, these lncRNAs and their correlated, paired PCGs are strong candidate loci for lncRNA-dependent *cis* regulation mechanisms active during first four hours – a phase in which lncRNA *cis* activity appears biased, with the potential to redirect the course of VSMCs towards a proliferative, pathogenic state.

LncRNA	PCG	LncRNA Type	PCG Type	Spearman's Rho	Spearman's p	TSS Distance (kbp)	LncRNA Max. FPKM	LncRNA Expression pattern	VSMC-related characterisation for PCG
VSMClnc6	*CXCL8*	Newly-assembled VSMC-enriched, elncRNA	Protein coding	0.59	9.62E-02	19.19	17.6	Transient induced, 4 hour peak	Chemokine and mitogen for VSMCs[39,256]
CD27-AS1	*NOP2*	LncRNA	Protein coding	0.80	1.00E-02	211.63	15.9	Repressed, 8 hour trough	Methylation of ICAM-1 mRNA to improve translation and aid leukocyte attachment[257]
MSTRG.12913 *MSTRG.12914*	*FOXL1*	Newly-assembled elncRNAs	TF	0.64 0.67	6.05E-02 4.72E-02	105.57 187.67	8.9 6.9	Sustained induction, 24 hour peak Transient induction, 4 hour peak	Capable of maintaining resident stem cell populations, Activates Wnt/β-catenin pathways[258,259]
PITPNA-AS1	*RILP*	LncRNA	Protein coding	-0.66	3.96E-02	133.15	8.2	Sustained induction, 24 hour peak	Identified risk for plaque rupture, involved in lysosomal transport[260]
RP11-221N13.3	*HMGA2*	ElncRNA	TF	0.77	1.31E-02	221.25	5.7	Sustained induction, 24 hour peak	Targets the TF Twist in epithelial-mesenchymal transitions, also associated with vasculomimicry in gastro carcinoma[261]
RP11-249C24.10	*MT1A* *MT1E* *MT1X*	LncRNA	Protein coding	0.70 0.69 0.67	2.66E-02 3.85E-02 4.72E-02	27.29 14.10 71.05	4.7	Transient induced, 4 hour peak	Enzymes using zinc metabolism to reduce oxidative stress, GWAS suggests MT1A as a CVD risk loci[262]
MSTRG.27167	*ZBED9*	Newly-assembled LncRNA	TF	0.65	5.68E-02	186.20	4.2	Transient induced, 4 hour peak	Hypertension GWAS hit, strong across 4 blood pressure traits, uncharacterised TF[263]
HOTAIR	*HOXC6* *HOXC8*	VSMC-enriched elncRNA	TFs	0.67 0.60	4.79E-02 9.00E-02	15.67 34.15	3.4	Repressed, 8 hour trough	Both TFs have described role in maturation of VSMC progenitors[264]
MSTRG.28277	*HIVEP2*	Newly-assembled LncRNA	TF	0.77	1.45E-02	10.24	3.1	Transient induced, 4 hour peak	Repression of NF-κB and TGF-β pathways, regulator of c-Myc[265]
AC002480.4	*IL6*	VSMC-enriched elncRNA	Protein coding	0.59	9.25E-02	136.18	1.8	Sustained induction, 24 hour peak	Activates VSMC proliferation/migration + osteoblast phenotype[37,43,266]
HOXA11-AS1	*HOXA6* *HOXA7* *HOXA10* *HOXA11*	LncRNA	TFs	0.62 0.84 0.84 0.86	7.30E-02 2.38E-03 2.38E-03 1.71E-03	34.81 27.47 5.15 0.19	1.4	Sustained induction, 24 hour peak	No direct links identified
BOLA3-AS1	*TET3*	LncRNA	TF	0.74	2.27E-02	161.64	1.3	Transient induced, 4 hour peak	A widely studied demethylation enzyme, promotes IL6 expression in rat atherosclerosis[267]

Table 4.1 All correlated early-response lncRNA-PCG pairs found within the first four hours of the SVSMC proliferation model where the PCG is either a TF or has characterisation in the literature which suggests relevance to VSMC pathology. The table is ordered from highest maximum lncRNA abundance to lowest.

4.4 Chapter 4 Discussion

4.4.1 An association of lncRNAs, enhancers and TFs in a crucial early phase in the induction of VSMC proliferation

We see that lncRNAs and TFs have a tendency for induction within the 0-4 hour timeframe of the SVSMC proliferation model, suggesting implementation of gene expression changes by TFs may be supported by lncRNA mechanisms and vice-versa. The particular association of lncRNAs and TFs relative to other PCGs has been noted from some of the earliest produced catalogues of lncRNAs in various contexts[270,271]. In these studies, the association was made through identifying a tendency for TFs and lncRNAs to be in close genomic proximity, sparking initial suggestions of widespread lncRNA-dependent *cis* regulation of TFs. We build on this observation here, showing a greater incidence of TF differential expression in the flanks of differentially expressed lncRNAs compared to genome wide. We also see signs suggesting that lncRNAs have greater tendency to *cis* regulate TFs than other PCGs. In the 0-4 hour window for instance, our analysis shows TFs are 2.3x more likely differentially expressed if in flanks of differentially expressed lncRNAs compared to genome wide, this is much higher than the rate for all PCGs (1.5x).

A possible confounding factor is that less genes per lncRNA are found when considering just TF neighbours rather than all PCG neighbours. Our method here effectively compares the rate of differential expression amongst different types of lncRNA-neighbours and this rate is dependent on total number of gene neighbours per lncRNA. This may lead to an inherently higher rate when considering TFs on their own as there are less TF neighbours per lncRNA than PCGs neighbours per lncRNA. We therefore do not definitively show lncRNA-dependent *cis*-regulation mechanisms are more concentrated around TFs than other PCGs - in part due to the small sample pool of TFs neighbouring early-response lncRNAs (30 total in 0-4 hours). However, when considered with prior observations of lncRNA-TF co-localisations along the genome[270,271], an increased tendency of lncRNAs to regulate TFs seems likely in the initiation of VSMC proliferation, particularly within the 0-4 hour timeframe. Exploring

lncRNA control of TF gene regulation and vice-versa would be an interesting future strategy to modify VSMC behaviour at an early stage.

We also see that elncRNAs, useful to study due to their potential functional input to enhancer regions[200,252] and specific expression profiles[136], are a group of lncRNAs which are particularly likely to be induced in the 0-4 hour period. Within this window, an increased incidence of differentially expressed PCGs near differentially expressed elncRNAs (2.3x more likely than genome-wide) relative to other differentially expressed lncRNAs (1.3x more likely than genome-wide) suggests greater presence of *cis*-acting mechanisms at elncRNA loci. In this comparison we see no significant difference in overall number of PCG neighbours for these different categories of lncRNA so this factor is unlikely to affect this conclusion. Our analysis in this chapter improves on the FC correlation method used to compare *cis*-regulation potential of elncRNAs and other lncRNAs done in section 3.4.3. This earlier method may be an ineffective way to observe differences in *cis* potential due to differences in sample size between the two lncRNA types and inherently high variabilities in FCs. In this chapter however, we find evidence suggesting a stronger potential for *cis* regulatory mechanisms to exist near elncRNAs than other lncRNAs. This suggests their requisite enhancer regions are being utilised in the SVSMC timecourse model and that elncRNA transcripts may play role in this activation - potentially working together to influence chromatin conformation or epigenetic factors in the surrounding area. This also tallies with prior observations that indicate transcriptional changes in the first few hours after stimuli response occur at enhancer sites predominately[190].

It is notable that on removing elncRNAs and VSMC-enriched lncRNAs from consideration the tendency for lncRNA induction within four hours was not seen - implicating these more cell-type specific cohorts of lncRNAs as of particular importance to the initial VSMC response to stimuli. This fits with previous analysis of early-responsive genes which has shown that the vast majority are stimuli-specific and cell-specific rather than a common group of ubiquitous genes applied across various stimuli and cell types[254]. This finding provides further justification to study lncRNAs during the early response phase as those which are specifically expressed could be used to modulate the more generic core early-response genes. For example, only one of over 700 cell

cycle core components shared across cell types appear to be lncRNAs[256] yet *SMILR* can be used to specifically regulate one such core component (*CENPF*) in VSMCs specifically[174]. We show that many more examples of specifically expressed lncRNAs controlling or fine-tuning more ubiquitous genes (such as early response genes) are likely to be found within the early phase.

4.4.2 Candidate cis-regulating lncRNAs driving VSMC proliferation

Observations on lncRNA-dependent *cis* activity acting upstream of later VSMC transitioning helped us to identify 13 candidate lncRNAs of interest with strong potential for further study through their association with TFs or PCGS which have relevant characterisation for VSMC pathology described in the literature. We see many newly-assembled lncRNAs from our expanded VSMC annotation within this population (5 of the 13) again underlining the value in our approach outlined in chapter 3 to widen the search for lncRNA targets. We provide candidate lncRNAs involved in anti-inflammatory mechanisms such as the induction of anti-oxidative stress enzymes in the *MT1* locus[264] (linked to *RP11-249C24.10*) or induction of the NF-kB/TGF-B pathway-repressing TF *HIVEP2*[267] (linked to *MSTRG.28277*). We also find candidate lncRNAs involved in pro-inflammatory mechanisms such as the methylation and stabilisation of ICAM-1 by methyltransferase activity of NOP[272] (linked to *CD27-AS1*) and - as with the lncRNA candidates identified in the manuscript in chapter 3 - production of *CXCL8* and *IL6*. We further identify a candidate lncRNA (*BOLA3-AS1*) linked to induction of the epigenetic demethylator enzyme *TET3* which has a described role in enhancing *IL6* production[269]. Other lncRNAs are correlated with PCGs with literature that suggests their involvement in the loss of VSMC identity which could be apparent in the SVSMC model (*ACTA2* expression is heavily repressed over the 24 hours). In particular, *HOXC6* and *HOXC8* are two TFs which are implicated in the maturation of resident vessel wall resident VSMC progenitors[266] so *cis*-regulation by the neighbouring VSMC-enriched lncRNA *HOTAIR* (a possibility initially explored in this locus in[273]) may be involved in VSMC maturation and dedifferentiation. Another example is the induced *FOXL1* TF which has been

shown to allow of intestinal stem cells[261], possibly linking to a potential role in loss of VSMC differentiation state in the SVSMC model.

4.4.3 Targeting lncRNAs acting upstream of proliferative pathways

Many of these highlighted candidate lncRNAs show rapid induction in expression after stimulus which peak at the four hour timepoint before rapid decreases. Genes with transient induction likely have a different role to genes with more sustained induction. Indeed genes displaying differing half lives at the RNA and protein level seem to form different functional groupings with those which rapidy degrade on the RNA and protein level being enriched with TFs[274]. Transient expression implies a need for precise control of such genes, for instance to prevent wide-ranging harmful effects from over-production. Targeting these underlying control mechanisms, for instance TF-regulating lncRNAs, could be an effective strategy to halt the redirection of VSMCs. In the context of SVSMC proliferation, we can transfect the vein graft with a therapeutic agent at the point where mitogenic stimuli initiate redirection of the cells from a quiescent to a proliferative or pro-inflammatory state. In our data, we identify lncRNAs paired to *FOXL1* or *HIVEP2* TFs that could represent upstream *cis* control mechanisms of these TFs which could be modified to prevent downstream activation of their associated proliferative pathways (Wnt/β-catenin[260] or NF-κB and TGF-β[267] respectively). Such targets may be less effective in other vascular remodelling circumstances where the initial redirection of VSMCs towards proliferation and dedifferentiation has already produced downstream consequences e.g. the development of an atheroma. In these situations, early-response genes may still be of interest, due to their potential to act upstream and negate multiple proliferative pathways but the focus would naturally shift to those which are regulated in a less transient manner for instance the early-response lncRNAs at the *CXCL8* and *IL6* locus which remain highly expressed after the initial four hour burst of induction and were also identified in the dataset of SVSMCs 72 hour dataset.

4.4.4 Chapter 4 Summary

In this chapter we harness the improved coverage of pathologically-associated lncRNAs in our expanded VSMC annotation to explore the regulation of lncRNAs in the initial phases of SVSMCs responding to cytokine/growth factor stimulus.

1. In keeping with other studies[254], we identify a strong enrichment of lncRNAs amongst genes induced in VSMCs in the initial burst of transcriptional activity after exposure to stimuli. Here we show this occurs within a four hour time frame that precedes induction of cell cycle genes and cell division but contains the largest transcriptional shifts seen in the 24 hour SVSMC timecourse.

2. LncRNAs which were VSMC-enriched or enhancer-transcribed appeared central to this enrichment, suggesting these lncRNAs may be particularly relevant in pushing VSMCs towards a proliferative state. In contrast to cell cycle genes, TF dynamics were co-localised to the initial four hour phase alongside lncRNAs, suggesting an interplay of lncRNA and TFs during initial gene regulation activity.

3. To explore the impact of lncRNAs within the four hour window, we looked at potential for *cis* effects, finding PCG and TFs in the flanks of lncRNAs had a high tendency for differential expression. This was particularly acute for those changing within four hours or those near lncRNAs transcribed from enhancers. Further, lncRNA-PCG neighbours that were both early-response genes also showed a stronger tendency to correlate across all timepoints if they were both differentially expressed in the first four hours.

Together this provides evidence suggesting that early lncRNA activity impacts on later VSMC cell division, in part through regulating expression of neighbouring PCGs, opening the door for further investigation of these pairings to find novel mechanisms that induce VSMC transitions upstream of activation of cell cycle genes.

Chapter 5: Final Discussion

5.1 The impact of this work

5.1.1 Providing a comprehensive annotation and characterisation of lncRNAs associated with pathological VSMC states

Despite a central contributing role in the most ubiquitous and fatal of CVDs, direct modulation of VSMC behaviour is not yet a clinical reality. To find new targets, and help understand our current the actions of any current candidate targets in development, requires thorough examination of the complex mechanisms controlling VSMCs behaviour across different vascular beds and stages of disease progression. In this thesis we contribute to these efforts through providing a timely cataloguing of lncRNA expression in this multifaceted cell type - focusing on lncRNAs expressed within pathological VSMC states. The specific expression profiles of lncRNAs across cell types, patients and temporal windows means reference annotation cannot hope to capture the full intricacies of their dynamics within such systems and focused annotation efforts in specific biological contexts of interest must be undertaken[227,228,231]. This specificity should be recognised and factored into experimental design of lncRNA studies using approaches such as ours.

Previous work has established various lncRNAs as drivers of VSMC-related disease using different *in vitro, in vivo* and *ex vivo* models (section 1.3). However, much remains to be done to provide a thorough transcriptomic overview of their activation within VSMCs, particularly those which are stimulated with typical triggers of pathological activity. This thesis addresses this concern as one of very few studies exploring lncRNA annotation beyond incomplete reference annotation in human VSMCs[160,165], the first to do this in proliferative, pathological VSMCs, and the first in SMCs derived from saphenous veins used in common vein grafting procedures. As the analysed RNA sequencing datasets are not derived from polyA-selected RNA libraries, this is also the first effort to improve coverage of non-polyadenylated lncRNAs in VSMCs. In doing so we substantially widen the scope of candidate lncRNAs driving VSMC pathology. Efforts like ours will spur the next phase of lncRNA studies in which the key examples already discovered justify and drive efforts to better understand gene regulation by the lncRNA class as a whole.

The pipeline we use to expand lncRNA annotation within VSMCs in chapter 3, alongside the application of this annotation to examine early lncRNA dynamics during VSMC response to proliferative stimuli in chapter 4 provides several immediate benefits for the study of VSMC-driven pathology:

1. Candidate lncRNA drivers of VSMC pathology can be selected for study with greater assurance that notable examples are not absent from analysis.

2. LncRNAs with biased expression to VSMCs are identified. They may have particular relevance for VSMC-specific pathways and function but also may feed into regulation of more ubiquitously used gene control mechanisms (as with *SMILR*[174]). They may be useful as therapeutic targets to reduce off-target effects on other cell types.

3. We identify cohorts of lncRNAs with particular traits that have been under-explored so far in the VSMC literature. The lncRNAs revealed through our pipeline showed a higher likelihood of being elncRNAs and VSMC-enriched lncRNAs than other lncRNAs. These lncRNA groups are thus partly obscured in most previous VSMC lncRNA studies yet here demonstrated particular relevance to VSMC pathology through showing a higher tendency than other lncRNAs to be a) stimuli-responsive and b) induced upstream of cell division.

4. Characterisation of gene/transcript structures of lncRNAs in VSMCs provides valuable information to understand their regulation and design future experiments in this cell type. We supported these efforts further through improving annotation of the 5' ends of lncRNA transcripts via integrating with FANTOM CAGEseq data.

We need as comprehensive an understanding of disease systems as possible to hope to modify them effectively. Annotation or cataloguing of gene regulators in different cell types and states such as presented in this thesis is

essential to aid this effort through firstly - and most obviously - providing new candidates for regulatory components that are previously missing from our models of disease systems. Examples of how previously unappreciated gene regulation components can shift our perspective on biological systems are many-fold in the decade or so since discovery of ncRNA function and have been widely documented already in section 1.3 of this thesis in the context of VSMCs. Secondly, annotation efforts such as ours can spur investigation of differing function between cell types. This is particularly important for lncRNAs which have a seemingly high capacity for cell-specific function[147]. Providing a comprehensive expression profile during the study of lncRNAs (and gene regulators more widely) is therefore crucial - aiding the formulation of initial hypotheses and enabling detailed conclusions to be drawn from later experimental characterisation studies *in vivo*. Thirdly, transcriptomic cataloguing efforts can also help reveal previously unappreciated functions for established gene regulators, e.g. through identifying novel isoforms for characterised genes or novel overlapping genomic features. For example, in this thesis we annotate enhancer regions with production of newly-assembled lncRNAs during entry of VSMCs into pathological states. Another example could be the future use of lncRNA annotation to reveal host transcripts which may have a separate (or even opposing) phenotypic effect to their derived miRNA or micropeptide product as demonstrated in recent documented cases such as *CARMN*/miR-143/miR-145[163,164], *H19*/miR-675[195] and *LINC00961*/SPAAR[152]. Expansion of lncRNA annotations is therefore not solely useful as a strategy to find new therapeutic targets, but to also provide new experimental avenues of investigation that can be used to build more detailed models of gene regulation underlying vasculoproliferative pathology.

Successful gene therapy approaches are still rare generally but if (or when) their take-up begins to accelerate, a need to refine and further develop their application will remain alongside trialling of various alterative targets with the joint aim of improving clinical outcomes. Developing a complete understanding of transcriptional changes underlying VSMC pathological transitions is and will be for the foreseeable future a key focus for studies hoping to build gene therapy solutions for vasculoproliferative disease. The more background and context provided for our models of vascular injury and regeneration the more

we can hope to properly direct studies towards robust conclusions and fruitful avenues of investigation.

5.1.2 Exploring novel candidate lncRNA drivers of VSMC pathology

Several approaches are taken in this thesis to highlight lncRNAs of interest to the field either through co-clustering with PCGs, expression profiling or analysing potential *cis* pairings in their local genomic neighbourhood. These cohorts are compiled chiefly to provide easily obtainable candidate lncRNAs and enable studies to build on this work that could take any number of these forward for phenotypic screening and characterisation. Our pipeline excels at finding stimuli-responsive, VSMC-enriched lncRNAs and in chapter 3 we identify 37 VSMC-enriched lncRNAs differentially expressed by either IL-1α/PDGF-BB or pathological stiffness (46% were newly-assembled genes). Of these, 13 (including *SMILR*) were induced between quiescence and IL-1α/PDGF-BB proliferation and detectable in atherosclerotic plaque VSMCs suggesting any identified function related to VSMC pathological transitions would be physiologically relevant. We also reveal for the first time that VSMC-enriched lncRNAs are particularly likely to be stimuli-responsive, as well as present amongst the first genes responding to mitogenic stimuli. Studies building on our work can explore whether they are involved in control of VSMC identity or like *SMILR*, represent cell-specific controls of more core mechanisms like the cell cycle.

Two loci of seemingly key importance to the SVSMC and IL-1α/PDGF-BB model are recurrently highlighted through this thesis, namely the loci upstream of *CXCL8* and *IL6* – two critical genes in VSMC pathology - corresponding to *VSMClnc6* and *AC002480.3* (or the proximal downstream gene *AC002480.4*). *VSMClnc6* is highlighted variously through notably high abundance and rate of induction with IL-1α/PDGF-BB which is apparent within the first four hours of the VSMC response. Further, the VSMC-enrichment of *VSMClnc6* could exceed that of nearly all other lncRNAs and was seen enriched within more VSMC subtypes than nearly all other lncRNAs. The lncRNA was also enhancer-transcribed and could be linked to regulation of the neighbouring

CXCL8 gene through GeneHancer interaction data as well as expression correlation. *AC002480.4* is also highlighted through early induction with IL-1α/PDGF-BB and VSMC-enrichment as well as a pairing to *IL6* via correlation whilst the upstream *AC002480.3* is also paired through GeneHancer interaction data.

Our analysis may help in exploring complex lncRNA-dependent regulatory mechanisms that appear to exist at these two cytokine-regulating loci. Both *AC002480.3* and the lncRNA *UMLILO,* which overlaps *VSMClnc6* on the opposite strand, have been shown through previous experimental manipulation to be involved in epigenetic priming of *IL6* and *CXCL8* - utilising a similar *cis* regulatory mechanism that deposits H3K4me3 activation marks at these loci when examined in endothelial cells or monocytes[275]. In FANTOM expression data, *AC002480.3* and *UMLILO* show no VSMC-enrichment compared to their overlapping VSMC-enriched lncRNAs *AC002480.4* and *VSMClnc6*. *UMLILO* is also absent in all our datasets. This suggests the potential existence of multiple lncRNA control mechanisms in the upstream enhancer locus of both *CXCL8* and *IL6* which may be used preferentially by differing cell types. Studies which explore the potential for *AC002480.4* to aid regulation of *IL6* expression alongside *AC002480.3* or similarly the role of *VSMCln6* in regulation of *CXCL8* expression alongside (or possibly instead of) *UMLILO* would be useful to untangle these potential control sites further. These loci are prime example of the need for studies like ours to tease out differing cell type specific mechanisms of control at key loci and should stimulate further studies into the other candidates highlighted through their potential *cis* regulation of neighbours or VSMC-enrichment.

5.2 Limitations of this work

5.2.1 Completeness of the VSMC lncRNA annotation

The conclusions offered in this thesis must be considered alongside limitations inherent to the sequencing technology used. As we rely on short-read (i.e. <200bp) RNA sequencing to expand VSMC lncRNA annotations, we are likely to have an unavoidable impact on transcript assembly accuracy considering the well described pitfalls in the identification of full transcript structures when using this technology in this way[209]. In particular, obtaining accurate information on splice junctions as well as transcript 5' and 3' limits relies on obtaining a sufficient level of read coverage across all these sites which may be difficult to obtain for lowly expressed isoforms. Additionally, the size of reads in such data means typically only a single exon-exon junction can be covered by a single read which complicates determination of isoform structure at complex loci where differing isoforms may share a proportion of their splice sites.

In chapter 3 we examine transcript completeness within the non-reference lncRNAs as unlike GENCODE lncRNAs, these were entirely derived from short-read sequencing data processed by our pipeline. An initial validation of structures was provided by use of CAGEseq data from FANTOM[136] to provide experimental validation of 5' end for 75% of all newly-assembled lncRNA transcripts. This indicates the majority of these transcript assemblies were largely complete at their 5' end. In terms of gene structure, we observed that 77% of newly-assembled transcripts found across more than one of the selected VSMC datasets are repeatedly assembled with the same intron chain, suggesting that transcript assemblies repeatedly produced for the same lncRNA across multiple datasets are highly consistent. Whilst together this provides reassurance of the promoter regions and exonic structures for our newly-assembed lncRNA structures, obtaining a complete impression of isoform complexity relevant to VSMCs may require use of long-read sequencing data at high depth in specific VSMC models of interest. For instance, application of the capture long-read sequencing technique - where probes are used to amplify read depth at loci of interest - can provide improved annotations at selected lncRNA loci and was used to improve annotation of

the highly complex *CARMN* loci[162]. Notably, use of this technique requires initial direction from studies like ours to provide the full range of lncRNAs of interest to target with probes.

5.2.2 Comprehensiveness of the VSMC lncRNA annotation

As exemplified throughout this thesis, studies relying solely on reference annotation to study lncRNAs in a particular biological context will likely contain a narrowed field of view. Hence, whilst we make our expanded VSMC lncRNA annotation available to be readily applicable into further work, it should not necessarily be used as the reference annotation for studies examining VSMC pathological states. Studies examining lncRNAs in other contexts of VSMC pathology may well be better served through use of our pipeline to reveal new intricacies present in other contexts. This construes another limitation in our study in that although we explore multiple VSMC subtypes in multiple states, complete accounting of lncRNA expression necessitates a level of profiling beyond the scope of this thesis. Indeed, considering the range of mechanical and biochemical stimuli VSMCs are exposed to and the estimated depth of 300 million reads per sample required to approach full identification of all lncRNA genes[226], such an effort cannot be readily achieved using present technological capacities. It would however be of use to do further targeted profiling of VSMC subtypes at increased depths, particularly to tease out the different proliferative capacities of arterial and venous cells (for instance within the IL-1α/PDGF-BB stimulus model[83]) to get perspective on differing lncRNA dynamics across vascular beds. *Ex vivo* models used to show physiological relevance of VSMC lncRNAs[174,178] would also be useful to profile further. Other VSMC contexts which would be interesting to profile with our pipeline and so provide greater coverage of lncRNA activity within vasculoproliferative pathology are highlighted in the manuscript discussion.

We also describe limitations in enhancer annotation via use of generic databases such as GeneHancer[243] and FANTOM CAT[136] rather than bespoke annotation within the VSMC model of interest in the manuscript (section 3.3.4) and additional work discussion (section 3.5.1) sections of chapter 3. An experimental plan to tackle this is discussed in 5.3.1.

5.2.3 Non-coding status of VSMC lncRNA annotation

Factors which affect the specificity of our pipeline to robustly assign coding status to transcripts should also be considered. We incorporate PLAR and the 3 coding prediction tools it utilises as a way of stringently separating genuine lncRNAs from transcripts which could be RNAseq artefacts (i.e. derived from misaligned reads), unannotated PCGs or fragments of annotated PCGs[139,238]. This stringency allows our conclusions to be made on a population of reliably non-coding transcripts but may also exclude certain lncRNAs which demonstrate features of PCGs. This particularly may apply particularly to the minority of conserved lncRNAs[222], as these are more likely to display alignment to PCGs or codon conservation (used by CPC[223] and RNAcode[221] respectively) and/or be derived from ancestral PCGs. The conserved lncRNA *H19* for example is erroneously excluded by PLAR in our hands through alignment to a putative ORF-containing ortholog in a rodent species in reference annotation. Comparison to a gold standard set of lncRNAs could be a potential solution as demonstrated by Chen et. al[222]. However, such a set of lncRNAs may be unreliable or hard to obtain; an ongoing debate in the lncRNA field concerns the extent to which those which are found in the cytoplasm are transcribed to produce micropeptides. None of the tools used within PLAR are optimised to reliably find micropeptides – particularly considering recent micropeptide catalogues which suggest conservation approaches will not identify the full range of such translated molecules[151]. Micropeptide-producing lncRNAs are unlikely misannotated PCGs - their peptide products being biochemically distinct and present at lower levels[153]. Hence, our pipeline stringently identifies non-artefactual, non-PCG transcripts which for ease we refer to as lncRNAs - though an unknown quantity may be capable of micropeptide production.

5.2.4 Functionality within the VSMC lncRNA annotation

In chapters 3 and 4 we use criteria such as abundance, fold change magnitude, early-responsiveness, VSMC-enrichment and potential for *cis* regulation of neighbouring PCGs to highlight candidate drivers of VSMC pathology worthy of investigation. However, these criteria – though valuable for context – cannot be used to suggest the functionality of these lncRNAs. Other criteria that can be used to suggest lncRNAs with greater functional potential do exist but cannot be relied upon to definitively select all functional lncRNAs - likely due to the diversity within the lncRNA class. For instance, conservation of lncRNA sequence, promoter sites, position relative to surrounding genes or any combination of these can be used to highlight lncRNAs under some level of evolutionary constraint – suggesting they are functionally important[138]. However key drivers of VSMC pathology such as *SMILR* and *SENCR* have no ortholog in key animal models such as mouse or pig[160,173], indicating use of this approach selects only a subset of functional lncRNAs. Previous authors have also suggested that lncRNAs with increased specificity of expression may be less likely conserved[138] – an observation which if true would mean exploring lncRNA conservation in this thesis would not be an approach that capitalises properly on the identification of our more specifically expressed, newly-assembled lncRNA cohort.

Another approach to suggest lncRNA functionality is the use of SNP data to link mutations at lncRNA loci to traits of interest (via GWAS) or to expression of neighbouring genes (via eQTL analysis) – two features which correlate with conservation and so can also be used as putative metrics of functional potential[136]. We utilise the second approach indirectly in chapter 3 through use of GeneHancer interaction data to highlight candidate lncRNA genes of interest. For instance, *MSTRG.10933* which contains an eQTL link to GLS within the first exon. Overlap of GWAS data to our expanded VSMC lncRNA annotation could also be fruitful. An issue is the necessity for GWAS approaches to use population level data to identify correlations between SNPs and disease traits or neighbouring gene expression which are apparent above background biological variation in human populations. Recent studies using increased sample size of biological replicates and tissue samples encouragingly reveal 800 lncRNAs explicitly linked to traits using GWAS

data[276]. The application of such tissue-level data here may be better suited to identifying functionality amongst more ubiquitously expressed, highly abundant lncRNAs rather than those which are cell-specific or lowly abundant. Nonetheless advances in GWAS and eQTL datasets should be routinely incorporated into studies focusing on identification of functional lncRNAs and could be fruitful to use in our expanded VSMC annotation.

As our focus here remained on providing a comprehensive accounting of lncRNAs within VSMC pathology we leave further analyses to identify subsets of lncRNAs with functional potential via conservation or SNP methods for future work by others.

5.3 Framework for future experiments

As a proper accounting of transcriptional activity is crucial for lncRNA study, our expanded VSMC lncRNA annotation can be used as a foundation for multiple strands of investigation to identify novel lncRNA contributions to pathology. Experimental plans to explore several of these strands are laid out here. As the earlier phases of SVSMC proliferation have been analysed in greater depth in chapter 4 and can be readily studied within our research group these plans focus on drawing further context for lncRNAs active in this model - aiming to outline those involved in early *cis*-regulation, miRNA or micropeptide production as well as those conserved in a large animal model.

5.3.1 Improving analysis of candidate cis*-acting lncRNAs through collection of bespoke enhancer annotation data*

Newly-assembled lncRNAs we uncover in VSMC pathological states tend towards low abundance and enhancer-transcription – traits representative of *cis*-acting lncRNAs[252]. We also identify strongest evidence for lncRNA-dependent *cis*-regulation in the first four hours of stimulus response in the SVSMC model and particularly focused around elncRNAs. Candidate *cis*-acting lncRNAs active in SVSMCs are identified in chapter 3 using GeneHancer and FANTOM interaction data and in chapter 4 using correlation to associate lncRNAs with neighbouring PCG loci within an earlier timeframe. These initial candidates provide a justification to obtain bespoke enhancer annotation data within the SVSMC model to obtain a more accurately assembled group of candidate *cis*-acting lncRNAs and their putative targeted PCGs.

A more accurate enhancer annotation could aim firstly to improve our method to annotate elncRNAs. We rely in this thesis on enhancer annotation data obtained from a wide range of cell types in the GeneHancer and FANTOM databases yet enhancer activity and elncRNA expression is known to be cell-type specific[136,255]. Obtaining ChIPseq data within the SVSMC timecourse experiment (ideally within the same biological replicates) for H3K27ac and H3K4me1 (as exemplified in Angiotensin-II stimulated VSMCs[188]) would be

sufficient to increase our annotation power and provide a more representative group of elncRNAs, including those which are more specifically expressed in SVSMCs or in response to IL-1α/PDGF-BB. Such data would also allow identifying early-response enhancers as well as super-enhancers – regions of particularly dense enhancer markings tied to cellular identity[277] and enhanced miRNA production[278] (see 5.3.2). We also note in chapter 4 increased evidence of *cis* effects around elncRNAs relative to other lncRNAs within the first four hours of stimuli but cannot compare to enhancers which do not produce lncRNAs. Identifying increased tendency for PCG differential expression at lncRNA-producing enhancers relative to other enhancers would associate lncRNA production with greater enhancer activity – in keeping with previous observations[200]. Together this data would refine (and possibly expand) our list of candidate elncRNA-PCG pairings whilst also highlighting enhancer dynamics in the SVSMC model and testing the hypothesis that an interplay of lncRNA and enhancer function drives VSMC pathology. This would help to justify (or rule out) further experimental characterisation at elncRNA loci of interest.

5.3.2 Exploring the interplay between lncRNA host genes and miRNAs within VSMC pathology

The role of lncRNAs as host genes for miRNAs is relatively unexplored at present. Though many host genes may act simply to produce simple precursor transcripts to be rapidly processed by miRNA biogenesis machinery, others may also have independent function. Key examples include *CARMN* as host of miR-143/145[163,164] and *H19* as host of miR-675[195] – both of which have been described as loci which seemingly produce lncRNA and miRNA as two distinct functional units (referred to hereon as dual functional loci). Investigation of such loci may be a viable strategy to identify functional lncRNAs driving VSMC pathology. The mechanism of action for miRNAs is well established with a high number characterised in VSMCs[279]. Therefore, the characterisation of associated host lncRNA isoforms, or non-host isoforms within the same gene, may help identify mechanisms which support or contrast with the function of the produced miRNA. Interrogation of miRNA host genes requires specific

knowledge of isoform structure within the biological context being probed. For instance, the lncRNA gene *MIR222HG* produces a miRNA hosting isoform and at least two non-miRNA hosting isoforms with the latter two exhibiting distinct miRNA-independent function[280]. Our bespoke VSMC annotation is therefore well suited to the task of identifying host lncRNAs producing either miRNAs, particularly with the use of CAGEseq data to improve accuracy in delineating the various TSSs within host genes.

To implement a study to identify and characterise candidate dual functional lncRNA-miRNA loci, we would use available small RNAseq data obtained from the same samples (again by Drs A.Mahmoud and M.Ballantyne) as were used to produce total RNAseq data of the early phases of IL-1α/PDGF-BB-induced SVSMC proliferation analysed in chapter 4. We therefore would have a small RNA library paired with a total RNA library for each sample at each timepoint, allowing identification of early-responsive miRNAs and their requisite lncRNA hosts whilst omitting any variation from using samples collected from different biological replicates from separate batches.

To identify lncRNA-miRNA loci of interest, we would firstly identify all loci where either a miRNA-hosting lncRNA or overlapping miRNA are differentially expressed within the initial 24 hours of IL-1α/PDGF-BB response. An initial strategy to then identify dual functional lncRNA-miRNA loci would focus on loci with both miRNA overlapping (host) isoforms and non-miRNA overlapping (non-host) isoforms. In these loci we would perform an analysis to highlight where host isoforms show signs of being regulated separately to non-miRNA host isoforms. If we observe host gene and miRNA expression negatively correlating over the 24 hours this provides initial indication that host isoforms and non-host isoforms are separately regulated at such loci providing a cohort of miRNA loci with potential for independent lncRNA host gene function. Notably, preliminary data shows downregulation of *CARMN* over 24 hours IL-1α/PDGF-BB exposure alongside stable expression of miR-143/miR-145, suggesting dual-function loci may also show no correlation within this time frame. If few or no negatively correlating lncRNA-miRNA examples are found we could examine all loci which do not have positive correlation to gain a wider candidate pool.

Future experimental framework

The above method aims to find lncRNA-miRNA loci where lncRNAs and miRNAs are separately regulated to identify potential dual functionality. However, lncRNA-miRNA loci with joint regulation (and so similar expression dynamics) could also be dual-functional loci (e.g. the miRs and lncRNAs at the *MIR222HG* locus originate from the same TSS[280]). These would likely positively correlate in a similar fashion to lncRNA-miRNA loci where lncRNA hosts have no independent function. This makes them hard to identify using integration of small and total RNAseq datasets. To include such loci, we could instead use an alternative method to highlight candidate lncRNA-miRNA loci – focusing on their potential importance to VSMC regulation rather than separate regulatory patterns. For instance, several lncRNA-miRNA loci appear to be super-enhancer based[163,281] - regions linked to control of cell identity[277], increased recruitment of miRNA processing enzymes and increased miRNA abundance[278]. Identifying miRNA-independent function for lncRNAs at such super-enhancer loci would therefore be of particular interest. The strategy laid out in section 5.3.1 could be used to obtain bespoke enhancer annotation to highlight these - focusing on newly-annotated super-enhancer regions which are found to be dynamically regulated during the first 24 hours of IL-1α/PDGF-BB exposure in the SVSMC model. Preliminary data suggests the lncRNA-miRNA loci *CARMN*/miR-143/145, *MIR3142HG*/miR-146a and *MIR155HG*/miR-155, *MIR222HG*/miR-221/miR-222 would be amongst those highlighted by such methodology within the SVSMC model - based on their differential expression or prior super-enhancer annotation available in the literature[164,188,281].

Once a pool of candidate dual functional lncRNA-miRNA loci are obtained, those loci of interest for further experimentation could be selected based on criteria which suggest the importance of the miRNA to the SVSMC model. These would include the availability of literature for the miRNA in VSMC pathology, the number of predicted mRNA targets which negatively correlate with the miRNA within the timecourse data and also the relative abundance of the miRNA within the overall miRNA pool. Experimental characterisation would focus on isolating distinct phenotypic effects from miRNA and lncRNA on IL-1α/PDGF-BB-induced SVSMC proliferation. An initial step would be to validate

the miRNA expression and use knockdown and overexpression to demonstrate a phenotypic influence within the model.

The experimental design to characterise an independent function for any candidate lncRNA host gene will be highly dependent on isoform complexity at the locus. Nascent transcript sequencing techniques could be utilised to properly determine TSS usage at the lncRNA-miRNA locus without interference from miRNA biogenesis and degradation of host isoforms[282]. These techniques could also be coupled to long-read sequencing to provide a more definitive exonic structures and quantifications of host and non-host isoforms at such loci[283]. If host isoforms are definitively shown to have a distinct exonic sequence or TSS to non-host isoforms then it may be possible to selectively target one or the other group of isoforms to examine potential for differing phenotypic effects. Otherwise, knockdown of the entire host locus through either ASO (if a common exon can be found between all isoforms) or CRISPR/Cas9 approaches may be necessary to look for any effect on IL-1α/PDGF-BB-induced SVSMC proliferation. After establishing that host knockdown leads to loss of miRNA expression, reintroduction of the miRNA using mimics can then help ascertain distinct phenotypic effects and establish or disprove dual functionality as demonstrated for *CARMN*[164] and *H19*[195]. This experimental plan could link together key classes of genetic regulators (lncRNAs, miRNAs and potentially enhancers) and is a promising route to obtain novel regulatory insights into important loci driving VSMC pathology.

5.3.3 Identifying lncRNA drivers of VSMC pathology applicable to a large animal model of cardiovascular disease

Identification of lncRNA orthologs in a closely related animal model resembling human disease is a viable route to select candidates that can be tested *in vivo* and so have particular translational potential. Pig models of atherosclerosis have closer resemblance to human plaque pathology than mouse[284] and so pig models are invaluable tool to study VSMC dynamics *in vivo*. Despite this they are so far underused for this purpose, with no published work yet attempting large-scale identification of human-pig conserved lncRNA drivers of VSMC pathology. Non-human species generally have particularly

incomplete lncRNA annotation that typically hampers ready identification of specific lncRNA orthologs but in pig we would have access to a recently published extensive transcript annotation dataset which combines high depth short-read and long-read sequencing methods across tissues[285]. This could be used as a foundational reference transcriptome to use in conjunction with our pipeline to create a bespoke pig VSMC annotation, adding in novel transcripts from RNAseq data of pig VSMCs treated with the analogous IL-1α/PDGF-BB treatment. This would provide a mirror to the SVSMC annotation created in chapter 3 of this thesis. Implementation of conservation comparisons could be done using the tool *slncky*[222], which allows ready identification of lncRNA orthologs and their categorisation by sequence or positional conservation. Positionally-conserved orthologs with lack of sequence homology are important to include in such a study as they can be used to infer more likely *cis*-acting lncRNAs. In these cases, conservation of the transcript sequence and structure (so *trans* binding capacity) is less important than conservation of the act of transcription at this site (for instance, leading to a change in chromatin conformation or blockage of a downstream TSS)[138]**.** This data could therefore also be useful to aid investigation of candidate *cis*-acting lncRNAs identified throughout this thesis.

References

1. Chamley JH, Campbell GR, Burnstock G. Dedifferentiation, redifferentiation and bundle formation of smooth muscle cells in tissue culture : the influence of cell number and nerve fibres. *J Embryol Exp Morph*. **1974**;32:297–323.

2. Alexander MR, Owens GK. Epigenetic Control of Smooth Muscle Cell Differentiation and Phenotypic Switching in Vascular Development and Disease. *Annu Rev Physiol*. **2012**;74:13–40.

3. Frismantiene A, Philippova M, Erne P, Resink TJ. Smooth muscle cell-driven vascular diseases and molecular mechanisms of VSMC plasticity. *Cell Signal*. **2018**;52:48–64.

4. Frösen J, Joutel A. Smooth muscle cells of intracranial vessels: From development to disease. *Cardiovasc Res*. **2018**;114:501–512.

5. Basatemur GL, Jørgensen HF, Clarke MCH, Bennett MR, Mallat Z. Vascular smooth muscle cells in atherosclerosis. Nat. Rev. Cardiol. **2019**;16:727–744.

6. Braun-Dullaeus RC, Mann MJ, Dzau VJ. Cell cycle progression: New therapeutic target for vascular proliferative disease. *Circulation*. **1998**;98:82–89.

7. Yahagi K, Kolodgie FD, Otsuka F, Finn A V., Davis HR, Joner M, Virmani R. Pathophysiology of native coronary, vein graft, and in-stent atherosclerosis. Nat. Rev. Cardiol. **2016**;13:79–98.

8. BHF. BHF CVD Statistics UK Factsheet. https://www.bhf.org.uk/what-we-do/our-research/heart-statistics[cited 2021 Sep 30]

9. de Vries MR, Simons KH, Jukema JW, Braun J, Quax PHA. Vein graft failure: from pathophysiology to clinical outcomes. *Nat Rev Cardiol*. **2016**;13:451–470.

10. Wang G, Jacquet L, Karamariti E, Xu Q. Origin and differentiation of vascular smooth muscle cells. *J Physiol*. **2015**;593:3013–3030.

11. Hirschi KK, Rohovsky SA, D'Amore PA. PDGF, TGF-β, and heterotypic cell-cell interactions mediate endothelial cell-induced recruitment of 10T1/2 cells and their differentiation to a smooth muscle fate. *J Cell Biol*. **1998**;141:805–814.

12. Yamashita J, Itoh H, Hirashima M, Ogawa M, Nishikawa S, Yurugi T, Naito M, Nakao K, Nishikawa SI. Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors. *Nature*. **2000**;408:92–96.

13. Yao Q, Renault M-A, Chapouly C, Vandierdonck S, Belloc I, Jaspard-Vinassa B, Daniel-Lamazì Ere J-M, Laffargue M, Merched A, Desgranges C, Gadeau A-P. Sonic hedgehog mediates a novel pathway of PDGF-BB–dependent vessel maturation. *Blood*. **2014**;123:2429–2437.

14. Scheppke L, Murphy EA, Zarpellon A, Hofmann JJ, Merkulova A, Shields DJ, Weis SM, Byzova T V., Ruggeri ZM, Iruela-Arispe ML, Cheresh DA. Notch promotes vascular maturation by inducing integrin-mediated smooth muscle cell adhesion to the endothelial basement membrane. *Blood*. **2012**;119:2149–2158.

15. Granata A, Bernard WG, Zhao N, Mccafferty J, Lilly B, Sinha S. Temporal and Embryonic Lineage-Dependent Regulation of Human Vascular SMC Development by NOTCH3. *Stem Cells Dev*. **2015**;24:846–856.

16. Cheung C, Bernardo AS, B Trotter MW, Pedersen RA, Sinha S. Generation of human vascular smooth muscle subtypes provides insight into embryological origin-dependent disease susceptibility. *Nat Biotechnol*. **2012**;

17. Dobnikar L, Taylor AL, Chappell J, Oldach P, Harman JL, Oerton E, Dzierzak E, Bennett MR, Spivakov M, Jorgensen HF. Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. *Nat Commun*. **2018**;9.

18. Trigueros-Motos L, Gonzalez-Granado JM, Cheung C, Fernandez P, Sanchez-Cabo F, Dopazo A, Sinha S,

Andrés V. Embryological-origin-dependent differences in homeobox expression in adult aorta: Role in regional phenotypic variability and regulation of NF-kB activity. *Arterioscler Thromb Vasc Biol*. **2013**;33:1248–1256.

19. Seidelmann SB, Lighthouse JK, Greif DM. Development and pathologies of the arterial wall. Cell. Mol. Life Sci. **2014**;71:1977–1999.

20. Greif DM, Kumar M, Lighthouse JK, Hum J, An A, Ding L, Red-Horse K, Espinoza FH, Olson L, Offermanns S, Krasnow MA. Radial Construction of an Arterial Wall. *Dev Cell*. **2012**;23:482–493.

21. Opitz F, Schenke-Layland K, Cohnert TU, Stock UA. Phenotypical plasticity of vascular smooth muscle cells - Effect of in vitro and in vivo shear stress for tissue engineering of blood vessels. *Tissue Eng*. **2007**;13:2505–2514.

22. Ichii T, Koyama H, Tanaka S, Kim S, Shioi A, Okuno Y, Raines EW, Iwao H, Otani S, Nishizawa Y. Fibrillar collagen specifically regulates human vascular smooth muscle cell genes involved in cellular responses and the pericellular matrix environment. *Circ Res*. **2001**;88:460–467.

23. Kelleher CM, McLean SE, Mecham RP. Vascular Extracellular Matrix and Aortic Development. *Curr Top Dev Biol*. **2004**;62:153–188.

24. WHO. Cardiovascular diseases (CVDs) Fact Sheet. http://www.who.int/mediacentre/factsheets/fs317/en/[cited 2021 Sep 30]

25. Nakashima Y, Chen YX, Kinukawa N, Sueishi K. Distributions of diffuse intimal thickening in human arteries: Preferential expression in atherosclerosis-prone arteries from an early age. *Virchows Arch*. **2002**;441:279–288.

26. Orekhov AN, R. Andreeva E, Mikhailova IA, Gordon D. Cell proliferation in normal and atherosclerotic human aorta: Proliferative splash in lipid-rich lesions. *Atherosclerosis*. **1998**;139:41–48.

27. DamiaÂn-Zamacona S, Toledo-Ibelles P, Ibarra-Abundis M, Uribe-Figueroa L, HernaÂndez-Lemus E, Macedo-Alcibia KP, Delgado-Coello B, Mas-Oliva J, Reyes-Grajeda JP. Early transcriptomic response to ldl and oxldl in human vascular smooth muscle cells. *PLoS One*. **2016**;11.

28. Tabas I, Williams KJ, Borén J. Subendothelial lipoprotein retention as the initiating process in atherosclerosis: Update and therapeutic implications. Circulation. **2007**;116:1832–1844.

29. Clarke MCH, Talib S, Figg NL, Bennett MR. Vascular smooth muscle cell apoptosis induces interleukin-1-directed inflammation: Effects of hyperlipidemia-mediated inhibition of phagocytosis. *Circ Res*. **2010**;106:363–372.

30. Cushing SD, Berliner JA, Valente AJ, Territo MC, Navab M, Parhami F, Gerrity R, Schwartz CJ, Fogelman AM. Minimally modified low density lipoprotein induces monocyte chemotactic protein 1 in human endothelial cells and smooth muscle cells. *Proc Natl Acad Sci U S A*. **1990**;87:5134–5138.

31. Laguna-Fernández A, Novella S, Bueno-Betí C, Marrugat J, Hermenegildo C. Endothelial transcriptomic changes induced by oxidized low density lipoprotein disclose an up-regulation of Jak-Stat pathway. *Vascul Pharmacol*. **2015**;73:104–114.

32. Nakagawa K, Nakashima Y. Pathologic intimal thickening in human atherosclerosis is formed by extracellular accumulation of plasma-derived lipids and dispersion of intimal smooth muscle cells. *Atherosclerosis*. **2018**;274:235–242.

33. Napoli C, D'Armiento FP, Mancini FP, Postiglione A, Witztum JL, Palumbo G, Palinski W. Fatty streak formation occurs in human fetal aortas and is greatly enhanced maternal, hypercholesterolemia. Intimal accumulation of low density lipoprotein and its oxidation precede monocyte recruitment into early atheroeclerotic lesions. *J Clin Invest*. **1997**;100:2680–2690.

34. Skålén K, Gustafsson M, Knutsen Rydberg E, Hultén LM, Wiklund O, Innerarity TL, Boren J. Subendothelial retention of atherogenic lipoproteins in early atherosclerosis. *Nature*. **2002**;417:750–754.

35. Sano H, Sudo T, Yokode M, Murayama T, Kataoka H, Takakura N, Nishikawa S, Nishikawa S-I, Kita T. Functional Blockade of Platelet-Derived Growth Factor Receptor-β but Not of Receptor-α Prevents Vascular Smooth Muscle Cell Accumulation in Fibrous Cap Lesions in Apolipoprotein E–Deficient Mice. *Circulation*. **2001**;103:2955–2960.

36. He C, Medley SC, Hu T, Hinsdale ME, Lupu F, Virmani R, Olson LE. PDGFRβ signalling regulates local inflammation and synergizes with hypercholesterolaemia to promote atherosclerosis. *Nat Commun*. **2015**;6:7770.

37. Ross R, Masuda J, Raines EW, Gown AM, Katsuda S, Sasahara M, Malden LT, Masuko H, Sato H. Localization of PDGF-B protein in macrophages in all phases of atherogenesis. *Science (80-)*. **1990**;248:1009–1012.

38. Mehrhof FB, Schmidt-Ullrich R, Dietz R, Scheidereit C. Regulation of vascular smooth muscle cell proliferation: Role of NF-κB revisited. *Circ Res*. **2005**;96:958–964.

39. Ikeda U, Ikeda M, Oohara T, Oguchi A, Kamitani T, Tsuruya Y, Kano S. Interleukin 6 stimulates growth of vascular smooth muscle cells in a PDGF-dependent manner. *Am J Physiol - Hear Circ Physiol*. **1991**;260:1713–1717.

40. Selzman CH, Miller SA, Zimmerman MA, Gamboni-Robertson F, Harken AH, Banerjee A. Monocyte chemotactic protein-1 directly induces human vascular smooth muscle proliferation. *Am J Physiol Circ Physiol*. **2002**;283:1455–1461.

41. Qin Y, Fan F, Zhao Y, Cui Y, Wei X, Kohama K, Gordon JR, Li F, Gao Y. Recombinant human CXCL8(3-72)K11R/G31P regulates smooth muscle cell proliferation and migration through blockage of interleukin-8 receptor. *IUBMB Life*. **2013**;65:67–75.

42. Libby P, Warner SJC, Friedman GB. Interleukin 1: A mitogen for human vascular smooth muscle cells that induces the release of growth-inhibitory prostanoids. *J Clin Invest*. **1988**;81:487–498.

43. Yu H, Clarke MCH, Figg N, Littlewood TD, Bennett MR. Smooth muscle cell apoptosis promotes vessel remodeling and repair via activation of cell migration, proliferation, and collagen synthesis. *Arterioscler Thromb Vasc Biol*. **2011**;31:2402–2409.

44. Wang Z, Newman WH. Smooth muscle cell migration stimulated by interleukin 6 is associated with cytoskeletal reorganization. *J Surg Res*. **2003**;111:261–266.

45. Nathe TJ, Deou J, Walsh B, Bourns B, Clowes AW, Daum G. Interleukin-1β inhibits expression of p21(WAF1/CIP1) and p27(KIP1) and enhances proliferation in response to platelet-derived growth factor-BB in smooth muscle cells. *Arterioscler Thromb Vasc Biol*. **2002**;22:1293–1298.

46. Chen CN, Li YSJ, Yeh YT, Lee PL, Usami S, Chien S, Chiu JJ. Synergistic roles of platelet-derived growth factor-BB and interleukin-1β in phenotypic modulation of human aortic smooth muscle cells. *Proc Natl Acad Sci U S A*. **2006**;103:2665–2670.

47. Jiang B, Xu S, Brecher P, Cohen RA. Growth factors enhance interleukin-1β-induced persistent activation of nuclear factor-κB in rat vascular smooth muscle cells. *Arterioscler Thromb Vasc Biol*. **2002**;22:1811–1816.

48. Fabunmi RP, Baker AH, Murray EJ, Booth RFG, Newby AC. Divergent regulation by growth factors and cytokines of 95 kDa and 72 kDa gelatinases and tissue inhibitors of metalloproteinases-1, -2 and -3 in rabbit aortic smooth muscle cells. *Biochem J*. **1996**;315:335–342.

49. Bond M, Fabunmi RP, Baker AH, Newby AC. Synergistic upregulation of metalloproteinase-9 by growth factors and inflammatory cytokines: An absolute requirement for transcription factor NF-κB. *FEBS Lett*. **1998**;435:29–34.

50. Benditt EP, Benditt JM. Evidence for a monoclonal origin of human atherosclerotic plaques. *Proc Natl Acad Sci U S A*. **1973**;70:1753–1756.

51. Murry CE, Gipaya CT, Bartosek T, Benditt EP, Schwartz SM. Monoclonality of smooth muscle cells in human

atherosclerosis. *Am J Pathol*. **1997**;151:697–706.

52. Misra A, Feng Z, Chandran RR, Kabir I, Rotllan N, Aryal B, Sheikh AQ, Ding L, Qin L, Fernández-Hernando C, Tellides G, Greif DM. Integrin beta3 regulates clonality and fate of smooth muscle-derived atherosclerotic plaque cells. *Nat Commun*. **2018**;9.

53. Feil S, Fehrenbacher B, Lukowski R, Essmann F, Schulze-Osthoff K, Schaller M, Feil R. Transdifferentiation of vascular smooth muscle cells to macrophage-like cells during atherogenesis. *Circ Res*. **2014**;115:662–667.

54. Shankman LS, Gomez D, Cherepanova OA, Salmon M, Alencar GF, Haskins RM, Swiatlowska P, Newman AAC, Greene ES, Straub AC, Isakson B, Randolph GJ, Owens GK. KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nat Med*. **2015**;21:628–637.

55. Rong JX, Shapiro M, Trogan E, Fisher EA. Transdifferentiation of mouse aortic smooth muscle cells to a macrophage-like state after cholesterol loading. *Proc Natl Acad Sci U S A*. **2003**;100:13531–13536.

56. Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, Kundu R, Nagao M, Coller J, Koyano TK, Fong R, Woo YJ, Liu B, Montgomery SB, Wu JC, Zhu K, Chang R, Alamprese M, Tallquist MD, Kim JB, Quertermous T. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med*. **2019**;25:1280–1289.

57. Cherepanova OA, Gomez D, Shankman LS, Swiatlowska P, Williams J, Sarmento OF, Alencar GF, Hess DL, Bevard MH, Greene ES, Murgai M, Turner SD, Geng Y-J, Bekiranov S, Connelly JJ, Tomilin A, Owens GK. Activation of the pluripotency factor OCT4 in smooth muscle cells is atheroprotective. *Nat Med*. **2016**;22:657–665.

58. Hu Y, Zhang Z, Torsney E, Afzal AR, Davison F, Metzler B, Xu Q. Abundant progenitor cells in the adventitia contribute to atherosclerosis of vein grafts in ApoE-deficient mice. *J Clin Invest*. **2004**;113:1258–1265.

59. Passman JN, Dong XR, Wu SP, Maguire CT, Hogan KA, Bautch VL, Majesky MW. A sonic hedgehog signaling domain in the arterial adventitia supports resident Sca1+ smooth muscle progenitor cells. *Proc Natl Acad Sci U S A*. **2008**;105:9349–9354.

60. Kramann R, Goettsch C, Wongboonsin J, Iwata H, Schneider RK, Kuppe C, Kaesler N, Chang-Panesso M, Machado FG, Gratwohl S, Madhurima K, Hutcheson JD, Jain S, Aikawa E, Humphreys BD. Adventitial MSC-like Cells Are Progenitors of Vascular Smooth Muscle Cells and Drive Vascular Calcification in Chronic Kidney Disease. *Cell Stem Cell*. **2016**;19:628–642.

61. Tang J, Wang H, Huang X, Li F, Zhu H, Li Y, He L, Zhang H, Pu W, Liu K, Zhao H, Bentzon JF, Yu Y, Ji Y, Nie Y, Tian X, Zhang L, Gao D, Zhou B. Arterial Sca1+ Vascular Stem Cells Generate De Novo Smooth Muscle for Artery Repair and Regeneration. *Cell Stem Cell*. **2020**;26:81–96.

62. Carrel T, Winkler B. Current trends in selection of conduits for coronary artery bypass grafting. *Gen Thorac Cardiovasc Surg*. **2017**;65:549–556.

63. Goldman S, Zadina K, Moritz T, Ovitt T, Sethi G, Copeland JG, Thottapurathu L, Krasnicka B, Ellis N, Anderson RJ, Henderson W. Long-term patency of saphenous vein and left internal mammary artery grafts after coronary artery bypass surgery: Results from a Department of Veterans Affairs Cooperative Study. *J Am Coll Cardiol*. **2004**;44:2149–2156.

64. Samano N, Geijer H, Liden M, Fremes S, Bodin L, Souza D. The no-touch saphenous vein for coronary artery bypass grafting maintains a patency, after 16 years, comparable to the left internal thoracic artery: A randomized trial. *J Thorac Cardiovasc Surg*. **2015**;150:880–888.

65. Wan S, George SJ, Berry C, Baker AH. Vein graft failure: Current clinical practice and potential for gene therapeutics. *Gene Ther*. **2012**;19:630–636.

66. Alexander JH, Hafley G, Harrington RA, Peterson ED, Ferguson TB, Lorenz TJ, Goyal A, Gibson M, Mack MJ, Gennevois D, Califf RM, Kouchoukos NT. Efficacy and safety of edifoligide, an E2F transcription factor decoy, for prevention of vein graft failure following coronary artery bypass graft surgery: PREVENT IV: A

randomized controlled trial. *J Am Med Assoc*. **2005**;294:2446–2454.

67. Friedl R, Li J, Schumacher B, Hanke H, Waltenberger J, Hannekum A, Stracke S. Intimal hyperplasia and expression of transforming growth factor-β1 in saphenous veins and internal mammary arteries before coronary artery surgery. *Ann Thorac Surg*. **2004**;78:1312–1318.

68. Owens CD, Wake N, Jacot JG, Gerhard-Herman M, Gaccione P, Belkin M, Creager MA, Conte MS. Early biomechanical changes in lower extremity vein grafts-distinct temporal phases of remodeling and wall stiffness. *J Vasc Surg*. **2006**;44:740–746.

69. Gasper WJ, Owens CD, Kim JM, Hills N, Belkin M, Creager MA, Conte MS. Thirty-day vein remodeling is predictive of midterm graft patency after lower extremity bypass. *J Vasc Surg*. **2013**;57:9–18.

70. Osgood MJ, Hocking KM, Voskresensky I V, Li FD, Komalavilas P, Cheung-Flynn J, Brophy CM. Surgical vein graft preparation promotes cellular dysfunction, oxidative stress, and intimal hyperplasia in human saphenous vein. *J Vasc Surg*. **2014**;60:202–211.

71. Khaleel MS, Dorheim TA, Duryee MJ, Durbin HE, Bussey WD, Garvin RP, Klassen LW, Thiele GM, Anderson DR. High-pressure distention of the saphenous vein during preparation results in increased markers of inflammation: A potential mechanism for graft failure. *Ann Thorac Surg*. **2012**;93:552–558.

72. Stigler R, Steger C, Schachner T, Holfeld J, Edlinger M, Grimm M, Semsroth S. The impact of distension pressure on acute endothelial cell loss and neointimal proliferation in saphenous vein grafts. *Eur J Cardio-thoracic Surg*. **2012**;42.

73. Dreifaldt M, Souza D, Bodin L, Shi-Wen X, Dooley A, Muddle J, Loesch A, Dashwood MR. The vasa vasorum and associated endothelial nitric oxide synthase is more important for saphenous vein than arterial bypass grafts. *Angiology*. **2013**;64:293–299.

74. Tseng CN, Karlöf E, Chang YT, Lengquist M, Rotzius P, Berggren PO, Hedin U, Eriksson EE. Contribution of endothelial injury and inflammation in early phase to vein graft failure: The causal factors impact on the development of intimal hyperplasia in murine models. *PLoS One*. **2014**;9:98904.

75. Westerband A, Crouse D, Richter LC, Aguirre ML, Wixon CC, James DC, Mills JL, Hunter GC, Heimark RL. Vein adaptation to arterialization in an experimental model. *J Vasc Surg*. **2001**;33:561–569.

76. Borin TF, Miyakawa AA, Cardoso L, De Figueiredo Borges L, Gonçalves GA, Krieger JE. Apoptosis, cell proliferation and modulation of cyclin-dependent kinase inhibitor p21cip1 in vascular remodelling during vein arterialization in the rat. *Int J Exp Pathol*. **2009**;90:328–337.

77. Cheng J, Wang Y, Ma Y, Chan BTY, Yang M, Liang A, Zhang L, Li H, Du J. The mechanical stress-activated serum-, glucocorticoid-regulated kinase 1 contributes to neointima formation in vein grafts. *Circ Res*. **2010**;107:1265–1274.

78. Conte MS, Owens CD, Belkin M, Creager MA, Edwards KL, Gasper WJ, Kenagy RD, Leboeuf RC, Sobel M, Clowes A. A single nucleotide polymorphism in the p27Kip1 gene is associated with primary patency of lower extremity vein bypass grafts. *J Vasc Surg*. **2013**;57:1179–1185.

79. Lau GT, Ridley LJ, Bannon PG, Wong LA, Trieu J, Brieger DB, Lowe HC, Freedman BS, Kritharides L. Lumen loss in the first year in saphenous vein grafts is predominantly a result of negative remodeling of the whole vessel rather than a result of changes in wall thickness. *Circulation*. **2006**;114:I-435-I–440.

80. Westerband A, Mills JL, Marek JM, Heimark RL, Hunter GC, Williams SK, Clowes AW, Westerband A, Stanley JC, Davies AH, Goldman MH, Sterpetti A, Hoch JR. Immunocytochemical determination of cell type and proliferation rate in human vein graft stenoses. *J Vasc Surg*. **1997**;25:64–73.

81. Owens CD, Rybicki FJ, Wake N, Schanzer A, Mitsouras D, Gerhard-Herman MD, Conte MS. Early remodeling of lower extremity vein grafts: Inflammation influences biomechanical adaptation. *J Vasc Surg*. **2008**;47:1235–1242.

82. Dardik A, Yamashita A, Aziz F, Asada H, Sumpio BE. Shear stress-stimulated endothelial cells induce

smooth muscle cell chemotaxis via platelet-derived growth factor-BB and interleukin-1α. *J Vasc Surg*. **2005**;41:321–331.

83. Turner NA, Ho S, Warburton P, O'Regan DJ, Porter KE. Smooth muscle cells cultured from human saphenous vein exhibit increased proliferation, invasion, and mitogen-activated protein kinase activation in vitro compared with paired internal mammary artery cells. *J Vasc Surg*. **2007**;45:1022–1028.

84. Frischknecht K, Greutert H, Weisshaupt C, Kaspar M, Yang Z, Lüscher TF, Carrel TP, Tanner FC. Different vascular smooth muscle cell apoptosis in the human internal mammary artery and the saphenous vein: Implications for bypass graft disease. *J Vasc Res*. **2006**;43:338–346.

85. Wu W, Wang C, Zang H, Qi L, Azhar M, Nagarkatti M, Nagarkatti P, Cai G, Weiser-Evans MCM, Cui T. Mature Vascular Smooth Muscle Cells, but Not Endothelial Cells, Serve as the Major Cellular Source of Intimal Hyperplasia in Vein Grafts. *Arterioscler Thromb Vasc Biol*. **2020**;40:1870–1890.

86. Cooley BC, Nevado J, Mellad J, Yang D, St. Hilaire C, Negro A, Fang F, Chen G, San H, Walts AD, Schwartzbeck RL, Taylor B, Lanzer JD, Wragg A, Elagha A, Beltran LE, Berry C, Feil R, Virmani R, Ladich E, Kovacic JC, Boehm M. TGF-β signaling mediates endothelial-to-mesenchymal transition (EndMT) during vein graft remodeling. *Sci Transl Med*. **2014**;6:227ra34.

87. Yazdani SK, Farb A, Nakano M, Vorpahl M, Ladich E, Finn A V., Kolodgie FD, Virmani R. Pathology of drug-eluting versus bare-metal stents in saphenous vein bypass graft lesions. *JACC Cardiovasc Interv*. **2012**;5:666–674.

88. Finn A V., Joner M, Nakazawa G, Kolodgie F, Newell J, John MC, Gold HK, Virmani R. Pathological correlates of late drug-eluting stent thrombosis: Strut coverage as a marker of endothelialization. *Circulation*. **2007**;115:2435–2441.

89. Kuramitsu S, Sonoda S, Ando K, Otake H, Natsuaki M, Anai R, Honda Y, Kadota K, Kobayashi Y, Kimura T. Drug-eluting stent thrombosis: current and future perspectives. *Cardiovasc Interv Ther*. **2021**;1:3.

90. Kandzari DE, Koolen JJ, Doros G, Garcia-Garcia HM, Bennett J, Roguin A, Gharib EG, Cutlip DE, Waksman R. Ultrathin Bioresorbable-Polymer Sirolimus-Eluting Stents Versus Thin Durable-Polymer Everolimus-Eluting Stents for Coronary Revascularization: 3-Year Outcomes From the Randomized BIOFLOW V Trial. *JACC Cardiovasc Interv*. **2020**;13:1343–1353.

91. Palmerini T, Benedetto U, Biondi-Zoccai G, Della Riva D, Bacchi-Reggiani L, Smits PC, Vlachojannis GJ, Jensen LO, Christiansen EH, Berencsi K, Valgimigli M, Orlandi C, Petrou M, Rapezzi C, Stone GW. Long-term safety of drug-eluting and bare-metal stents: Evidence from a comprehensive network meta-analysis. *J Am Coll Cardiol*. **2015**;65:2496–2507.

92. Yla-Herttuala S, Baker AH. Cardiovascular Gene Therapy: Past, Present, and Future. *Mol Ther*. **2017**;25:1095–1106.

93. Lan N, Massam B, Kulkarni S, Lang C. Pulmonary Arterial Hypertension: Pathophysiology and Treatment. *Diseases*. **2018**;6:38.

94. Stenmark KR, Meyrick B, Galie N, Mooi WJ, Mcmurtry IF. Animal models of pulmonary arterial hypertension: the hope for etiological discovery and pharmacological cure. *Am J Physiol Lung Cell Mol Physiol*. **2009**;297:1013–1032.

95. Humbert M, Guignabert C, Bonnet S, Dorfmüller P, Klinger JR, Nicolls MR, Olschewski AJ, Pullamsetti SS, Schermuly RT, Stenmark KR, Rabinovitch M. Pathology and pathobiology of pulmonary hypertension: state of the art and research perspectives. *Eur Respir J*. **2019**;53.

96. Wohrley JD, Frid MG, Moiseeva EP, Orton EC, Belknap JK, Stenmark KR. Hypoxia selectively induces proliferation in a specific subpopulation of smooth muscle cells in the bovine neonatal pulmonary arterial media. *J Clin Invest*. **1995**;96:273–281.

97. Frid MG, Moiseeva EP, Stenmark KR. Multiple phenotypically distinct smooth muscle cell populations exist in the adult and developing bovine pulmonary arterial media in vivo. *Circ Res*. **1994**;75:669–681.

98. Stenmark KR, Frid MG, Graham BB, Tuder RM. Dynamic and diverse changes in the functional properties of vascular smooth muscle cells in pulmonary hypertension. *Cardiovasc Res*. **2018**;114:551–564.

99. Taraseviciene-Stewart L, Kasahara Y, Alger L, Hirth P, Mahon GM, Waltenberger J, Voelkel NF, Tuder RM. Inhibition of the VEGF receptor 2 combined with chronic hypoxia causes cell death-dependent pulmonary endothelial cell proliferation and severe pulmonary hypertension. *FASEB J*. **2001**;15:427–438.

100. Majka SM, Skokan M, Wheeler L, Harral J, Gladson S, Burnham E, Loyd JE, Stenmark KR, Varella-Garcia M, West J. Evidence for cell fusion is absent in vascular lesions associated with pulmonary arterial hypertension. *Am J Physiol - Lung Cell Mol Physiol*. **2008**;295:L1028–L1039.

101. Rich S, Pogoriler J, Husain AN, Toth PT, Gomberg-Maitland M, Archer SL. Long-term effects of epoprostenol on the pulmonary vasculature in idiopathic pulmonary arterial hypertension. *Chest*. **2010**;138:1234–1239.

102. Burke DL, Frid MG, Kunrath CL, Karoor V, Anwar A, Wagner BD, Strassheim D, Stenmark KR. Sustained hypoxia promotes the development of a pulmonary artery-specific chronic inflammatory microenvironment. *Am J Physiol Cell Mol Physiol*. **2009**;297:L238–L250.

103. Savai R, Al-Tamari HM, Sedding D, Kojonazarov B, Muecke C, Teske R, Capecchi MR, Weissmann N, Grimminger F, Seeger W, Schermuly RT, Pullamsetti SS. Pro-proliferative and inflammatory signaling converge on FoxO1 transcription factor in pulmonary hypertension. *Nat Med*. **2014**;20:1289–1300.

104. Tuder RM, Davis LA, Graham BB. Targeting energetic metabolism: A new frontier in the pathogenesis and treatment of pulmonary hypertension. *Am J Respir Crit Care Med*. **2012**;185:260–266.

105. Bertero T, Oldham WM, Cottrill KA, Pisano S, Vanderpool RR, Yu Q, Zhao J, Tai Y, Tang Y, Zhang YY, Rehman S, Sugahara M, Qi Z, Gorcsan J, Vargas SO, Saggar R, Saggar R, Wallace WD, Ross DJ, Haley KJ, Waxman AB, Parikh VN, De Marco T, Hsue PY, Morris A, Simon MA, Norris KA, Gaggioli C, Loscalzo J, Fessel J, Chan SY. Vascular stiffness mechanoactivates YAP/TAZ-dependent glutaminolysis to drive pulmonary hypertension. *J Clin Invest*. **2016**;126:3313–3335.

106. Shi N, Chen SY. Mechanisms simultaneously regulate smooth muscle proliferation and differentiation. *J Biomed Res*. **2014**;28:40–46.

107. Hungerford J, Little C. Developmental Biology of the Vascular Smooth Muscle Cell: Building a Multilayered Vessel Wall. *J Vasc Res*. **1999**;36:2–27.

108. Wang Z, Wang DZ, Pipes GCT, Olson EN. Myocardin is a master regulator of smooth muscle gene expression. *Proc Natl Acad Sci U S A*. **2003**;100:7129–7134.

109. Wang DZ, Chang PS, Wang Z, Sutherland L, Richardson JA, Small E, Krieg PA, Olson EN. Activation of cardiac gene expression by myocardin, a transcriptional cofactor for serum response factor. *Cell*. **2001**;105:851–862.

110. Horita H, Wysoczynski CL, Walker LA, Moulton KS, Li M, Ostriker A, Tucker R, McKinsey TA, Churchill MEA, Nemenoff RA, Weiser-Evans MCM. Nuclear PTEN functions as an essential regulator of SRF-dependent transcription to control smooth muscle differentiation. *Nat Commun*. **2016**;7:1–17.

111. Mack CP, Thompson MM, Lawrenz-Smith S, Owens GK. Smooth Muscle α-Actin CArG Elements Coordinate Formation of a Smooth Muscle Cell–Selective, Serum Response Factor–Containing Activation Complex. *Circ Res*. **2000**;86:221–232.

112. McDonald OG, Wamhoff BR, Hoofnagle MH, Owens GK. Control of SRF binding to CArG box chromatin regulates smooth muscle gene expression in vivo. *J Clin Invest*. **2006**;116:36–48.

113. Cao D, Wang Z, Zhang C-L, Oh J, Xing W, Li S, Richardson JA, Wang D-Z, Olson EN. Modulation of Smooth Muscle Gene Expression by Association of Histone Acetyltransferases and Deacetylases with Myocardin. *Mol Cell Biol*. **2005**;25:364–376.

114. Yao F, Yu P, Li Y, Yuan X, Li Z, Zhang T, Liu F, Wang Y, Wang Y, Li D, Ma B, Shu C, Kong W, Zhou B, Wang L. Histone Variant H2A.Z Is Required for the Maintenance of Smooth Muscle Cell Identity as Revealed

by Single-Cell Transcriptomics. *Circulation*. **2018**;138:2274–2288.

115. Qiu P, Ritchie RP, Fu Z, Cao D, Cumming J, Miano JM, Wang DZ, Li HJ, Li L. Myocardin enhances Smad3-mediated transforming growth factor- β1 signaling in a CArG box-independent manner: Smad-binding element is an important cis element for SM22α transcription in vivo. *Circ Res*. **2005**;97:983–991.

116. Miller CL, Pjanic M, Wang T, Nguyen T, Cohain A, Lee JD, Perisic L, Hedin U, Kundu RK, Majmudar D, Kim JB, Wang O, Betsholtz C, Ruusalepp A, Franzen O, Assimes TL, Montgomery SB, Schadt EE, Bjorkegren JLM, Quertermous T. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat Commun*. **2016**;7.

117. Liu R, Jin Y, Tang WH, Qin L, Zhang X, Tellides G, Hwa J, Yu J, Martin KA. Ten-eleven translocation-2 (TET2) is a master regulator of smooth muscle cell plasticity. *Circulation*. **2013**;128:2047–2057.

118. Boettger T, Beetz N, Kostin S, Schneider J, Krüger M, Hein L, Braun T. Acquisition of the contractile phenotype by murine arterial smooth muscle cells depends on the Mir143/145 gene cluster. *J Clin Invest*. **2009**;119:2634–2647.

119. Xin M, Small EM, Sutherland LB, Qi X, McAnally J, Plato CF, Richardson JA, Bassel-Duby R, Olson EN. MicroRNAs miR-143 and miR-145 modulate cytoskeletal dynamics and responsiveness of smooth muscle cells to injury. *Genes Dev*. **2009**;23:2166–2178.

120. Regan CP, Adam PJ, Madsen CS, Owens GK. Molecular mechanisms of decreased smooth muscle differentiation marker expression after vascular injury. *J Clin Invest*. **2000**;106:1139–1147.

121. Wang Z, Wang DZ, Hockemeyer D, McAnally J, Nordheim A, Olson EN. Myocardin and ternary complex factors compete for SRF to control smooth muscle gene expression. *Nature*. **2004**;428:185–189.

122. Yoshida T, Gan Q, Owens GK. Krüppel-like factor 4, Elk-1, and histone deacetylases cooperatively suppress smooth muscle cell differentiation markers in response to oxidized phospholipids. *Am J Physiol - Cell Physiol*. **2008**;295:1175–1182.

123. Liu Y, Sinha S, McDonald OG, Shang Y, Hoofnagle MH, Owens GK. Kruppel-like factor 4 abrogates myocardin-induced activation of smooth muscle gene expression. *J Biol Chem*. **2005**;280:9719–9727.

124. Zheng J-P, He X, Liu F, Yin S, Wu S, Yang M, Zhao J, Dai X, Jiang H, Yu L, Yin Q, Ju D, Li C, Lipovich L, Xie Y, Zhang K, Li HJ, Zhou J, Li L, Yang Y. YY1 directly interacts with myocardin to repress the triad myocardin/SRF/CArG box-mediated smooth muscle gene transcription during smooth muscle phenotypic modulation. **2020**;10:21781.

125. Nagao M, Lyu Q, Zhao Q, Wirka RC, Bagga J, Nguyen T, Cheng P, Kim JB, Pjanic M, Miano JM, Quertermous T. Coronary disease-associated gene TCF21 inhibits smooth muscle cell differentiation by blocking the myocardin-serum response factor pathway. *Circ Res*. **2020**;126:517–529.

126. Cordes KR, Sheehy NT, White MP, Berry EC, Morton SU, Muth AN, Lee TH, Miano JM, Ivey KN, Srivastava D. MiR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature*. **2009**;460:705–710.

127. Marais R, Wynne J, Treisman R. The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. *Cell*. **1993**;73:381–393.

128. Hirschi KK, Lai L, Belaguli NS, Dean DA, Schwartz RJ, Zimmer WE. Transforming growth factor-β induction of smooth muscle cell phenotype requires transcriptional and post-transcriptional control of serum response factor. *J Biol Chem*. **2002**;277:6287–6295.

129. Horita HN, Simpson PA, Ostriker A, Furgeson S, Van Putten V, Weiser-Evans MCM, Nemenoff RA. Serum response factor regulates expression of phosphatase and tensin homolog through a MicroRNA network in vascular smooth muscle cells. *Arterioscler Thromb Vasc Biol*. **2011**;31:2909–2919.

130. Tandon P, Miteva Y V., Kuchenbrod LM, Cristea IM, Conlon FL. Tcf21 regulates the specification and maturation of proepicardial cells. *Dev*. **2012**;140:2409–2421.

131. Elia L, Kunderfranco P, Carullo P, Vacchiano M, Farina FM, Hall IF, Mantero S, Panico C, Papait R,

Condorelli G, Quintavalle M. UHRF1 epigenetically orchestrates smooth muscle cell plasticity in arterial disease. *J Clin Invest*. **2018**;128:2473–2486.

132. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue CH, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu YB, Ruan XA, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan YJ, Wold B, Carninci P, Guigo R, Gingeras TR. Landscape of transcription in human cells. *Nature*. **2012**;489:101–108.

133. ENCODE Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. **2012**;489:57–74.

134. Palazzo AF, Lee ES. Non-coding RNA: what is functional and what is junk? *Front Genet*. **2015**;6.

135. Rands CM, Meader S, Ponting CP, Lunter G. 8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage. *PLoS Genet*. **2014**;10:e1004525.

136. Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJL, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J, Lizio M, Kawaji H, Kasukawa T, Itoh M, Burroughs AM, Noma S, Djebali S, Alam T, Medvedeva YA, Testa AC, Lipovich L, Yip CW, Abugessaisa I, Mendez M, Hasegawa A, Tang D, Lassmann T, Heutink P, Babina M, Wells CA, Kojima S, Nakamura Y, Suzuki H, Daub CO, de Hoon MJL, Arner E, Hayashizaki Y, Carninci P, Forrest ARR. An atlas of human long non-coding RNAs with accurate 5 ' ends. *Nature*. **2017**;543:199–204.

137. Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, Sisu C, Wright JC, Armstrong J, Barnes I, Berry A, Bignell A, Boix C, Carbonell Sala S, Cunningham F, Di Domenico T, Donaldson S, Fiddes IT, García Girón C, Gonzalez JM, Grego T, Hardy M, Hourlier T, Howe KL, Hunt T, Izuogu OG, Johnson R, Martin FJ, Martínez L, Mohanan S, Muir P, Navarro FCP, Parker A, Pei B, Pozo F, Riera FC, Ruffier M, Schmitt BM, Stapleton E, Suner MM, Sycheva I, Uszczynska-Ratajczak B, Wolf MY, Xu J, Yang YT, Yates A, Zerbino D, Zhang Y, Choudhary JS, Gerstein M, Guigó R, Hubbard TJP, Kellis M, Paten B, Tress ML, Flicek P. GENCODE 2021. *Nucleic Acids Res*. **2021**;49:D916–D923.

138. Ulitsky I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet*. **2016**;17:601–614.

139. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Rep*. **2015**;11:1110–1122.

140. Penny GD, Kay GF, Sheardown SA, Rastan S, Neil Brockdorfft B. Requirement for Xist in X chromosome inactivation. *Nature*. **1996**;379:131–137.

141. Leighton PA, Ingram RS, Eggenschwiler J, Efstratiadis A, Tilghman SM. Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature*. **1995**;375:34–39.

142. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell*. **2007**;129:1311–1323.

143. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. **2010**;142:409–419.

144. Congrains A, Kamide K, Oguro R, Yasuda O, Miyata K, Yamamoto E, Kawai T, Kusunoki H, Yamamoto H, Takeya Y, Yamamoto K, Onishi M, Sugimoto K, Katsuya T, Awata N, Ikebe K, Gondo Y, Oike Y, Ohishi M, Rakugi H. Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and

CDKN2A/B. *Atherosclerosis*. **2012**;220:449–455.

145. Congrains A, Kamide K, Katsuya T, Yasuda O, Oguro R, Yamamoto K, Ohishi M, Rakugi H. CVD-associated non-coding RNA, ANRIL, modulates expression of atherogenic pathways in VSMC. *Biochem Biophys Res Commun*. **2012**;419:612–616.

146. Ramilowski JA, Yip CW, Agrawal S, Chang J-C, Ciani Y, Kulakovskiy I V, Mendez M, Li J, Ooi C, Ouyang JF, Parkinson N, Petri A, Roos L, Severin J, Yasuzawa K, Abugessaisa I, Akalin A, Antonov I V, Arner E, Bonetti A, Bono H, Borsari B, Brombacher F, Cameron CJF, Cannistraci CV, Cardenas R, Cardon M, Chang H, Dostie J, Ducoli L, Favorov A, Fort A, Garrido D, Gil N, Gimenez J, Guler R, Handoko L, Harshbarger J, Hasegawa A, Hasegawa Y, Hashimoto K, Hayatsu N, Heutink P, Hirose T, Imada EL, Itoh M, Kaczkowski B, Kanhere A, Kawabata E, Kawaji H, Kawashima T, Kelly ST, Kojima M, Kondo N, Koseki H, Kouno T, Kratz A, Kurowska-Stolarska M, Tae A, Kwon J, Leek J, Lennartsson A, Lizio M, López-Redondo F, Luginbühl J, Maeda S, Makeev VJ, Marchionni L, Medvedeva YA, Minoda A, Müller F, Muñoz-Aguirre M, Murata M, Nishiyori H, Nitta KR, Noguchi S, Noro Y, Nurtdinov R, Okazaki Y, Orlando V, Paquette D, Parr CJC, Rackham OJL, Rizzu P, Fernando Sánchez Martinez D, Sandelin A, Sanjana P, Semple CAM, Shibayama Y, Sivaraman DM, Suzuki T, Szumowski SC, Tagami M, Taylor MS, Terao C, Thodberg M, Thongjuea S, Tripathi V, et al. Functional annotation of human long noncoding RNAs via molecular phenotyping. *Genome Res*. **2020**;30:1060–1072.

147. Liu SJ, Horlbeck MA, Cho SW, Birk HS, Malatesta M, He D, Attenello FJ, Villalta JE, Cho MY, Chen Y, Mandegar MA, Olvera MP, Gilbert LA, Conklin BR, Chang HY, Weissman JS, Lim DA. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science (80-)*. **2017**;355.

148. Volders PJ, Anckaert J, Verheggen K, Nuytens J, Martens L, Mestdagh P, Vandesompele J. Lncipedia 5: Towards a reference set of human long non-coding rnas. *Nucleic Acids Res*. **2019**;47:D135–D139.

149. Monteiro JP, Bennett M, Rodor J, Caudrillier A, Ulitsky I, Baker AH. Endothelial function and dysfunction in the cardiovascular system: the long non-coding road. *Cardiovasc Res*. **2019**;115:1692–1704.

150. Makarewich CA, Olson EN. Mining for Micropeptides. Trends Cell Biol. **2017**;27:685–696.

151. van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, Kirchner M, Maatz H, Blachut S, Sandmann CL, Kanda M, Worth CL, Schafer S, Calviello L, Merriott R, Patone G, Hummel O, Wyler E, Obermayer B, Mücke MB, Lindberg EL, Trnka F, Memczak S, Schilling M, Felkin LE, Barton PJR, Quaife NM, Vanezis K, Diecke S, Mukai M, Mah N, Oh SJ, Kurtz A, Schramm C, Schwinge D, Sebode M, Harakalova M, Asselbergs FW, Vink A, de Weger RA, Viswanathan S, Widjaja AA, Gärtner-Rommel A, Milting H, dos Remedios C, Knosalla C, Mertins P, Landthaler M, Vingron M, Linke WA, Seidman JG, Seidman CE, Rajewsky N, Ohler U, Cook SA, Hubner N. The Translational Landscape of the Human Heart. *Cell*. **2019**;178:242-260.e29.

152. Spencer HL, Sanders R, Boulberdaa M, Meloni M, Cochrane A, Spiroski A-M, Mountford J, Emanueli C, Caporali A, Brittan M, Rodor J, Baker AH. The LINC00961 transcript and its encoded micropeptide, small regulatory polypeptide of amino acid response, regulate endothelial cell function. *Cardiovasc Res*. **2020**;116:1981–1994.

153. Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MAS, Brocard M, Couso JP. Extensive translation of small open reading frames revealed by poly-ribo-seq. *Elife*. **2014**;3:1–19.

154. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. **2011**;25:1915–1927.

155. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A*. **2008**;105:716–721.

156. Kurian L, Aguirre A, Sancho-Martinez I, Benner C, Hishida T, Nguyen TB, Reddy P, Nivet E, Krause MN, Nelles DA, Esteban CR, Campistol JM, Yeo GW, Belmonte JCI. Identification of Novel Long Noncoding RNAs Underlying Vertebrate Cardiovascular Development. *Circulation*. **2015**;131:1278–1290.

157. Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Perkins AC, Mattick JS. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res*. **2008**;18:1433–1445.

158. Breschi A, Muñoz-Aguirre M, Wucher V, Davis CA, Garrido-Martín D, Djebali S, Gillis J, Pervouchine DD, Vlasova A, Dobin A, Zaleski C, Drenkow J, Danyko C, Scavelli A, Reverter F, Snyder MP, Gingeras TR, Guigó R. A limited set of transcriptional programs define major cell types. *Genome Res*. **2020**;30:1047–1059.

159. Viereck J, Bührke A, Foinquinos A, Chatterjee S, Kleeberger JA, Xiao K, Janssen-Peters H, Batkai S, Ramanujam D, Kraft T, Cebotari S, Gueler F, Beyer AM, Schmitz J, Bräsen JH, Schmitto JD, Gyöngyösi M, Löser A, Hirt MN, Eschenhagen T, Engelhardt S, Bär C, Thum T. Targeting muscle-enriched long non-coding RNA H19 reverses pathological cardiac hypertrophy. *Eur Heart J*. **2020**;41:3462–3474.

160. Bell RD, Long XC, Lin MY, Bergmann JH, Nanda V, Cowan SL, Zhou Q, Han Y, Spector DL, Zheng DY, Miano JM. Identification and Initial Functional Characterization of a Human Vascular Cell-Enriched Long Noncoding RNA. *Arterioscler Thromb Vasc Biol*. **2014**;34:1249–1259.

161. He X, Lian Z, Yang Y, Wang Z, Fu X, Liu Y, Li M, Tian J, Yu T, Xin H. Long Non-coding RNA PEBP1P2 Suppresses Proliferative VSMCs Phenotypic Switching and Proliferation in Atherosclerosis. *Mol Ther - Nucleic Acids*. **2020**;22:84–98.

162. Lagarde J, Uszczynska-Ratajczak B, Carbonell S, Pérez-Lluch S, Abad A, Davis C, Gingeras TR, Frankish A, Harrow J, Guigo R, Johnson R. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat Genet*. **2017**;49:1731–1740.

163. Dong K, Shen J, He X, Wang L, Hu G, Bunting KM, Dixon-Melvin R, Zheng Z, Xin H, Xiang M, Vazdarjanova A, Zhou J. Discovery of an evolutionarily conserved smooth muscle cell-specific lncRNA CARMN. *bioRxiv*. **2020**;2020.01.30.927335.

164. Vacante F, Rodor J, Lalwani MK, Mahmoud AD, Bennett M, De Pace A, Miller E, van Kuijk K, de Bruijn JB, Gijbels M, Williams TC, Clark MB, Scanlon JP, Doran AC, Montgomery R, Newby DE, Giacca M, O'Carroll D, Hadoke PW, Denby L, Sluimer JC, Baker AH. CARMN Loss Regulates Smooth Muscle Cells and Accelerates Atherosclerosis in Mice. *Circ Res*. **2021**;128:1258–1275.

165. Zhao JJ, Zhang W, Lin MY, Wu W, Jiang PT, Tou E, Xue M, Richards A, Jourd'heuil D, Asif A, Zheng DY, Singer HA, Miano JM, Long XC. MYOSLID Is a Novel Serum Response Factor-Dependent Long Noncoding RNA That Amplifies the Vascular Smooth Muscle Differentiation Program. *Arterioscler Thromb Vasc Biol*. **2016**;36:2088–2099.

166. Wang S, Zhang X, Yuan Y, Tan M, Zhang L, Xue X, Yan Y, Han L, Xu Z. BRG1 expression is increased in thoracic aortic aneurysms and regulates proliferation and apoptosis of vascular smooth muscle cells through the long non-coding RNA HIF1A-AS1 in vitro. *Eur J Cardio-Thoracic Surg*. **2015**;47:439–446.

167. Zhou J, Zhang M, Fang H, El-Mounayri O, Rodenberg JM, Imbalzano AN, Herring BP. The SWI/SNF chromatin remodeling complex regulates myocardin-induced smooth muscle-specific gene expression. *Arterioscler Thromb Vasc Biol*. **2009**;29:921–928.

168. Broadbent HM, Peden JF, Lorkowski S, Goel A, Ongen H, Green F, Clarke R, Collins R, Franzosi MG, Tognoni G, Seedorf U, Rust S, Eriksson P, Hamsten A, Farrall M, Watkins H, Consortium P. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Hum Mol Genet*. **2008**;17:806–814.

169. Zhang C, Ge S, Gong W, Xu J, Guo Z, Liu Z, Gao X, Wei X, Ge S. LncRNA ANRIL acts as a modular scaffold of WDR5 and HDAC3 complexes and promotes alteration of the vascular smooth muscle cell phenotype. *Cell Death Dis*. **2020**;11.

170. Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K, Finstermeier K, Stahringer A, Wilfert W, Beutner F, Gielen S, Schuler G, Gäbel G, Bergert H, Bechmann I, Stadler PF, Thiery J, Teupser D. Alu Elements in ANRIL Non-Coding RNA at Chromosome 9p21 Modulate Atherogenic Cell Functions through Trans-Regulation of Gene Networks. *PLoS Genet*. **2013**;9:1003588.

171. Broadbent HM, Peden JF, Lorkowski S, Goel A, Ongen H, Green F, Clarke R, Collins R, Franzosi MG, Tognoni G, Seedorf U, Rust S, Eriksson P, Hamsten A, Farrall M, Watkins H. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Hum Mol Genet*. **2008**;17:806–814.

172. Holdt LM, Stahringer A, Sass K, Pichler G, Kulak NA, Wilfert W, Kohlmaier A, Herbst A, Northoff BH, Nicolaou A, Gäbel G, Beutner F, Scholz M, Thiery J, Musunuru K, Krohn K, Mann M, Teupser D. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun*. **2016**;7.

173. Ballantyne MD, Pinel K, Dakin R, Vesey AT, Diver L, Mackenzie R, Garcia R, Welsh P, Sattar N, Hamilton G, Joshi N, Dweck MR, Miano JM, McBride MW, Newby DE, McDonald RA, Baker AH. Smooth Muscle Enriched Long Noncoding RNA (SMILR) Regulates Cell Proliferation. *Circulation*. **2016**;133:2050–2065.

174. Mahmoud AD, Ballantyne MD, Miscianinov V, Pinel K, Hung J, Scanlon JP, Iyinikkel J, Kaczynski J, Tavares AS, Bradshaw AC, Mills NL, Newby DE, Caporali A, Gould GW, George SJ, Ulitsky I, Sluimer JC, Rodor J, Baker AH. The Human-Specific and Smooth Muscle Cell-Enriched LncRNA SMILR Promotes Proliferation by Regulating Mitotic CENPF mRNA and Drives Cell-Cycle Progression Which Can Be Targeted to Limit Vascular Remodeling. *Circ Res*. **2019**;125:535–551.

175. Wu G, Cai J, Han Y, Chen J, Huang ZP, Chen C, Cai Y, Huang H, Yang Y, Liu Y, Xu Z, He D, Zhang X, Hu X, Pinello L, Zhong D, He F, Yuan GC, Wang DZ, Zeng C. LincRNA-p21 regulates neointima formation, vascular smooth muscle cell proliferation, apoptosis, and atherosclerosis by enhancing p53 activity. *Circulation*. **2014**;130:1452–1465.

176. Tang R, Mei X, Wang YC, Cui XB, Zhang G, Li W, Chen SY. LncRNA GAS5 regulates vascular smooth muscle cell cycle arrest and apoptosis via p53 pathway. *Biochim Biophys Acta - Mol Basis Dis*. **2019**;1865:2516–2525.

177. Wang YNZ, Shan K, Yao M Di, Yao J, Wang JJ, Li X, Liu B, Zhang YY, Ji Y, Jiang Q, Yan B. Long noncoding RNA-GAS5. *Hypertension*. **2016**;68:736–748.

178. Zehendner CM, Valasarajan C, Werner A, Boeckel JN, Bischoff FC, John D, Weirick T, Glaser SF, Rossbach O, Jae N, Demolli S, Khassafi F, Yuan K, De Jesus Perez VA, Michalik KM, Chen W, Seeger W, Guenther A, Wasnick RM, Uchida S, Zeiher AM, Dimmeler S, Pullamsetti SS. Long noncoding RNA TYKRIL plays a role in pulmonary hypertension via the p53-mediated regulation of PDGFRb. *Am J Respir Crit Care Med*. **2020**;202:1445–1457.

179. Tang R, Zhang G, Wang YC, Mei X, Chen SY. The long non-coding RNA GAS5 regulates transforming growth factor β (TGF-β)-induced smooth muscle cell differentiation via RNA Smad-binding elements. *J Biol Chem*. **2017**;292:14270–14278.

180. Li K, Cui M, Zhang K, Wang G, Zhai S. LncRNA CRNDE affects the proliferation and apoptosis of vascular smooth muscle cells in abdominal aortic aneurysms by regulating the expression of Smad3 by Bcl-3. *Cell Cycle*. **2020**;19:1036–1047.

181. Zhou Y, He X, Liu R, Qin Y, Wang S, Yao X, Li C, Hu Z. LncRNA CRNDE regulates the proliferation and migration of vascular smooth muscle cells. *J Cell Physiol*. **2019**;234:16205–16214.

182. Lino Cardenas CL, Kessinger CW, Cheng Y, MacDonald C, MacGillivray T, Ghoshhajra B, Huleihel L, Nuri S, Yeri AS, Jaffer FA, Kaminski N, Ellinor P, Weintraub NL, Malhotra R, Isselbacher EM, Lindsay ME. An HDAC9-MALAT1-BRG1 complex mediates smooth muscle dysfunction in thoracic aortic aneurysm. *Nat Commun*. **2018**;9.

183. Yu CK, Xu T, Assoian RK, Rader DJ. Mining the Stiffness-Sensitive Transcriptome in Human Vascular Smooth Muscle Cells Identifies Long Noncoding RNA Stiffness Regulators. *Arterioscler Thromb Vasc Biol*. **2018**;38:164–173.

184. Brock M, Schuoler C, Leuenberger C, Bühlmann C, Haider TJ, Vogel J, Ulrich S, Gassmann M, Kohler M, Huber LC. Analysis of hypoxia-induced noncoding RNAs reveals metastasis-associated lung adenocarcinoma transcript 1 as an important regulator of vascular smooth muscle cell proliferation. *Exp Biol*

Med. **2017**;242:487–496.

185. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, König IR, Cazier JB, Johansson Å, Hall AS, Lee JY, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lyytikäinen LP, Rafelt S, Shungin D, Strawbridge RJ, Thorleifsson G, Tikkanen E, Van Zuydam N, Voight BF, Waite LL, Zhang W, Ziegler A, Absher D, Altshuler D, Balmforth AJ, Barroso I, Braund PS, Burgdorf C, Claudi-Boehm S, Cox D, Dimitriou M, Do R, Doney ASF, El Mokhtari NE, Eriksson P, Fischer K, Fontanillas P, Franco-Cereceda A, Gigante B, Groop L, Gustafsson S, Hager J, Hallmans G, Han BG, Hunt SE, Kang HM, Illig T, Kessler T, Knowles JW, Kolovou G, Kuusisto J, Langenberg C, Langford C, Leander K, Lokki ML, Lundmark A, McCarthy MI, Meisinger C, Melander O, Mihailov E, Maouche S, Morris AD, Müller-Nurasyid M, Nikus K, Peden JF, Rayner NW, Rasheed A, Rosinger S, Rubin D, Rumpf MP, Schäfer A, Sivananthan M, Song C, Stewart AFR, Tan ST, Thorgeirsson G, Van Der Schoot CE, Wagner PJ, Wells GA, Wild PS, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. **2013**;45:25–33.

186. Ahmed ASI, Dong K, Liu J, Wen T, Yu L, Xu F, Kang X, Osman I, Hu G, Bunting KM, Crethers D, Gao H, Zhang W, Liu Y, Wen K, Agarwal G, Hirose T, Nakagawa S, Vazdarjanova A, Zhou J. Long noncoding RNA NEAT1 (nuclear paraspeckle assembly transcript 1) is critical for phenotypic switching of vascular smooth muscle cells. *Proc Natl Acad Sci U S A*. **2018**;115:E8660–E8667.

187. Leung A, Trac C, Jin W, Lanting L, Akbany A, Sætrom P, Schones DE, Natarajan R. Novel long noncoding RNAs are regulated by angiotensin II in vascular smooth muscle cells. *Circ Res*. **2013**;113:266–278.

188. Das S, Senapati P, Chen Z, Reddy MA, Ganguly R, Lanting L, Mandi V, Bansal A, Leung A, Zhang S, Jia Y, Wu X, Schones DE, Natarajan R. Regulation of angiotensin II actions by enhancers and super-enhancers in vascular smooth muscle cells. *Nat Commun*. **2017**;8:1–19.

189. Das S, Zhang E, Senapati P, Amaram V, Reddy MA, Stapleton K, Leung A, Lanting L, Wang M, Chen Z, Kato M, Oh HJ, Guo Q, Zhang X, Zhang B, Zhang H, Zhao Q, Wang W, Wu Y, Natarajan R. A novel angiotensin II-induced long noncoding RNA giver regulates oxidative stress, inflammation, and proliferation in vascular smooth muscle cells. *Circ Res*. **2018**;123:1298–1312.

190. Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drabløs F, Lennartsson A, Rönnerblad M, Hrydziuszko O, Vitezic M, Freeman TC, Alhendi AMN, Arner P, Axton R, Baillie JK, Beckhouse A, Bodega B, Briggs J, Brombacher F, Davis M, Detmar M, Ehrlund A, Endoh M, Eslami A, Fagiolini M, Fairbairn L, Faulkner GJ, Ferrai C, Fisher ME, Forrester L, Goldowitz D, Guler R, Ha T, Hara M, Herlyn M, Ikawa T, Kai C, Kawamoto H, Khachigian LM, Klinken SP, Kojima S, Koseki H, Klein S, Mejhert N, Miyaguchi K, Mizuno Y, Morimoto M, Morris KJ, Mummery C, Nakachi Y, Ogishima S, Okada-Hatakeyama M, Okazaki Y, Orlando V, Ovchinnikov D, Passier R, Patrikakis M, Pombo A, Qin XY, Roy S, Sato H, Savvi S, Saxena A, Schwegmann A, Sugiyama D, Swoboda R, Tanaka H, Tomoiu A, Winteringham LN, Wolvetang E, Yanagi-Mizuochi C, Yoneda M, Zabierowski S, Zhang P, Abugessaisa I, Bertin N, Diehl AD, Fukuda S, Furuno M, Harshbarger J, Hasegawa A, Hori F, Ishikawa-Kato S, Ishizu Y, Itoh M, Kawashima T, Kojima M, Kondo N, Lizio M, Meehan TF, Mungall CJ, Murata M, Nishiyori-Sueki H, Sahin S, Nagao-Sato S, Severin J, De Hoon MJL, Kawai J, et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science (80-)*. **2015**;347:1010–1014.

191. Liu X, Cheng Y, Zhang S, Lin Y, Yang J, Zhang C. A Necessary role of miR-221 and miR-222 in vascular smooth muscle cell proliferation and neointimal hyperplasia. *Circ Res*. **2009**;104:476–486.

192. Sun C, Fu Y, Gu X, Xi X, Peng X, Wang C, Sun Q, Wang X, Qian F, Qin Z, Qu W, Piao M, Zhong S, Liu S, Zhang M, Fang S, Tian J, Li C, Maegdefessel L, Tian J, Yu B. Macrophage-Enriched lncRNA RAPIA: A Novel Therapeutic Target for Atherosclerosis. *Arterioscler Thromb Vasc Biol*. **2020**;40:1464–1478.

193. Lv J, Wang L, Zhang J, Lin R, Wang L, Sun W, Wu H, Xin S. Long noncoding RNA H19-derived miR-675 aggravates restenosis by targeting PTEN. *Biochem Biophys Res Commun*. **2018**;497:1154–1161.

194. Voellenkle C, Garcia-Manteiga JM, Pedrotti S, Perfetti A, De Toma I, Da Silva D, Maimone B, Greco S,

Fasanaro P, Creo P, Zaccagnini G, Gaetano C, Martelli F. Implication of Long noncoding RNAs in the endothelial cell response to hypoxia revealed by RNA-sequencing. *Sci Rep*. **2016**;6.

195. Li DY, Busch A, Jin H, Chernogubova E, Pelisek J, Karlsson J, Sennblad B, Liu S, Lao S, Hofmann P, Bäcklund A, Eken SM, Roy J, Eriksson P, Dacken B, Ramanujam D, Dueck A, Engelhardt S, Boon RA, Eckstein HH, Spin JM, Tsao PS, Maegdefessel L. H19 induces abdominal aortic aneurysm development and progression. *Circulation*. **2018**;138:1551–1568.

196. Latos PA, Pauler FM, Koerner M V., Şenergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, Aumayr K, Pasierbek P, Barlow DP. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science (80-)*. **2012**;338:1469–1472.

197. Bassett AR, Akhtar A, Barlow DP, Bird AP, Brockdorff N, Duboule D, Ephrussi A, Ferguson-Smith AC, Gingeras TR, Haerty W, Higgs DR, Miska EA, Ponting CP. Considerations when investigating lncRNA function in vivo. *Elife*. **2014**;3:1–14.

198. Gao F, Cai Y, Kapranov P, Xu D. Reverse-genetics studies of lncRNAs-what we have learnt and paths forward. *Genome Biol*. **2020**;21:1–23.

199. Lyu Q, Xu SW, Lyu YY, Choi M, Christie CK, Slivano OJ, Rahman A, Jin ZG, Long XC, Xu YW, Miano JM. SENCR stabilizes vascular endothelial cell adherens junctions through interaction with CKAP4. *Proc Natl Acad Sci U S A*. **2019**;116:546–555.

200. Gil N, Ulitsky I. Production of Spliced Long Noncoding RNAs Specifies Regions with Increased Enhancer Activity. *Cell Syst*. **2018**;7:537–547.

201. Leung A, Trac C, Jin W, Lanting L, Akbany A, Saetrom P, Schones DE, Natarajan R. Novel Long Noncoding RNAs Are Regulated by Angiotensin II in Vascular Smooth Muscle Cells. *Circ Res*. **2013**;113:266–278.

202. Uszczynska-Ratajczak B, Lagarde J, Frankish A, Guigó R, Johnson R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat Rev Genet*. **2018**;19:535–548.

203. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Manuel Rodriguez J, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigo R, Hubbard TJ. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res*. **2012**;22:1760–1774.

204. Liu SJ, Nowakowski TJ, Pollen AA, Lui JH, Horlbeck MA, Attenello FJ, He D, Weissman JS, Kriegstein AR, Diaz AA, Lim DA. Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol*. **2016**;17.

205. Seiler J, Breinig M, Caudron-Herger M, Polycarpou-Schwarz M, Boutros M, Diederichs S. The lncRNA VELUCT strongly regulates viability of lung cancer cells despite its extremely low abundance. *Nucleic Acids Res*. **2017**;45:5458–5469.

206. Chang TC, Pertea M, Lee S, Salzberg SL, Mendell JT. Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. *Genome Res*. **2015**;25:1401–1409.

207. Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol*. **2011**;12:R16.

208. Hayer KE, Pizarro A, Lahens NF, Hogenesch JB, Grant GR. Benchmark analysis of algorithms for determining and quantifying full-length mRNA splice forms from RNA-seq data. *Bioinformatics*. **2015**;31:3938–3945.

209. Steijger T, Abril JF, Engström PG, Kokocinski F, Akerman M, Alioto T, Ambrosini G, Antonarakis SE, Behr J, Bertone P, Bohnert R, Bucher P, Cloonan N, Derrien T, Djebali S, Du J, Dudoit S, Gerstein M, Gingeras TR, Gonzalez D, Grimmond SM, Guigó R, Habegger L, Harrow J, Hubbard TJ, Iseli C, Jean G, Kahles A, Lagarde J, Leng J, Lefebvre G, Lewis S, Mortazavi A, Niermann P, Rätsch G, Reymond A, Ribeca P, Richard

H, Rougemont J, Rozowsky J, Sammeth M, Sboner A, Schulz MH, Searle SMJ, Solorzano ND, Solovyev V, Stanke M, Stevenson BJ, Stockinger H, Valsesia A, Weese D, White S, Wold BJ, Wu J, Wu TD, Zeller G, Zerbino D, Zhang MQ. Assessment of transcript reconstruction methods for RNA-seq. *Nat Methods*. **2013**;10:1177–1184.

210. Zhang C, Zhang BH, Lin LL, Zhao SR. Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC Genomics*. **2017**;18.

211. Man HSJ, Sukumar AN, Lam GC, Turgeon PJ, Yan MS, Ku KH, Dubinsky MK, Ho JJD, Wang JJ, Das S, Mitchell N, Oettgen P, Sefton M V, Marsden PA. Angiogenic patterning by STEEL, an endothelial-enriched long noncoding RNA. *Proc Natl Acad Sci U S A*. **2018**;115:2401–2406.

212. Ehsani R, Drablos F. Measures of co-expression for improved function prediction of long non-coding RNAs. *BMC Bioinformatics*. **2018**;19.

213. Yu CK, Xu T, Assoian RK, Rader DJ. Mining the Stiffness-Sensitive Transcriptome in Human Vascular Smooth Muscle Cells Identifies Long Noncoding RNA Stiffness Regulators. *Arterioscler Thromb Vasc Biol*. **2018**;38:164–173.

214. Jeong G, Kwon DH, Shin S, Choe N, Ryu J, Lim YH, Kim J, Park WJ, Kook H, Kim YK. Long noncoding RNAs in vascular smooth muscle cells regulate vascular calcification. *Sci Rep*. **2019**;9.

215. Leung A, Trac C, Jin W, Lanting L, Akbany A, Sætrom P, Schones DE, Natarajan R. Novel long noncoding RNAs are regulated by angiotensin II in vascular smooth muscle cells. *Circ Res*. **2013**;113:266–278.

216. Chen J, Guo J, Cui X, Dai Y, Tang Z, Qu J, Raj JU, Hu Q, Gou D. Long Non-Coding RNA LnRPT is Regulated by PDGF-BB and Modulates Proliferation of Pulmonary Artery Smooth Muscle Cells. *Am J Respir Cell Mol Biol*. **2017**;

217. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. **2010**;28:511–515.

218. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. **2015**;33:290–295.

219. Ulitsky I, Bartel DP. lincRNAs: Genomics, Evolution, and Mechanisms. *Cell*. **2013**;154:26–46.

220. Wang L, Park HJ, Dasari S, Wang SQ, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res*. **2013**;41.

221. Washietl S, Findeiss S, Muller SA, Kalkhof S, von Bergen M, Hofacker IL, Stadler PF, Goldman N. RNAcode: Robust discrimination of coding and noncoding regions in comparative sequence data. *Rna*. **2011**;17:578–594.

222. Chen J, Shishkin AA, Zhu XP, Kadri S, Maza I, Guttman M, Hanna JH, Regev A, Garber M. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol*. **2016**;17.

223. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. **2007**;35:W345–W349.

224. Eddy SR. Profile hidden Markov models. *Bioinformatics*. **1998**;14:755–763.

225. Wucher V, Legeai F, Hedan B, Rizk G, Lagoutte L, Leeb T, Jagannathan V, Cadieu E, David A, Lohi H, Cirera S, Fredholm M, Botherel N, Leegwater PAJ, Le Beguec C, Fieten H, Johnson J, Alfoldi J, Andre C, Lindblad-Toh K, Hitte C, Derrien T. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res*. **2017**;45.

226. Ounzain S, Burdet F, Ibberson M, Pedrazzini T. Discovery and functional characterization of cardiovascular long noncoding RNAs. J. Mol. Cell. Cardiol. **2015**;89:17–26.

227. Alvarez-Dominguez JR, Hu W, Yuan B, Shi J, Park SS, Gromatzky AA, Van Oudenaarden A, Lodish HF.

Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood*. **2014**;123:570–581.

228. Hudson WH, Prokhnevska N, Gensheimer J, Akondy R, McGuire DJ, Ahmed R, Kissick HT. Expression of novel long noncoding RNAs defines virus-specific effector and memory CD8 + T cells. *Nat Commun*. **2019**;10.

229. Roux BT, Heward JA, Donnelly LE, Jones SW, Lindsay MA. Catalog of Differentially expressed long non-coding RNA following activation of human and mouse innate immune response. *Front Immunol*. **2017**;8.

230. Verma A, Jiang Y, Du W, Fairchild L, Melnick A, Elemento O. Transcriptome sequencing reveals thousands of novel long non-coding RNAs in B cell lymphoma. *Genome Med*. **2015**;7:110.

231. Tsoi LC, Iyer MK, Stuart PE, Swindell WR, Gudjonsson JE, Tejasvi T, Sarkar MK, Li B, Ding J, Voorhees JJ, Kang HM, Nair RP, Chinnaiyan AM, Abecasis GR, Elder JT. Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin. *Genome Biol*. **2015**;16:24.

232. Ounzain S, Micheletti R, Beckmann T, Schroen B, Alexanian M, Pezzuto I, Crippa S, Nemir M, Sarre A, Johnson R, Dauvillier J, Burdet F, Ibberson M, Guigo R, Xenarios I, Heymans S, Pedrazzini T. Genome-wide profiling of the cardiac transcriptome after myocardial infarction identifies novel heart-specific long non-coding RNAs. *Eur Heart J*. **2015**;36:353–368.

233. Alloza I, Goikuria H, Idro JL, Triviño JC, Fernández Velasco JM, Elizagaray E, García-Barcina M, Montoya-Murillo G, Sarasola E, Vega Manrique R, Freijo MDM, Vandenbroeck K. RNAseq based transcriptomics study of SMCs from carotid atherosclerotic plaque: BMP2 and IDs proteins are crucial regulators of plaque stability. *Sci Rep*. **2017**;7:1–12.

234. Andrews S. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc[cited 2021 Sep 30]

235. Krueger F. Trim Galore! https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/[cited 2021 Sep 30]

236. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **2013**;29:15–21.

237. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. **2011**;12.

238. Ulitsky I. Pipeline for lncRNA annotation from RNA-seq data (PLAR). http://www.weizmann.ac.il/Biological_Regulation/IgorUlitsky/pipeline-lncrna-annotation-rna-seq-data-plar[cited 2021 Sep 30]

239. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*. **2006**;34:D590–D598.

240. Pertea M, Pertea G. GFF Utilities: GffRead and GffCompare. *F1000Research*. **2020**;9:304.

241. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **2014**;15.

242. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*. **2010**;11.

243. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, Lancet D, Cohen D. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*. **2017**;

244. Diedenhofen B, Musch J. cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS One*. **2015**;10:e0121945.

245. Wang X, Fei F, Qu J, Li C, Li Y, Zhang S. The role of septin 7 in physiology and pathological disease: A systematic review of current status. *J Cell Mol Med*. **2018**;22:3298.

246. Bernard K, Logsdon NJ, Benavides GA, Sanders Y, Zhang J, Darley-Usmar VM, Thannickal VJ. Glutaminolysis is required for transforming growth factor-β1–induced myofibroblast differentiation and activation. *J Biol Chem*. **2018**;293:1218–1228.

247. Cheng HS, Besla R, Li A, Chen Z, Shikatani EA, Nazari-Jahantigh M, Hammoutène A, Nguyen M-A, Geoffrion M, Cai L, Khyzha N, Li T, MacParland SA, Husain M, Cybulsky MI, Boulanger CM, Temel RE, Schober A, Rayner KJ, Robbins CS, Fish JE. Paradoxical Suppression of Atherosclerosis in the Absence of microRNA-146a. *Circ Res*. **2017**;121:354–367.

248. WZ G, LM G, TQ X, YH Y, F J. Identification of a multidimensional transcriptome signature for survival prediction of postoperative glioblastoma multiforme patients. *J Transl Med*. **2018**;16.

249. Chung I-H, Lu P-H, Lin Y-H, Tsai M-M, Lin Y-W, Yeh C-T, Lin K-H. The long non-coding RNA LINC01013 enhances invasion of human anaplastic large-cell lymphoma. *Sci Reports 2017 71*. **2017**;7:1–10.

250. Y S, D L, Y Z, K W. MicroRNA 4323 induces human bladder smooth muscle cell proliferation under cyclic hydrodynamic pressure by activation of erk1/2 signaling pathway. *Exp Biol Med (Maywood)*. **2017**;242:169–176.

251. Bogunovic N, Meekel JP, Micha D, Blankensteijn JD, Hordijk PL, Yeung KK. Impaired smooth muscle cell contractility as a novel concept of abdominal aortic aneurysm pathophysiology. *Sci Rep*. **2019**;9.

252. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet 2019 212*. **2019**;21:102–117.

253. Vangala P, Murphy R, Quinodoz SA, Gellatly K, McDonel P, Guttman M, Garber M. High-Resolution Mapping of Multiway Enhancer-Promoter Interactions Regulating Pathogen Detection. *Mol Cell*. **2020**;80:359–373.

254. Aitken S, Magi S, Alhendi AMN, Itoh M, Kawaji H, Lassmann T, Daub CO, Arner E, Carninci P, Forrest ARR, Hayashizaki Y, Consortium the F, Khachigian LM, Okada-Hatakeyama M, Semple CA. Transcriptional Dynamics Reveal Critical Roles for Non-coding RNAs in the Immediate-Early Response. *PLOS Comput Biol*. **2015**;11:e1004217.

255. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jørgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhata E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Müller F, Forrest ARR, Carninci P, Rehli M, Sandelin A. An atlas of active enhancers across human cell types and tissues. *Nature*. **2014**;507:455–461.

256. Giotti B, Chen S-H, Barnett MW, Regan T, Ly T, Wiemann S, Hume DA, Freeman TC. Assembly of a parts list of the human mitotic cell cycle machinery. *J Mol Cell Biol*. **2019**;11:703.

257. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. The Human Transcription Factors. *Cell*. **2018**;172:650–665.

258. Kim HY, Kang YJ, Song IH, Choi HC, Kim HS. Upregulation of interleukin-8/CXCL8 in vascular smooth muscle cells from spontaneously hypertensive rats. *Hypertens Res*. **2008**;31:515–523.

259. Tao Y, Zhang J, Chen L, Liu X, Yao M, Zhang H. LncRNA CD27-AS1 promotes acute myeloid leukemia progression through the miR-224-5p/PBX3 signaling circuit. *Cell Death Dis*. **2021**;12.

260. N P, JP K, SD S, KH K. Foxl1 controls the Wnt/beta-catenin pathway by modulating the expression of proteoglycans in the gut. *J Biol Chem*. **2001**;276:43328–43333.

261. Aoki R, Shoshkes-Carmel M, Gao N, Shin S, May CL, Golson ML, Zahm AM, Ray M, Wiser CL, Wright CVE, Kaestner KH. Foxl1-Expressing Mesenchymal Cells Constitute the Intestinal Stem Cell Niche. *Cell Mol*

Gastroenterol Hepatol. **2016**;2:175–188.

262. Xu B-F, Liu R, Huang C-X, He B-S, Li G-Y, Sun H-S, Feng Z-P, Bao M-H. Identification of key genes in ruptured atherosclerotic plaques by weighted gene correlation network analysis. *Sci Reports 2020 101*. **2020**;10:1–10.

263. Sun J, Sun B, Sun R, Zhu D, Zhao X, Zhang Y, Dong X, Che N, Li J, Liu F, Zhao N, Wang Y, Zhang D. HMGA2 promotes vasculogenic mimicry and tumor aggressiveness by upregulating Twist1 in gastric carcinoma. *Sci Reports 2017 71*. **2017**;7:1–13.

264. Giacconi R, Kanoni S, Mecocci P, Malavolta M, Richter D, Pierpaoli S, Costarelli L, Cipriano C, Muti E, Mangialasche F, Piacenza F, Tesei S, Galeazzi R, Theodoraki E V., Lattanzio F, Dedoussis G, Mocchegiani E. Association of MT1A haplotype with cardiovascular disease and antioxidant enzyme defense in elderly Greek population: comparison with an Italian cohort. *J Nutr Biochem*. **2010**;21:1008–1014.

265. Goodarz K, Siamak S, Mahdi A, Bahareh S-K, Kamran G, Saeid RD, Mahmoud AR, Farzad H, Fereidoun A, Maryam D. Genome-wide association study on blood pressure traits in the Iranian population suggests ZBED9 as a new locus for hypertension. *Sci Rep*. **2021**;11.

266. Klein D, Benchellal M, Kleff V, Jakob HG, Ergün S. Hox genes are involved in vascular wall-resident multipotent stem cell differentiation into smooth muscle cells. *Sci Reports 2013 31*. **2013**;3:1–12.

267. Iwashita Y, Fukuchi N, Waki M, Hayashi K, Tahira T. Genome-wide Repression of NF-κB Target Genes by Transcription Factor MIBP1 and Its Modulation by O-Linked β-N-Acetylglucosamine (O-GlcNAc) Transferase. *J Biol Chem*. **2012**;287:9887–9900.

268. Kurozumi A, Nakano K, Yamagata K, Okada Y, Nakayamada S, Tanaka Y. IL-6 and sIL-6R induces STAT3-dependent differentiation of human VSMCs into osteoblast-like cells through JMJD2B-mediated histone demethylation of RUNX2. *Bone*. **2019**;124:53–61.

269. Zhang P, Chen X, Zhang Y, Su H, Zhang Y, Zhou X, Sun M, Li L, Xu Z. Tet3 enhances IL-6 expression through up-regulation of 5-hmC in IL-6 promoter in chronic hypoxia induced atherosclerosis in offspring rats. *Life Sci*. **2019**;232:116601.

270. Igor U, Alena S, Calvin H J, Hazel S, David P B. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*. **2011**;147:1537–1550.

271. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. **2009**;458:223–227.

272. Luo Y, Feng J, Xu Q, Wang W, Wang X. NSun2 Deficiency Protects Endothelium From Inflammation via mRNA Methylation of ICAM-1. *Circ Res*. **2016**;118:944–956.

273. Amândio AR, Necsulea A, Joye E, Mascrez B, Duboule D. Hotair Is Dispensible for Mouse Development. *PLOS Genet*. **2016**;12:e1006232.

274. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. *Nat 2011 4737347*. **2011**;473:337–342.

275. Fanucchi S, Fok ET, Dalla E, Shibayama Y, Borner K, Chang EY, Stoychev S, Imakaev M, Grimm D, Wang KC, Li GL, Sung WK, Mhlanga MM. Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. *Nat Genet*. **2019**;51:138–150.

276. Goede OM de, Nachun DC, Ferraro NM, Gloudemans MJ, Rao AS, Smail C, Eulalio TY, Aguet F, Ng B, Xu J, Barbeira AN, Castel SE, Kim-Hellmuth S, Park Y, Scott AJ, Strober BJ, Anand S, Gabriel S, Getz GA, Graubert A, Hadley K, Handsaker RE, Huang KH, Li X, MacArthur DG, Meier SR, Nedzel JL, Nguyen DT, Segrè A V., Todres E, Balliu B, Bonazzola R, Brown A, Conrad DF, Cotter DJ, Cox N, Das S, Dermitzakis ET, Einson J, Engelhardt BE, Eskin E, Flynn ED, Fresard L, Gamazon ER, Garrido-Martín D, Gay NR, Guigó R, Hamel AR, He Y, Hoffman PJ, Hormozdiari F, Hou L, Jo B, Kasela S, Kashin S, Kellis M, Kwong A, Li X,

Liang Y, Mangul S, Mohammadi P, Muñoz-Aguirre M, Nobel AB, Oliva M, Park Y, Parsana P, Reverter F, Rouhana JM, Sabatti C, Saha A, Stephens M, Stranger BE, Teran NA, Viñuela A, Wang G, Wright F, Wucher V, Zou Y, Ferreira PG, Li G, Melé M, Yeger-Lotem E, Bradbury D, Krubit T, McLean JA, Qi L, Robinson K, Roche N V., Smith AM, Tabor DE, Undale A, Bridge J, Brigham LE, Foster BA, Gillard BM, Hasz R, Hunter M, Johns C, et al. Population-scale tissue transcriptomics maps long non-coding RNAs to complex disease. *Cell*. **2021**;184:2633-2648.e19.

277. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell*. **2013**;153:307–319.

278. Suzuki HI, Young RA, Sharp PA. Super-Enhancer-Mediated RNA Processing Revealed by Integrative MicroRNA Network Analysis. *Cell*. **2017**;168:1000-1014.e15.

279. Wang D, Atanasov AG. The microRNAs Regulating Vascular Smooth Muscle Cell Proliferation: A Minireview. *Int J Mol Sci*. **2019**;20.

280. Sun T, Du S-Y, Armenia J, Qu F, Fan J, Wang X, Fei T, Komura K, Liu SX, Lee G-SM, Kantoff PW. Expression of lncRNA MIR222HG co-transcribed from the miR-221/222 gene promoter facilitates the development of castration-resistant prostate cancer. *Oncog 2018 73*. **2018**;7:1–13.

281. Duan Q, Mao X, Xiao Y, Liu Z, Wang Y, Zhou H, Zhou Z, Cai J, Xia K, Zhu Q, Qi J, Huang H, Plutzky J, Yang T. Super enhancers at the miR-146a and miR-155 genes contribute to self-regulation of inflammation. *Biochim Biophys Acta - Gene Regul Mech*. **2016**;1859:564–571.

282. Bouvy-Liivrand M, Hernández de Sande A, Pölönen P, Mehtonen J, Vuorenmaa T, Niskanen H, Sinkkonen L, Kaikkonen MU, Heinäniemi M. Analysis of primary microRNA loci from nascent transcriptomes reveals regulatory domains governed by chromatin architecture. *Nucleic Acids Res*. **2017**;45:9837.

283. Reimer KA, Neugebauer KM. Preparation of Mammalian Nascent RNA for Long Read Sequencing. *Curr Protoc Mol Biol*. **2020**;133:e128.

284. Suzuki Y, Yeung AC, Ikeno F. The Representative Porcine Model for Human Cardiovascular Disease. *J Biomed Biotechnol*. **2011**;2011.

285. Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Flicek P, Girón CG, Griffin DK, Hall R, Hannum G, Hourlier T, Howe K, Hume DA, Izuogu O, Kim K, Koren S, Liu H, Manchanda N, Martin FJ, Nonneman DJ, O'Connor RE, Phillippy AM, Rohrer GA, Rosen BD, Rund LA, Sargent CA, Schook LB, Schroeder SG, Schwartz AS, Skinner BM, Talbot R, Tseng E, Tuggle CK, Watson M, Smith TPL, Archibald AL. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience*. **2020**;9:1–14.