LONDON'S GLOBAL UNIVERSITY

REGULATION OF NKX2-5 IN BLOOD VESSELS

Athina Dritsoula

April 2017

Thesis submitted to the University College London

for the degree of

Doctor of Philosophy (PhD)

DECLARATION

I, Athina Dritsoula, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Athina Dritsoula Date:

ABSTRACT

NKX2-5 is a transcription factor required for the formation of the heart and vessels during development. Postnatal expression is significantly downregulated, and then re-activated in diseased conditions characterised by vascular remodelling. However, the mechanisms regulating NKX2-5 activation in diseased vessels remain unknown. The aim of this thesis is to identify these mechanisms and provide information on how the gene contributes to cardiovascular pathologies, such as scleroderma-associated pulmonary hypertension.

A case-control genetic association study was performed in two independent cohorts of scleroderma patients. Associated SNPs located in the *NKX2-5* genomic region were cloned into reporter vectors, and transcriptional activity was assessed by reporter-gene assays. Associated SNPs were further evaluated through protein-DNA binding assays, chromatin immunoprecipitation and RNA silencing. Signalling mechanisms activating NKX2-5 expression were investigated in vascular endothelial and smooth muscle cells using a panel of selective inhibitors.

Meta-analysis across the two independent cohorts revealed that rs3131917 was associated with scleroderma. Rs3132139, downstream of *NKX2-5*, was significantly associated with pulmonary hypertension in both cohorts. The region containing rs3132139 and rs3131917 was shown to be a novel functional enhancer, which increased *NKX2-5* transcriptional activity through the binding of GATA6, c-JUN, and MEF-2c. An activator TEAD/YAP1 complex was shown to bind at rs3095870, another functional SNP upstream of *NKX2-5* transcription start site, which showed marginal association with scleroderma. Signalling mechanisms, involving TGF-β, ERK5, AKT and hypoxia, stimulated NKX2-5 expression during phenotypic modulation of vascular endothelial and smooth muscle cells.

Overall, the data showed that *NKX2-5* is genetically associated with scleroderma and pulmonary hypertension. Functional evidence revealed a regulatory mechanism, activated by TGF-β, which results in *NKX2-5* transcription in human vascular smooth muscle cells through the interaction of an upstream promoter and a novel downstream enhancer. These regulatory mechanisms can act as a model for NKX2-5 activation in cardiovascular disease characterised by vascular remodelling.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my primary supervisor Dr Markella Ponticos for giving me the opportunity to undertake my PhD in her research group, for her support and guidance throughout my project. I would also like to thank my second supervisor Prof David Abraham for his contribution to this project, and Dr Jill Norman for proof reading this thesis. I know it must have been tedious, and for that, I am grateful. Also, Prof Chris Denton and Dr Carmen Fonseca for the contribution of their clinical knowledge in this thesis.

This thesis would not have been possible without the help and support of my friends and colleagues. I would first like to thank Dr Sandra Guerra, who has always been there for me physically and mentally even when she was away, and Dr Ioannis Papaioannou for his crazy ideas and critical thinking. Johanna Verneau and Bahja Ahmed Abdi, who made the unlimited time spent in the lab cosy and fun! Rob, Sarah, Grace and Nyra for our laughs, lunches and English lessons, and Xu for his endless stories.

I would like to thank every member of the Centre for Rheumatology that contributed to this project. Also, British Heart Foundation for funding my PhD studentship and Professors Steve Humphries and Philippa Talmud, to whom I owe this journey.

Finally, I would like to thank Greg, who has been there for me day and night, holding strong by my side and eventually becoming an expert in molecular biology! Also, my cousins and friends, for their patience and understanding over the past four years. This thesis is dedicated to my family, mum, dad and sister, who have always believed in me and given me the support and strength I needed, especially when things got hard. I love you all very much and thank you for everything.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

COMMON ABBREVIATIONS

ACA	anti-centromere antibody
AID	autoimmune diseases
AKT	serine/threonine protein kinase 1
ALK5	or TGFBR1 transforming growth factor beta receptor 1
ARA	anti-RNA polymerase antibody
ATA	anti-topoisomerase I antibody
BMP	bone morphogenetic protein
BMPR2	bone morphogenetic protein receptor type 2
CAD	coronary artery disease
CHD	congenital heart disease
COL1A2	collagen 1 alpha 2
CTD	connective tissue diseases
CTGF	connective tissue growth factor
CVD	cardiovascular disease
DNA	deoxyribonucleic acid
ECE	endothelin converting enzyme
ECM	extracellular matrix
EndoMT	endothelial to mesenchymal transition
ERK	extracellular signal-regulated kinases
ET-1	endothelin-1
FGF	fibroblast growth factor
GWAS	genome-wide association study
HIF-1	Hypoxia-induced factor 1
HLA/MHC	human leukocyte antigen/ major histocompatibility complex
HPASMC	human pulmonary artery smooth muscle cells

IL (eg. -6, -9)	interleukin
JNK	c-Jun N-terminal kinase
KLF-4,5	kruppel-like factors 4,5
LD	linkage disequilibrium
MAPK	Mitogen-activated kinase-like protein
miR	microRNA
MMP	matrix metalloproteinase protein
mPAP	mean pulmonary arterial pressure
mRNA	messenger RNA
NKX2-5	NK-2 homeobox 5 containing gene
NO	nitric oxide
PAD	peripheral arterial disease
PAEC	pulmonary artery endothelial cells
PAH	pulmonary arterial hypertension
PCR	polymerase chain reaction
PDGF	platelet derived growth factor
PF	pulmonary fibrosis
PH	pulmonary hypertension
PI3K	phosphatidylinositol 3-kinase
RA	rheumatoid arthritis
RC	renal crisis
RHC	right heart catheterisation
RNA	ribonucleic acid
RT-qPCR	real-time quantitative polymerase chain reaction
RVSP	right ventricular systolic pressure
siRNA	silencing RNA

SLE	systemic lupus erythematosus
SM-22	smooth muscle protein 22 or Transgelin
SM-MHC	smooth muscle myosin heavy chain
SMC	smooth muscle cells
SNP	single nucleotide polymorphism
SRF	serum response factor
SSc	scleroderma
Tag SNP	tagging SNP
TAK-1	TGF-β-activated kinase 1
TEAD	transcriptional enhancer activator domain
TGF-β	transforming growth factor beta
TIMP	tissue inhibitor of matrix metalloproteinase
TNF-α	tumour necrosis factor alpha
VSMC	vascular smooth muscle cells
WHO	world health organisation
YAP	Yes-associated protein
α-SMA	alpha smooth muscle actin

PUBLISHED DATA ARISING FROM THIS THESIS

Articles:

- *Functional analysis of scleroderma and pulmonary hypertension associated polymorphisms within the NKX2-5 genomic locus reveals mechanism that activates expression in vascular remodelling*. A. Dritsoula, I. Papaioannou, S. Guerra, C. Fonseca, J. Martin, A. Herrick, D. Abraham, C. Denton, M. Ponticos. Submitted at A&R, December 2016.
- *NKX2-5 is a critical regulator of smooth muscle cell differentiation in vessel remodelling and vascular disease*. M. Ponticos, A. Dritsoula, R. Baliga, S. Trinder, I. Papaioannou, J. Santos Cade, X. Shiwen, A. Hobbs, C. Jackson, C. Denton, D. Abraham. Ready for re-submission to Nature Communications.

Abstracts presented in conferences and meetings:

- *"PH and Scleroderma associated SNPs reveal a transcriptional mechanism for NKX2-5 in diseased vessels".* A. Dritsoula, I. Papaioannou, S. Guerra, C. Fonseca, J. Martin, A. Herrick, D. Abraham, C. Denton, M. Ponticos. Poster presentation at London Vascular Biology Forum, London, UK, November 20th, 2016.
- *"NKX2-5 is associated with pulmonary arterial hypertension"*. A. Dritsoula, C. Fonseca, J. Martin, C. Denton, M. Ponticos. Poster presentation at the American Thoracic Society (ATS) International Conference 2016, San Francisco, USA, May 12-18, 2016.
- "*NKX2-5: A novel locus genetically associated with Scleroderma"*. A. Dritsoula, C. Fonseca, J. Martin, A. Herrick, C. Denton, M. Ponticos. Poster presentation at the 4th Systemic Sclerosis World Congress, Lisbon, Portugal, February 18-20, 2016.
- "*Genetic links between NKX2-5 and Scleroderma".* A. Dritsoula, C. Fonseca, J. Martin, A. Herrick, C. Denton, M. Ponticos. Poster presentation at the 14th International Workshop on Scleroderma Research, Cambridge, UK, August 1-5, 2016. Oral presentation at UCL Post-Graduate School Day, London, UK, July 9th, 201

CHAPTER 1 - INTRODUCTION

This thesis focuses on the regulation of the *NKX2-5* gene at the genetic, transcriptional and post-transcriptional levels in human adult blood vessels. In particular, expression and regulation of the *NKX2-5* gene has been studied in the context of pulmonary pathologies characterised by vascular remodelling, such as pulmonary hypertension and scleroderma.

In this chapter, I will briefly describe the cardiovascular system, the pulmonary vasculature and the pulmonary vascular smooth muscle cells. I will then introduce cardiovascular diseases and the mechanisms that are involved in the disease pathogenesis, including vascular smooth muscle cell de-differentiation, vascular remodelling and endothelial-to-mesenchymal transition. Next, I will provide a brief background on pulmonary hypertension and scleroderma, the focus of the experimental work found in this thesis. Subsequently, I will discuss the genetics of cardiovascular diseases particularly, the genetics of pulmonary hypertension and scleroderma. Finally, I will close this chapter with a detailed account of *NKX2-5*, and will describe the evidence which leads me to believe that it has a major role in the remodelling of diseased blood vessels.

1.1 Cardiovascular system

The cardiovascular system is the engine of life, and its ultimate purpose is to allow the vital exchange of gases, fluids, electrolytes, and other large molecules between the cells and the outside environment. The heart and the vasculature are the main components of the cardiovascular system and ensure that adequate blood flow is delivered to organs and tissues to facilitate gas exchange.

1.1.1 The heart and the blood vessels

The cardiovascular system consists of two major organs: the heart and the blood vessels including arteries, arterioles, capillaries, venules and veins (1). The venules and the veins are part of the pulmonary circulation that sends deoxygenated blood to the lungs to receive oxygen and unload carbon dioxide. The rest of the vessels are all part of the systemic circulation that sends oxygenated blood and nutrients to the body while removing wastes. The heart can be considered as a pump that receives blood from the peripheral veins at low pressure, contracts and provides the organs and tissues with blood flow at high pressure.

The performance of the heart is usually expressed in terms of the cardiac output (1). Any factor that alters the heart rate or volume of the injected blood will alter the cardiac output. The heart rate is determined by groups of cells within the heart that act as electrical pacemakers, and their activity is increased or decreased by autonomic nerves and hormones. In recent years, it has been established that the heart also retains a role as a secretory organ, synthesising and releasing several hormones (2), such as the atrial natriuretic peptide (ANP), which is critical for the regulation of blood volume and pressure (3).

Blood vessels contract and dilate to regulate arterial blood pressure, alter blood flow within organs, regulate capillary blood pressure, and distribute blood volume within the body. Changes in vascular diameters are carried though the vascular smooth muscle cells (VSMCs) within the vascular wall upon cell activation by autonomic nerves, metabolic and biochemical signals from the outside of the blood vessel, and vasoactive substances released by cells lining the vessels. Other functions include vascular homeostasis and modulation of the vascular tone.

A third component of the cardiovascular system is the lymphatic circulation. The lymphatic system does not contain blood and is not involved in the vital gas exchange function, but contributes to the collection of excess fluids and their transport back into the venous circulation.

1.1.2 The blood vessel wall

The arterial wall is composed of three layers known as the *tunics*: the *interna*, the *media*, and the *externa* (Figure 1.1). The anatomy and physiology of the blood vessel wall is described in detail in (4).

The *tunica interna* or *intima* (Figure 1.1) lines the lumen of the blood vessel and is exposed to the blood. It consists of endothelial cells that form a selective permeable barrier for materials entering or leaving the blood flow. Endothelial cells secrete chemokines to stimulate dilation or constriction of the vessel, and repel blood cells and platelets from attaching to the wall (5). When the endothelium is damaged, the endothelial cells produce cell-adhesion molecules that induce leukocytes and platelets to adhere to the surface of the wall and initiate a defensive action (6).

The middle layer, *tunica media* (Figure 1.1) consists of VSMCs, collagen fibres, elastins and other components of the extracellular matrix (ECM). VSMCs in the *tunica media* provide structural support, strength and elasticity, and control blood pressure and blood flow through highly regulated contractile mechanisms. The

tunica media retains the important role of contraction of the vessel and influences its structure and function to accommodate changes in the environment during pathological processes. The ratio of smooth muscle, collagen, and elastin, each of which has different elastic properties, determines the overall mechanical properties of the vessel (1). For example, the aorta has a large amount of elastin, which enables it to passively expand and contract as blood is pumped into it from the heart. This mechanism enables the aorta to dampen the arterial pulse pressure. In contrast, smaller arteries and arterioles have a relatively large amount of smooth muscle cells, which is required for these vessels to contract and thereby regulate arterial blood pressure and organ blood flow.

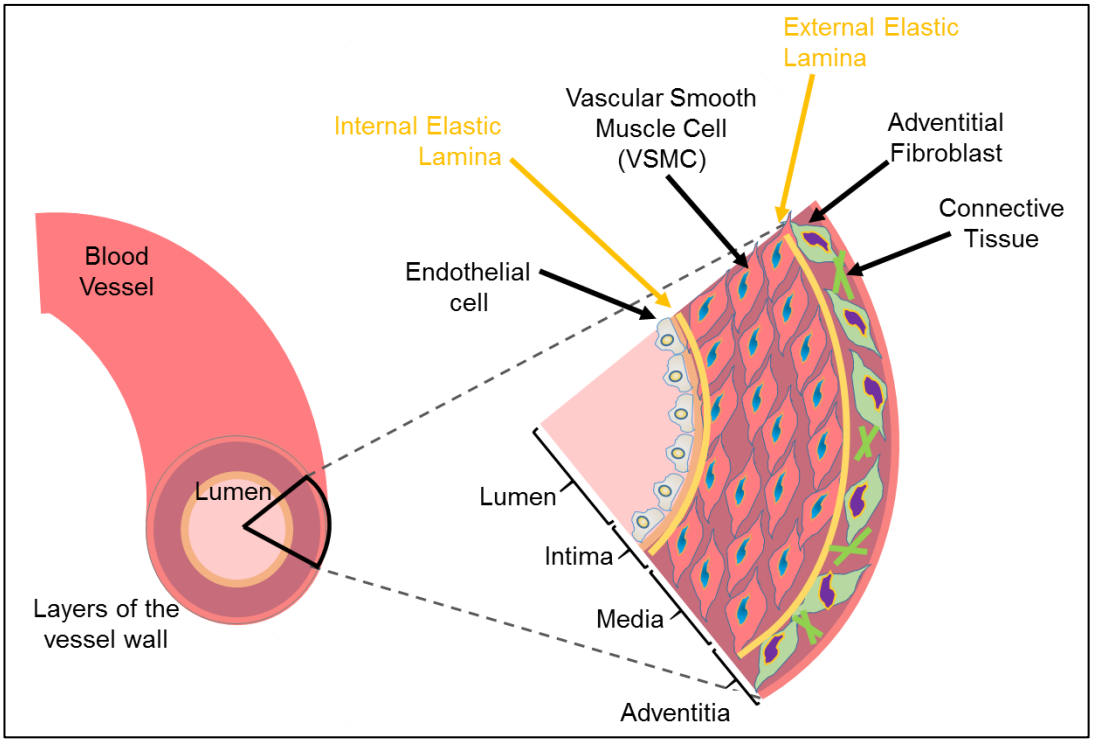


Figure 1.1 The blood vessel wall. *The blood vessel wall consists of three layers, also known as tunicas: the intima, the media, and the externa or adventitia. The tunica intima consists of a thin layer of endothelial cells that form a selective permeable barrier. The endothelial cells face the lumen and communicate with material travelling through the blood flow. The tunica media mainly consists of VSMC and provides structural support, strength and elasticity, as well as control of blood pressure and blood flow. The adventitia consists of fibroblasts, loose connective tissue and perivascular nerves. It functions as a dynamic compartment for cell trafficking, and participates in growth and repair of blood vessels.*

The external layer, *tunica externa* or *adventitia* (Figure 1.1) consists of fibroblasts, loose connective tissue and perivascular nerves. The *adventitia* functions as a dynamic compartment for cell trafficking, it participates in growth and repair of blood vessels and mediates the communication with VSMCs and endothelial cells (7).

Also, it anchors the vessel and provides passage for small nerves, lymphatic vessels, and smaller blood vessels that supply the tissues of the larger vessels. Moreover, it contains resident populations of macrophages, T-cells, B-cells, mast cells, and dendritic cells that carry out important surveillance and innate immune functions in response to foreign antigens (8-10). The exact role of the *adventitia* in the formation of tissue lesions and in tissue repair is still under scrutiny.

1.1.3 Vascular endothelial cells

The vascular endothelium is a single layer of cells that line all blood vessels. Depending on the type of vessel and tissue location, endothelial cells are joined together by different types of intercellular junctions. Some of these junctions are very tight, whereas others have gaps between the cells that enable and facilitate blood cells to move in and out of the capillary. Endothelial cells have several important functions with the most important being the regulation of vascular tone. In addition, endothelial cells serve as a barrier for the exchange of fluid, electrolytes, macromolecules, and cells between the intravascular and extravascular space (11). Further, they are able to regulate smooth muscle function and modulate platelet aggregation.

The vascular endothelium exerts its functions through the production and secretion of several vasoactive factors [reviewed in (5, 12)]. The importance of normal endothelial function has been established through studies that associated endothelial damage and dysfunction with cardiovascular disease (CVD) including atherosclerosis, pulmonary arterial hypertension (PAH) and organ fibrosis (13-17). Damage of the endothelium at the capillary level increases capillary permeability which leads to increased capillary fluid filtration and tissue oedema.

1.1.4 Vascular smooth muscle cells (VSMC)

The *tunica media* consists of VSMCs. Depending on the size of the vessel, there may be several layers of VSMCs, some arranged circumferentially and others arranged helically along the longitudinal axis of the vessel. Contractile proteins such as actin and myosin are present and involved in the cell structure and cytoskeleton. However, VSMCs exhibit a different structural arrangement compared to cells found in the cardiac or skeletal muscle (1). In detail, bands of actin filaments are joined together and anchored by dense bodies within the cell or dense bands on the inner surface of the sarcolemma. Each myosin filament is surrounded by several actin filaments. Similar to cardiac myocytes, VSMC are electrically connected by gap

junctions. These low-resistance intercellular connections allow propagated responses along the length of the blood vessels.

VSMCs carry out the vital process of contraction that is responsible for the structural strength and sustainability of blood vessels. The contraction of VSMC in a relaxation state is usually slow and sustained, and determines the diameter and the resting tone of the vessel (1). Contraction is affected by various activating and inhibitory stimuli, such as sympathetic adrenergic nerves, circulating hormones, substances released by the endothelium, and vasoactive substances released by the tissue surrounding the blood vessel.

In addition, contraction can be initiated by electrical, chemical, and mechanical stimuli [reviewed in (18, 19)]. Electrical depolarisation occurs through changes in ion concentrations or by the receptor-coupled opening of ion channels. In particular, it elicits contraction primarily by opening voltage-dependent calcium channels, causing an increase in the intracellular concentration of calcium. Mechanical stimuli that can cause contraction (stretch) usually originate from the smooth muscle itself, and this is known as the myogenic response (20), which also results from activation of ion channels leading to calcium influx. Several signal transduction mechanisms modulate intracellular calcium concentration and therefore the state of vascular tone [reviewed in (18)]. These mechanisms involve: a) the inositol triphosphate (IP_3) pathway through the activation of phospholipase C, b) the cAMP mechanism through the activation of adenylyl cyclase, and c) the NO-cGMP system. Chemical stimuli include norepinephrine, epinephrine, angiotensin II, vasopressin, and endothelin that can all bind to specific receptors on the surface of VSMCs.

1.1.4.1 VSMC contractile phenotype

Cellular differentiation is the process by which multi-potent cells in a developing organism acquire specific functions and properties that distinguish them from other cell types. The mature VSMC is a highly differentiated *contractile* cell that is responsible for the vital processes of contraction and relaxation of the blood vessel, the maintenance of structural strength and sustainability of homeostasis. The contractile VSMCs directly regulate lumen calibre, and thus, arterial and venous tone and vascular resistance. These, in turn, control the distribution of blood flow throughout the body. Mechanistically, VSMCs contain many thin actin filaments and relatively few thick myosin filaments, which are arranged along the long axis of the cells (21, 22). Taking advantage of this arrangement, VSMC produce prolonged forces with low energy consumption to maintain vascular tone and blood flow.

Contractile cells that grow on collagen do not proliferate (23), and are able to retain a high proportion of myofilaments (24). However, their morphology and shape change, and they acquire “stress fibres” that provide the ability to contract (25). Stress fibres are composed of bundles of approximately 10-30 actin filaments, which are held together by the actin-crosslinking protein α-actinin (26). Non-muscle myosin and tropomyosin are also structural components of the fibres (26).

Under homeostatic unstimulated conditions, adult medial VSMC are predominantly quiescent, proliferate at an extremely low rate (27), and produce very low amounts of ECM (28, 29). However, the expression of a repertoire of contractile proteins, agonist receptors, ion channels, and signal-transducing molecules is required to perform contraction. Gene and protein expression is meticulously regulated, and although well studied, some parts still remain unclear. The panel of contractile proteins includes alpha smooth muscle actin (α-SMA), smooth muscle myosin heavy and light chains (SM-MHC and SM-MLC, respectively), calponin, SM22, vinculin, tropomyosin, and other intermediate filaments and integrins. A thorough review of the differentiation of VSMC and the contractile proteins was written by G. Owens (30). A few of the most abundant and important proteins for VSMC contractility are discussed below.

α-SMA is the most abundant protein found in VSMCs, comprising almost 40% of total cell protein. It is the earliest known marker of differentiated VSMCs expressed during development of the vasculature and is essential for the VSMC function. The protein is expressed in all mesodermal-derived cells in development, tissue repair and neoplastic growth. It is also expressed in early stages of differentiation of both cardiac and skeletal muscle, as well as in fibroblasts in wound healing and in tumours (31). The gene is differentially regulated at the transcriptional level in a tissue-specific manner (31).

SM-MHC is an essential component of the contractile system and is expressed in muscle and non-muscle cells. Multiple isoforms of MHC have been found and their expression is differentially regulated in tissue-specific and developmental stage-specific ways. Mature VSMCs express at least three smooth muscle and two non-muscle isoforms. SM-MHC is a highly specific marker of the SMC lineage. Similarly to MHC, two isoforms of SM-MLC are expressed in VSMCs, and are also differentially regulated.

In contrast to α-SMA and myosin, expression of calponin is exclusively restricted to VSMCs, whereas expression of SM22 is found in both VSMC and in myofibroblasts.

Both calponin and SM22 regulate contraction, but the exact mechanism remains unclear [reviewed in (32)]. The expression of VSMC protein markers appears to be relatively uniform between VSMCs in mature blood vessels. However, during development, contractile protein expression appears to be heterogeneous between different VSMCs, indicating that the developmental timing of expression is not synchronous and depends on the embryonic origin (32). In addition, contractile protein expression may increase in response to increased needs of tissues for contractility and other functions, explaining the number of different isoforms that have evolved for each protein. For instance, α-SMA and SM-MHC/MLC must be present in almost all cells, since they are required for cytokinesis and cell motility apart from contractility. However, the high mechanical forces produced by VSMCs require much higher expression levels of the contractile proteins compared to non-muscle cells (32).

1.1.4.2 VSMC plasticity

Unlike cardiac and skeletal muscle that undergo terminal and irreversible differentiation, VSMCs retain remarkable plasticity, and can undergo profound and reversible changes in phenotype in response to changes in the local environment (32). Plasticity provides the cell with the ability to reverse the differentiated phenotype to a less differentiated stage that is known as the *synthetic* phenotype. This procedure is called *de-differentiation.* Plasticity is considered necessary for the SMC differentiation and maturation program, and it is believed that it evolved in higher organisms because it conferred a survival advantage (33).

Phenotype switching can occur under physiological conditions. Striking examples of plasticity are seen in vascular development, when VSMCs are directly involved in the morphogenesis of blood vessels and exhibit high rates of proliferation, migration, and production of ECM consisting of collagens, elastins and proteoglycans (30). Similarly, in vascular injury VSMCs acquire a synthetic phenotype required for the physiological process of repair (30). However, a high degree of plasticity may predispose the cell to abnormal environmental signals that can lead to adverse phenotypic switching contributing to development of vascular disease. Indeed, VSMC de-differentiation is profound in vascular remodelling in major diseases.

It is important to note that the phenotypes of VSMCs are not mutually exclusive. Consequently, plasticity has confounded efforts to understand the cellular and molecular mechanisms that control VSMC differentiation. The extent of phenotypic

modulation and its reversibility appears to be dependent on many factors, and this has undoubtedly contributed to many of the controversies that exist in the literature regarding VSMC differentiation.

The VSMC synthetic phenotype will be further discussed in detail in section 1.2.3.2.

1.1.4.3 Embryonic origin of VSMC

Numerous studies focused on VSMC lineage have revealed that extensive heterogeneity exists among VSMC isolated not only from different arteries, but also from different compartments of the same vessel (34, 35). Heterogeneity is found in cell morphology, organisation at confluence, growth control, functional properties such as production of peptide growth factors, sensitivity to heparin-mediated growth inhibition, ability to proliferate in serum-free medium, as well as gene and protein expression patterns. These findings suggest that VSMCs from different sources can exhibit a wide range of diverse phenotypes, possibly due to the different embryonic origins of VSMCs (32).

A thorough review by Majesky *et al* (36) showed that VSMC in the cardiovascular system arise from eight different origins. However, the majority of them derive from the neural crest, the secondary heart field (SHF), the somites and various stem cells. In detail, neural crest-derived VSMCs populate a small segment of the aorta and other great arteries in close proximity to the heart. A second, clearly distinct type of VSMCs is found in the walls of coronary arteries and is derived from the proepicardial mesothelium, as are all the progenitors of the coronary vessels (37). The SHF also contributes cells to the cardiovascular system. Specifically, cells of the SHF migrate to the outflow tract, and then enter the aortic sac and differentiate into VSMCs that form the base of the aorta and pulmonary trunk.

Studies focusing on lineage tracing have showed that other cell types such as progenitor cells also contribute to the development of the cardiovascular system. Progenitor cells are able to differentiate to VSMCs within the vessels, and this procedure is not limited to embryogenesis, but also occurs in the quiescent adult arterial wall (38).

The differentiation of embryonic origin-specific VSMCs from human pluripotent stem cells has been well-established and published by Cheung *et al* (39). Different combinations of growth factors and small molecules were added to the culture medium to induce three intermediate tissue lineages (neuroectoderm, lateral plate mesoderm, and paraxial mesoderm) that give rise to the majority of SMCs in the body (39). Based on this study, it was recently shown that the embryological origin

of human VSMCs affects their ability to support endothelial growth formation (40). Specifically, it is demonstrated that lateral mesoderm-derived VSMC provide superior support to endothelial network formation compared to VSMC originating from the paraxial mesoderm and neuroectoderm (40).

1.1.4.4 NKX2-5 in embryogenesis

Nkx2-5 is one of the earliest markers of the cardiac lineage and its role in embryonic development has been studied extensively [reviewed in (41, 42)]. It is expressed in the primary heart field (PHF) as well as the SHF, the pharyngeal mesoderm and the pharyngeal endodermal cells underlying the SHF (43). Prall *et al* showed that *Nkx2-5* regulates SHF cell proliferation and outflow tract morphology, and demonstrated that *Nkx2-5* also orchestrates the transition between periods of heart induction, SHF progenitor cell proliferation, and outflow tract morphogenesis via a Smad1-dependent negative feedback loop (44). *Nkx2-5* alone or in co-operation with other factors such as Isl1 is also involved in the regulation and specification of cardiac progenitors towards the different myocardial lineages, and ensures proper acquisition of myocyte subtype identity (45).

Currently, the specific requirements of *Nkx2-5* expression in the different embryonic tissues with regards to heart development are incompletely understood. However, it has been shown that *Nkx2-5* expression is essential in the mesoderm, while endodermal expression is dispensable for early heart formation in mammals (46). In addition, *Nkx2-5* expression is found in the precardiac mesoderm and the pharyngeal endoderm at 7.8 dpc (days post-coitum) in the mouse embryo (43).

Targeted disruption of *Nkx2-5* in mouse embryos causes arrest in heart development after the initial stage of looping and embryos die by 9.0-11.0 dpc (41) due to growth retardation and abnormalities of the heart, including a failure of ventricular chamber development (47). Commitment to the cardiac lineage is not altered, however, expression of essential heart-specific genes is affected, such as *eHAND*, which is down-regulated in *Nkx2-5* deficient mice and is essential for the differentiation of embryonic ventricular myocardium (48).

The role of Nkx2-5 has also been studied at later stages of heart development using *Nkx2-5* conditional null mice. The animals exhibited chamber dilatation and progressive heart failure (49). After birth, *Nkx2-5* continues to be expressed in the post-natal heart, and its expression is significantly increased in hypertrophied hearts (43). In addition, mutations in the human *NKX2-5* gene lead to congenital heart disease (50).

1.1.5 Pulmonary vascular bed

The term “vascular bed” is used to describe a local vascular system of an organ or a part of the general cardiovascular circulation. Well-studied vascular beds include the pulmonary, cerebral, and renal beds. In addition, different vessels such as the coronary artery, the carotid, and the femoral artery can be also considered as different vascular beds. Although the arterial tree and the circulatory system look continuous, there is substantial heterogeneity among the different vascular beds resulting from significant anatomical and developmental differences, genetic factors and the local environment. This heterogeneity and phenotypic variability can explain the differential local responses to systemic risk factors and the propensity of certain vascular beds to develop CVD, such as atherosclerosis and hypertension. Indeed, clinical studies have demonstrated that abnormalities in both protein and cellular haemostatic regulatory elements are associated with specific risks in different vascular beds (51). In addition, Vanderlaan *et al* showed that differential gene expression occurs in different vascular beds in response to the local flow patterns (52).

The pulmonary vascular bed is the main focus of this thesis. It is a highly specialised vascular system entirely responsible for the delivery of oxygen to organs and tissues and the high volume of gas exchange (1). The pulmonary circulation receives the entire cardiac output during each cardiac cycle, and maintains a low blood pressure and low vessel resistance throughout the lung. These properties are critical to allow a delicate gas exchange and optimise the efficiency of the right ventricle. There are additional challenges that pulmonary circulation faces occasionally, and those include: the maintenance of low pressure in response to a dramatic increase in cardiac output, and the conversion of the pulsatile blood flow from the right heart into steady-state flow in the capillary bed (1). In order to perform these vital and orchestrated procedures, the pulmonary vasculature has developed complex structural properties and functions.

It was assumed that VSMCs residing in the pulmonary vasculature and specifically in the pulmonary arteries, are uniform in phenotype throughout the pulmonary circulation. It has now become apparent that these cells are phenotypically and functionally heterogeneous at a single anatomical site and along the pulmonary vascular bed. The heterogeneity is believed to rise from differences in origin, environmental factors, and spatiotemporal variability (53). At least four different and distinct populations of pulmonary artery smooth muscle cells (PASMC) were isolated from the media of the bovine main pulmonary artery (54). The sub-

populations were distinct in regards to morphology, expression profile, differentiation state, proliferation potential, and responses to growth factors and hypoxia (55). It has been suggested that these sub-populations would also exhibit different functions in health and disease. Heterogeneity has also been observed in human PASMCs isolated from different anatomical locations in the lung (56).

1.1.5.1 Pulmonary vessels

The main blood vessels of the pulmonary circulation are the right and the left pulmonary arteries, the arterioles, the capillaries, and the pulmonary veins [reviewed in (57)]. In embryonic development, pulmonary arteries derive from the truncus arteriosus. In the developed heart, the pulmonary artery (or also known as the trunk) originates from the right ventricle, and it expands into two branches: the left and right pulmonary arteries that deliver deoxygenated blood to the lung. Pulmonary veins are large blood vessels that receive oxygenated blood from the pulmonary capillaries of the lungs and return it back to the left atrium.

Pulmonary vessels exhibit a different structure and histology compared to the rest of the vascular beds, which is required to maintain the low pressures of the pulmonary blood circulation and low stiffness of the pulmonary vessels. The pulmonary arterial wall is generally thinner compared to the diameter of the vessel, with large amounts of elastin and smaller amounts of smooth muscle tissue. The pulmonary veins and the capillaries have very thin walls, and the capillaries are lined with endothelial cells. In contrast, the histology of the bronchial arteries is the same as that of other systemic arteries.

1.1.5.2 Vascular work load and disease

Scientists study the vascular load imposed on the right side of the heart by the pulmonary circulation as a cause of heart failure. There are several factors that contribute to the vascular load with the vast majority of it resulting from two events: the pulmonary vascular resistance (PVR) imposed by the arterioles and capillaries of the lung, and the pulmonary vascular stiffness (PVS) or the elasticity of the arteries that determines their capacitance. PVR is defined as the ratio of the drop in mean pulmonary arterial pressure (mPAP) to cardiac output. Elevated PVR leads to increased mPAP due to distal vasoconstriction, vascular remodelling, and thrombosis (58), and it is considered a defining cause of pulmonary hypertension (PH) (59). The contribution of vascular stiffening to the pulmonary vascular function has not been well studied (60). New imaging techniques such as cardiac MRI aim to correct this.

1.1.5.3 Control of the pulmonary vascular tone

There are various mechanisms that control pulmonary vascular tone that can be either “passive” or “active”. The passive mechanisms involve the recruitment and expansion of blood vessels (arterioles and capillaries) (61). Under normal conditions, not all of the pulmonary vessels are active. However, as pressure increases above the normal levels, the previously inactive vessels are now recruited to conduct blood. This response is immediate and leads to decreased PVR and normalisation of pulmonary circulation (61).

The active mechanisms involve two regulatory cell types, VSMCs and endothelial cells, and vasoactive stimuli that directly affect the vessel contraction and relaxation. The vasoactive stimuli can exert vasoconstrictor or vasodilator effects.

1.1.5.4 VSMC in the regulation of pulmonary vascular tone

The mechanisms that VSMCs regulate contraction and the pulmonary vascular tone in normal conditions were discussed in 1.1.4 and are reviewed in (18, 62). However, upon injury or in disease, the vasoactive stimuli travel through ion channels or membrane receptors on the endothelial barrier and reach VSMCs. Consequently, changes in calcium concentration in the intracellular and extracellular space are dynamic and highly regulated in a spatio-temporal manner (63). A large chemical gradient is required in order for calcium to enter the cytoplasm of VSMCs. Calcium is then removed from cells by two basic mechanisms: an ATP-dependent calcium pump and a non-ATP sodium-calcium exchanger (1). Other vasoactive substances can influence vascular tone by increasing the production of the second messengers, cAMP and cGMP, which can cause an increase in intracellular calcium. These second messengers are short-lived because they are degraded by phosphodiesterase (PDE) (64). Inhibition of PDE in VSMCs is a powerful tool to reduce vascular resistance, by prolonging the half-life of cAMP and cGMP.

1.1.5.5 Endothelial cells in the regulation of pulmonary vascular tone

Endothelial cells can produce various vasodilators [e.g. nitric oxide (NO), prostacyclin, endothelium-derived hyperpolarising factor (EDHF)], and vasoconstrictors [e.g. endothelin (ET)] in response to physiological or pathological stimuli. A few of the major vasoactive substances are discussed below.

Nitric oxide. NO is a vasodilator produced and released by the endothelium, and acts through cGMP to regulate potassium channels [reviewed in (65)]. NO is synthesized from L-arginine by NO synthase (NOS), which exists in 3 isoforms:

inducible NOS (iNOS) expressed mainly by macrophages, neuronal NOS (nNOS) expressed by neurons, and endothelial (eNOS). Synthesis of eNOS depends on calcium and although it is widely expressed in the endothelium, its action is strictly localised in the vascular bed where is produced. Thus, eNOS synthase is restricted to the pulmonary endothelium.

Prostaglandins. Prostaglandins are released from lung tissue and they participate in the regulation of pulmonary vascular resistance as vasodilators (66). Physiologically, prostacyclin, a member of prostaglandin family, is a local not circulating hormone, and its release causes relaxation of the underlying VSMCs and prevents platelet aggregation within the bloodstream (67). Prostacyclin exerts its effects through activation of cAMP-dependent pathways.

EDHF. Expression of potential EDHFs varies in species and in vascular beds, but all members act in a similar way by increasing potassium conductance resulting in the subsequent propagation or depolarisation and relaxation of VSMCs (68). Experimental studies suggest that the contribution of EDHF increases as the vessel size decreases, with predominant EDHF activity in the resistance vessels and a compensatory upregulation of EDHF when NO is not available (69). Although the role of EDHF has been studied and reported in many studies, the identity of these factors is not clear. Members of arachidonic acid derivatives, hydrogen peroxide, potassium channels, and the C-type natriuretic peptide have been proposed as EDHFs (70-72).

Endothelin. ET is a potent vasoconstrictor peptide that plays an important role in the regulation of pulmonary vascular tone. Three isopeptides have been identified: ET-1, -2, and -3, and all are synthesised by endothelin-converting enzyme (ECE). ET-1 is the predominant isoform in the cardiovascular system, and is produced by ECE-1. ET-1 is synthesised by a variety of different cell types, including endothelial cells, vascular and airway SMCs, leukocytes, macrophages, cardiomyocytes and mesangial cells. ET-1 is abundantly expressed in the pulmonary vasculature and exerts its major vascular effects through activation of two distinct G-protein coupled receptors, ETA and ETB (73). ETA receptors are located on VSMCs where they cause vasoconstriction, while ETB receptors are found on both endothelial cells and VSMCs (74). The activation of ETB receptors on endothelial cells causes vasodilation in contrast to the receptors on VSMCs that induce vasoconstriction. Overall, the effect of ET-1 depends mainly on its abundance and activity of the receptors on the responding cell, and usually there is a balance between production and clearance.

1.2 Cardiovascular disease (CVD)

1.2.1 Overview

CVD is a broad spectrum of conditions that involve complications of the cardiovascular system. Based on data published by the World Health Organisation (WHO) in 2012 and revised on September 2016, CVD is one of the leading causes of death representing 31% of all deaths worldwide (WHO, 2012) (75). In the UK specifically, CVD accounts for 26-28% of premature deaths based on BHF published data of 2015 (76).

Both genetic and environmental factors are implicated in CVD. Genetic composition is a significant risk factor, but the precise magnitude of the role of inheritance varies in different diseases and is not yet completely understood. For instance, in the general population, a history of premature atherosclerotic CVD in a parent confers ~3.0-fold increase in CVD risk to the offspring (77). In recent years, scientists have developed genetic testing to determine the risk of CVD using information about specific DNA variations that have been associated with the disease (78).

These tests also use information regarding environmental and lifestyle factors that have been proved to be risk factors for CVD. The most important lifestyle risk factors are the diet, smoking, alcohol, and inactivity. These risk factors directly affect the health output of individuals, and can be easily monitored through quantitative measurements of blood pressure, blood glucose and lipids, and obesity. These quantitative traits can indicate an increased risk of heart attack, stroke, heart failure and other complications. Indeed, smoking cessation, regular exercise and a healthier diet can decrease CVD risk. Such an example provides the study by Khan *et al*, where the effect of ApoE genotype in combination with various cardiovascular biomarkers was assessed on the risk of ischaemic stroke (79).

In addition, the male gender, the age as well as the ethnic background of an individual are determining factors for CVD. In particular, CVD is more common in people from south and middle Asia, African or Caribbean origin due to the prevalence of other CVD-associated conditions such as diabetes and high blood pressure (80). Other environmental and socioeconomic factors that affect CVD risk are stress (81) and poverty (82). More than 75% of CVD deaths occur in low and middle income countries (WHO, 2012). In these countries, people at high risk often do not have the benefit of primary health care, early detection and treatment compared to people in high-income countries. Consequently, people at high risk are

detected late and die younger of CVD during their most productive years. In such circumstances, education and counselling are absolutely vital, as are the early detection, management and treatment of disease.

Overall, the impact of CVD in health and socioeconomic status needs to be the motivation for the research community to tackle the disease by focusing on two goals: prevention and improved therapy.

1.2.2 Cell types that contribute to CVD

The vascular wall constitutes a major element in the pathogenesis of CVD. Major cell types implicated in CVD include endothelial cells, VSMC, and fibroblasts. Other cell types with distinct roles in CVD are the macrophages, other inflammatory cells such as lymphocytes, platelets, and vascular wall progenitor and multipotent stem cells such as pericytes and mesenchymal stem cells [reviewed in (83, 84)]. In this thesis, the primary focus is on the role of VSMCs in the development of CVD. The role of vascular endothelial cells is also examined at a lesser extent.

Findings propose that dysfunctional VSMCs is the primary cause behind CVD, and new therapies target VSMCs to control and eliminate CVD (85). In disease, VSMC are vastly associated with vascular remodelling, a process of structural changes that underlies most vascular diseases. Vascular endothelial cell dysfunction is recently considered as the trigger for many forms of CVD. Endothelial dysfunction is closely related with endothelial-to-mesenchymal transition (EndoMT), a process that gains increasing attention, as one of the earliest processes implicated in many vascular diseases, and will be further discussed in section 1.2.4.

1.2.3 Vascular remodelling underlying CVD

The term “remodelling” refers to the responses of vasculature to potentially noxious haemodynamic, metabolic, and inflammatory stimuli. Under physiological conditions, remodelling maintains functional, compensatory and adaptive characteristics of the vasculature. Although remodelling can be regarded as a mechanism that naturally occurs with aging, early vascular remodelling is associated with cardiovascular morbidity and mortality. Processes involved in remodelling include fibrosis, hyperplasia of the arterial intima and media, changes in vascular collagen and elastin, endothelial dysfunction, and arterial calcification. Vascular remodelling is a common underlying mechanism found in many CVD including PH, atherosclerosis, and peripheral artery disease.

1.2.3.1 The mechanism of vascular remodelling

Vascular remodelling is a dynamic process that describes structural and functional changes of the vascular wall that occur in response to disease, injury, or aging. The process involves interactions among four different processes: cell growth, cell death, cell migration and changes in ECM (86). Vascular remodelling is set into motion by a variety of complex pathophysiological mechanisms that are closely related, and that influence both the cellular and non-cellular components of the vascular wall. Structural alterations include changes in the diameter of the lumen in resistance vessels, as well as changes in the dimension of the vessel wall itself (Figure 1.2). Clearly, these alterations directly affect vascular tone, blood pressure, and vessel compliance, the disruption of which leads to disease.

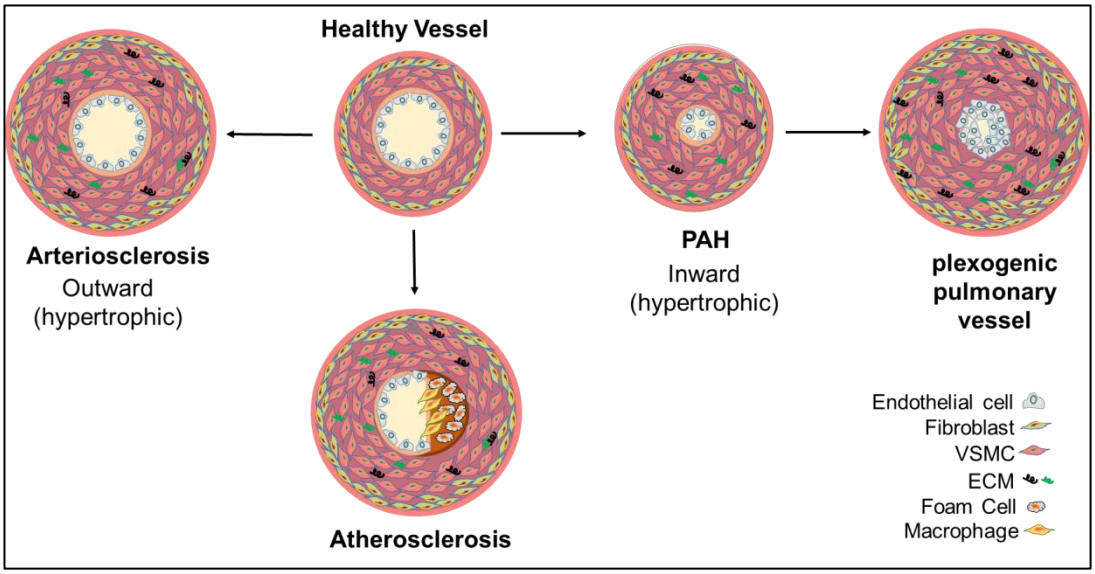


Figure 1.2 Structural changes in vascular remodelling. *Vascular remodelling describes structural alterations in the healthy vessel wall (centre) that affect the diameter of the lumen, as well as the dimension of the vessel wall itself. Although endothelial cells and adventitial fibroblasts, as well as many types of circulating cells are implicated in vascular remodelling, phenotypically de-differentiated VSMC are the key players. Different forms of vascular remodelling are shown in the figure mainly focused on hypertrophic remodelling. Remodelling in arteriosclerosis and in aortic aneurysms is mostly outward (left), described by medial expansion but with no change in the lumen of the vessel. In atherosclerosis (lower image), both inward and outward remodelling take place. During atherogenesis, the vessel adapts and is remodelled in a way that the lumen retains its diameter even after the formation of a plaque (expanding or outward remodelling). Constrictive remodelling is also seen in obstructive disease during plaque growth or shrinkage of the local vessel segment (87). Inward remodelling is often seen in hypertensive vessels, and in CAD. In early stages of PAH or in milder disease, inward vascular remodelling is caused by medial thickening leading to vasoconstriction and increased blood pressure in the lung. However, at later stages of disease, adventitial fibroblast hypertrophy and the development of plexiform lesions also occur. Plexiform lesions are caused due to non-structured hypertrophy of the endothelial cells causing almost absolute occlusion of the lumen. The development of plexiform lesions leads to disrupted blood flow and consequently tissue damage and necrosis in the lung.*

The morphological changes due to disease pathogenesis can change over time. Scientists have studied the structural changes in the context of disease, and particularly in atherosclerosis and PH. Vascular remodelling can either be inward or outward (Figure 1.2) (86). Inward or constrictive remodelling causes a narrowing of the lumen that decreases the maximal blood flow rate, whereas outward growth describes the abluminal expansion preserving the lumen and maintaining the blood flow. Inward remodelling contributes to atherosclerosis, hypertensive vascular disease, restenosis after angioplasty, and coronary artery disease (CAD). Outward remodelling mainly contributes to the formation of aortic aneurysm; however, it can also compensate for the atherosclerotic plaque growth (Figure 1.2). The changes in the amount or the characteristics of the material within the vessel wall is another feature of remodelling, with hypertrophic remodelling exhibiting an increase in the amount of material, and hypotrophic remodelling exhibiting a reduction (86).

1.2.3.2 VSMC phenotypic modulation

During vascular remodelling, the contractile VSMC de-differentiate to a disease-associated *synthetic* phenotype (Figure 1.3). The phenotypic modulation was originally based on morphological criteria, but alterations in functional and structural properties in response to environmental cues are profound. Subtle changes in gene and protein expression and signalling mechanisms have been extensively studied and well documented (32, 88). Synthetic VSMCs in vascular remodelling are characterised by loss of contractility, hypertrophy and hyperplasia, increased proliferation and migration, and excessive production and deposition of ECM within the vessel wall, as well as altered susceptibility to apoptosis (Figure 1.3) (89).

Hypertrophy describes an increase in cell size and DNA content with or without DNA synthesis. The dominant hypertrophic mechanism seems to be cell enlargement as a consequence of increased intracellular protein and increased intracellular water content, without DNA synthesis (53). Increased intracellular protein amount occurs due to activated protein synthesis and inhibition of protein degradation (90). Hypertrophy can be reversible, but only if DNA synthesis is not activated. Hypertrophy can be induced by G protein-coupled receptors (GPCR) agonists, such as angiotensin II (ANGII), ET-1, thromboxane A2 (TXA2), and through many autocrine growth factor mechanisms (91).

Proliferation is an ordered sequence of tightly regulated events during which synthetic cells duplicate and divide in response to growth factors. For most of the

cells in homeostasis, there is no need for proliferation (92). However, in vascular remodelling the need is significantly increased.

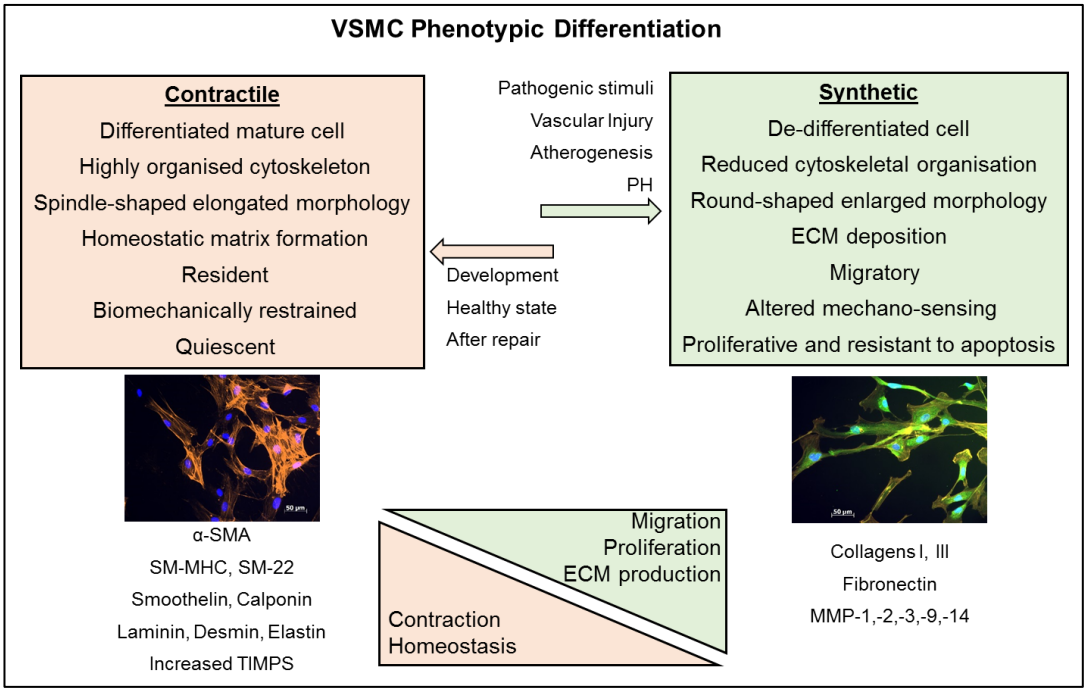


Figure 1.3 VSMC Phenotypic Differentiation. *VSMC are characterised by remarkable plasticity that allows the cell to acquire different phenotypes depending on the developmental stage, the environmental stimuli, and the needs and requirements of the cell. During development and in health, VSMC exhibit their contractile function and contribute to homeostasis of the normal blood vessels. However, in disease or following vascular injury, VSMC de-differentiate from a mature contractile cell to a less differentiated synthetic phenotype. During this phenotypic switch, the morphology, the expression profile and the functions of the cell change. The contractile cell displays a spindle-like elongated shape with an organised cytoskeleton, which is a non-migratory resident and quiescent cell with high expression of contractile proteins, such as α-SMA, SM-MHC, and calponin. Upon de-differentiation, synthetic cells proliferate, migrate and produce ECM, which is deposited within the vessel wall causing medial thickening. Synthetic cells are bigger than the contractile cells with reduced cytoskeletal organisation. The phenotypic switch is also followed by an altered expression profile, with reduced expression of contractile proteins and increased expression of ECM components and MMPs. Contractile cells stained for α-SMA expression (orange) and a synthetic cells stained for COL1A2 expression (green) are shown in the figure.*

Migration of VSMCs occurs in development, injury and vascular remodelling. In the normal vessel wall, VSMCs are non-migratory. But in vascular remodelling, different stimuli induce cell migration, and a panel of structural proteins that are required for cell migration is expressed. When a cell is stimulated to migrate, actin polymerises and creates a structure within the cell that is used as a guide for cell organisation to lead migration (93). The presence of heparin and tissue inhibitors of matrix metalloproteinases (MMP) in the ECM strongly inhibits cell migration. However,

many peptide growth factors, cytokines and components of ECM have been identified as pro-migratory molecules, including platelet-derived growth factor (PDGF), epidermal growth factor (EGF), FGF2, IL-6, collagens I and IV, and fibronectin (91). Other physical factors such as blood flow, shear stress and matrix stiffness can also influence cell migration.

In vascular remodelling, VSMC are also characterised by increased resistance to apoptosis. A proper balance between cell proliferation and cell death is required for the normal development and function of tissues. In disease state though such as PH, this balance is disturbed through unregulated function of ion channels. For instance, inhibition of potassium channels blocks apoptosis in PH (94).

Morphological changes are correlated with a switch in the cytoplasmic apparatus that includes a highly organised cytoskeleton with defined F-actin filaments, nuclear hypertrophy, and enlarged Golgi (95). Contractile and synthetic VSMCs express different marker proteins specific to each phenotypic state, as part of the mechanism that regulates the de-differentiation procedure (Figure 1.3). There is decreased expression of contractile-associated proteins including α-SMA, SM-MHC, calponin, and desmin. The search for synthetic-specific protein markers has been really challenging and disappointing, although expression of non-muscle isoforms and ECM components are increased (29, 96). A useful “definitive” marker of the synthetic state of VSMC expressed in vascular injury, atherosclerosis and during development is the non-muscle myosin heavy chain or SMemb (97).

Sometimes though, the changes observed in protein levels do not correlate well with the morphological changes of the phenotype. Indeed, there is evidence suggesting that VSMCs balance between the contractile and synthetic states and the cells retain characteristics of both phenotypes, and whether one state predominates over the other is under strict and complex regulation. However, Owens *et al* has proposed that VSMCs are able to acquire a wide spectrum of possible phenotypes under different physiological and pathological circumstances that are difficult to define. In vascular development for example, VSMCs retain contractile properties but simultaneously participate in vessel growth and remodelling (30).

Although VSMC phenotypic switching in vascular remodelling is best described in atherosclerosis and PH, additional examples of diseases associated with SMC function include asthma, gastrointestinal and reproductive disorders. However, it is important to remember that phenotypic modulation is applicable to all SMC or SMC-like cells irrespective of their origins or location in the body, and depends on the

environmental cues that influence the behaviour of all SMC under different circumstances (30).

Defective SMC differentiation is also found in many forms of cancer. Although it is well known that solid tumours require angiogenesis and development of circulatory supply, it is less well discussed that the newly developed blood vessels are immature or defective, greatly enlarged and leaky with limited support from VSMCs (98). In addition, in many cases the VSMCs present appear to be abnormal and to lack the ability to express the appropriate repertoire of specific markers (99). The overall mechanisms responsible for defective SMC-pericyte investment in tumours are very poorly understood, but it seems that the problem relates to abnormal recruitment and differentiation of VSMCs and precursor cells.

Work in our lab has recently identified the *NKX2-5* gene as an important modulator of vascular remodelling in mouse and human. Our findings are further discussed in detail in section 1.4.

1.2.3.3 Regulation of VSMC phenotypic modulation

Despite the volume of evidence that vascular remodelling is a key component in the pathogenesis of many diseases, information regarding the molecular and signalling mechanisms and the environmental cues that activate the phenotypic modulation remain elusive. A literature review reveals that complex interactions between growth factors, inflammatory cytokines, vasoactive substances and haemodynamic stimuli in the vessel wall control remodelling.

I*n vitro* studies have revealed a large number of factors controlling the VSMC phenotypic switch, including mechanical forces, contractile agonists, ECM components such as collagen I and IV, neuronal factors, ROS, endothelial-SMC interactions, thrombin, interleukins, and growth factors (PDGF, FGF, EGF, IGF, TGF-β) [reviewed in (30, 100)]. In addition, gene knock-out studies in mice have implicated a number of factors and pathways mostly focused on components of TGF-β signalling pathway (Figure 1.4). Most of the gene deletions were embryonic lethal, however, it is not clear whether the lethality was caused due to defective vascular maturation or inability of VSMCs to differentiate. Since the main focus of this study is the pulmonary vasculature and pathology, growth factors and mechanisms that have been implicated in abnormal VSMC proliferation and migration contributing to pulmonary vascular remodelling are discussed in more detail.

Platelet-derived growth factor (PDGF). Peptide growth factors, and in particular PDGF, elicit their signals through highly selective tyrosine kinase receptors and seem to play a prominent role in vascular remodelling. PDGF levels are elevated in PAH (101), and this factor has also been studied in atherosclerosis and restenosis. PDGF was initially identified as a highly specific and selective inhibitor of VSMC de-differentiation, as treatment of rat aortic SMC with PDGF-BB was associated with rapid downregulation of contractile markers including α-SMA, and SM-MHC (102). PDGF is now regarded as the most potent mitogen for VSMCs. In fact, it was recently shown that genetic ablation of PDGF receptor β (PDGFRβ) in mice limited PDGF-dependent downstream effects including extracellular signal-regulated kinases 1/2 (ERK1/2) and Akt phosphorylation, cyclin D1 induction, proliferation, migration, and protection against apoptosis, ultimately blocking vascular remodelling and PAH (103). PDGFs bind their tyrosine kinase receptors to activate the downstream signalling pathway, which involves the PI3K and ERK1/2 kinases. Inhibition of PDGFR signalling may be achieved by tyrosine kinase inhibitors, such as Imatinib. Imatinib was developed to target the Bcr/Abl oncogene, c-kit, and the PDGFRα and β, and it has been effective in the treatment of chronic myeloid leukaemia and PAH (104).

Epidermal growth factor (EGF). The EGF receptor (EGFR) signalling pathway is one of the most important pathways that regulate growth, survival, proliferation, and differentiation in mammalian cells. Fifteen members of the endogenous EGF ligand family have been identified that bind the 4 ErbB family receptors (EGFR; ErbB1-4) and induce their homo- and hetero- dimerisation. EGF is a growth factor that mediates pulmonary vascular remodelling through PASMC proliferation, migration and resistance to apoptosis (105). Heparin-binding EGF, mainly expressed in the airway epithelium, is a potent mitogen and chemotactic factor for SMC, and it is significantly upregulated in asthmatic airways (106).

Fibroblast growth factor 2 (FGF2). FGF2 is a member of a large family of heparin-binding growth factors involved in proliferation and differentiation of various cells and tissues. It is produced by various cells types including fibroblasts, endothelial cells and macrophages. FGFs mediate their biological activity by binding to four cell surface receptor tyrosine kinases, designated FGF receptors (FGFR1-4). Ligand-receptor interaction results in assembly of a complex that involves Frs2, a FGFR adaptor protein, Grb2 and Sos, and leads to the activation of Ras/mitogen-activated protein kinase (MAPK) signalling cascade. FGFs promote SMC conversion to the proliferative phenotype and stimulate growth and migration (107). FGF2 has also

been implicated in the initiation and progression of PH by promoting PASMC proliferation (108).

Transforming growth factor-β (*TGF-β) signalling pathway.* The TGF-β superfamily is a large group of cytokines that control many cellular processes with respect to vascular remodelling (Figure 1.4). The superfamily consists of three different TGF-β isoforms and more than twenty BMPs (Figure 1.4). Downstream signalling requires the partnership of type I (TGF-βR2 and BMPR2) and type II receptors (ALK1-7), followed by the activation of SMAD dependent and independent cascades (Figure 1.4). The role of the BMPR2 receptor has been extensively studied in vascular disease, and especially in PAH. Genetic mutations have been found in the *BMPR2* gene that cause loss-of-function and reduced TGF-β downstream signalling, and are associated with familial and idiopathic cases of PAH. Dysregulated BMPR2 and TGF-β signalling contribute to the pathogenesis of PAH (109).

Under defined circumstances, TGF-β signalling regulates SMC proliferation and apoptosis, cellular differentiation and activation, inflammation, and synthesis of ECM. The role of TGF-β in disease is reviewed in (110). Although TGF-β is a critical component of vascular remodelling, its exact roles remain vague with controversial published data. In contrast to PDGF and FGF, it has been proposed that TGF-β promotes VSMC switch towards the contractile phenotype (111). Although the pro-fibrotic role of TGF-β has been well established, there is an active debate regarding the proliferative/migratory roles in vascular disease. A good example is atherosclerosis, where although some studies suggest a pro-atherosclerotic effect, other studies propose that TGF-β inhibits atherosclerosis and atherosclerotic progression (110, 112-120).

TGF-β activates the expression of VSMC-specific transcription factors such as serum response factor (SRF) and myocardin through phosphorylation of Smad2/3 (121), to further modulate transcription of contractile specific markers such as SM22 (122). Overexpression of TGF-β1 stimulates neo-intimal hyperplasia, fibrosis and remodelling (123). The TGF-β pathway promotes VSMC differentiation, which is critical for the maintenance of the normal adult vasculature (124), the balance between inflammation and fibrous plaque growth in atherosclerosis (125), effective wound healing and reduced scarring (126). In vascular injury, TGF-β upregulation increases intimal thickening and the effect is Smad-mediated with the Smad mediators being overexpressed at the sites of injury, and also associated with intimal thickening (127). It has also been shown that TGF-β in co-operation with BMP2/4/7 inhibits PASMC proliferation and this effect is Smad1-dependent (128),

whereas its anti-mitogenic effect is ALK5-mediated through Smad3 and p38 kinase (129). In a similar pattern, the TGF-β-Smad2/3 signalling cascade interferes with EndoMT *in vivo* and decreases both the neointimal formation and the relative contribution of endothelial lineage-derived cells to the neointima (130). However, work from a different group has established the profibrotic functions of TGF-β in the microenvironment of a neointimal tissue with enhanced matrix synthesis coupled with diminished MMP-mediated matrix degradation (131). Genetic deletion of various components of TGF-β signalling cascade has led to disease, such as aneurysm formation (124), atherosclerosis (132), restenosis and vascular injury (127), and PH (91).

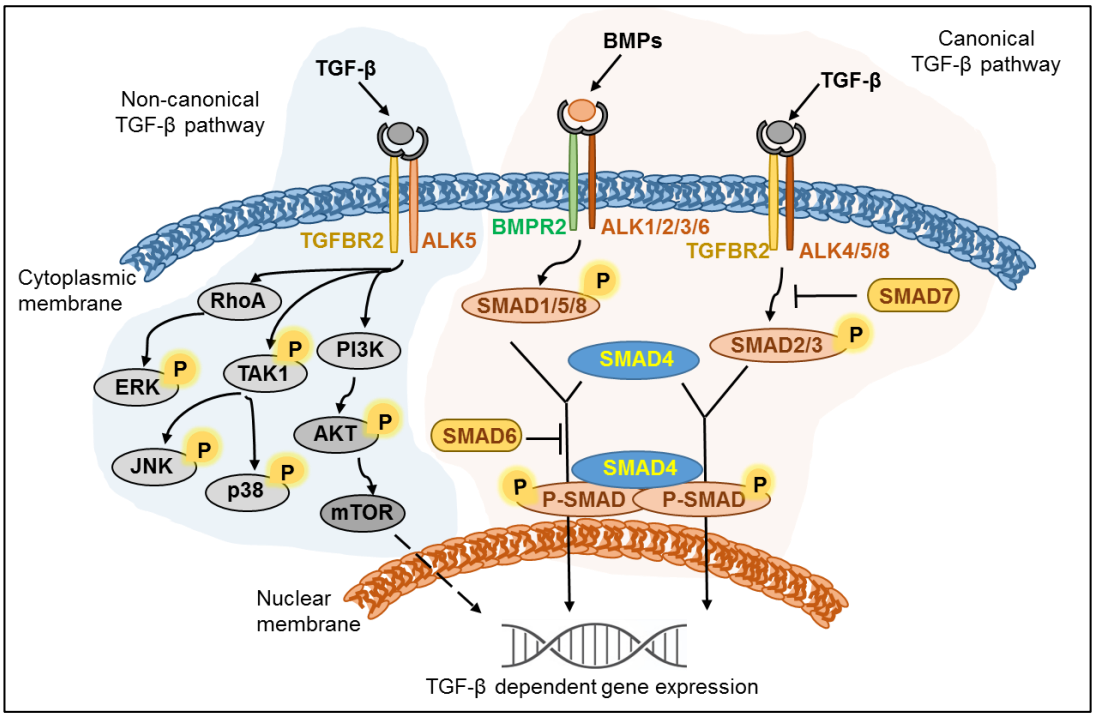


Figure 1.4 TGF-β-mediated signalling pathways. *The TGF-β superfamily consists of 3 TGF-β isoforms (TGF-β1/2/3) and more than 20 BMPs. Downstream signaling requires the partnerships of transmembrane serine/threonine kinase receptors that are divided into two classes: type I (TGF-βR2 and BMPR2) and type II (ALK1-7). TGF-β usually exerts its functions and regulates gene expression through hetero-tetrameric complexes of receptors type I and II. The complexes lead to activation of type I receptors that recruit and subsequently phosphorylate and activate Smad signalling molecules. ALK1/2/3/6 interact with Smad1/5/8, whereas ALK4/5/7 interact with Smad2/3. Upon phosphorylation and activation, Smad molecules interact with Smad4, and complexes migrate to the nucleus to regulate the transcription of downstream target genes. Apart from Smad-dependent or canonical TGF-β pathways, TGF-β induces SMC differentiation through Smad-independent pathways transmitting signalling through the MAPK cascade, involving the kinase PI3K, AKT, ERK, TAK1, JNK, and p38 (110, 124).*

FGFs have been reported to antagonise the TGF-β-mediated induction of smooth muscle markers expression in pericytes and SMCs (133), by repressing TGF-β

signalling (132). *In vitro* inhibition of FGF signalling upregulates TGF-βR1 and downstream Smad2/3 activity, whereas *in vivo* reduces SMC proliferation (132). These observations suggest that FGF signalling regulates the TGF*β*-dependent proliferative SMC phenotype modulation towards the contractile phenotype (119).

Studies in our lab have shown that TGF-β has a pro-fibrotic role and it promotes proliferation and migration of SMCs and fibroblasts in fibrotic conditions such as scleroderma, PAH, and fibrotic kidney disease (134-136). In addition, in fibroblasts TGF-β activates expression of *Col1a2* gene through a non-Smad-dependent signalling pathway (137), and expression of *Ctgf* gene through a TGF-β responsive element in the gene promoter (134). A part of this thesis aims to explore the effects of TGF-β signalling in PASMC in vascular remodelling in regards to: 1) the phenotypic switch of PASMC, and 2) the downstream target and signalling cascade leading to vascular remodelling.

Angiotensin II (AngII). The renin-angiotensin-aldosterone System (RAAS) is a central component of the physiological and pathological responses of the cardiovascular system, and its role has been studied extensively (138-140). AngII is a critical vasoactive factor that stimulates VSMC contraction under physiological circumstances, however, persistent elevation of AngII levels drives VSMC growth (140). In cultured aortic SMCs, AngII transactivates EGFR signalling through ADAM17, a plasma membrane metalloproteinase (141).

Notch signalling pathway. The Notch signalling pathway retains a prominent role in the regulation of vascular development and maintenance, by affecting the expression of SMC differentiation markers in a cell context-dependent manner. For example, the Notch downstream transcription regulator RBPL binds the *α-SMA* promoter to activate its expression (142). However, in a different setting of the activation of the canonical Notch pathway, its target genes HERP/HEY were shown to interact with the SRF/Myocardin complex to inhibit expression of the SMC differentiation marker (143). These data suggest the involvement of Notch signalling pathway in SMC phenotypic switch, but the exact regulatory effects remain obscure.

Redox regulation. Reduction-oxidation (redox) signalling retains a major regulatory role in the maintenance of cellular homeostasis, and in physiological adaptive responses of endothelial cells and VSMCs. Disruption in redox homeostasis leads to redox stress and ultimately to maladaptive vascular remodelling [reviewed in (144)]. Dysregulation of redox signalling impairs endothelial cell proliferation, promotes apoptosis and induces vascular hypertrophy leading to vascular

remodelling. In VSMCs, redox signalling controls the phenotypic de-differentiation of the cells through the transcription of contractile proteins, modulation of adhesion, migration and proliferation.

Transcriptional Regulation. The VSMC phenotypic switch is also regulated at the gene expression level. Work in transgenic mice has revealed that the transcription of SMC-specific genes depends on CArG DNA binding elements. CArG elements have been found in promoters and enhancers in nearly 200 muscle-specific genes that are involved in the formation and regulation of the cytoskeleton and contractile apparatus (145). At least one CArG element has been found in the genes specific for the contractile phenotype including α-SMA, SM22, calponin, and SM-MHC (146). SRF, activated by serum stimulation, bind CArG elements as a dimer. Apart from smooth muscle genes, SRF also regulates transcription of skeletal and cardiac muscle-specific genes.

Many of the genes identified to date to be under control of CArG/SRF, contain more than one CArG element in their promoter/enhancer sites, suggesting functional interactions between multiple elements. These interactions may allow for spatio-temporal regulation of the expression of smooth muscle-specific genes. The ability of SRF to exert its function depends on the presence of program-specific co-activators or co-repressors. Myocardin is an extremely potent SRF co-activator that is selectively expressed in cardiac and differentiated SMCs (147). In addition, myocardin has been shown to selectively induce expression of all CArG-dependent SMC markers including α-SMA, SM22α, calponin, and SM-MHC (30, 122, 147-150). Myocardin knock-out mice are embryonic lethal by E10.5, with no evidence of vascular differentiation, whereas *in vitro* knock-down of myocardin expression significantly reduced expression of multiple SMC marker genes [reviewed in (151)]. Although evidence suggest that endogenous myocardin expression is required for regulation of expression of multiple smooth muscle-specific genes, more studies are required to elucidate its exact role.

1.2.4 Endothelial to mesenchymal transition (EndoMT)

Endothelial cells are the interface between the blood vessels and the tissues and their main functions include sensing of the environment and the signal modulations of vascular function, maintenance of homeostasis, and defence against injury. Many studies provide evidence that the pulmonary endothelium, specifically, is a critical local source of several key mediators for vascular remodelling, including growth factors (FGF-2, serotonin, AngII), and vasoactive peptides (NO, PGI2, ET-1),

cytokines (IL-1, IL-6), chemokines, and adipokines [reviewed in (152)]. Inappropriate signalling from vascular endothelial cells contributes to CVD, and endothelial dysfunction is thus a hallmark of human diseases.

Endothelial dysfunction is characterised by reduced vasodilation, increased endothelium-dependent contraction, cell proliferation, platelet activation, vascular permeability, and a pro-inflammatory and pro-thrombotic phenotype, including leucocyte-endothelial interactions that participate in vascular inflammation and increased adhesion and aggregation of platelets (153). Endothelial dysfunction occurs in association with several cardiovascular risk factors, such as hypercholesterolemia, hypertension, and insulin resistance (154). Overall, it promotes vascular inflammation by inducing the production of vasoconstrictor agents, adhesion molecules, and growth factors, as well as vascular remodelling.

1.2.4.1 The mechanism of EndoMT

Like VSMCs, endothelial cells are also capable of changing their phenotype in response to various stimuli during a process known as EndoMT. EndoMT is a multipotent procedure during which vascular endothelial cells are stimulated by various growth factors, inflammatory cytokines, and signalling molecules to undergo a phenotypic transition towards a mesenchymal-like cell. Endothelial-derived mesenchymal cells take on properties of multipotent stem cells and can differentiate into synthetic VSMCs, myofibroblasts, pericytes, skeletal muscle, cardiac muscle, bone, cartilages, and fat cells (155). A loss of cell-cell contact seems to be a triggering step in the development of EndoMT (156). During EndoMT, the basal lamina of the endothelium is replaced by ECM composed of type I and III collagens and fibronectin (157). The new matrix promotes motility, and triggers endothelial cells to detach and migrate. As cells migrate away from the monolayer, their cortical cytoskeleton is rearranged to enable cell motility through the development of actin-rich projections such as lamellipodia or filopodia. They also contract away from the adjacent cells increasing gap junctions causing “leaky” barrier function. In addition, their morphological characteristics change, as they lose their “cobblestone”-like phenotype and they become more elongated (158).

Gene and protein expression profile is also changed. The cells lose expression of specific endothelial markers CD31/PECAM-1, von Willebrand (vWF), and VE-cadherin, and at the same time they initiate expression of mesenchymal specific markers including α-SMA, vimentin, type I collagen (158). Loss of VE-cadherin

expression, a cell-cell adhesion glycoprotein, consistently precedes the downregulation of the endothelial phenotype and SM-like transformation (156).

1.2.4.2 EndoMT in disease

Although EndoMT has been demonstrated in normal physiological processes during development, various *in vitro* and *in vivo* studies have shown that endothelial cells undergo EndoMT and contribute to vascular remodelling (159). EndoMT has been studied extensively in PH and pulmonary vascular remodelling, where it contributes significantly to smooth muscle-like cells that migrate to the site of vascular injury (159). *In vitro*, smooth muscle-like cells derived from EndoMT are indistinguishable from cells of other origins, since they exhibit similar morphological characteristics and express abundant mesenchymal markers. Numerous studies have reported EndoMT in various animal models of fibrosis, as well as human conditions associated with tissue fibrosis [reviewed in (158)], where it was shown that EndoMT provides another source of activated myofibroblasts contributing to the progression of fibrosis. Human fibrotic conditions with an established role of EndoMT include pulmonary and cardiac fibrosis, idiopathic PH and PH associated with scleroderma [reviewed in (158)].

Despite these observations, however, the case of EndoMT participating in the development and progression of human fibrotic diseases is considered controversial (160). In this regard, it is important to note that not-complete but partial trans-differentiation of endothelial cells to fibroblasts/SM-like cells could be sufficient for the initiation and progression of pathological fibrogenesis (161).

1.2.5 CVD characterised by vascular remodelling

Vascular remodelling is a common pathological feature underlying various CVD and it has been extensively studied in many occasions including atherosclerosis, CAD, peripheral artery disease (PAD), PH, PAH, and connective tissue diseases (CTD) such as scleroderma (SSc), systemic lupus erythematosus (SLE), and rheumatoid arthritis (RA) [reviewed in (162-167)]. This thesis is primarily focused on PH, which will be discussed in detail in section 1.2.5.3. But first, I will briefly introduce atherosclerosis and PAD as examples of conditions underlined by vascular remodelling and as reference to following chapters.

1.2.5.1 Atherosclerosis

Atherosclerosis is a gradual and progressive disease, and possibly the most common occlusive disorder of CVD. Major clinical consequences are myocardial

infarction and stroke caused by thrombotic events due to rupture of unstable plaques. Disease pathogenesis is characterised by acute inflammatory responses, dysfunctional metabolism, vascular and arterial remodelling. Inflammation is often the first trigger in atherosclerosis, followed by remodelling which leads to the formation of plaques and lesions (168), a process known as pathological intimal thickening (169). Macrophages, macrophage-like cells, as well as VSMCs and VSMC-derived cells are the key players in atherosclerosis, and their preponderance determines the fate of the plaque and therefore the progression of the disease (170). The role of VSMCs in atherosclerosis is complex and studies report contradictory data. For years, it was thought that VSMC phenotypic switch drove atherogenesis contributing to remodelled vessels and plaque formation. In contrast, VSMCs also retain an athero-protective role as they support plaque stability and repair and inhibit rupture. Among transcription factors that regulate VSMC phenotypic switching in atherosclerosis are myocardin, which activates the expression of contractile marker genes (171), and KLF4, which mediates the transition of VSMC to macrophage-like cells (172).

The degree and the type of arterial remodelling in atherosclerosis (Figure 1.2) depends on the size of the forming plaque (173). Arterial remodelling appears to represent a co-ordinated stress response with controlled and transient activation of proinflammatory signalling pathways. A primary signal for arterial remodelling is shear stress, which is a haemodynamic force that under normal conditions promotes expression of vasodilator and anti-thrombotic factors, suppresses growth and pro-inflammatory factors, and generally maintains a state of vascular health. However, a macroscopic increase in blood flow increases local shear stress and stimulates arterial expansion and remodelling. The endothelium responds to shear stress through mechanisms that include PI3K/Akt cascade, NO synthesis, and the MAPK signalling cascade (174).

Other signalling mechanisms involved in remodelling in atherosclerosis include the Notch and TGF-β signalling pathways. TGF-β is critical for maintaining the balance between inflammation and fibrous plaque growth in atherosclerosis (125). Simons *et al* have recently published important data regarding the signalling regulation of atherosclerosis (13, 114, 119, 132). Briefly, they demonstrated the FGF-dependent regulation of TGF-β, and they showed that FGF regulates VSMC phenotypic modulation by controlling TGF-β signalling and the contribution of VSMC proliferation to the growth of atherosclerotic plaque. This effect is mediated by let-7 microRNA, which is decreased due to the loss of FGF causing prolongation of TGF-

β-dependent signalling. Increased TGF-β signalling leads to EndoMT that in turn accelerates the progression of atherosclerosis (132).

1.2.5.2 Peripheral arterial disease (PAD)

PAD is a progressive atherosclerotic condition that causes stenosis and occlusion of non-cerebral and non-coronary arteries (175), including those found in the extra-cranial carotid circulation, mesenteric circulation, renal circulation and the upper and lower extremities (176). The progressive reduction in arterial blood flow can lead to claudication, rest pain in the leg or foot, tissue loss, non-healing wounds or ulcers, infection and gangrene (176). Complications may result in chronic limb ischemia, which is a severe form of PAD predominantly caused by atherosclerosis in the peripheral arterial system, amputation and an increased risk of death. Vascular remodelling in muscular peripheral vessels in PAD is more often inwardly hypertrophic, probably reflecting sustained vasoconstriction of vessels.

The major risk factors for PAD have been determined from large epidemiological studies and are consistent with the risk factors for cerebrovascular disease and ischaemic heart disease including, but not limited to, advanced age, smoking, diabetes, hypertension and hyperlipidaemia (177). The disease progresses very slow, and although some patients follow a gradual progression from asymptomatic PAD to intermittent claudication and critical limb ischaemia, others can progress to limb amputation without exhibiting any symptoms prior to surgery (175). Due to the high prevalence of asymptomatic disease and the small percentage of PAD patients presenting classic claudication, PAD is frequently underdiagnosed and thus undertreated. Therefore, early clinical diagnosis remains a challenge. Evaluation is performed with non-invasive physiological vascular studies such as segmental arterial pressures, pulse volume recordings, and Doppler echocardiography.

PAD patients also suffer depression, a reduced quality of life, and a significantly higher risk of cardiovascular events. Based on epidemiological projections, 27 million people in Europe and North America (16% of the population aged ≥55 years) have PAD (178). The 5-year rate of non-fatal cardiovascular events, including myocardial infarction and stroke, for patients with symptomatic PAD is approximately 20%, and the 5-year mortality is 15%–30% (179).

1.2.5.3 Pulmonary hypertension (PH)

1.2.5.3.1 Introduction and classification

PH contributes significantly to CVD, and is usually associated or caused by a variety of underlying conditions. Vascular remodelling and PVR are common features in PH (91). Other processes contributing to pulmonary vascular remodelling involve endothelial dysfunction and perivascular inflammation (159).

Clinical classification criteria of PH were re-considered during the 5th World Symposium held in 2013 in Nice, France (180, 181) to individualise different categories of PH sharing similar pathological findings, similar haemodynamic characteristics and similar management (Table 1.1). Five groups of disorders that cause PH were defined: Group 1: PAH, Group 2: PH due to left heart disease, Group 3: PH due to lung diseases and/or hypoxia, Group 4: Chronic thromboembolic PH, Group 5: PH with unclear multifactorial mechanisms.

1.2.5.3.2 Pulmonary arterial hypertension (PAH)

PAH is a complex life-threatening progressive disorder with poor prognosis associated with high morbidity and mortality due to abnormally elevated pulmonary pressure and right heart failure. The disease prevalence in the general Caucasian population is 15 cases per million, and the median age of those affected is 37–50 years (182), although recent reports suggest identification of significant PAH in elderly patients which is probably associated with aging (183). Both children and adults can be affected, and the disease is observed more commonly in females than males, at a ratio of 1.9-4 females: 1 male (184).

PAH is often slow to be diagnosed with an estimated mean time of ~2years between the onset of symptoms and diagnosis, and can be fatal with a mean survival of untreated PAH patients of ~2.8 years (185).

1. Pulmonary arterial hypertension	
1.1 Idiopathic	
1.2 Heritable	1.2.1 BMPR2 mutations 1.2.2 Other mutations 1.2.3 Unknown
1.3 Drug- and toxin- induced	
1.4 Associated with	1.4.1 Connective tissue disease (SSc, SLE, RA) 1.4.2 HIV infection 1.4.3 Portal hypertension 1.4.4 Congenital heart disease 1.4.5 Schistosomiasis
1'. Pulmonary veno-occlusive disease and/or pulmonary capillary haemangiomatosis	
1'.1 Idiopathic	
1'.2 Heritable	1'.2.1 EIF2AK4 mutation 1'.2.2 Other mutations
1'.3 Drug, toxin and radiation induced	
1'.4 Associated with	1'.4.1 Connective tissue disease 1'.4.2 HIV infection
1''. Persistent PH of the newborn	
2. Pulmonary hypertension due to left heart disease	
2.1 Left ventricular systolic dysfunction	
2.2 Left ventricular diastolic dysfunction	
2.3 Valvular disease	
2.4 Congenital/acquired left heart inflow/outflow tract obstruction and congenital cardiomyopathies	
2.5 Other	
3. Pulmonary hypertension due to lung disease and/or hypoxia	
3.1 Chronic obstructive pulmonary disease	
3.2 Interstitial lung disease	
3.3 Other pulmonary diseases with mixed restrictive and obstructive pattern	
3.4 Sleep-disordered breathing	
3.5 Alveolar hypoventilation disorders	
3.6 Chronic exposure to high altitude	
3.7 Developmental lung diseases	
4. Chronic thromboembolic pulmonary hypertension and other pulmonary artery obstruction	
4.1 Chronic thromboembolic pulmonary hypertension	
4.2 Other pulmonary artery obstructions	
5. Pulmonary hypertension with unclear and/or multifactorial mechanisms	
5.1 Haematological disorders	
5.2 Systemic disorders	
5.3 Metabolic disorders	
5.4 Others	

Table 1.1 Classification of pulmonary hypertension. *Sub-categorisation of Groups 1-5 of pulmonary hypertension based on classification criteria defined during the 5th World Symposium of PH, Nice, France in 2013 (180). Table adapted from (181).*

PH is defined as a resting mPAP >25mm Hg, or a mPAP with exercise >30mm Hg. PAH is clinically diagnosed when mPAP is maintained at ≥25 mm Hg with normal pulmonary arterial wedge pressure (PAWP) ≤15mm Hg and elevated PVR (180). Screening and diagnostic protocols involve both non-invasive and invasive techniques. Doppler echocardiogram has been used successfully for the approximate evaluation of mPAP, and its routine use has significantly facilitated the diagnosis of PAH. New modalities in non-invasive diagnostics include cardiac magnetic resonance that may offer more reliable data both at rest and during exercise, and screening for biomarkers such as the brain natriuretic peptide (BNP) and N-terminal-proBNP (59). However, the right heart catheterisation (RHC) remains the gold standard for the diagnosis of PAH (186). RHC is used for the accurate measurement of haemodynamic parameters such as mPAP, cardiac output, right atrial pressure, and PAWP.

1.2.5.3.3 Vascular remodelling in PAH

Vascular remodelling in PAH involves structural and functional changes in the normal architecture of the wall of the pulmonary arteries. It is associated with thickening of the vessel wall, increased muscularisation of the muscular and non-muscular arteries and formation of intimal and plexiform lesions (Figure 1.2). Phenotypic de-differentiated PASMCs underlie these changes through hypertrophy, proliferation, migration, and resistance to apoptosis. The changes affect almost all the arteries of the pulmonary arterial tree. In the proximal muscular arteries, vascular remodelling leads to reduction in the diameter of the lumen of the vessel, whereas in the large conducting pulmonary arteries there is often a dilatation and loss of elastic properties. Extensive production of ECM and cell proliferation lead to formation of neointima within the vessel wall. In addition to PASMCs, activated fibroblasts/myofibroblasts, also contribute to neointima formation (187). Plexiform lesions are often seen in severe forms of PAH. These structures are formed by endothelial channels initiating from the intima in the simultaneous presence of α-SMA expressing cells that produce matrix (Figure 1.2) (188).

Extensive vascular remodelling in PAH leads to right heart failure, which is the leading cause of death in PAH patients (189). The right side of the heart is responsible for pumping venous blood into the pulmonary circulation for oxygenation. The remodelled vessels cause increased PVR leading to elevated pulmonary pressure. This process interferes with the normal function of the right ventricle, which deteriorates as the disease progresses, ultimately leading to right

heart failure. Current therapies include vasodilator drugs that reduce PVR, relieve symptoms, and enhance exercise capacity (190). Although these therapies ameliorate the symptoms, they do not provide a cure and cannot stop disease progression, leaving heart or lung transplantation as the only option for end-stage patients. Ongoing efforts are focused on designing drugs that would potentially block vascular remodelling, revert the pathological mechanisms, and possibly regenerate microvessels in the lung (191).

1.2.5.3.4 Endothelial dysfunction and EndoMT in PAH

Endothelial dysfunction is thought to play a key role in PAH and to initiate the disease due to the loss of barrier integrity. Endothelial cells control vascular function and dysfunction through a fine balance of secretion and production of vasoactive substances, thrombotic mediators and inflammatory cytokines that affect PASMCs (192). When barrier function is compromised, these circulatory factors can reach the vascular media and drive PASMCs de-differentiation leading to vascular remodelling and disease progression (193). Indeed, when PASMCs were exposed to media conditioned by pulmonary artery endothelial cells (PAECs), cell proliferation and migration were enhanced, and the effect was greater when the conditioned media derived from PAECs of PAH patients (194).

The Rho family of GTPases are key regulators of the pulmonary endothelial barrier function, and their role in pulmonary vascular disorders is reviewed in (195, 196). In particular, Rho-GTPase signalling regulates cell motility and cytoskeleton, endothelial permeability, angiogenesis, NO production, smooth muscle contractility, proliferation, differentiation and apoptosis. Members of Rho family (RhoA RhoB, Rac1) are activated in the pulmonary vasculature of PH patients and animal models in response to hypoxia and stress, and they consist targets for prospective therapies (197).

The role of EndoMT in PAH has recently gained increasing attention. Studies from different groups have provided convincing experimental evidence that EndoMT occurs in PAH patients and animal models (rat and mouse) of induced PAH (14, 198, 199). Good *et al* demonstrated the presence of transitional EndoMT cells in the pulmonary vasculature of patients with SSc-PAH and in the hypoxia/SU5416 model of PAH in a small proportion of vessels (14). Similarly, Ranchoux *et al* also showed EndoMT is implicated in vascular remodelling in human PAH, and in rat models of severe induced PH (monocrotaline and SuHx) (199).

EndoMT is regulated by growth factors, inflammatory signalling, and the mechanical *in situ* environment [reviewed in (200, 201)]. At present, the best-described inducers of EndoMT are members of the TGF-β superfamily. Inflammation, oxidative stress, and changes in the mechanical factors imposed on the blood vessel, are also seen often in PAH. The inflammatory microenvironment observed in PAH contains a myriad of signalling proteins such as TNF-α, IL-1β, IL-6, and ROS, each of which may have individual effects on the EndoMT (201).

EndoMT is an important pathological process that may exacerbate vascular leak, inflammatory infiltration, and vascular remodelling in PAH. However, it is critical to remember that EndoMT can be detected only in a narrow time window, during the transitional stage when endothelial cells express both endothelial and mesenchymal markers. Once endothelial cell markers are lost, the cells are hardly distinguishable from other mesenchymal-like cells.

1.2.5.3.5 Signalling and regulation in PAH

PAH is a complex disorder and can occur due to genetic mutations, environmental stimuli and/or exposure to toxins. Approximately 10% of PAH patients have a family history of disease which in some cases has been observed to be transmitted through multiple generations (202). The genetic background in PAH is very important and will be reviewed in section 1.3.2. Briefly, almost 80% of familial PAH cases and 10-40% of idiopathic PAH (IPAH) are diagnosed with mutations in the *BMPR2* gene (203).

PAH predominantly affects women, but the molecular basis of this phenomenon is poorly understood. Studies have been focused on estrogen production and metabolism, and mostly estradiol 2, which is the most prevalent estrogen in non-pregnant women. Mutations have been found in the *CYP1B1* gene, a regulator of estrogen metabolism that lead to overexpression of a metabolite able to trigger cell growth and potentially promote the development of vascular lesions in PAH (202). However, contradictory data from animal studies do not support this hypothesis and provide evidence of a protective effect of estrogen production in PAH. Indeed, recently it has been shown that agonist stimulation of surface estrogen receptors reduced pulmonary hypertension and improved right ventricular function (204). This finding is known as the “estrogen paradox” and requires further research to elucidate the link between PAH and estrogen production.

Most of the molecular and cellular mechanisms that are implicated in PAH have been discussed in detail in section 1.2.3.3 of the regulation of vascular remodelling.

These include growth factors that affect vascular remodelling such as PDGF and TGF-β, AngII and the redox axis, as well as the Notch signalling pathway. In addition, hypoxia and inflammation retain a distinct role in PAH.

Exposure to chronic hypoxia induces changes in the structure of pulmonary arteries, as well as in the biochemical and functional phenotypes of all vascular cell, leading to vasoconstriction and vascular remodelling. Hypoxia-induced changes are site-specific, such that the remodelling process in the large vessels differs from that in the smallest vessels. The mechanisms involved in hypoxia-induced PAH are reviewed in (205, 206). Heterogeneity occurs to a great extent in the cell phenotypes and responses under hypoxic conditions, and this is the reason that many studies report contradictory data regarding the effect of hypoxia on proliferation of cultured PASMCs. It has been shown that PASMCs that respond to chronic hypoxia with extensive proliferation, exhibit a less differentiated or an undifferentiated phenotype compared to all other PASMCs found in the pulmonary vascular bed (205).

The mechanisms underlying this proliferative response of PASMCs to hypoxia remain elusive. However, it has been shown that hypoxia reduces the production of prostacyclin and NO and increases levels of ET-1, serotonin, PDGF, and IL-6 and other vasoactive factors by PAECs and platelets, and this imbalance might promote PASMCs proliferation (91). In addition, hypoxia-induced proliferative PASMCs respond to GPCR agonists with stimulation of protein kinase C (PKC), whose activation in turn leads to proliferation via the ERK1/2 signalling pathway (207). In addition, hypoxia affects voltage-gated channels causing intrinsic changes in potassium and calcium levels (208).

HIF-1α is a heterodimeric transcription factor comprised of an oxygen-regulated α subunit that mediates the transcriptional responses to hypoxia (209). During hypoxia, HIF-1α accumulates and dimerises with its partner HIF-1β to activate hypoxia-specific genes. Cells lacking HIF-1α demonstrate impaired up-regulation of cellular proliferation and VEGF expression during hypoxia (210), and knock-down of HIF-1α inhibits hypoxia-induced proliferation (211). Recently, it was shown that KLF5 contributes to hypoxia-induced pulmonary vascular remodelling in a HIF-1α-dependent way (212). KLF5 is a zinc finger-containing transcription factor, required for AngII-induced VSMC differentiation and proliferation, and cardiac hypertrophy. KLF5 levels positively correlated with disease severity in lung tissues of patients with PH (213). Normoxic activation of HIF-1α also occurs in PAH PAECs, where it

correlates with reduced number of mitochondria and a shift towards the glycolytic metabolism (214).

Inflammation has long been considered a critical regulator of initiation and progression of vascular remodelling in PH, influenced by an imbalance of pro- and anti-inflammatory activities (86). Patients with IPAH and other types of PH exhibit higher circulating levels of inflammatory chemokines and cytokines such as VCAM-1, IL-6, IL-1β, IL-10, RANTES, MCP-1, TNF-α, and NF-κB (215-217), and their release contributes to PASMC proliferation and migration. Currently, none of the PAH drugs targets inflammation, and there is no evidence that immunosuppression is effective as a therapy, except perhaps in the cases of CTD- associated PAH, such as SLE and scleroderma (218).

Components of ECM have also been shown to regulate PASMC phenotype and proliferation, such as collagens I and IV and laminin that inhibit proliferation and promote the VSMC contractile phenotype, whereas fibronectin has the opposite effect [reviewed in (91)]. MMPs are zinc containing proteins responsible for ECM production and the maintenance of extracellular structures. MMPs contribute to disease progression and PASMC migration in several animal models of PH (91). Activity of MMPs is regulated by the tissue inhibitors of MMPs (TIMPs). *In vitro* studies showed increased TIMP-1 and decreased MMP-2 and MMP-3 expression in PASMC isolated from IPAH patients compared to healthy individuals (219). Although altered levels of MMPs and TIMPs have been shown in PAH, further studies are required to elucidate their effect.

Often, exposure to environmental triggers such as drugs, toxins, bacteria and viruses is enough to initiate PAH. For instance, the use of anorexigen, a known neuro-stimulant, has been shown to cause PAH (220). The effect of serotonin has also been studied, and animal models overexpressing the serotonin transporter were prone to the development of PAH (221). Other drugs such as methamphetamines and cancer drugs have also been associated with PAH, but the exact molecular mechanisms contributing to vascular remodelling is not well understood (191).

1.2.5.3.6 Treatment of PAH

The goals of PAH treatment include alleviating patients symptoms, improving their functional status and quality of life, halting or reversing disease progression including right ventricular dysfunction, and improving overall survival.

Pathways of prostacyclin, ET-1, and NO have been targets for therapy. Prostacyclin is a powerful vasodilator that inhibits proliferation and platelet aggregation (222). Three prostacyclin analogues have been commonly used in PAH treatments: epoprostenol, prostinil, and iloprost. The analogues have been reported to alleviate haemodynamic abnormalities in patients, and improve the mean arterial pressures, exercise capacity and PVR. ET-1 blocking agents have been used in PAH patients to block its effects. Of those most prominently used are the bosentan, sitaxsentan, and ambrisentan. Bosentan is an ET antagonist that is capable of blocking both ETA and ETB receptors. Its use improves symptoms, haemodynamic feature and exercise capacity, and also confers significantly improved survival (223). Other therapies target downstream components of NO signalling pathway by inhibiting phosphodiesterase-type 5 (PDE-5), the enzyme that catalyses cGMP to GMP (224), which has not been effective in all PAH patients (225).

EndoMT could potentially constitute a new strategy for targeted therapy in which specific therapeutic agents would be used to abrogate the process and improve the status of the PAH patients. Identification of the key modulators of EndoMT could significantly enhance this effort.

1.2.6 Connective tissue disease-associated PH (CTD-PH)

Based on the most recent classification criteria of PH (180), there is a sub-category of PAH associated with CTD. CTD is a heterogeneous group of systemic inflammatory disorders characterised by the presence of circulating autoantibodies and autoimmune-mediated organ damage (226). The lung is a frequent target, and more than one thoracic compartments can be involved, including the airways, lung parenchyma, pulmonary vasculature, and pericardium (226). The CTDs that often affect the respiratory system are scleroderma, RA, SLE, mixed CTD (MCTD), Sjogren's syndrome (SS), and undifferentiated CTD (UCTD) (226).

Although PAH is rare in the general population (IPAH has an incidence of 1–2/million/year) (227), it is more common in CTD. Indeed, several lines of evidence support a role for autoimmunity and inflammation in the development of the pulmonary vascular changes, including the presence of circulating autoantibodies and pro-inflammatory cytokines (IL-1 and IL-6) (228). Despite the similarities in disease pathogenesis and haemodynamic perturbations, outcomes in patients with CTD-associated PAH differ significantly from other forms of PAH.

1.2.6.1 Scleroderma (SSc)

Scleroderma is a complex rheumatic disease characterised by autoimmunity and inflammation, vasculopathy and fibrosis of the skin and internal organs. The pathogenesis of the disease remains unknown, although it is known to involve genetic and environmental factors, and complex organ-based complications (Figure 1.5). Scleroderma is a rare condition with the highest case-specific mortality among all autoimmune rheumatic diseases with more than half patients dying of the disease (229). Scleroderma occurs worldwide and is represented in all ethnic groups. The prevalence of scleroderma in the USA is 276/million adults compared to 88/million in the UK (230), however, the incidence and prevalence can vary among populations. Women are affected more often than men with a reported ratio of 4-5:1 (230). The usual age of onset is 30-50 years, and although it can affect any age group, it is rare in children.

Primary classification criteria were issued and reviewed in 2013 by the American College of Rheumatology/European League against Rheumatism collaborative initiative (ACR/EULAR) (231). Clinical diagnosis of scleroderma is defined by the presence of Raynaud's phenomenon, hardening of the skin and visceral organ involvement. Depending on the clinical symptoms, patients can be diagnosed as scleroderma or as an overlap with other AIDs such as RA, SLE, polymyositis (232, 233). Two different disease subsets have been characterised: the limited cutaneous (lcSSc) and the diffuse cutaneous (dcSSc). In lcSSc, the hardening of the skin occurs at the distal joints (elbows, knees) and the face, compared to dcSSc, where proximal and distal extremities, face and trunk are affected. The two subsets have different organ involvement and autoantibody-specific overproduction. LcSSc is more common than dcSSc and affects almost 70% of scleroderma patients. Scleroderma is characterised by an overproduction of autoantibodies (234). Identification of autoimmune antibodies specific to the disease facilitates diagnosis and disease classification. Whether these antibodies play a direct role in the pathogenesis of the disease or they are an epiphenomenon of disturbed autoimmunity is not clear yet. However, they have been used as a diagnostic tool for the disease, as well as a prognosis for specific organ involvement.

Organs affected by scleroderma include the heart, lungs, vessels, kidneys, gastrointestinal tract, with severity and symptoms varying amongst patients (Figure 1.5). The complexity and multi-organ involvement means it is extremely important to identify and treat the organ-based complications. Renal crisis and pulmonary fibrosis can occur at an early disease stage, while PAH occurs later. Pulmonary

fibrosis is the most common form of interstitial lung disease in scleroderma. Apart from the lungs, the gastrointestinal tract, oesophageal, stomach, small intestine, and large bowel can all be affected causing a range of symptoms. Involvement of small intestine generally occurs in patients with established scleroderma and is a major cause of mortality. Scleroderma-associated renal crisis typically causes accelerated hypertension and acute renal failure which may or may not resolve.

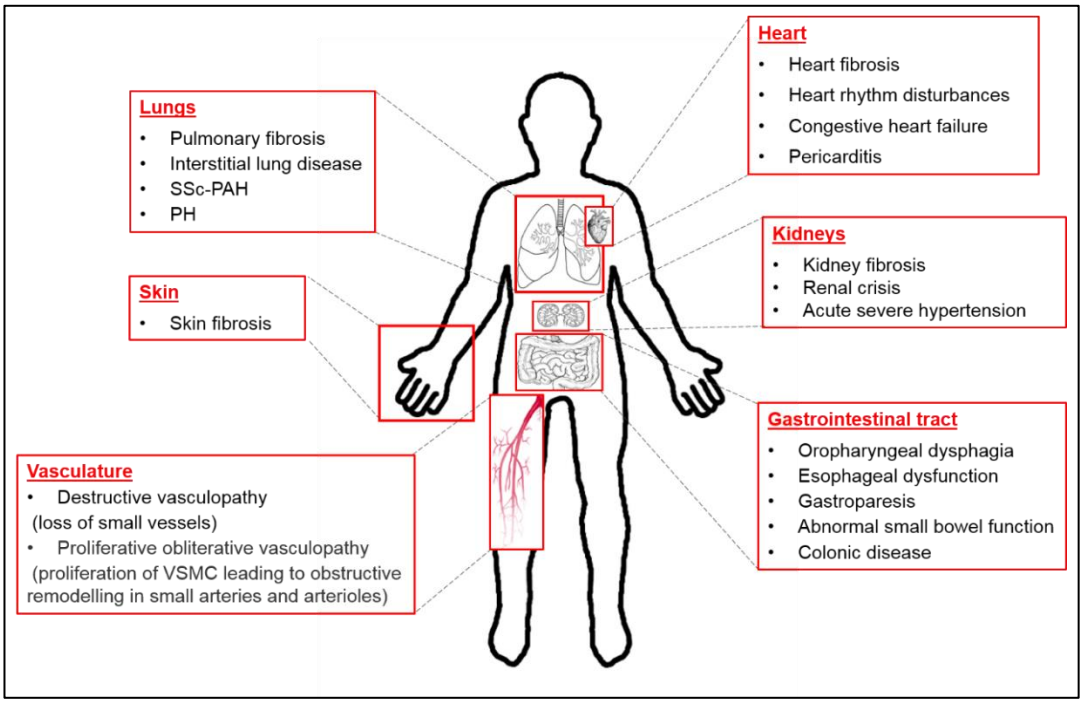


Figure 1.5 Organ involvement in scleroderma. *Scleroderma is a complex rheumatic disease characterised by autoimmunity and inflammation, vasculopathy and fibrosis of the skin and internal organs. Affected organs include the heart, lungs, kidneys, gastrointestinal tract, skin and vasculature.*

1.2.6.2 SSc-associated PH

Scleroderma is considered a susceptible phenotype for the development of PH. Pulmonary vasculopathy is common among scleroderma patients and may lead to PH or remain subclinical in ~15% of the patients. At the Centre for Rheumatology and Connective Tissue Disease in the Royal Free NHS Trust Foundation Hospital, as well as in other clinical centres in the world, scleroderma patients are on active follow-up screening programmes from diagnosis and onwards. The screening programmes include lung biopsies and assessment of the physiology of patients every year as well as autopsy studies. Worsening of a patient's condition would point towards a RHC. Patients with subclinical pulmonary vasculopathy exhibit borderline mPAP, and half of them will develop PH at some point in time (235).

Patients that undergo RHC and found to have mPAP≥25mm Hg are diagnosed as PH. Based on the clinical criteria reviewed in Nice 2013 (231), 60% of those patients have PAH and this is categorised as Class I of PH (236). A sub-group of this category is Class I', which is PH due to pulmonary veno-occlusive disease (PVOD). Of the rest of PH patients, 15% exhibit post-capillary heart failure with PAWP>15mm Hg, and this is categorised as Class II. Class III consists of PH patients (15%) presenting arterial vascular remodelling and significant lung fibrosis >20% diagnosed with high resolution CT (HRCT), and/or forced vital capacity (FVC) <70% (237). RHC is essential for SSc-PAH diagnosis, since echocardiography is limiting due to the inaccuracy of the Doppler signal in assessing true RV systolic pressure (238).

SSc-PAH exhibits the worst prognosis among CTD-associated PAH (239). Despite advances in PAH therapies, the 3-year survival of SSc-PAH is ~50% (240). SSc-PAH can occur in both lcSSc and dcSSc forms, and can remain asymptomatic until a quite advanced stage of disease progression. Initial symptoms include exertional breathlessness, chest pain or syncope. SSc-PAH is characterised by vascular remodelling intimal hyperplasia, medial hypertrophy and adventitial fibrosis and a lower number of plexiform lesions compared to IPAH, leading to vessel obliteration (241). Other vascular changes include endothelial activation with expression of cell adhesion molecules, endothelial dysfunction, apoptosis, and inflammatory cell recruitment (242). The extent of these vascular changes in vital organs such as the lungs, kidneys and heart defines the prognosis of patients with scleroderma. Increased levels of VCAM-1 and VEGF in lung tissues of SSc-PAH patients reveal extensive endothelial injury and increased angiogenesis, respectively (242).

PAH-specific treatments have been used in SSc-PAH patients including ET-1 receptor antagonists, prostacyclin analogues and receptor agonists, and molecules targeting the NO pathway. Combinatorial therapies with PAH-specific drugs have also been used in the management of CTD-PAH, however, patients in this population have been observed to have a lower response to PAH-specific therapy compared to IPAH (239). Possible reasons include an increased prevalence of PVOD lung disease in SSc-PH patients, or more severe vascular lesions affecting not only the proximal but also the distal pulmonary vessels, as well as the heart (such as inflammatory myocarditis) in CTD. The poor outcome of SSc-PAH compared to IPAH, together with a lack of effective combination treatment render the treatment and management of the disease a major challenge. Longitudinal observational studies embedded in clinical practice have provided important

evidence that the systematic screening of SSc-PAH may be associated with improved outcomes (243).

1.3 Genetics of CVD

1.3.1 Introduction

CVD are complex conditions that involve both gene-gene and gene-environment interactions. Apart from the field of pharmacogenetics, very little progress has been made in our understanding of gene-environment interactions. This is partly due to the difficulty of accurate measurement of most environmental factors as compared to the genetic factors, and the low power to detect and analyse combinatorial effects (244). A growing part of basic and clinical research is focused on the identification of genetic factors that contribute risk susceptibility to disease. In comparison to traditional risk factors, genetic markers associated with disease are expected to exhibit better clinical relevance in the prediction and diagnosis of CVD. Observations form genetic studies also help to elucidate inter-individual differences in cardiovascular protection in order to develop well-defined strategies leading to clinical genetic testing, genetic counselling and ultimately personalised therapies.

Genetic inheritance has also been proved complicated in complex disease, with combinations of numerous susceptibility-conferring alleles at several loci to interact in a particular individual. Some of them may affect the risk of CVD in a way that cannot be predicted from the separate effect of each variant. In addition, rare variations might also confer susceptibility with smaller effects, and these effects cannot be easily captured. This is the major obstacle for the characterisation of the genetics of complex traits proposing the exploration of genomic systems rather than single genes.

1.3.1.1 Genetic studies and design

Although, most of CVDs are not single-gene disorders, there are cases where a faulty gene is enough to cause disease. In single-gene disorders, tremendous efforts were made from 1990 to 2010, a period referred to as “golden era” of single-gene disorders. Of the estimated 7000 single-gene disorders, a gene has been discovered for over 3000 genetic conditions (245). The first CVD to be mapped was hypertrophic cardiomyopathy and then dilated cardiomyopathy followed by many others such as Wolff–Parkinson–White syndrome (WPW), atrial fibrillation, Long QT syndrome, and Brugada syndrome (246). These are rare disorders occurring with a frequency of <1%.

A basic representation of the designs in genetic studies is shown in Figure 1.6. Initially, disease-causing genes were identified in families with several affected members over different generations using linkage studies, which examined whether particular alleles are co-transmitted with the disease at a higher frequency than expected by chance. Genomic regions that carry a disease-causing gene can be identified by testing of the co-segregation of the disease with genetic markers that "tag" or label specific regions of the genome. The success of linkage studies depends significantly on the efficient phenotypic characterisation of the family members. Originally, microsatellites were used as genetic markers in the linkage studies. However, the use of single nucleotide polymorphisms (SNPs) was later established. SNPs are the most common variation across the human genome with 5.9 million already identified of the 11 million expected to exist (247). Each SNP has two alleles, and depending on the minor allele frequency (MAF), SNPs are divided in common (>5% MAF), low frequency (0.5%-5% MAF), and rare (<0.5% MAF). Other forms of genetic variation are copy number variations, including deletions, insertions and duplications of parts of DNA.

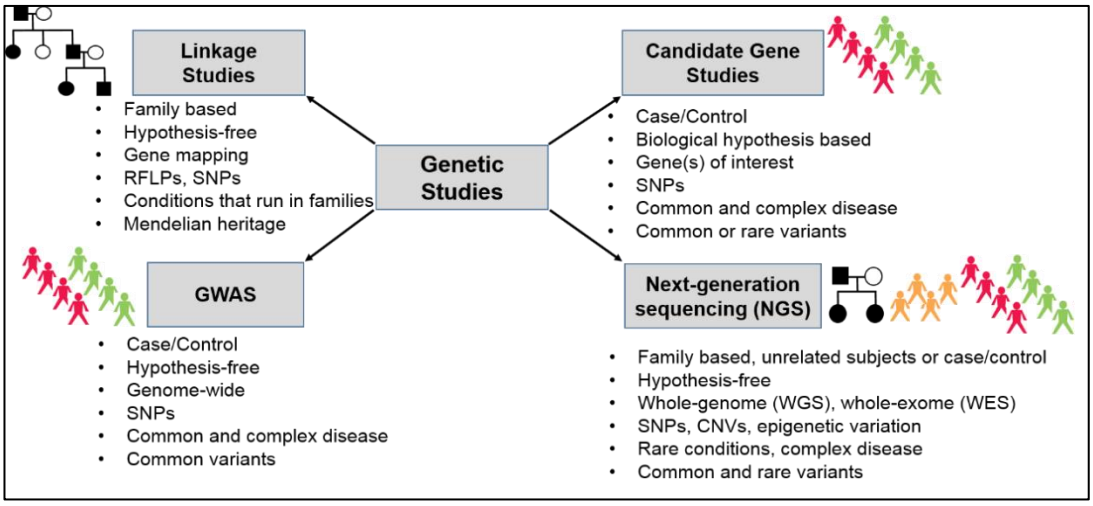


Figure 1.6 Design of genetic studies. *Description of different designs of genetic studies. Although scientific interest has been shifted from linkage studies to GWAS and next-generation sequencing through the years, designing a genetic study is important and depends on the needs and the objectives of the study.*

Occasionally, alleles of neighbouring SNPs can be inherited together, a phenomenon that is called linkage disequilibrium (LD) and is caused by the epistatic natural selection, mutation, random drift, genetic "hitchhiking", or genetic flow. LD is responsible for the non-random association between alleles of different genomic loci and even genes. The associated alleles form haplotypes, and segments of the genome with increased LD are called haplotype blocks, and they are considered the

outcome of homologous recombination occurring over generations (248). The International HapMap project was the first portal to collect and publish genotype, linkage and frequency data of different human populations, and was launched in 2005 and ran until June 2016, when it was retired by NCBI. The 100,000 Genomes Project is a new powerful database providing current and best data of genotypes, sequences and genome mapping.

Since complex diseases do not follow a clear pattern of Mendelian inheritance, the strategy used to identify predisposing genes is usually not based on family studies but in case-control association studies. Genetic association studies can either focus on a specific gene locus (candidate gene studies) or target the whole genome (genome-wide association studies or GWAS) (Figure 1.6). Candidate gene association studies are focused on genes that exhibit biological significance for the pathogenesis of the disease based on experimental evidence or a working hypothesis. When strong experimental evidence is lacking, data mining software that have recently advanced substantially, can meet the needs of growing research and provide intelligent tools for candidate gene selection and study design (249).

GWAS represent an unbiased hypothesis-free approach, whereby markers distributed throughout the human genome are tested for association with no preference in terms of their selection. The GWAS strategy is devoted to the discovery of novel genetic biomarkers and their impact on the likelihood of disease onset, progression, prediction and management. The first GWAS was published in 2005 and used a micro-array that contained probes against $5x10^5$ SNPs (250). Today, GWAS are conducted using chip assays with genome-wide SNP genotyping coverage (Illumina). Moreover, high-throughput innovative advances provide whole-exome and whole-genome sequencing with competitive costings (Figure1.6). Data obtained from genetic association studies are often tested in independent replication cohorts of the same or different descent, to further validate significant associations.

In genetic association studies, DNA is genotyped in two cohorts: a case/patient cohort and a control/healthy cohort of unrelated individuals. In this approach, the allele or genotype frequencies of the genotyped SNPs in cases are compared to those of controls. An allele that is found to be more common in cases compared to controls will be considered as disease-associated as it increases predisposition to the disease, and is usually called a risk allele. On the contrary, an allele that is found in lower frequency in cases compared to controls is protective over the disease. Various genetic models (genotypic, recessive, dominant, etc.) are used to

address different types of questions regarding the effect that an allele or a genotype exerts on disease phenotype.

The statistics applied in genetic tests is challenging, as stringent criteria need to be considered and applied to avoid false positive data. Also, a multiple comparison issue due to accounting for the same SNP in various tests can lead to false discovery rates (FDR). This is usually addressed in two ways: by applying Bonferroni corrections or performing permutation analysis. Candidate gene studies often use less stringent criteria compared to GWAS, where Bonferroni correction is a gold standard due to the infinite numbers of tests performed. The reason is that a conservative threshold such as Bonferroni would be overly stringent particularly in the context of a disorder with no major gene effects (249).

1.3.1.2 Epigenetic studies

The term epigenetics refers to chromatin or DNA-based mechanisms important in the regulation of gene expression, which do not involve changes in the DNA sequence. Epigenetic research unravels the complex relationships between disease pathology and genetics, and provides novel insight into the mechanisms of cardiovascular health and disease. It aims to explore the role of genetic heritability and environmental interactions, as well as to explain discrepancies caused by environmental insults. The genetic heritability of CVD can vary significantly, and studies report a 40%-80% of genetic contribution to CVD (251). Indeed, studies have shown that monozygotic twins that lived significantly different lifestyles or were exposed to different environments, exhibited substantially different epigenetic patterns (252). Other factors known to affect the prevalence of CVD include gene-gene and gene-environment interactions potentially mediated by epigenetics, gene imprinting, and other factors (252).

Although the field of epigenetics is relatively new, it has gained a lot of attention in the context of CVD, and recent reviews discuss its impact in CVD pathogenesis and therapeutics (253-257). Epigenetic patterns can differ among different cell types in the body, and among cells of the same tissue. These patterns serve as cellular memory of exposure to abnormal environmental stimuli early in life, and they can affect health later in life as well as be passed to future generations (257). Epigenetic mechanisms can affect gene expression at post-transcriptional and post-translational levels and can be divided into three distinct but interrelated processes: DNA methylation, RNA-based mechanisms, and histone modifications.

DNA methylation plays a key role in embryonic development, cell type lineage specification, X-chromosome inactivation, and genomic imprinting (254). Methylation changes are mediated by DNA methyltransferases (DNMTs), which catalyse the addition of methyl groups to the C5 position of cytosine residues (5mC) at repetitive CpG dinucleotides, known in mammals as CpG islands. Methylation usually occurs in gene regulatory regions such as promoters and enhancers, and it is typically a repressive mark associated with inhibition of transcriptional initiation and suppression of gene expression. Changes in DNA methylation patterns have been observed in CVD, atherosclerosis, congenital heart disease, autoimmune diseases (AIDs), infection and cancer. Athero-protective genes, including estrogen receptors, have been found constitutively hyper-methylated and thus silenced in human atherosclerotic tissue and plaques (258). This excessive methylation could interfere with the athero-protective effects of estrogen in women and also reveals the influence of methylation in vascular aging and atherosclerosis (258).

RNA-based epigenetic mechanisms involve two post-transcriptional regulators: the microRNAs (miRs) and the long non-coding RNAs (lncRNAs). MicroRNAs are short non-coding RNAs 20–22 nucleotides long that have emerged as important regulators of gene expression. MicroRNAs modulate gene expression through binding at 3'UTRs of target genes and inhibition of mRNA translation or of other post-transcriptional events, and transcript degradation. Estimates based on computational approaches currently find more than 60% of human genes to be targeted by microRNA, with many of these interactions being highly conserved throughout evolution (259). In mammals, more than 1000 different microRNAs have been identified, with different tissue-specific and process-specific expression. The role of microRNA has been studied extensively in CVD [reviewed in detail in (253)]. For instance, miR-126 and miR-145 are downregulated in CAD, and miR-1, miR-133b, and miR-499 are downregulated in myocardial infarction. Also, miR-21 and miR-29 are two of the microRNAs mostly studied in cardiac fibrosis.

MicroRNAs have also been used efficiently as highly specific diagnostic markers in the circulation of CVD patients. Moreover, innovative drug therapies have been developed using microRNA as therapeutic targets, either by blocking them and restoring normal gene expression, or by using microRNA mimic molecules to inhibit deleterious disease-causing gene expression. Although significant insights have been provided on the role of several microRNA in CVD, a deeper understanding of microRNA functions in the cardiovascular system is required for future strategies for CVD prevention, diagnosis, and therapy.

LncRNAs constitute a large and diverse class of transcribed RNA molecules that do not encode proteins and are primarily located in the nucleus (260). LncRNAs function either by binding to DNA or RNA in a sequence-specific manner or by binding to proteins (260). Their expression is developmentally regulated and can be tissue- and/or cell type- specific (260). Epigenetic modifications of the histone code, including lysine acetylation and methylation and serine phosphorylation, influence significantly the chromatin structure, and modify the accessibility of transcriptional regulators to DNA-binding elements.

1.3.2 Genetics of PH

PH is a relatively rare disorder that can occur within families in a hereditary manner, it can be sporadic, idiopathic, or it can even be associated with other diseases such as Scleroderma, or syndromes such as Down's and DiGeorge (261). The disease course is very severe and progressive, although there is currently no cure, the disease can be managed and monitored to increase the life expectancy of the patients. However, diagnosis is often delayed for years, which severely impacts on outcome. In this regard, genetic studies have attempted to decode the genetic background of the disease and offer new ways of diagnosis and potential treatments.

The *BMPR2* gene was the first to be directly linked to PAH over 20 years ago (262-264) . To date, more than 300 independent *BMPR2* mutations have been identified accounting for >75% of familial PAH and 10-25% of sporadic cases of PAH (261). Mutations in the *BMPR2* gene are haploinsufficient with most patients carrying only one mutated allele. Another very interesting finding is that not all the *BMPR2* mutation carriers develop the disease, suggesting an incomplete penetrance. In detail, the overall penetrance of the *BMPR2* mutations is 27%, which is approximately three times higher in women (42%) compared to men (14%) (265). Although the mechanism for the incomplete penetrance remains unclear, researchers suggest that female hormone metabolism may contribute significantly to this mechanism (202, 266). Additional causal mutations have been identified in other genes of the TGF-β superfamily (*ACVRL1, ENG, SMAD1, SMAD4*, and *SMAD9*), however, these are believed to account for only 1%-3% of PAH cases.

Whole-exome sequencing in PAH patients without mutations in the *BMPR2* gene or any other TGF-β family member unravelled new rare genetic variants in *CAV1, KCNK3*, and *EIF2AK4* genes (267). Six mutations have also been found in the *KCNK3* gene within highly conserved protein domains, and it was hypothesised that

loss-of-function mutations in the gene is a hallmark for idiopathic and heritable PAH (268). Indeed, KCNK3 expression and activity are strongly reduced in PASMCs and PAECs, and its inhibition promoted increased proliferation, vasoconstriction, and inflammation (268). Mutations found in the *EIF2AK4* gene were identified as the cause of two previously unlinked rare types of PAH, pulmonary capillary proliferation and occlusive lesions of small pulmonary veins (PVOD) (269, 270). Other genes with novel PAH-associated mutations that recently evoked from GWAS (271) are *Cerebellin 2* (*CBLN*2) (270) and the potassium channel *KCNA5* (272).

Although these mutations and variants have been linked to PAH by affecting pathways relevant to pulmonary vascular homeostasis, they are not sufficient to explain the entire genetic background of PAH. In addition, it has been shown that *BMPR2* mutations are not found in CTD-associated PAH, portal hypertension, and HIV infection. Taken together, the above imply that apart from the pathogenesis and regulation of disease, differences in the genetic profile of the individual must also exist. Thus, other genetic and epigenetic mechanisms must be operating and their elucidation might help increase the understanding of PH.

Indeed, the role of epigenetics in PAH is a fast-growing area of research, and modifications have been studied extensively (273, 274). Many miRNAs have been linked functionally with PAH including miR-21 (275, 276), miR-124 (277), miR-17-92 (278), miR-145 (279), and miR-204 (280). DNA methylation and histone modifications have also been studied in PAH and recent findings are reviewed in (273). Findings highlight a clear epigenetic mechanism for PASMCs proliferation, important for the pathophysiology of PAH, through the methylation of SOD2 and the interaction with HIF-1α (281). Histone modification studies have been focused on HDAC-mediated control of PAH, and HDAC inhibitors have been used successfully to reduce cardiac hypertrophy and fibrosis (282).

1.3.3 Genetics of SSc-PH

As described previously, scleroderma is a multi-organ disease with very complex pathogenesis and manifestation [summarised in (283)], and an important interplay between genes and environment. Epidemiological studies have shown a significant increase in the risk of scleroderma among first-degree relatives (1.6%) compared to the general population (0.026%) (284). The prevalence is increased even more between monozygotic twins (4.7%) (285). Although family history of scleroderma has been identified as the highest risk factor, the likelihood of developing the disease among offspring of patients is <1% (286). There is also strong evidence of

familial clustering of patients, and in these clusters relatives tend to have the same disease-associated autoantibodies (285). These data imply a genetic susceptibility to scleroderma overall and an inherited susceptibility to develop specific sub-phenotypes of disease. Thus, unravelling the genetic profile of scleroderma would enhance the understanding of disease pathogenesis and help in designing novel therapies.

However, the design of a genetic study in complex and heterogeneous diseases such as scleroderma is not trivial. A multistep genetic association analysis and sophisticated statistical analysis should be carried to ensure that studies are strong with sufficient power to identify true and significant associations, and that heterogeneity will not obscure the results. In addition, it is important to study and correlate genotypes with individual phenotypes as outcomes of the disease to further explore pathogenesis. To achieve this in scleroderma, it is essential that all clinical sub-phenotypes are studied including disease subsets (limited or diffuse), autoantibodies (ATA, ACA, and ARA), SSc-PAH, PH, PF, and renal crisis. This type of analysis provides the opportunity to detect genetic influences on specific phenotypes within the disease spectrum. The structure of genomic locus under question should also be examined through a haplotype association analysis in order to evaluate whether a combination of SNPs rather than individual markers confer susceptibility to disease.

To date, candidate-gene approaches as well as GWAS have revealed many susceptibility loci along the genome, but each one of them explains only a small proportion of the disease. Most of the genetic findings are summarised in recent reviews (287, 288). The first GWAS in scleroderma was published in 2009 in the Korean population (289), followed by another in 2010, which identified the *CD247* gene as a new susceptibility locus in scleroderma (290). CD247 is a mediator of T-cell receptor signalling, previously seen to be involved in the pathogenesis of SLE, and retains a critical role in the imbalanced immune response. The association of *CD247* with scleroderma has been replicated in an independent study (291). A meta-analysis of two GWAS was published in 2013 and identified *KIAA0319L*, *PXK* and *JAZF1* as novel susceptibility loci for scleroderma and SLE, increasing significantly the knowledge of the genetic basis of autoimmunity (292). Another GWAS follow-up reported a significant association of rs310746 in the *PPARG* gene ($p=5x10^{-7}$, OR=1.25) and scleroderma suggesting a role of this gene in disease pathogenesis (293).

The strong autoimmune component of scleroderma, the dysregulation of the immune system, and the over-production of autoantibodies suggest a solid pathogenic background shared by AIDs. Indeed, different polymorphisms within the human leukocyte antigen (HLA) region of the major histocompatibility complex (MHC) have consistently been associated with various AIDs including scleroderma, SLE, and RA (290, 294). Modest associations have been found between given haplotypes and the disease overall, but stronger evidence for associations have been found between specific HLA polymorphisms and autoantibody production in scleroderma patients (295). However, due to the extensive LD within the HLA/MHC locus, the identification of causal variants remains challenging, requiring studies in bigger cohort and the use of novel high-throughput sequencing technologies. In order to further explore scleroderma risk loci common with AIDs and to fine map these areas, an analysis was conducted using the Immunochip (Illumina), which contains 186 known autoimmunity risk loci (296). The HLA/MHC associations with scleroderma, ACA, and ATA were confirmed (296). In addition, 4 new non-HLA loci were associated with scleroderma at a genome-wide significant level including *DNASE1L3*, an intergenic SNP between *SCHIP1* and *IL12A*, an intronic SNP within *ATG5* gene, and another SNP between *TREH* and *DDX6* (296).

Beyond the HLA/MHC region, various genes associated with scleroderma are involved in adaptive and innate immunity, and in particular in pathways related with cell signalling regulation and T-cell differentiation. More specifically, genes involved in the innate immune response include the interferon regulatory factor 5 (*IRF5)* (286, 290, 297), macrophage inhibitory factor *(MIF)* that acts as a proinflammatory and immune-regulatory cytokine (298, 299), and *ITGAM* gene that is important in activation adherence and migration of leukocytes, phagocytosis and neutrophil apoptosis (300, 301). Genes involved in the adaptive immune response include the *STAT4* gene that was a top associated gene at GWAS-level (290, 302), *BANK1* gene that expresses a scaffold protein exclusively expressed in B-cells (302-304), *BLK* gene encodes a BCR-associated transducing molecule with a key-role in B-cell development (305, 306), tumour necrosis factor ligand superfamily 4 *TNFSF4* (299, 305), *TNFAIP3* that participates in the NF-kB pathway and the B-cell survival (297), and *PTPN22* gene that negatively regulates TCR-signalling (307).

Scleroderma is a highly heterogeneous disease, and this complicates significantly the efforts to dissect the genetic component. Other than loci and genes related to immunity and inflammation, not many genes in other molecular and cellular pathways have been associated with scleroderma disease pathogenesis. *CTGF* is

probably the most studied example that bridges genetics and function. In particular, Fonseca *et al* revealed that rs6918698 (-945C/G) located within the promoter region of *CTGF* is significantly associated ($p<0.001$) with scleroderma-associated pulmonary fibrosis, as well as with the presence of ATA and ACA antibodies in the UK scleroderma population (308). The association failed to replicate in a meta-analysis among 6 independent case-control studies (309). However, a study in a Japanese population showed that rs6918698 CC/CG genotype greatly decreased the susceptibility of scleroderma in a dominant model analysis ($p=0.005$, OR=0.632) (310). CTGF is significantly up-regulated in the skin and lung fibroblasts of scleroderma patients, with an established role in the pathogenesis of disease (134).

In 2016, a whole-exome sequencing study was conducted in patients with dcSSc, and 70 genes that were enriched with deleterious variants were identified (311). Of these, two (*BANK1* and *TERT*) were known susceptibility loci previously implicated in scleroderma pathogenesis, five were newly identified genes (*COL4A3, COL4A4, COL5A2, COL13A1*, and *COL22A1*) that are significant components of ECM associated with fibrosis, and one gene (*XRCC4*) is involved in the DNA repair pathway (311). In the same year, another study identified rs58905141 in the *TNFAIP3* gene to be strongly associated with the silica-induced profibrotic response in lung fibroblasts (312). The SNP was consistently associated with time-course and dose-response expression of MMP3 and MMP1 in fibroblasts stimulated with silica particles in Caucasian subjects (312). *In silico* analysis using ENCODE data revealed that rs58905141 might affect the binding of TNFAIP3 transcription facto.

Although studies have been successful in identifying susceptibility loci, genetics alone is unable to fully account for scleroderma risk. Epigenetic mechanisms have also been reported to contribute in scleroderma. Various microRNAs have been linked to the disease, including let-7a, let-7g, miR-29a, miR-29b, miR-30, miR-125b, miR-129-5p, miR-145, miR-150, miR-196a, miR-206, miR-7, miR-21, miR-92a, miR-142-3p [reviewed in (287, 313)]. Some of the microRNAs are downregulated while others are increased in the skin, serum and fibroblasts of scleroderma patients. Verified targets of scleroderma-associated microRNAs include genes important in ECM deposition in fibrosis such as *COL1A1, COL1A2,* and *COL3A1*, as well as genes involved in signalling such as *SMAD3/7/5, TGFB1/2/R*, and *PDGFRβ*. MicroRNA therapies are considered to be effective against AIDs and novel therapies exploiting research finding are expected to be developed in scleroderma

soon. However, since microRNAs can have different targets, extra caution should be taken to avoid unforeseen adverse effects.

Many studies have described abnormalities in DNA methylation in scleroderma fibroblasts, lymphocytes, and endothelial cells, with the *FLI1* gene to be extensively studied (313). In particular, *Fli1* negatively regulates collagen transcription, and so abundant methylation of the *Fli1* promoter induces collagen transcription and excessive ECM production (314). DNA methylation could also explain the female preponderance of scleroderma through the demethylation of the *CD40* ligand gene. The *CD40* ligand gene is located in the X chromosome, and the binding of CD40 on $CD4^+$ T cells causes the maturation of B cells into plasma cells and memory B cells. Demethylation of *CD40* ligand promoter has been found in the inactive X chromosome in female scleroderma patients, leading to CD40 ligand overexpression (315).

No lncRNAs are yet recognised as modifiers of scleroderma pathogenesis, however there are a few that are reported to regulate immune responses, which could also affect scleroderma. Finally, histone modification mechanisms have been explored to some degree in scleroderma and are vital for regulation of gene expression. However, more studies are required to dissect the exact role of epigenetics in pathogenesis and manifestation of scleroderma.

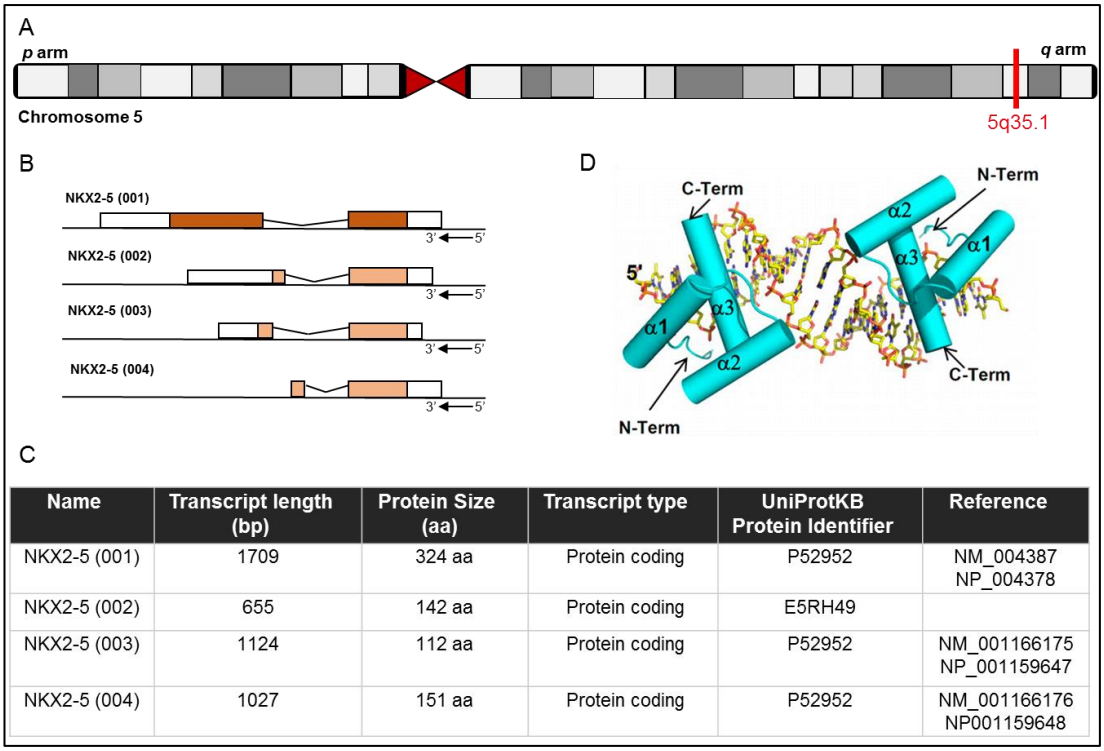
1.4 The *NKX2-5* gene

1.4.1 Introduction

NKX2-5 (NK-2 homolog E) is a transcription factor that belongs to the family of homeobox DNA binding transcription regulators, which are structurally and functionally conserved through evolution (316). *NKX2-5* is the human homolog of the drosophila *tinman* gene, and other homologs have been identified in many vertebrates including mouse, chicken, xenopus and zebrafish (317). NKX2-5 appears to be highly conserved among species (Figure 1.7), in terms of both primary protein sequence and mRNA expression pattern (317). The conserved protein structure consists of four distinct domains: a N-terminal TN-domain, a homeodomain, NK-2 domain, and a conserved C-terminal peptide (318). The homeodomain recognises and binds to a DNA consensus sequence (TNAAGEGG) through a helix-turn-helix motif (Figure 1.8D) and interacts with other transcription factors to regulate gene transcription of downstream targets (319). Less is known about the function(s) of the other domains, although it is thought that the NK-2

domain has the ability to repress transcriptional activity through protein-protein interactions.

Figure 1.7 Multiple sequence alignment of NKX2-5 protein sequence in different homologs. *NKX2-5 protein is an evolutionarily conserved protein and a member of the NK2-family of transcription factors. NKX2-5 protein sequences of mouse, chicken, xenopus, zebrafish and the drosophila homolog tinman are aligned against the human homolog. Amino acids shown in red are highly conserved amongst all homologs, whereas those in blue are less conserved. Amino acids shown in grey are not present in all homologs. The human NKX2-5 protein is 324 amino acids long, and contains a centrally located conserved homeodomain that is involved in nuclear translocation, interaction with other transcription factors, and DNA binding. The sequence within the grey box encodes the NK family homeodomain, and it is highly conserved in evolution from tinman to the human homolog. Sequences harbouring the homeodomain are also highly conserved and they encode the N- and C-terminal Alanine/Proline rich regulatory domains. The percentages in parentheses indicate the similarity of each protein sequence compared to the human sequence, thus mouse and human NKX2-5 protein sequences are 87% similar. FASTA protein sequences were obtained from Uniprot (http://www.uniprot.org/), and the sequence alignment was constructed in Protein Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome).*



C

Name	Transcript length (bp)	Protein Size (aa)	Transcript type	UniProtKB Protein Identifier	Reference
NKX2-5 (001)	1709	324 aa	Protein coding	P52952	NM_004387 NP_004378
NKX2-5 (002)	655	142 aa	Protein coding	E5RH49	
NKX2-5 (003)	1124	112 aa	Protein coding	P52952	NM_001166175 NP_001159647
NKX2-5 (004)	1027	151 aa	Protein coding	P52952	NM_001166176 NP001159648

Figure 1.8 Characteristics of NKX2-5 gene and protein. *A. Human NKX2-5 gene is located on the q arm of chromosome 5:173,232,109-173,235,357. B-C. The gene can be transcribed in four splice variants 001-004, which are all protein coding. The transcripts differ in size, and each has an individual reference number (NM). Transcribed proteins also differ in size, and three of them are recognised by the name identifier in UniProt (P52952). The longest transcript (1.7Kb) identified by NM_004387 reference number encodes a 342aa protein and is considered the most common NKX2-5 transcript. D. A high-resolution picture of the crystal structure of NKX2-5 in complex with the promoter of ANF gene, one of its downstream targets (320).The homeodomain is shown in cyan, and the DNA is shown in stick representation with carbon in yellow, oxygen in red, nitrogen in blue, and phosphorus in orange. Interactions between the DNA and homeodomain are mediated through residues from three regions of the homeodomain: N-terminal extension, the loop connecting α1 and α2, and helix α3.*

NKX2-5 is one of the earliest markers of the cardiac lineage, and studies have established an essential and non-redundant role in the cardiovascular system during embryogenesis (41, 42). *Tinman* was originally identified as a gene essential for the heart and vessel formation. Expression was found in mesoderm and endoderm that give rise to the heart, the visceral organs, and the pharyngeal structures respectively (321). Although NKX2-5 was not previously thought to be implicated in blood vessel formation in mammals, more recently studies in mouse and zebrafish showed that *Nkx2-5*-expressing mesoderm gives rise to the heart muscle as well as the endothelium of the pharyngeal arch arteries, which are transient embryonic blood vessels contributing to the carotid arteries and the great vessels of the heart, the aorta and pulmonary arteries (322, 323). After birth,

NKX2-5 expression is significantly downregulated in all organs, and its expression is only found in cardiomyocytes and hypertrophied hearts (324).

It has been proposed lately that many biological mechanisms that govern normal embryonic and foetal development are central to postnatal repair and injury responses, and that the same pathways seem to be recruited in disease states, with aberrant expression and activity of key mediators (325, 326). Therefore, it is hypothesised that *NKX2-5* might play a substantial role in pathological conditions in adulthood, where the heart and the blood vessels are involved. Such conditions may involve congenital heart defects, wound healing, and vascular remodelling.

According to the GTEx portal and based on mRNA differential expression (RNAseq Illumina) in normal tissues, expression of *NKX2-5* in adulthood is only found in the heart, in the left ventricle and atrial appendage, in the spleen, and in the coronary artery at very low levels (http://www.gtexportal.org/home/gene/NKX2-5). At the protein level, NKX2-5 is only expressed in the heart.

Data generated in the UCL Centre for Rheumatology and Connective Tissue Disease have revealed important insights and compelling evidence that NKX2-5 is a key and major regulator of vascular remodelling. It is the first time that NKX2-5 gene is implicated in different pathologies in adulthood, and these data provided the basis for the research conducted during this PhD.

1.4.2 NKX2-5 in vascular remodelling

A distinct role of NKX2-5 was identified by Ponticos *et al* showing that the mouse Nkx2-5 activates expression of the *Col1a2* gene in VSMC (Figure 1.9). Specifically, a far-upstream enhancer of *Col1a2* gene was identified approximately 17Kb from the transcription start site that contains the necessary regulatory elements to confer tissue-specific expression in the majority of collagen-producing cells, including blood vessels (327). The far-upstream enhancer is conserved between mouse and human (327). Within this enhancer, a sequence element ~100bp long was identified, which is specific to VSMC and is able to regulate collagen expression exclusively in VSMC (328). The expression is activated through the binding of Nkx2-5 at the VSMC-specific element and it is further potentiated in the presence of Gata6 (328). Specifically, a mechanism has been proposed that involves competition for binding between two transcription factors: Nkx2-5 and the repressor, Zeb1 (Figure1.9) (328). Nkx2-5 binds the specific sequence element within the upstream enhancer and activates *Col1a2* transcription competing out Zeb1 (328).

Collagen type I is the most abundant component of ECM and the most abundant fibril in all three tunicae of the blood vessels [reviewed in (329)]. Collagen is found around VSMC of the media, where it provides the necessary mechanical strength and contractility, and also in adventitia. Little is known about the regulation of collagen type I expression in blood vessels and especially the transcriptional control of this process during development and in adult vessels, where a precise transcriptional regulation is required to prevent disease caused by ECM deposition.

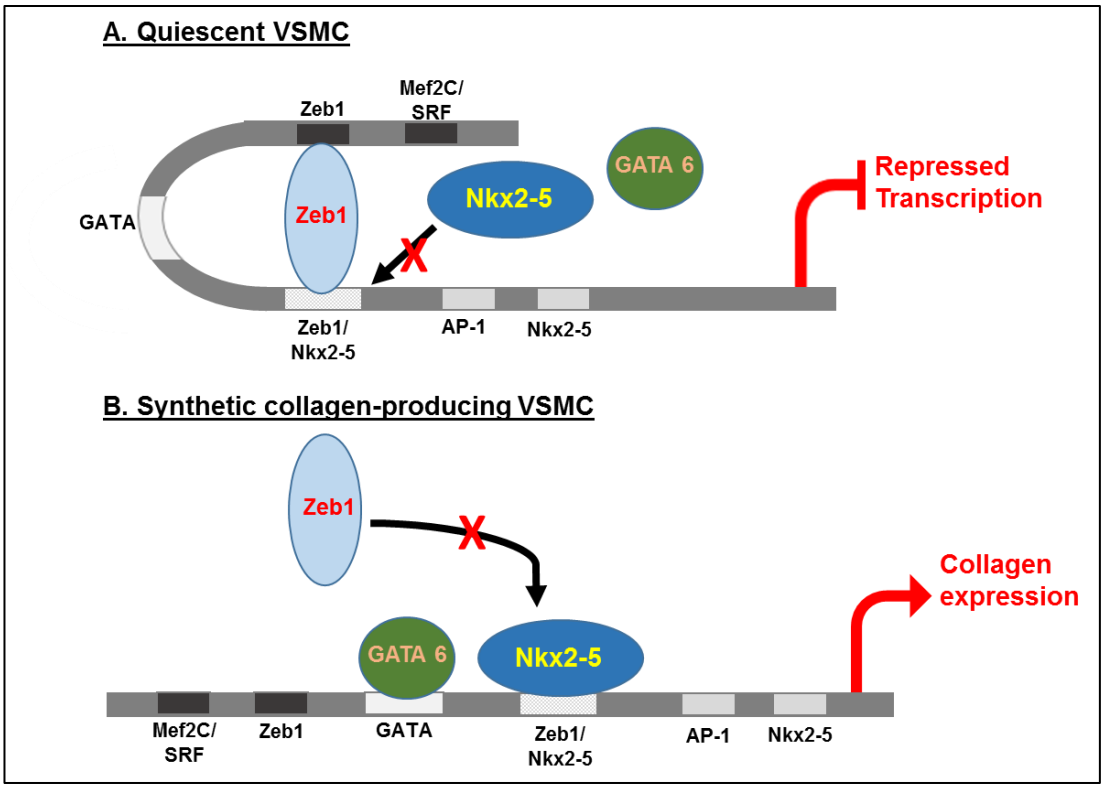


Figure 1.9 Nkx2-5 controls collagen expression in VSMC. *A mechanism has been identified and proposed by Ponticos et al (327), whereby collagen expression is controlled by Nkx2-5 in VSMC. A. Contractile quiescent VSMC do not produce collagen, and the δEF1/Zeb1 homeodomain repressor is bound on two adjacent sites in a VSMC-specific element located at approximately 16.6Kb upstream of the transcriptional start site. The 3' δEF1/Zeb1 DNA binding site overlaps a Nkx2-5 homeodomain activator binding site. When the δEF1/Zeb1 repressor is occupying both of its sites, Nkx2-5 is unable to access its binding motif. Gata6 is also unable to access its binding site when δEF1/Zeb1 is bound to the DNA. These conditions result in a repressed/silenced element. B. When VSMC are activated and collagen type I is synthesised, δEF1/Zeb1 is displaced from its binding site by Nkx2-5, opening the DNA and allowing GATA6 to bind and synergise with Nkx2-5, leading to transcriptional activation of pro-col1a2. Figure adapted from (328).*

Since the *Col1a2* enhancer is not expected to be engaged and active under normal conditions in most human adult tissues, it was hypothesised that Nkx2-5-dependent transcriptional activation of *Col1a2* would only occur under conditions where

synthesis of collagen and ECM is required, such as in response to injury or in disease. It is therefore expected that NKX2-5 itself is upregulated under similar conditions.

Indeed, we have recently demonstrated that NKX2-5 is expressed in human vascular pathologies including atherosclerosis, PAD and SSc-PAH (Figure 1.10), and it drives VSMC phenotypic modulation in vascular remodelling *in vivo* and *in vitro* (Figure 1.11) (330).

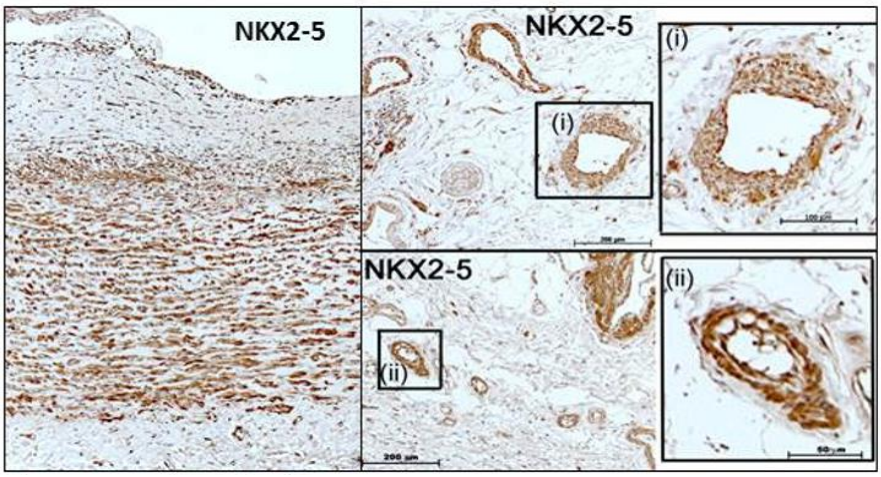


Figure 1.10 NKX2-5 is expressed by VSMC in human vascular pathology. *Sections of pulmonary tissue from SSc-PAH patients (n=3) were immunostained for NKX2-5 (brown) and counterstained with H&E. Expression is observed in large (>100μm, i), medium (40-70μm, ii) and small (20-40μm, iii) pulmonary arteries. Figure adapted from (330).*

We showed that NKX2-5 expression is increased in the synthetic phenotype, where it controls the de-differentiation of VSMC (Figure 1.11A). Specifically, when *NKX2-5* was knocked down *in vitro* in synthetic human pulmonary artery SMC (HPASMC) using siRNA, the phenotype was reversed to the contractile state (Figure 1.11B). Phenotypic modulation was arrested as shown by the expression of marker proteins specific to either the contractile or the synthetic phenotypes (Figure 1.11B) (330). In the same study, *Nkx2-5* was conditionally deleted in two mouse models of vascular disease: the chronic hypoxia model of pulmonary hypertension and the carotid ligation model of vascular injury. In both mouse models, deletion of *Nkx2-5* resulted in a significant decrease of vascular remodelling and a reduction of associated symptoms including decreased pulmonary pressures, decreased vascular resistance and muscularisation, and inhibition of injury-induced neointima formation (Figure 1.11C) (330).

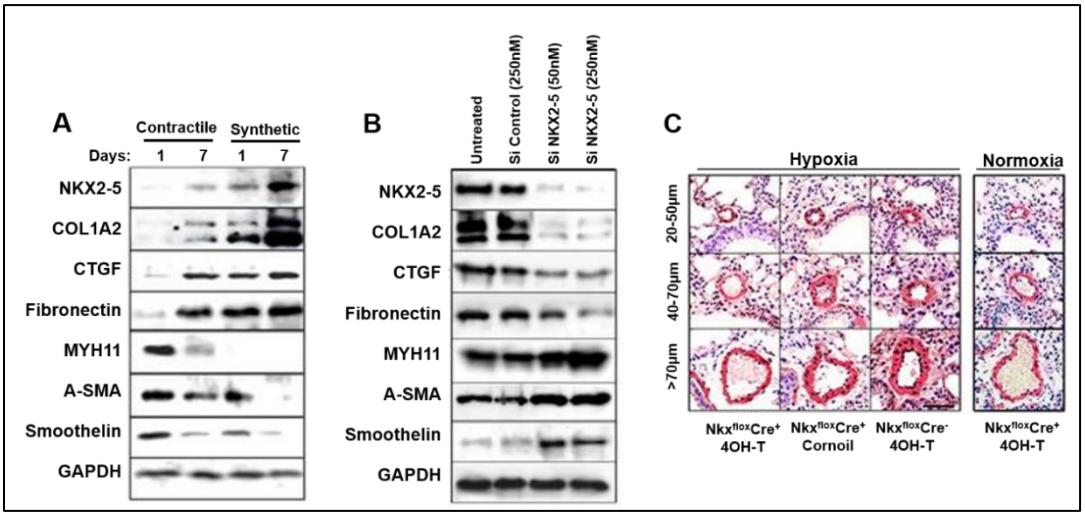


Figure 1.11 NKX2-5 regulates vascular remodelling. *A. HASMC were cultured in vitro under conditions favouring the contractile or synthetic phenotype, over a time course of 7 days. Protein expression levels of NKX2-5, synthetic protein markers COL1A2, CTGF and Fibronectin, and contractile protein markers MYH11 (myosin heave chain), α-SMA, and Smoothelin were analysed. NKX2-5 was not expressed in contractile HASMC, but expression increased in the synthetic state. B. NKX2-5 knockdown in synthetic HASMC (day 7) was carried out using siNKX2-5 (50nM or 250nM) or scrambled siControl (250nM) oligonuclueotides and expression of contractile and synthetic marker proteins was measured. Synthetic HASMC revert to a contractile phenotype after NKX2-5 knockdown compared to untreated or siControl-treated cells exhibiting decreased expression of synthetic markers and an increase in the levels of myosin heavy chain, α-SMA and Smoothelin. C. NKX2-5 null (NKX2-5^{flox} Cre^{+} 4OH-T) or control male mice were used in the chronic hypoxia model of pulmonary hypertension or under normoxia for 21days. Sections of the entire left lobe of the mouse lungs were immunostained for α-SMA to visualise vessels. Representative images of small (20-50μm), medium (40-70μm) and large arteries (>70μm) are shown. Figure adapted ftom (330).*

In this study, it was shown that NKX2-5 expression is activated in different adult vascular beds in disease, and mediates key repair and pathogenic processes. The data reveal a critical role for NKX2-5 in vascular disease. However, very little is known about the signalling mechanisms that activate NKX2-5 in disease, and understanding these mechanisms is very important. Dissection of these mechanisms is a key objective of this PhD, and considerable findings are presented in Chapter 4.

1.4.3 NKX2-5 is genetically associated with disease

In addition to a functional role in vascular remodelling, mutations and SNPs have been identified along the *NKX2-5* gene leading to various congenital heart defects. Apart from its involvement in the heart defects, NKX2-5 has not been studied in the genetic context of any other disease.

Of particular interest is a GWAS in SLE in a Japanese cohort, which reported a genetic association with a functional SNP located at the binding site of ITP3 (an inositol second messenger receptor) on the *NKX2-5* gene (331). Extending the analysis to RA and Grave's disease, they showed that rs3095870 was associated with susceptibility to common AIDs through functional interactions between ITRP3 and NKX2-5 (331). Rs3095870 is located upstream of *NKX2-5* transcription start site. However, this association has not yet been replicated in any other study cohort.

AIDs such as SLE and scleroderma exhibit similar disease manifestation and common pathological features such as inflammation and fibrosis. Although, the SNP association noted above has not been replicated, the study overall provides compelling evidence that *NKX2-5* gene might be implicated in the pathogenesis of AIDs. The functional data regarding the role of NKX2-5 in vascular remodelling further corroborates this hypothesis, since PH resulting from vascular remodelling is a common feature in both SLE and scleroderma.

It is, thus, vital to further explore the potential that the *NKX2-5* gene locus is genetically associated with AIDs in different study cohorts and different origins. This aspect is indeed another objective of this thesis.

1.4.4 The role of NKX2-5 in the cardiovascular system

1.4.4.1 Congenital heart disease (CHD)

CHD is a heterogeneous group of diseases characterised by structural and/or functional defects of the heart and the great vessels due to abnormal cardiac development. CHD is the most prevalent type of birth defect, with an estimated incidence of 4–50 in every 1000 live births, and is the leading non-infectious cause of infant death worldwide (332). Therefore, in recent years, there has been a considerable interest in studying the genetic determinants of CHD. Even though significant progress has been achieved in diagnostic and therapeutic strategies, the aetiology of CHD is not well understood. However, advanced sequencing techniques have provided increasing evidence that defects in single genes cause various kinds of CHD, with *NKX2-5* found to be the most commonly mutated gene (333). Over 41 missense and nonsense mutations within the *NKX2-5* gene have been identified in CHD patients, and abnormal expression levels of *NKX2-5* were also found to be associated with multiple cardiac malformations (50, 334-336). CHD caused by mutations in *NKX2-5* include atrial septal defect (ASD), atrial ventricular block (AVB), tetralogy of Fallot, and ventricular septal defect (VSD).

1.4.4.2 Cardiac conduction system (CCS)

The CCS is a specialised structure responsible for the coordinated contraction of the heart by establishing and maintaining electrophysiological activities. The development of CCS is a highly complex process and closely associated with cardiac development. Purkinje fibres are specialised cardiomyocytes whose role is to co-ordinate the rapid spread of action potential in the ventricular myocardium, and their differentiation and maturation requires precise regulation by Nkx2-5 (337). Thus, Nkx2-5 expression is elevated and correlates with the recruitment of cells to the development of the ventricular conduction system, but the exact mechanism remains unknown (338).

Disorders of the CCS occur often and result in arrhythmias, a condition that can be life-threatening. Dominant mutations in the mouse *Nkx2-5* gene result in electrophysiological abnormalities of the conduction system independent of the presence of CHD (49). In addition, mutations have also been identified in the human *NKX2-5* gene that result in AVB (339).

1.4.4.3 Vasculature

NKX2-5 is not normally expressed in the adult vasculature. However, studies in mice suggest that the gene retains an important role in vessel formation. *Nkx2-5* null mice have poorly developed blood vessels in addition to heart malformations. More specifically, *Nkx2-5* is upstream of *Tbx1*, a gene important for the morphogenesis of the outflow tract (OFT), and when *Tbx1* was deleted from Nkx2-5-expressing cells, mutant mice displayed a defective aortic arch phenotype (340).

Nkx2-5 has also been shown to be expressed and to regulate the formation of the pulmonary myocardium, a myocardial layer which forms a sheath around the pulmonary vessels (341). The pulmonary myocardium is an important source of electrical activity that initiates atrial fibrillation and expresses the gap-junction protein Connexin-40, a downstream target of Nkx2-5 essential for fast atrial conduction (341).

1.4.4.4 Adult tissues

In adulthood, NKX2-5 is only expressed in the heart and the spleen. The overall role in the heart tissue is further confirmed by the activation of NKX2-5 expression in heart hypertrophy, induced by agonists and right ventricular pressure overload (342). However, NKX2-5 is not directly causal and alone is not sufficient to induce hypertrophy (343). Recently, studies in transgenic mice overexpressing a dominant

negative mutant of *Nkx2-5* driven by the *α-SMA* promoter showed impaired cardiac function and degeneration of cardiomyocytes (344). Overall, this study suggested that Nkx2-5 retains a protective role in the adult heart against stress or cytotoxic damage.

1.4.5 The regulation of NKX2-5

The structure of the human *NKX2-5* gene is complex and not well studied, and most information comes from *in vivo* and *in vitro* studies in mice. Since NKX2-5 is not expressed in normal healthy blood vessels, the regulatory mechanism responsible for the transcriptional regulation of NKX2-5 in vessels or in VSMC has not been defined.

1.4.5.1 Transcriptional regulation of the mouse Nkx2-5 gene

Similar to the human *NKX2-5*, the mouse gene consists of two exons and can be spliced to alternative isoforms (Figure 1.8B-C). Studies of the mouse gene have revealed a very complex structure, with a regulatory region of 23Kb surrounding the gene, which contains at least 7 activating and 3 possible inhibitory regions [reviewed in (345)]. The activating regions provide tissue specificity, since they are active in certain locations of the embryo with distinctive regulatory elements specific to the heart, thyroid, spleen, pharynx, and stomach. A study by Chi and Schwartz identified 3 distal enhancers over 20Kb upstream of the transcriptional start site arranged in a modular manner and responsible for later cardiac chamber specification and expression in tongue (346).

In the mouse *Nkx2-5* gene, a regulatory element located between 3-2.5Kb upstream of the transcriptional start site was identified as sufficient for the initial expression of the gene during heart development (347, 348). This enhancer region contains essential GATA binding sites, as well as a cluster of Smad binding sites, which are targets of BMP signalling. Another enhancer specific to the heart tissue contained GATA/Smad binding sites and was a direct target of Smad4 (349). In addition, Brown *et al* identified a novel enhancer region that was also rich in GATA and Smad binding sites, and they specifically showed that BMP signalling activates Nkx2-5 directly through Smad1/4 (350), whereas there is no co-dependence on Gata4 *in vivo*. In the same study, it was proposed that the TGF-β signalling pathway may regulate Nkx2-5 activity via different members of TGF-β superfamily.

Apart from GATA and Smads, Mef-2c has also been shown to regulate Nkx2-5 expression, but this relationship is reciprocal since the two factors regulate each

other (351). Furthermore, Hif-1α was also proposed as an upstream activator of *Nkx2-5* in Xenopus, however, whether the effect is direct through Hif-1α binding upstream of *Nkx2-5* or indirect is not clear (352).

1.4.5.2 Signalling pathways involved in Nkx2-5 activation

The pathways that activate Nkx2-5 expression are complex. Indeed, signalling mechanisms that result in the activation of Nkx2-5 are still under investigation, and form a substantial part of this thesis.

The TGF-β signalling pathway has been implicated in many cellular processes with a panel of members of the superfamily including BMPs and SMADs controlling heart development, cell differentiation, and vascular remodelling. Experiments conducted in chicken embryos have shown that Bmp2 is sufficient to induce *Nkx2-5* in the stomach, and Bmp4 also controls *Nkx2-5* expression (353). Conserved binding sites for Smads have been identified in the regulatory region upstream of the murine *Nkx2-5* gene, through which BMPs can control its expression depends on the spatio-temporal conditions (354). Indeed, Smad4 binds those elements and is required for *Nkx2-5* transcription (349). Apart from positive regulation, TGF-β signalling has been implicated in a negative feedback loop, during which Nkx2-5 controls cell proliferation of cardiac progenitor *in vivo* through the repression of Bmp2/Smad1 signalling (44).

We have previously shown that TGF-β is able to activate NKX2-5 expression in human VSMC (330), and similar patterns are also seen in stem cell studies. TGF-β upregulates *Nkx2-5* expression in skeletal muscle derived primitive cells (113). In addition, in mouse embryonic teratocarcinoma stem cells, the use of a TGF-β neutralising antibody inhibited the induction of *Nkx2-5* and prevented cardiomyocyte differentiation (355).

Another signalling pathway proposed to regulate *Nkx2-5* expression is the Wnt signalling pathway. The Wnt signal transduction cascade controls a myriad biological phenomena throughout development and adult life of all animals, as well as a wide range of pathologies in humans [reviewed in (356, 357)]. Wnt1 and Wnt3α, members of the Wnt canonical pathway, inhibit *Nkx2-5* expression in the anterior mesoderm (358). In addition, use of Wnt inhibitors in mouse embryonic stem cells prevented Nkx2-5 and Gata4 activity (359). In a recent study, it was also shown that Nkx2-5 leads to attenuation of the Wnt/β-catenin pathway and that pharmacological activation of Wnt signalling can significantly ameliorate the phenotype in the conditional *Nkx2-5* mutants (360).

A study by Patel and Kos showed that ET-1 and neuregulin-1 upregulate Nkx2-5 expression in murine embryonic cardiomyocytes (361). Neuregulin-1 (NRG-1) is a multifunctional regulator that acts through receptor tyrosine kinases of the EGF/ErbB receptor family in diverse tissues. ErbB receptors are expressed in development, and the importance of the NRG-1/ErbB signalling axis in heart development has been investigated but is still poorly described. In a study by Wang *et al*, NRG-1 up-regulated the expression of Nkx2-5 and Gata-4, which was later blocked by inhibitors of PI3K and the ErbB receptor (362). Furthermore, it was recently reported that ET-1 is a downstream target of Nkx2-5 in H9c2 cardiomyoblasts (363). These findings suggest a potential positive feedback loop involving Nkx2-5 and the ET-1 cascade.

1.4.5.3 Post-translational regulation of Nkx2-5

Very little is known about the post-translational regulation of Nkx2-5. Kazahara and Izumo explored the phosphorylation of Nkx2-5 protein and found that Nkx2-5 is phosphorylated *in vitro* and *in vivo,* and that cytoplasmic and nuclear proteins are differentially phosphorylated (364). In addition, they showed that casein kinase 2 (CK2) is one of the kinases that phosphorylates Nkx2-5 *in vivo* at a highly conserved serine residue (serine 163) of the homeobox leading to its nuclear translocation of the protein (364). Based on this findings, inhibitors of CK2 prevent Nkx2-5 nuclear translocation and therefore its activation, leading to down-regulation of processes controlled by Nkx2-5 such as VSMC de-differentiation, proliferation and migration (365).

Nkx2-5 is also regulated by sumoylation, another post-translational modification. Sumoylation has been studied in terms of sub-nuclear localisation and genome integrity. Small-ubiquitin like modifiers are small molecules that can be covalently and reversibly conjugated to specific lysine residues localised in SUMO-targeted sequences. Nkx2-5 is sumoylated on lysine 51, a modification that substantially increases its activity (366). Mutation of lysine 51 to arginine (K51R) suppressed SUMO binding and the activity of Nkx2-5 (366).

To add to the already complex regulation, Nkx2-5 is also modified by glycosylation, and excessive O-GlcNAcylation results in the down-regulation of Nkx2-5 protein (367). Regulatory mechanisms of NKX2-5 are summarised in Table 1.2.

NKX2-5 Regulation		
Cytokines, Growth Factors, Transcription Factors	**Signalling pathways**	**Post-transcriptional/ post-translational modifications**
• BMPs • GATA • Smads • Mef-2c • HIF-1α • ET-1	• TGF-β pathway • WNT pathway (Wnt1, Wnt3α) • MAPK cascade	• Phosphorylation (364) • Sumoylation (368) • Glycosylation (367) • Acetylation (369)

Table 1.2 Regulatory mechanisms that affect expression of the mouse NKX2-5 gene. *The table summarises growth factors, signalling pathways and mechanism that are known to regulate the mouse Nkx2-5 gene and protein expression. The information reported in this table is mostly reviewed in (370). Where information is based elsewhere, the reference is given on the table.*

1.4.6 Nkx2-5 downstream targets

Most of the available data regarding Nkx2-5-dependent regulation arise from developmental studies and gene arrays conducted in genetically-engineered animals. Reporter assays, binding assays, immunoprecipitation and mutational analysis have provided significant insights regarding Nkx2-5 downstream targets. Available published data has been carefully collated by Ponticos *et al* (371) (Table 1.3). However, the list of Nkx2-5 downstream targets increases rapidly. Both the human and the mouse NKX2-5 proteins recognise the NKE consensus DNA sequences, where it binds with high specificity to regulate expression. These sites are generally found in promoters and enhancers of downstream targets that are usually genes crucial in development, VSMC differentiation and production of ECM.

NKX2-5 exerts its function alone or interacts with other proteins to achieve positive or negative regulation of downstream targets. Identification and mutational analysis of NKE consensus DNA binding sites in the promoters and enhancers of various genes provided evidence that many transcription factors and transcriptional regulators are direct targets of Nkx2-5. In detail, Riazi *et al* showed that Nkx2-5 binds NKE sites in promoters of *β-catenin* and *Gata4* and it is essential for the suppression of the first and upregulation of the second (372). β-catenin is a part of Wnt signalling pathway that has emerged as a key regulator of cardiac progenitor cell specification. Findings provide compelling evidence that the Wnt/β-catenin pathway plays a positive regulatory role in precardiac and cardiac mesoderm, and

promotes committed cardiac cell proliferation and differentiation (373). Recently, a novel pathway was identified by which Nkx2-5 upregulates Wnt signalling and promotes cardiac cell growth by regulating directly the expression of R-spondin3 (360).

Target Gene Name	Function/Process	References
β-catenin, Gata4	Human cardiac myocytes maturation	(372)
Mef-2c	Cardiac, skeletal muscle, smooth muscle determination	(351)
Myocardin	Embryonic development and VSMC differentiation	(374)
Tbx1, HOP, Id2, eHand, Pitx2	Embryonic development (cardiac development, ventricular conduction system, cardiac outflow tract, left-right symmetry organogenesis, formation of pulmonary myocardium, left ventricular development)	(340, 341, 375-378)
Connexins (-40,-43,-45)	Gap junction proteins involved in cardiac conduction system	(379, 380)
Col1a2, Plod1	Structural component of ECM involved in embryogenesis, adult wound repair, and fibrosis	(327)
Plod1	Stabilisation of collagens	(376)
Cardiac α-actin	Contractile protein	(381)
Ece-1	Pharyngeal artery patterning	(382)
Stat4	Proliferation of endothelial precursor of mesoderm	(383)

Table 1.3 Selected downstream target genes of the mouse Nkx2-5 protein. *Table adapted from (371).*

In addition, Skerjanc *et al* have shown that Mef-2c and Nkx2–5 upregulate each other's expression, induce cardiomyogenesis, and activate the promoters of cardiac muscle-specific genes, which contain NKE and MEF2 binding sites in their promoters (351). Another example is *myocardin* that was found to be downregulated in *Nkx2-5* null mouse hearts (374). Indeed, 5 NKE elements within the *myocardin* promoter were identified, one of which was necessary for the full activation of the gene. In the same study, it was shown that BMP/TAK1 signalling augmented *myocardin* expression through the SRE sites.

Another group of proteins that are directly regulated by Nkx2-5 are the Connexins. Connexins are gap-junction trans-membrane proteins specialised in cell–cell communication that directly link the cytoplasm of neighbouring cells (384). They mediate the direct transfer of metabolites and ions from one cell to another.

Therefore, it has long been hypothesised that Connexins retain a crucial role in the maintenance of homeostasis, morphogenesis, cell differentiation, and growth control in multicellular organisms. In cardiomyocytes specifically, Connexins are responsible for co-ordinated contraction. Connexins 40, 43 and 45 are expressed in the heart and CCS and they are all regulated by Nkx2-5 [reviewed in (385)]. Multiple NKE binding site have been identified in promoters of *Connexin*s, where Nkx2-5 can act either as an activator or repressor of transcription.

Pitx2 is a transcriptional marker of vertebrate heart development, and *ANF* promoter is one of its downstream targets. Ganga *et al* showed that Pitx2c isoform can synergistically activate the *ANF* promoter in the presence of Nkx2-5. In addition, they showed that *Plod1* promoter is also regulated by Nkx2-5 (376). Mechanistically, Pitx2c and Nkx2-5 co-operate to regulate *ANF* and *Plod1* transcription through binding at their respective DNA elements.

As mentioned already, the *Col1a2* gene, which is a direct target of Nkx2-5 in VSMC, is an important component of the ECM that is activated in physiological processes such as tissue injury and repair, but also during disease in fibrosis and vascular remodelling. Another Nkx2-5 downstream target involved in ECM is the Ece-1, which is a zinc metalloprotease that cleaves and activates ET-1 (363).

Apart from the important role of NKX2-5 in development that is well established, some of its downstream targets are genes encoding proteins implicated in the deposition and regulation of ECM, such as collagens and Plod1, and smooth muscle-specific structural proteins, such as α-SMA and myocardin. Dysregulation of these genes by Nkx2-5 could lead to de-differentiation of VSMCs towards the synthetic phenotype that is related with CVD, and also extreme deposition of ECM within the vascular wall, which could lead to increased medial thickening and vascular resistance in the pulmonary vasculature, leading ultimately to PAH.

1.5 Aims and objectives

My hypothesis is that the *NKX2-5* gene is genetically associated with vascular disease, and regulated at transcriptional, post-transcriptional and epigenetic levels through mechanisms that increase its expression during vascular remodelling.

The overall aim of this thesis is to explore these mechanisms in adult human vessels. The individual aims and objectives are the following:

i. **Genetic association of *NKX2-5* with vascular disease**: I will perform a candidate gene genetic association study of the *NKX2-5* genomic locus in scleroderma patients. Case-control, sub-phenotype and haplotype analysis will be used to test the hypothesis. Any associated SNPs will be further studied for potential functionality.

ii. **The signalling mechanisms that activate NKX2-5 expression:** I will investigate the signalling pathways that increase NKX2-5 expression in HPASMCs by treating the cells with a panel of selective inhibitors of signalling pathways and protein kinases.

iii. **The regulation of the *NKX2-5* gene**: I will investigate the transcriptional activity of the disease-associated SNPs. I will also study the transcriptional activation of *NKX2-5* through the binding of activators and/or repressors on promoter or enhancer regions. In addition, I will explore whether NKX2-5 expression is regulated via microRNAs or DNA methylation.

iv. **The expression of NKX2-5 in the endothelium in disease**: The final aim is to explore if NKX2-5 is expressed in two different types of endothelial cells under normal or disease-associated conditions such as EndoMT.

Ultimately, this study aims to provide a better understanding of the functions and regulation of NKX2-5, an established master regulator of vascular remodelling. It could eventually contribute in designing strategies to block NKX2-5 expression, which would prevent the progression of disease pathogenesis and ameliorate the symptoms in patients with vascular diseases.

CHAPTER 2 - METHODS AND MATERIALS

2.1 Study cohorts

Two independent cohorts were used for the genetic analysis in this study. Each cohort is described in detail below. All participants (scleroderma patients and healthy individuals) gave written informed consent to participate in the study. The study was approved by the local ethics committees.

2.1.1 Description of discovery (UK) cohort

The discovery cohort consisted of a total of 1334 scleroderma patients; 899 patients presenting to the Centre for Rheumatology, Royal Free NHS Foundation Trust Hospital, London and 435 patients from the Institute of Inflammation and Repair, University of Manchester. A total of 901 control DNA samples were included in the discovery cohort; 487 samples were collected from healthy donors of UK/Caucasian origin at the Centre for Rheumatology, 192 samples of random human DNA donors were purchased from Sigma/ECACC, and 222 samples were kindly provided by Professor Steve Humphries, Institute of Cardiovascular Genetics, UCL. The control samples were matched for ethnic origin, age and sex.

2.1.2 Description of replication (Spanish) cohort

The replication cohort consisted of 1736 scleroderma patients and 1753 healthy individuals of Spanish origin. DNA samples were collected from different Rheumatology clinics across Spain and were sent to the Institute of Parasitology and Biomedicine Lopez-Neyra, Granada, Spain. The replication cohort was genotyped in Spain by Dr Lara Bossini-Castillo and Ms Aurora Serrano Lopera under the guidance of Professor Javier Martin.

2.1.3 Clinical diagnosis

All patients diagnosed with scleroderma fulfilled the 2013 classification criteria for scleroderma issued by the collaboration of the American College of Rheumatology (ACR) and European League Against Rheumatism (EULAR) (231). The joint committee determined that skin thickening of the fingers extending proximal to the metacarpophalangeal joints is sufficient for the patient to be classified as having scleroderma. Otherwise, diagnosis can be based on the presence of the following seven features: skin thickening of the fingers, fingertip lesions, telangiectasia,

abnormal nail fold capillaries, interstitial lung disease or PAH, Raynaud's phenomenon, and scleroderma-related auto-antibodies.

2.1.4 Sub-phenotypes

Scleroderma patients were categorised into groups based on:

i) The disease subset: lcSSc (limited) and dcSSc (diffuse);

ii) The presence of auto-antibodies verified by immunoassays: ACA (anti-centromere), ATA (anti-topoisomerase I), and ARA (anti-RNA polymerase III);

iii) The presence of organ complications: PF that was assessed by high resolution computed tomography (HRCT) and a restrictive pattern on the lung function test, PAH that was defined as a mPAP of >25mmHg with normal pulmonary arterial wedge pressure of <15mmHg on RHC, PH when PAP>25mmHg was observed in the presence of PF at the same time, and RC when rapidly progressive renal failure and new onset accelerated hypertension occurred.

No ARA and RC data were available for the scleroderma patients from Manchester or for the replication cohort from Spain.

2.2 Selection of tagging SNPs

A tagging SNP is a representative SNP that captures the genetic variation of a genomic region with high LD, which is known as a haplotype block. The use of tagging SNPs eliminates the need to study every individual SNP reducing the time and expense of mapping genomic areas associated with disease. Tagging SNPs are also useful in GWAS in which hundreds of thousands of SNPs across the entire genome are genotyped.

2.2.1 Application of online tools

For the purpose of this study, I focused on a genomic region of 13.2Kb that centred *NKX2-5* and overhung the gene upstream and downstream by 5Kb (chr5:172,654,107-172,667,315, GRCh37/hg19 assembly, UCSC Genome Browser). After selecting the area, genotype data for the CEU cohort (cohort of 180 Utah residents with Northern and Western European ancestry) were downloaded from the International HapMap Project and used in different online tools to select the tagging SNPs at the genomic region. I used Tagger (386), a tool for the selection and evaluation of tagging SNPs from genotype data. Tagger is a user-friendly tool

that combines the simplicity of pairwise tagging methods with the efficiency benefits of multimarker haplotype approaches. As output, Tagger produces a list of tagging SNPs and corresponding statistical tests to capture all variants of interest, and a summary coverage report of the selected tagging SNPs. Tagger has also been implemented in Haploview (387), which gave similar results when used.

2.2.2 Selection of best candidate tagging SNPs

After obtaining a list of candidate tagging SNPs, I aimed to prioritise them based on functional evidence available from an *in silico* analysis. The original list consisted of 12 tagging SNPs. However, due to the extensive LD in the area, any SNP could be chosen as a successful tagger. The *in silico* analysis was conducted in order to select 6 tagging SNPs across the *NKX2-5* genomic locus. For the analysis, I used data from the ENCODE project (Encyclopaedia of DNA elements) available through the UCSC Genome Browser and the HaploReg software (3). The ENCODE project provides functional annotation of gene elements that is accomplished primarily by sequencing a diverse range of RNA sources, comparative genomics, integrative bioinformatics methods, and human curation. Regulatory elements are typically investigated through DNA hypersensitivity assays, assays of DNA methylation, and immunoprecipitation of proteins that interact with DNA and RNA, i.e., modified histones, transcription factors, chromatin regulators, and RNA-binding proteins, followed by sequencing. HaploReg is a tool for exploring annotations of the non-coding genome at variants in haplotype blocks, such as candidate regulatory SNPs at disease-associated loci. Using LD information from the 1000 Genomes Project, linked SNPs and small indels can be visualised along with the chromatin state and protein binding annotations from the Roadmap Epigenomics and ENCODE projects, sequence conservation across mammals, the effect of SNPs on regulatory motifs, and the effect of SNPs on expression from eQTL studies.

2.3 Extraction of genomic DNA

2.3.1 Extraction from blood samples

DNA was extracted from blood samples using a simple salting out method as described by Miller *et al* (4). Briefly, the blood was collected into tubes containing anticoagulant (EDTA) and spun down at 1000xg for 10 minutes at 4^{o}C. The white blood cells were washed twice in 0.144M NH_4Cl, 1mM NaHCO3, pH 7.4 and spun down at 1000xg for 10 minutes at 4^{o}C. The white blood cell pellet was resuspended in 10mM Tris-HCl, 400mM NaCl, 2mM Na_2EDTA, pH 8.2 and incubated for 60

minutes at 37°C. Saturated NaCl (1ml) was added to the lysate followed by vigorous vortexing for 15 seconds. After centrifugation, the supernatant containing the DNA was transferred to a clean tube and the DNA was ethanol precipitated.

2.3.2 Extraction from cells

There are a number of different procedures for the preparation of genomic DNA. They all start with a basic cell lysis step, followed by deproteinisation and recovery of DNA. The main differences between various approaches lie in the extent of deproteinisation and in the molecular weight of the recovered DNA. The isolation protocol I used combines the powerful proteolytic activity of proteinase K and the denaturing ability of the ionic detergent SDS. The cell pellet is first loosened in ice-cold TE (10mM Tris pH 8.0, 1mM EDTA) buffer before the cells are lysed in lysis buffer (TE, 20µg/ml RNase A, 0.5% SDS). EDTA prevents the enzymatic activity of DNases in the lysis buffer. The lysate is incubated at 37°C for 1 hour for RNA digestion. Next, 100µg/ml Proteinase K is added and the samples are incubated at 50°C for 3 hours to overnight. Next day, DNA is recovered through phenol-chloroform extraction and ethanol precipitation.

2.3.3 Phenol-chloroform extraction and ethanol precipitation

Phenol extraction is a commonly used method for removing proteins from a DNA sample during genomic DNA preparation. The procedure is described in steps in Table 2.1. After the addition of phenol: chloroform: isoamyl alcohol two phases are formed: an aqueous phase on top that contains the DNA, and the phenol phase at the bottom where all the proteins have been trapped. The upper aqueous phase is carefully transferred to a new tube. If needed, the procedure is repeated. A last chloroform extraction following the same steps should be done to completely remove the phenol. Next, the salts are removed from the sample and the DNA is concentrated by ethanol precipitation (Table 2.1). The DNA pellet is reconstituted in water or TE.

During the ethanol precipitation, carriers and ammonium or sodium acetate can be used. Carriers or co-precipitants are substances used during alcohol precipitations to facilitate the recovery of target nucleic acids. They are insoluble in ethanol or isopropanol solutions, and they form a precipitate that helps to trap nucleic acids. During centrifugation, carriers form a visible pellet, which aids in removing the supernatant without perturbing the nucleic acid pellet. Common carriers are the yeast tRNA, salmon sperm DNA, and glycogen (used at 50-150 µg/ml final concentration). Addition of sodium or ammonium acetate (NH_4OAc) can be added to

enhance the removal of DNA-binding proteins and free unincorporated dNTPs. In the past, sodium acetate was more widely used, but it has now been shown that ammonium acetate performs better.

Procedure	**Reagents**	**Centrifuge**
Phenol/Chloroform precipitation	1x Phenol: Chloroform: Isoamyl alcohol (25:24:1)	16000xg, 10', RT
	1x Chloroform	16000xg, 10', RT
Ethanol precipitation	2.5x Ice-cold 100% Ethanol, 0.5x Ammonium acetate, 50-150µg/ml Glycogen	16000xg, 30', 4^{o}C
	1x Ice-cold 70% Ethanol	16000xg, 10', 4^{o}C

Table 2.1 Phenol/chloroform extraction and ethanol precipitation. *A short description of the protocol is given including the materials required, the centrifugation timing and temperatures per step.*

2.3.4 Extraction from cells or vessels using a commercial kit

Many of the traditional protocols that have been used in the labs for decades have been exchanged for simple and straight-forward commercial kits. The DNeasy Blood & Tissue Kit (Qiagen, Cat. no: 69506) is designed for rapid purification of total DNA from various sources. The protocol is based on a filter spin-column technology, it does not need any phenol/chloroform extraction or ethanol precipitation and the DNA is free of contaminants and suitable for high-throughput downstream applications such as sequencing. In this study, the kit has been used occasionally when the source of DNA was limited or when required for a sensitive downstream application.

2.3.5 Measurement of DNA concentration and purity.

The quantity and the purity of the DNA were measured using the Nanodrop 2000. This instrument is a scanning spectrophotometer which uses fibre-optic technology. Nucleic acids have a peak absorbance wavelength at 260nm, while a variety of contaminants such as phenol and protein absorb at 280nm. Therefore, the ratio 260/280 defines the purity of the sample. A ratio of ~1.8-~2.0 for DNA and ~2.0 for RNA is generally accepted as pure. All the DNA samples that were used for downstream applications such as genotyping had a 260/280 ratio between of ~1.8-2.0.

2.4 SNP genotyping

Two methods of SNP genotyping were used in this study, both described in detail below.

2.4.1 TaqMan SNP genotyping

The TaqMan SNP genotyping assays exploit the chemical properties of the 5'-nuclease that provides a fast and simple way to distinguish SNP genotypes. Each predesigned assay includes two allele-specific probes containing distinct fluorescent dyes and a PCR primer pair to detect specific SNP targets. These probes and primer sets uniquely align with the genome to identify the allele of interest with high specificity. For the genotyping of SNPs rs703752, rs3131917 and rs3132139 TaqMan pre-designed assays (Thermo Fisher Scientific, Cat. no: 4351379) were used. TaqMan custom assays were specifically designed for SNPs rs12514371 and rs2277923.

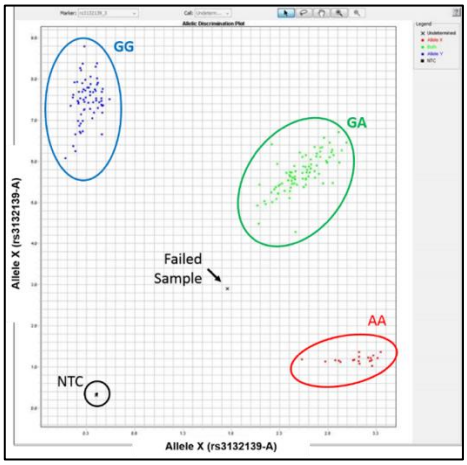


Figure 2.1 TaqMan-Allele Discrimination plot. *An allele discrimination plot, also known as a "cluster plot", for SNP rs3132139 is shown in the figure. Three clusters formed showing the three different genotypes (AA, AG, GG). The genotypes in each cluster are grouped closely together, and each cluster is well separated from the others. No-template control (NTC) samples and samples that failed in the assay appear near the origin of the plot. Reasons for failed genotyping include absence of the sample in the reaction, low DNA concentration, low purity etc.*

All the DNA samples were plated in 96-well plates at a concentration of 20ng/µl and stored at -20^{o}C. For the PCR amplification, 20ng of DNA per sample were added in a 5µl final volume of reaction in a 384-well plate and amplified on an ABI GeneAmp PCR System 9700 according to the manufacturer's instructions (Thermo Fisher Scientific) (Table 2.2 & 2.3). For the post-read and the allelic discrimination

protocols an ABI Prism 7900HT Real-Time Thermocycler was used. The genotypes were then analysed using SDS 2.3 software (Applied Biosystems).

Reagents	1x	Final Concentration
DNA	1µl	20ng
TaqMan Universal PCR Master Mix	2.5µl	1x
SNP genotyping assay	0.25µl	1x
Water	1.25µl	
Total volume	5µl	

Table 2.2 TaqMan-PCR reaction set-up.

Step	Temperature and Time	No of Cycles
Hold	95 ^{o}C, 10'	1 Cycle
Denature	92 ^{o}C, 15''	40 Cycles
Anneal/Extend	60 ^{o}C, 1'	

Table 2.3 TaqMan-PCR cycling conditions.

2.4.2 High resolution melting (HRM) genotyping

SNP rs3095870 is surrounded by a highly repetitive sequence and a long string of A and T nucleotides that did not allow the design of a successful TaqMan assay. Instead, an alternative genotyping method was used for this SNP known as high resolution melting (HRM) analysis (Type-it HRM PCR Kit, Qiagen, Cat. no: 206542).

HRM analysis is a relatively new, post-PCR analysis method used to identify variations in nucleic acid sequences. The method is based on detecting small differences in dsDNA-binding fluorescent dyes (such as the EvaGreen dye used in the Type-it HRM kit) during the dissociation of DNA molecules. The dye is incorporated to DNA during a first PCR amplification step with SNP-specific primers. For the genotyping, the amplified DNA molecules are heated at high temperature to denature and the fluorescent dye fades away as the double stranded DNA separates, generating a melting curve. Because different genetic sequences melt at slightly different rates, the dye will fade at different times based on the alleles of the SNP. The intensity of the dye is monitored with real-time PCR instrumentation and software designed specifically for HRM analysis.

For the genotyping of rs3095870, an 85bp long genomic region surrounding the SNP was amplified using the following primers: forward: 5'- GACTCCTGAA TTGTAAGCAA-3', and reverse: 5'- GGGAGGTCTGATGAAAGC-3'. 10ng of DNA was added in a 10µl total reaction with 0.7µM of each primer (Table 2.4). The DNA amplification was done in a Corbett Rotor Gene 6000 cycler following the manufacturer's instructions (Table 2.5).

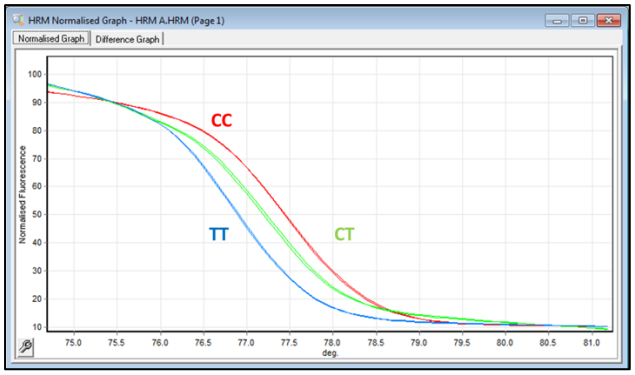


Figure 2.2 High Resolution Melting (HRM) analysis for rs3095870. *The plot shows the melting curve analysis for the genotyping of 6 samples. Three different curves appear for each different genotype (CC, CT, TT) that display a different melting temperature shown in the x axis. The melting curve for the heterozygote genotype always appears between the other two curves.*

Reagents	1x	Final Concentration
DNA	1µl	10ng
HRM Master Mix	5µl	1x
Forward Primer	0.7µl	0.7µM
Reverse Primer	0.7µl	0.7µM
Water	2.6µl	
Total volume	10µl	

Table 2.4 HRM-PCR reaction set-up.

Step	Temperature and Time	No of Cycles
Hold	95 ^{o}C 5'	1 Cycle
Denature	95 ^{o}C 5''	40 Cycles
Anneal/Extend	55 ^{o}C 30'' and acquire on green	
HRM	from 65^{o}C to 95^{o}C, 0.1^{o}C per step 90'' pre-melt, 2'' per step	

Table 2.5 HRM-PCR amplification.

2.5 Genetic association study

2.5.1 Quality control and association analysis in Plink

For the genetic association analysis, I used Plink (388), a free open-source whole genome association analysis toolset. The tagging SNPs included in this study had a minor allele frequency (MAF) >0.01 [plink --maf 0.01]. All the SNPs had a successful genotyping call rate ≥90% (more than 90% of the samples were genotyped per SNP) [plink --geno 0.1]. Only individuals that were successfully genotyped for at least 4 SNPs were included in the study [plink --mind 0.4]. Due to the fact that the gender was not known for a portion of the samples and hypothesizing that the control samples matched at best the scleroderma cases in age and sex, I excluded the sex from the analysis [plink --allow-no-sex].

An overall genetic association study was conducted between scleroderma patients and controls in Plink using a basic association analysis and a model analysis. The association analysis [plink --assoc] examines the basic allelic test (A_1 vs A_2). The model analysis [plink --model] examines 3 different models of association: the dominant (A_2A_2 +A_1A_2 vs A_1A_1), the recessive (A_2A_2 vs A_1A_1 + A_1A_2), and the additive (A_1A_1 + A_2A_2 vs A_1A_2). For the meta-analysis, an association test was conducted across the two independent cohorts [plink --meta-analysis].

2.5.2 Sub-phenotype analysis

A sub-phenotype association analysis was performed to test the associations between the SNPs and the scleroderma-associated phenotypes described in 2.1.4. The same method of analysis with both the allele test and the model analysis was performed. The sub-phenotype analysis was structured as shown below:

1) Phenotype-positive vs healthy controls: examines whether there is a significant association or not
2) Phenotype-positive vs phenotype-negative: examines whether the association found is scleroderma-dependent or specific to the phenotype
3) Phenotype-negative vs healthy controls: examines whether the association is random when the apparent associated phenotype is removed.

2.5.3 LD, haplotype blocks and analysis

The concepts of LD and the haplotype blocks, and their fundamental importance in the study of the human genome and genetic diseases are reviewed in detail in (248, 389). The genome is divided into haplotype blocks, which are regions of high LD

separated by recombination hotspots that occur at random times during evolution. SNPs that reside within haplotype blocks are in high LD with each other and can be inherited together by the next generations. The alleles of these SNPs can co-occur in certain combinations more often than expected by chance, which is called a haplotype.

Patterns of LD are of fundamental importance in gene mapping, and also important for the understanding of the evolutionary history of humans, identification of demographic effects in population growth, and the detection of natural selection. Examining haplotypes can lead to the identification of patterns of genetic variation that are associated with health or disease. Haplotype-based association methods are generally regarded as being more powerful than methods based on single markers since they fully exploit LD information from multiple markers (390).

HAPLOVIEW software was used for the haplotype association analysis, where defined haplotypes (by the software or by the user) were tested for genetic association with scleroderma or the sub-phenotypes. Genotype data from the two cohorts were used to build the LD plots separately, and an association analysis was performed between the SNPs and the phenotypes found in the populations.

2.5.4 Hardy-Weinberg equilibrium

The law of Hardy–Weinberg states that a genetic marker with two alleles and allele frequencies p and $q=1-p$, is in equilibrium if and only if the proportion of subjects with genotypes AA, Aa, and aa will be $\pi_0 = p^2$, $\pi_1 = 2pq$, and $\pi_2 = q^2$. Departure from Hardy-Weinberg equilibrium (HWE) can be caused by factors such as inbreeding caused by consanguinity, assortative mating, non-random mating, selection, or migration (391). However, the first standard source for deviation from HWE therefore is genotyping error. Secondly, if the entire population is in perfect HWE, the presence of a genetic association, i.e., a difference in genotype frequencies between cases and controls implies that neither cases nor controls can be in HWE (392). Because the proportion of affected subjects in a population is small, the degree of deviation from HWE is expected to be stronger in cases than in controls.

Therefore, as an indicator of genotyping quality, compatibility with HWE should be investigated in control groups only (393). However, it could be argued that this approach can create bias and lead to false discoveries due to genotyping error. Thus, careful design and appropriate handing is required depending on the needs of each individual study.

In this study, the genotyping scale was small with only few SNPs genotyped in each cohort, and the data were handled with care, so that potential genotyping errors could be identified and excluded from the study. Thus, HWE was assessed in control groups only in both discovery and replication cohorts. All 6 tagging SNPs complied with the Hardy-Weinberg equilibrium [plink --hardy] (Plink cut-off p-value for HWE: 0.001).

2.5.5 P-value, statistical analysis and correction for multiple testing

Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data (e.g., the sample mean difference between two compared groups) would be equal to or more extreme than its observed value. In genetic association studies, p-value is used as a parameter of statistical significance to determine the certainty of an association. A p-value provides the probability that a given result from a test is due to chance. A common cut-off for each statistical test is 0.05, which claims a 95% certainty that the result is not a coincidence. When a p-value is lower than 0.05, the null hypothesis of no association is rejected, and the result can be interpreted as a significant association. The level of statistical significance is subject to the study design. In GWAS, a p-value $< 5x10^{-8}$ is considered as statistically significant to identify common causal variants.

However, genetic association studies usually test multiple genetic markers; therefore false positive data occurring simply due to chance will accumulate. In fact, a common problem in genetic studies is the poor replication of genetic associations. This is partly due to misuse or misinterpretation of the p-values, and inadequate knowledge to perform the correct statistical analysis. Recently, the American Statistical Association published a very interesting article addressing the scientific community, where it clarified several widely agreed principles underlying the proper use and interpretation of the p-value, the importance of correct statistical measurements, and principles on how to use them wisely (394). Among others, it is mentioned that scientific conclusions and decisions should not be based only on whether a p-value passes a specific threshold.

In order to account for false positive findings, the statistical p-values need to be corrected for multiple testing. Various statistical methods can be used, with the most popular being the Bonferroni correction, false discovery rate (FDR) correction, and permutation analysis (395, 396). Permutation analysis is considered the gold standard for multiple correction adjustment (397) and was used throughout this

study. The permutation procedure is a robust but computationally intensive. To calculate permutation-based p-values the case–control (or phenotype) labels are randomly shuffled (which assures that the null hypothesis holds, as there can be no relationship between phenotype and genotype), and all tests are recalculated on the reshuffled data set, with the smallest p-value of these tests being recorded (398). The procedure is repeated many times to construct an empirical frequency distribution of the smallest p-values. For the statistical analysis of the genetic association studies 1000 permutation tests were performed in Plink [plink --mperm 1000].

2.5.6 Odds ratio (OR)

The OR is used as a measurement of association between an exposure, which in this case is the given genotype or the allele, and an outcome which is often a disease status. In case-control studies, the OR represents the odds that the disease will occur in the presence of the associated allele or genotype. The 95% CI is used to estimate how precise the OR is, with narrower CI indicating a highly precise OR. If OR=1 or 95%CI includes the null value (=1) means that the genotype or allele under question does not affect the disease.

2.6 Cloning

2.6.1 Cloning of the minimal promoter and the 3' UTR of NKX2-5

A 578bp fragment (chr5: 172,662,217-172,662,794) upstream of the *NKX2-5* transcription start site encompassing the minimal promoter of the gene was amplified from a DNA sample of a healthy donor and cloned into the pGL4.10 reporter vector (Promega) upstream of the firefly luciferase gene (Figure 2.3A). pGL4.10 is a basic vector that does not contain any promoter and is designed for high expression. The vector is genetically engineered so it contains fewer consensus regulatory sequences and a synthetic gene, which has been codon optimised for mammalian expression.

The 3' UTR of *NKX2-5*, encompassing a 473bp long genomic region (chr5:173,232,104-173,232,576), was cloned into the pmirGLO vector (Promega, Cat. no: E1330) after the firefly luciferase gene (luc2) (Figure 2.3B). The A allele of the SNP was introduced by site-directed mutagenesis (SDM). The pmirGLO vector is designed to quantitatively evaluate microRNA activity by the insertion of

microRNA target sites downstream or 3´ of the firefly luciferase gene. The vector contains both the firefly and the renilla luciferase genes. Firefly is the primary reporter gene, whereas renilla acts as an internal control reporter for normalization and selection. The 3'UTR-microRNA target sequence is cloned downstream of the firefly gene, and the constructs are transfected into the cells together with microRNA mimic or inhibitor molecules. Reduced firefly luciferase expression indicates the binding of endogenous or introduced microRNAs to the cloned microRNA target sequence.

Both constructs were made by former research stuff prior to my arrival in the lab.

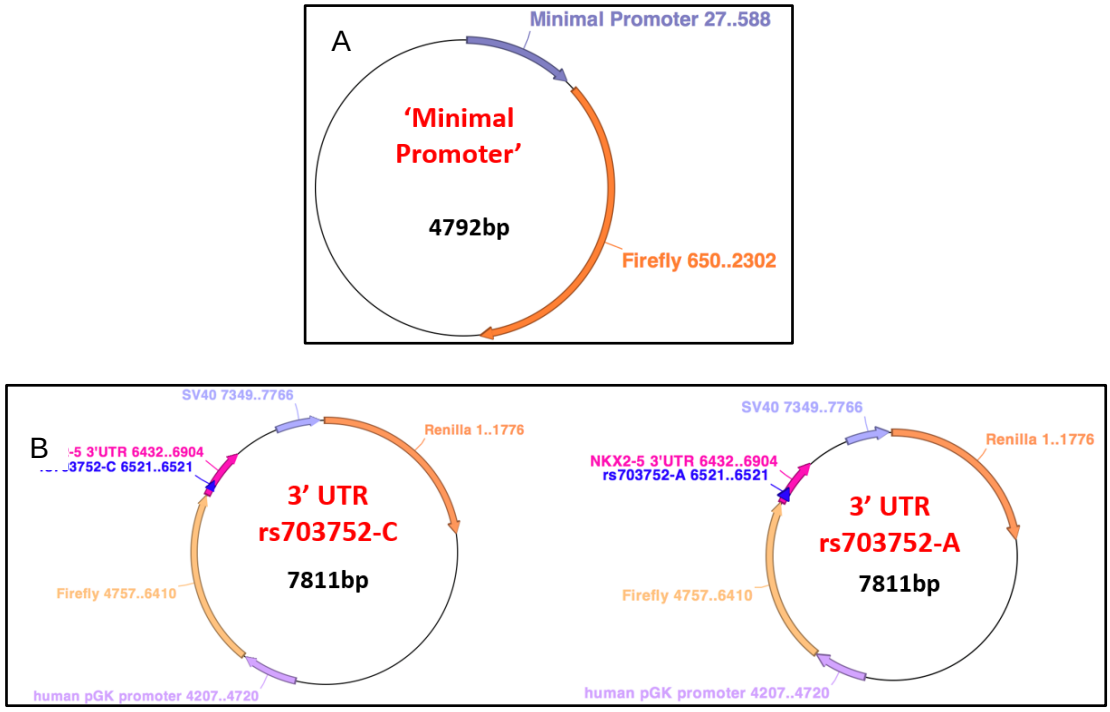


Figure 2.3 Minimal promoter and 3'UTR constructs. *A schematic representation of the cloning of the minimal promoter into the pGL4.10 reporter vector (A) and the 3'UTR (B) in the pmirGLO reporter vector. For the 3'UTR two constructs were made based on the different rs703752 alleles.*

2.6.2 Cloning of the NKX2-5 upstream promoter

The genomic region containing the rs3095870 was cloned into the pGL4.10 reporter vector. The SNP is located 1.3Kb upstream of the *NKX2-5* transcription start site, in a region that proved to be challenging to clone due to an extensive GC-rich and repetitive DNA sequence around the locus. Many attempts at cloning were performed, each time changing variables such as increasing and decreasing the length of the cloning fragment, trying different reporter vectors, numerous sets of cloning primers and restriction enzyme digestions and ligations. Finally, successful

cloning of a 1.6Kb (chr5:172,662,792-172,664,444) long genomic region upstream of the minimal promoter into the pGL4.10 vector was achieved (Figure 2.4).

The successful final steps included the insertion of restriction enzyme sites (SacI and EcoRV) into the cloning fragment by PCR amplification that created hanging cloning sites at the two ends of the cloning fragment. The same enzymes were also used to set up restriction enzyme digestions to linearise the pGL4.10-minimal promoter vector. Then, the linearised vector and the cloning fragment were ligated together. To amplify the cloning fragment DNA from a patient homozygous for the rs3095870-C allele was used. For the PCR amplification the Phusion High-Fidelity DNA polymerase (NEB, Cat. no: M0530) was used, and SURE2 competent cells were used for the transformation reactions (Stratagene, Cat. no: 200152).The T allele was introduced by SDM (Quick Change II, Agilent, Cat. no: 200523).

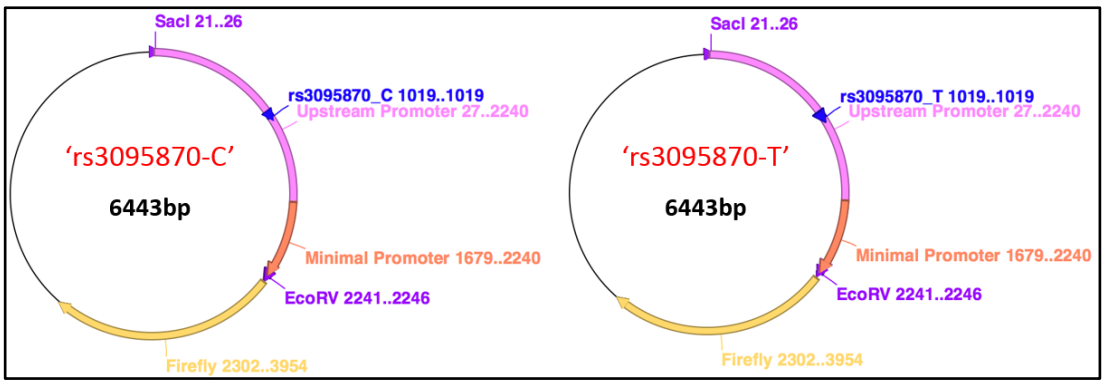


Figure 2.4 Upstream promoter constructs. *A schematic representation of the cloning of the upstream promoter including the rs3095870 SNP into the pGL4.10 reporter vector. Two different constructs were made based on the different alleles of rs3095870. In both cases the cloning fragments were inserted prior to the minimal promoter and the luciferase gene.*

2.6.3 Cloning of the NKX2-5 downstream enhancer

For the analysis of the putative enhancer, a genomic locus of 1.6Kb (chr5: 173,228,601-173,230,244) downstream of *NKX2-5* that spanned two associated polymorphisms (rs3132139, rs3131917) was amplified in a similar way and cloned into the pGL4.10 reporter vector upstream of the minimal promoter (Figure 2.5). In detail, a 2.5Kb long piece of DNA that contained the region of interest was amplified using the Phusion High-Fidelity DNA polymerase. A second nested PCR was then performed to amplify the target region using the In-Fusion HD Cloning Plus CE (Clontech, Cat. no: 638916). The pGL4.10 vector was linearised using the Phusion High-Fidelity DNA polymerase. The linearised vector and the purified product were

cloned together using the In-Fusion HD Cloning Plus CE, and the DNA was used to transform Stellar competent cells (Clontech, Cat. no: 636763). Primers and cycling conditions are given in the Table 2.7. DNA was isolated from bacteria colonies, and sent for sequencing at Source Bioscience to verify its integrity. Then, SDM was performed to change the SNP alleles, creating two identical constructs with different sets of alleles for the desired SNPs (Table 2.6). The selection of the alleles was performed based on the haplotype analysis.

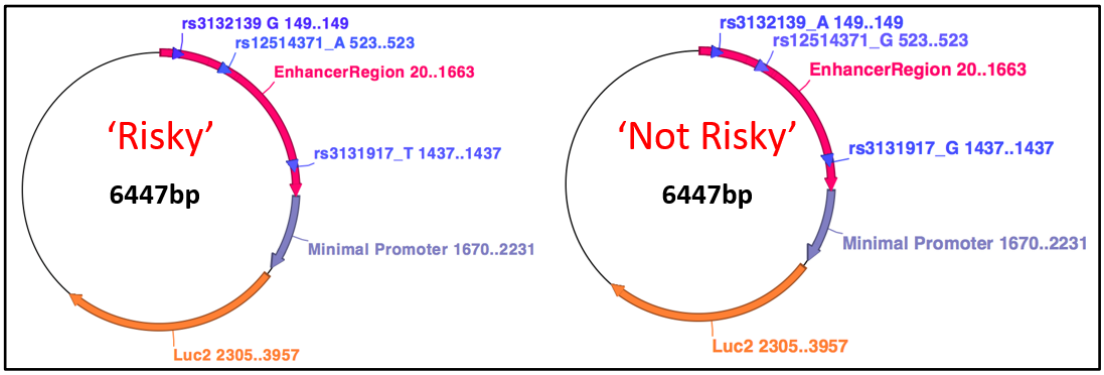


Figure 2.5 Enhancer constructs. *A schematic representation of the cloning of the downstream enhancer into the pGL4.10 reporter vector. Two different constructs were made with each having a different set of alleles. The cloning fragments were inserted prior to the minimal promoter and the luciferase gene. Construct called "risky" contains the disease-associated alleles, as will be further discussed. The "not risky" construct contains the alternative alleles of the SNPs.*

2.6.4 Useful programmes and tools

Sequencher 5.1 software was used to read and verify the integrity of the sequences. ApE.exe is a useful plasmid editor that I used during the cloning experiments.

'Risky' Enhancer Construct	'Not-Risky' Enhancer Construct
rs3132139-G	rs3132139-A
rs12514371-A	rs12514371-G
rs3131917-T	rs3131917-G

Table 2.6 SNP genotypes at the enhancer constructs.

PCR	Primers	Cycling conditions
Region amplification	Forward: 5'- CCAACCTGCCAAATGATGAGAAT- 3' Reverse: 5'- ATCCGAAATGACCCGTATTTGCT- 3'	1x 95^{o}C 5' 30x 96^{o}C 5'', 53^{o}C 5'', 68^{o}C 1' 15'' 1x 72^{o}C 1' 4^{o}C for ever
Nested PCR	Forward: 5'- TAACTGGCCGGTTACCTGGGCAACGTA- 3' Reverse: 5' –GCTCTGGGTGAGCTCTGGGCCTTTTTC- 3'	1x 98^{o}C 30'' 35x 98^{o}C 10'', 55^{o}C 15'', 72^{o}C 9'' 1x 72^{o}C 5' 4^{o}C for ever
pGL4.10 vector linearisation	Forward: 5'- GAGCTCACCCAGAGCC- 3' Reverse: 5'- GGTACCGGCCAGTTAG- 3'	1x 98^{o}C 1' 35x 98^{o}C 5'', 55^{o}C 10'', 72^{o}C 2' 25'' 1x 72^{o}C 10' 4^{o}C for ever

Table 2.7 Cloning of the enhancer. *The table describes the three PCR reactions performed to clone the enhancer region into the pGL4.10 reporter vector. The primers and the PCR cycling conditions are given.*

2.7 Cell culture

A variety of different cell lines and culture conditions were used for the experiments described in this thesis.

2.7.1 Primary human pulmonary artery smooth muscle cells (HPASMC)

HPASMC were purchased from Promocell (Cat. no: C-12521). The company provided information regarding the donor's age, sex and ethnicity, as well as viability and growth characteristics such as the population doubling time. In addition, phenotypic characterisation of the cells was conducted by immunohistochemistry within the 2 passages, and expression of α-SMA was confirmed. The cells were negative for CD90 (specific to fibroblasts) and vWF (von Willebrand factor, specific to endothelial cells). The primary cells were cultured in Smooth Muscle Cell Growth Medium (Promocell, Cat. no: C-22062) at 37^{o}C and 5%CO_2 under different serum conditions. To mimic the de-differentiation of VSMC *in vitro*, contractile cells were cultured in medium containing 5% foetal calf serum (FCS) (LabTech, cat. no: SA/500) at medium to high confluence (70-80%), while synthetic cells were cultured in 10% FCS and were passaged after reaching 70-80% confluence. Cells at passages 3-9 were used for the experiments unless otherwise stated.

2.7.2 Immortalised HPASMC (ImHPASMCs)

It is known that primary cells can only undergo a limited number of cell divisions in culture, before they reach a state where they can no longer divide, known as replicative senescence (399). In order to have a consistent supply of material throughout a research project, primary cells with an extended replicative capacity are required. For this reason, the technology of immortalising mammalian cells has been well-established for years. Immortalised cells retain a similar or identical genotype and phenotype to the primary cells of origin. They also acquire the ability of continuous cell-division.

Several methods exist for immortalising mammalian cells in culture conditions. One method is to use viral genes, such as the simian virus 40 (SV40) T antigen (400). SV40 T antigen has been shown to be the simplest and most reliable agent for the immortalisation of many different cell types and the mechanism of SV40 T antigen in cell immortalisation is relatively well understood (401).

For the purposes of this study, immortalised human pulmonary artery smooth muscle cells (ImHPASMCs) were purchased from ABM Good (Cat. no: T0558). The cells were immortalised using the SV40. ImHPASMCs were cultured in DMEM (Life Technologies, Cat. no: 31966-047)) supplemented with 5% FCS (unless otherwise stated) on collagen-coated vessels. For the coating, rat tail collagen type I was diluted in 0.1N acetic acid to a final concentration of 1mg/ml under sterile conditions. The collagen was added into the flasks or plates and left to polymerise in the incubator for 1-3 hours, followed by a 30' drying incubation under sterile conditions in a Class II hood prior to use.

2.7.3 Primary endothelial cells

For the EndoMT experiments, primary human HPAEC were used. The cells were purchased from Promocell (Cat. no: C-12241) and cultured in endothelial cell growth media (Promocell, Cat. no: C-22010) that contained 2% FCS. HPAECs were cultured in vessels coated with gelatin as an attachment factor that enhances the growth of vascular endothelial cells (Invitrogen, Cat. no: S-006-100).

For preliminary experiments, human umbilical vein endothelial cells (HUVECs) were cultured in the same way as HPAEC.

2.7.4 Isolation of VSMC from diseased vessels

Another source of primary cells was tissue from lung transplants of PAH patients and from healthy individuals, as well as parts of vessels from diabetic and PAD patients undergoing vascular surgery. Written informed consent was given prior to transplantation or surgery by the patients or their next of kin relatives. The following protocol was adapted from a protocol developed in Prof Lucie Clapp's lab, Institute of Cardiovascular Sciences, and I was taught by Dr Rijan Gurung. The tissue was kept in physiological salt solution (112mM NaCl, 5mM KCl, 1.8mM $CaCl_2$, 1mM $MgCl_2$, 25mM $NaHCO_3$, 0.5 KH_2PO_3, 0.5mM $NaHPO_3$, 10mM Glucose) while the arteries were dissected out. The outer surface of the vessel and the connective tissue were removed. The arteries were decontaminated in 3% penicillin/streptomycin and 1.5µg/ml gentamicin in PBS for 45 minutes at 4^{o}C. Then, the arteries were minced using scalpels and enzymatic dissociation at 37^{o}C for 30 minutes. To prepare the dissociation cocktail 5mg elastase (Sigma, Cat. no: E7885), 10mg collagenase (Sigma, Cat. no: C9722), 2.5mg trypsin inhibitor (Sigma, Cat. no: T6522), 150mg BSA (Sigma, Cat. no: A9647), and 100µl MEM vitamins (sigma, Cat. no: M6895) were mixed in 10ml DMEM/F12 media (Thermo Fisher Scientific, Cat. no: 11320074).The dissociated fragments were forced into a sieve (cell strainer,

40µm, Nylon BD Falcon, Cat. no: 352340) and then enzyme activity was neutralised with DMEM/F12 supplemented with pen/strep and FCS. The suspension was spun down at 100xg for 5 minutes at room temperature. The pelleted VSMC were resuspended in HPASMC specific media supplemented with FCS and plated into flasks.

In addition to VSMC, fibroblasts were also efficiently isolated. The vessels and tissues were washed in DMEM supplemented with pen/strep to remove the blood. The tissue was then fragmented into small parts using scalpels. The parts were plated sparsely into flasks and left to dry before media was added and placed in the incubators.

Cell Type	ID	Gender	Age	Vessel	Population Doubling (h/PD)	Viability (%)
Promocell (PASMC)	**HPASMC1**	Male	23	Pulmonary artery	35	79
	HPASMC1	Male	27	Pulmonary artery	36.7	84
	HPASMC1	Female	23	Pulmonary artery	21.1	73
	HPASMC1	Male	15	Pulmonary artery	38.6	73
	HPASMC1	Male	30	Pulmonary artery	49.2	81
	HPASMC1	Male	77	Pulmonary artery	38	72
Donors (PASMC and VSMC)	**N1**	NA	NA	Pulmonary artery	~	~
	N2	NA	NA	Pulmonary artery	~	~
	N2	NA	NA	Pulmonary artery	~	~
	PAH1	NA	5	Pulmonary artery	~	~
	PAH2	NA	5	Pulmonary artery	~	~
	PAH3	NA	14	Pulmonary artery	~	~
	PAH4	NA	NA	Pulmonary artery	~	~
	PAH5	NA	4	Pulmonary artery	~	~
	PAD1	NA	NA	Peripheral Artery	~	~
	PAD2	NA	NA	Peripheral Artery	~	~
	DIAB1	NA	NA	Peripheral Artery	~	~

Table 2.8 Characteristics of the cells used in the study. *HPASMC: primary human pulmonary artery smooth muscle cells commercially bought from Promocell, N: human pulmonary artery cells smooth muscle cells isolated from pulmonary artery of lung tissue from post-mortem samples, PAH: human pulmonary artery cells smooth muscle cells isolated from pulmonary artery of lung tissue from PAH patients undergoing lung transplant, PAD: vascular smooth muscle cells isolated from peripheral arteries of patients with peripheral arterial disease, DIAB: vascular smooth muscle cells isolated from peripheral arteries of patients with diabetes, NA: not available information.*

2.7.5 Immunofluorescence to characterise isolated VSMC

To phenotype the isolated primary cells, immunofluorescence was used to confirm expression of protein markers specific to VSMC and fibroblasts. α-SMA and CD90/Thy1 were considered as VSMC- and fibroblast- specific markers respectively, aligned to the commercially used protocols. VSMCs isolated from vessels or tissues were plated in 8-well chamber slides (20.000 cells/ well) (Falcon, Cat. no: 354118). Next day, the cells were fixed in 2% PFA in PBS for 15', treated with PBS/ 0.2% TritonX-100 for 10 minutes and blocked in normal serum for 10 minutes. Primary antibody for CD90/Thy1 (Abcam, Cat. no: ab133350) was added to the cells overnight at 4^{o}C. Primary antibody for α-SMA-Cy3 conjugated (Sigma, Cat. no: C6198) was incubated for one hour in the dark. Secondary antibodies conjugated to Alexa Fluor dyes (Life Technologies) were incubated for one hour in the dark at room temperature. Mounting medium containing 4,6-diamidino-2-phenylindole (DAPI) (Vector laboratories, Cat. no: H-1200) was used to mount the slides. Cells were viewed and photographed on an Axioscope Z fluorescence microscope with an Axiocam digital camera in combination with Axiovision software (Carl Zeiss). Three different fields were photographed for each condition, and the number of the cells (nuclei and expressing cells) was assessed and counted by eye.

2.8 Cell treatments, stimulations, agonists, antagonists

Gene and protein expression are dependent on the state of the cell. Therefore, experimental results might vary and be inconsistent if the cells studied are in different stages of cell cycle. Various physiological processes within the cell including proliferation, survival, and apoptosis are controlled by the presence of serum and growth factors. Specifically, the effect of serum has been well studied in the survival and the mechanical properties of VSMCs (402). In addition, VSMC differentiation is directly affected by serum through the SRF-dependent transcription of contractile and synthetic genes (403). To control these limitations, a serum starvation/deprivation procedure is used to synchronise the cells. This classic method of cell synchronisation causes a cell cycle arrest between G0 and G1 phases, and it is proposed to initiate quiescence induced by growth restriction (404).

Serum starvation conditions. Depending on the cell type and the experimental procedure, the induction of quiescence requires different periods of time. Thus, careful optimisation of both the amount of serum withdrawn and the length of starvation is necessary. In the literature, studies indicate that human and animal VSMCs are serum-starved for varying periods of time ranging from 24 to 72 hours in

the complete absence of serum (405, 406). In my studies, complete absence of serum depleted NKX2-5 expression indicating a direct role of serum on *NKX2-5* transcription explained by the SRF-dependent regulation, and also affected the viability of the cells when the starvation was extended for longer periods of time. For the purposes of this study, cells were always serum-starved in 0.1% FCS in DMEM. The period of starvation varied depending on the experiment and is carefully reported in the figure legends. For the experiments with the selective inhibitors reported in Chapter 4, a serum starvation of ~6-7 hours was conducted before TGF-β stimulation. For the rest of the experiments that required serum-starvation, a longer overnight period of ~16-20 hours of serum-deprivation was followed. The main reason based on which these conditions were selected was the overall time of treatment, ranging from 24-48 hours. Cells used for binding experiments (pull down assays and EMSAs) and chromatin immunoprecipitation assays were not serum starved before TGF-β stimulation.

TGF- β stimulation conditions. A recombinant human TGF-β1 protein (R&D, Cat. no: 240-B-010) was used to stimulate the cells at a concentration of 2ng/ml, unless otherwise stated. TGF-β stimulation was almost exclusively performed for 24 hours, unless otherwise stated, and is reported in the figure legend of each individual experiment.

Cells used to perform chromatin immunoprecipitation assays were stimulated with TGF-β for 16 hours, without being subjected to serum-starvation. This condition was selected in order to ensure that most cells were transcriptionally active with the transcription factors bound on the DNA at the time of cross-linking.

Other stimulation conditions. Apart from TGF-β1, other cytokines, growth factors and proteins were also used to stimulate cells and study their effect on NKX2-5 protein and gene expression: BMP2, BMP4, ET-1, FGF2, TNF-α and IL-1β. The final concentration of each individual factor varied between experiments, and is therefore given in each individual experiment. The cells were treated with the above factors for 24 hours.

Selective inhibitors. To examine the molecular and signalling pathways upstream of *NKX2-5* that lead to its activation, a panel of agonists and antagonists were used as shown in Table 2.9. For the experiments with inhibitors, ImHPASMCs were grown on collagen-coated flasks in DMEM supplemented with 5% FCS. The cells were plated in 6-well plates and serum-starved in 0.1%FCS/DMEM for ~6-7 hours after reaching 60%-70% confluence. Then, the media was changed to 5%FCS/DMEM

and the cells were treated with 2ng/ml TGF-β for 24 hours. After TGF-β treatment, the media was changed again to 5%FCS/DMEM and the cells were treated with the inhibitors in the concentrations shown below for 24 hours before harvesting for protein and RNA.

Molecule Name	Pathway/ Kinase	Company, Cat. Number	Final [c]	Ref
FR180204	ERK-1,2	Tocris	30μM	Optimised*
ERK5-in-1	ERK5	Selleckhem	10μM	Optimised*
BI78D3	JNK	Tocris	1μM	(407)
GSK2126458	PI3K/mTORC1/2	Selleckhem	1μM	Selleckhem
GSK690693	pan-AKT	Selleckhem	1μM	Optimised*
SD208	TGF-βR1/ALK5	Sigma Aldrich	5μM	Optimised*
TC-ASK10	ASK1	Tocris- 4825	10μM	Optimised*
SB202190	p38	Sigma Aldrich	10μM	(408)
CX-4945	CK2	Selleckhem	5μM	Optimised*
OXO	TAK1	Cayman Chemical	100nM	Optimised*
BAY	Hypoxia/HI1α	Stratech-87-2243	100nM	(409)
Bosentan	ET-1 pathway	BQ788	1μM	Optimised*
Imatinib	PDGF	Cayman Chemical	2μM	Optimised*

Table 2.9 Pathway agonists and antagonists. *Final [c]: Final concentration used in the experiment. Ref: the source of information based on which the final concentration of each inhibitor was selected.* Optimised doses were empirically determined.*

In EndoMT experiments, the endothelial cells were treated with a “cocktail” of cytokines that consisted of TGF-β (5ng/ml), TNF-α (5ng/ml), and IL-1β (0.1ng/ml) (14).

2.9 Cell transfections

2.9.1 General principle

Transfection is the process of introducing nucleic acids (plasmid DNA, mRNA, siRNA, shRNA, and microRNA) into eukaryotic cells. There are various methods divided into two categories: viral (also known as infection) and non-viral. Non-viral transfections can either be transient, where the foreign genetic material is expressed but not integrated into the host genome, or stable, where the foreign genetic material is integrated into the host genome. The non-viral methods are based on chemical compounds that are used as carriers of DNA, and include

liposomes, non-liposome lipids, and other macromolecules. There are also mechanical methods such as electroporation, where an electric current is used to create pores in the cell membrane to allow genetic material to insert into the cell. Transient transfections are used to investigate a short-term impact of the inserted material on gene and protein expression.

Transient transfection methods were exclusively used in this thesis, with a selection of different transfection reagents: FuGENE-HD for plasmid DNA, Lipofectamine 2000 for microRNAs and plasmid DNA, and Oligofectamine for siRNAs. All three transfection reagents use especially designed cationic lipids. The basic structure of cationic lipids consists of a positively charged head group and one or two hydrocarbon chains. The charged head group governs the interaction between the lipid and the phosphate backbone of the nucleic acid, and facilitates nucleic acid condensation. The positive surface charge of the liposomes mediates the interaction of the nucleic acid and the cell membrane, allowing for fusion of the liposome/nucleic acid transfection complex with the negatively charged cell membrane. The transfection complex is thought to enter the cell through endocytosis.

2.9.2 Optimisation of transfection

Transfection experiments usually involve a balance between efficient delivery of the nucleic acid and cellular toxicity. The extent of cellular toxicity caused by transfection is influenced by the reagent and the nature of the cells. In particular, cell types including primary cells, toxicity can be directly visualised under the microscope a few hours after transfections with large numbers of floating cells. However, some cell types and usually transformed cell lines are more resilient. Some of the key-points that affect the cell viability after transfection are the transfection reagent itself, the concentration of the nucleic acid, the confluency of the cells at the time of transfection and the time-course of the experiment. With all these in mind and taking into consideration that primary HPASMC are difficult to transfect, I performed extensive optimisation experiments to identify the least cytotoxic conditions.

Optimisation involved two transfection reagents and a GFP plasmid so that when efficiently expressed after transfection, the transfected cells fluoresced green. Transfection efficiency was assessed by microscopy to detect the GFP fluorescence emitted from transfected cells, and, the cell viability was assessed by cell viability assays. FuGENE-HD gave higher transfection efficiency than Lipofectamine 2000 in

the primary cells. The amount of DNA transfected into the cells did not affect cell viability, however, a 3:1 transfection reagent: DNA ratio gave the highest transfection efficiency. The best conditions selected for the primary HPASMC were also used for ImHPASMCs without further optimisation. In general, ImHPASMCs were easier to transfect and more resilient to transfection. However, both cell types should be plated at higher confluency compared to other cell lines, especially for longer periods of incubation. That is thought to be due to VSMC requirement of cell-cell contact to maintain normal growth and proliferation.

2.9.3 Cell viability assays

Cell viability was measured using the CellTiter 96 AQueous One Solution Cell Proliferation Assay (Promega, Cat. no: G3582), a colorimetric method that determines the number of viable cells. The reagent contains a novel tetrazolium compound, MTS. The MTS is bio-reduced by cells into a coloured formazan product that is soluble in tissue culture medium. This conversion is presumably accomplished by NADPH or NADH produced by dehydrogenase enzymes in metabolically active cells. Assays were performed by adding a small amount of the reagent directly to culture wells, and after 1 hour incubation the absorbance at 490nm was read in a Mithras LB 940 Plate Reader. Higher absorbance indicates more in number viable cells.

2.10 Luciferase reporter assays

Primary HPASMC were seeded in 24-well plates at 7-8x10^4 cells/well and transfected with constructs as described below. Cells were cultured for 48 hours, lysed and analysed for luciferase expression. Luciferase activity was measured using the Dual-Glo Luciferase Assay System (Promega, Cat. no: E2920) in a Mithras LB 940 Plate Reader. Each experiment was repeated at least 3 times. The graphs represent the ratio of firefly/renilla luciferase expression.

2.10.1 Transfection of reporter vectors

For the analysis of the luciferase activity of the promoter and the enhancer, the DNA vectors were complexed with FuGENE-HD transfection reagent (Promega, Cat. no: E2311) in Opti-Mem I reduced media serum (Gibco, Cat. no: 31985-062) at a ratio of 3:1 (transfection reagent: DNA). A maximum of 825ng DNA was added per well; 700ng of the reporter vector (minimal promoter alone, rs3095870-C, rs3095870-T, enhancer-risky, enhancer-not risky), 100ng of the TEAD1 or TEAD3 expression vectors, and 25ng of the renilla luciferase vector driven by the CMV promoter. As a

negative control, the empty pGL4.10 vector was used. All the DNA vectors were complexed at the same time. To control for a potential effect of the DNA load on cell viability, pGL4.10 vector was co-transfected whenever the required total DNA was less than 825ng (eg. in the cells with TEAD1 or TEAD3 were not co-transfected).

2.10.2 Transfection of microRNA mimic and inhibitor molecules

For the analysis of the 3'UTR, the rs703752-C and rs703752-A constructs were co-transfected together with microRNA mimic and inhibitor molecules. For this set of experiments, Lipofectamine 2000 (Thermo Fisher Scientific. Cat. no: 11668027) was used and the complexes were prepared in Optimem I. The microRNA mimics were used at a final concentration of 10nM and the inhibitor molecules at 50nM.

2.11 RNA silencing

2.11.1 General principle and experimental design

RNA interference represents a natural mechanism to protect the genome. In recent years the field has evolved at a surprisingly high pace. The underlying molecular mechanism of gene silencing provides us with short interfering RNA (siRNA) molecules that allow the targeting of any gene with high specificity and efficiency. siRNAs can now be obtained in various ways allowing for numerous *in vitro* and *in vivo* applications. Successful knock-downs of disease-related genes indicate that siRNA technology is also a promising novel therapeutic mechanism.

Using siRNA specific to TEAD1, TEAD3, and YAP1, I aimed to investigate the effect of these factors on *NKX2-5* transcription. ON-TARGETplus SMARTpool siRNA molecules were purchased from Dharmacon (TEAD1: Cat. no: L-012603, TEAD3: Cat. no: L-012604, YAP1: Cat. no: L-012200). As a negative control, a non-targeting siRNA pool (Cat. no: D-001210-10) that consisted of 4 different oligonucleotides was used. However, the negative control was subsequently found to have an effect on NKX2-5 gene and protein levels, and further investigation revealed that 1 oligo was off-targeting the *NKX2-5* gene. To avoid this side-effect, a new individual non-targeting siRNA was ordered (Cat. no: D-001210-01) with the following sequence 'UGGUUUACAUGUCGACUAA' that did not affect NKX2-5 levels under the conditions used in the ImHPASMCs. SiRNAs were transfected into the cells using Oligofectamine transfection reagent (Thermo Fisher Scientific. Cat. no: 12252011) in Opti-Mem I.

ImHPASMCs were plated in 12-well plates at 60% confluence and serum-starved overnight (~16-20 hours). The following day, the cells were transfected with the siRNA in serum free medium. Four hours after transfection, serum-supplemented medium was added to the cells to give a final serum concentration of 10%. A series of optimisation steps identified the best conditions for the experiments for each individual siRNA. The optimisation tested a range of concentrations (50-200nM final) and 2 different time-points (48, and 72 hours). The optimal conditions are described below.

2.11.2 siRNA for TEAD1 and TEAD3

siRNAs for TEAD1 and TEAD3 were used at 100nM final concentration, and the cells were lysed after 72 hours of transfection.

2.11.3 siRNA for YAP1

siRNA for YAP1 was used at 130nM final concentration, and the cells were lysed after 48 hours of transfection.

2.12 Protein-DNA binding assays

Proteins interact with nucleic acids (DNA and RNA) in several processes essential to normal cell function. As with protein-protein interactions, disruption of protein-nucleic acid interactions can have profound consequences for the cells. In this study, I have only focused on protein-DNA interactions.

Protein-DNA interactions are integrated into several key cellular processes, including transcription, translation, regulation of gene expression, replication and repair, etc. The common property of DNA-binding proteins is their ability to recognise and manipulate DNA structures. Chromatin remodelling, transcription complex formation, initiation of transcription and translation of mRNA to protein all involve formation of protein-DNA complexes. These complexes play a role in the regulation of protein expression. Depending on the nature of the complex, proteins bind to nucleic acids in either a sequence-specific or secondary structure-dependent manner, often inducing drastic structural changes in the nucleic acid. Defining sequence-specific interactions can aid the study of gene regulation.

2.12.1 Experimental design

Several methods for detecting and identifying protein-DNA interactions have been established and used in research. These methods provide specific information regarding the identity of the DNA-binding protein, the exact genomic locus of the binding, and whether the protein acts alone or in a complex with other factors. Three different assays were designed and performed for this work, and described in detail below.

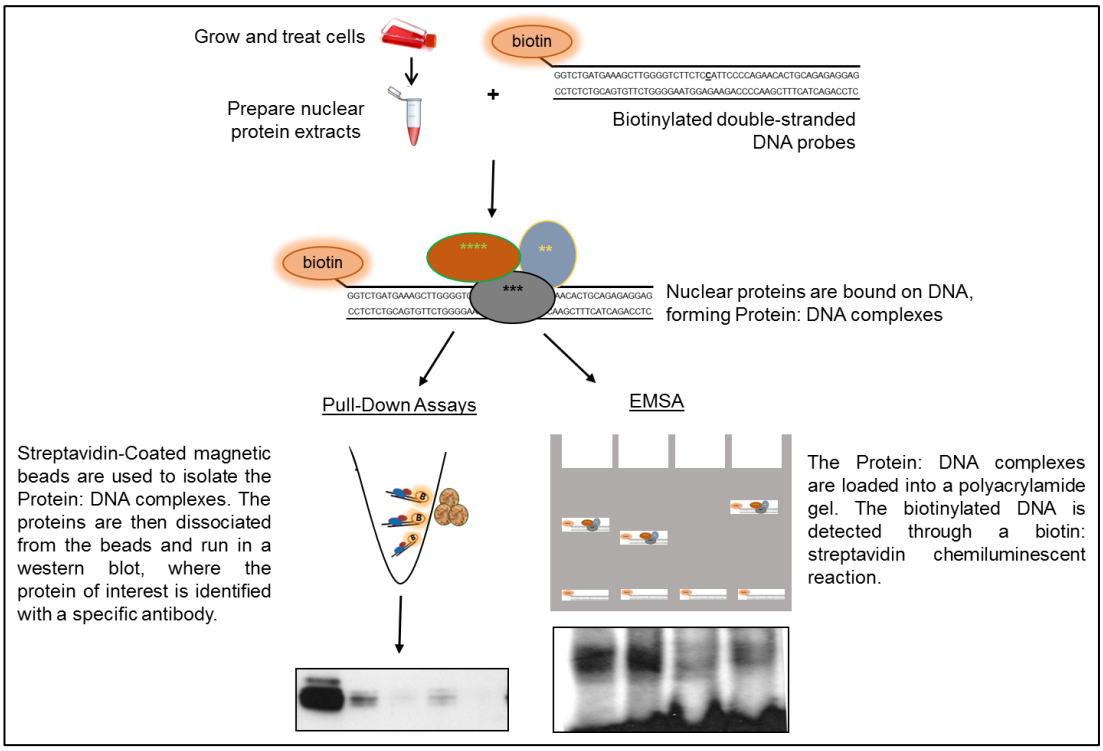


Figure 2.6 Basic design of the binding assays. *Nuclear protein extracts were prepared from ImHPASMCs and mixed with biotinylated double-stranded DNA probes specific to rs3095870-C or T alleles. The protein: DNA complexes were isolated and analysed either by pull-down assay or EMSA. In the pull-down assay, the protein: DNA complexes are isolated using streptavidin magnetic beads. Then, the protein is eluted from the beads and analysed by Western Blotting with the antibody of interest. In the EMSA, the protein: DNA complexes are run in a native acrylamide gel, where the biotinylated DNA is detected with HRP-conjugated streptavidin through a chemiluminescent reaction.*

For the protein-DNA binding assays, two 54bp-biotinylated double-stranded DNA probes were designed spanning the rs3095870 site (Table 2.10). Each probe was specific to either the C or the T allele of the rs3095870. The probes were tagged with biotin at the 5' end, and mixed with nuclear protein extracts prepared from ImHPASMCs after 16 hours of TGF-β stimulation. Complexes of DNA and protein were formed, and analysed by pull-down assays and EMSAs (Figure 2.6). The two different approaches address different aspects of the same question: the pull-down

assay identifies the proteins that bind the DNA and the EMSA the DNA bound by the proteins.

2.12.2 Electrophoretic mobility shift assays (EMSA)

EMSAs have been used extensively for studying protein-DNA interactions. The assay is based on the slower migration of protein-DNA complexes than unbound DNA, through a native polyacrylamide gel. The individual protein-DNA complexes form discreet bands within a native acrylamide gel that are detected based on a chemiluminescent reaction. In super-shift assays, antibodies are used to identify proteins involved in the protein-DNA complex. The formation of an antibody-protein-DNA complex further reduces the mobility of the complex within the gel resulting in a “super-shift”.

ImHPASMCs were used to prepare nuclear extracts (NE-PER extraction kit, Thermo Scientific, Cat. no: 78833). Nuclear extract (4µg) and 20fmol of each biotinylated DNA probe specific to the C or T allele of rs3095870 were added in a total volume of 20µl binding reaction (10mM Tris, 150mM KCl, 1mM DTT, 10% glycerol, 15mM MgCl2, 200ng dI-dC, 0.05% NP-40). For the super-shifts, 1-2µg of the antibodies were used per reaction (TEAD1, Abcam, Cat. no: ab133533; YAP1, Cell Signalling 4912). The reactions were incubated for 20 minutes at room temperature and then loaded into a 6% native acrylamide gel. The gel was run in 0.5x TBE at 120V for 3 hours at 4^{o}C. The protein: DNA complexes were transferred to a positive nylon membrane for 30 minutes at 15V using a Trans-blot semi-dry apparatus (Biorad). The membrane was UV-light cross-linked for 1 minute (UV Stratalinker 2400, Stratagene). Then, the membrane was probed with streptavidin-HRP conjugate, incubated with the substrate and developed following the LightShift™ Chemiluminescent EMSA protocol (Thermo Fisher Scientific, Cat. no: 20148).

2.12.3 Pull-down assays

Similar to the EMSAs, protein-DNA pull down assays were performed. In addition to the 2 probes described earlier, 2 more probes were designed for the pull-down assays and used as negative controls: a 54bp-biotinylated double stranded DNA probe with scrambled DNA sequence and a non-biotinylated probe specific to the C allele of rs3095870 (Table 2.10).

ImHPASMCs were treated with 2ng/ml TGF-β for 16 hours. The cells were washed twice in ice-cold PBS and scraped with cell scrapers in ice-cold Hank’s balanced

salt solution (Invitrogen, Cat. no: 14170-088). The cell suspension was spun down at 16000xg at 4^{o}C and the cell pellet was used to prepare nuclear extracts according to the manufacturer's instructions (NE-PER extraction kit, Thermo Scientific, Cat. no: 78833). 0.5nmol of each DNA probe was added in 1ml of binding buffer (10mM HEPES, 0.5mM EDTA, 0.5mM DTT, 10% glycerol, 200ug salmon sperm DNA) together with 100µl Streptavidin MagneSpher Paramagnetic Particles (Promega, Cat. no: Z5481). The beads had previously been washed 3 times in PBS and blocked in 200µg BSA in PBS for 1 hour at room temperature. The DNA and the beads were mixed by rotation at 4^{o}C for 1 hour. Then, ~60-70µg of nuclear extract supplemented with protease and phosphatase inhibitors were added to each binding reaction and rotated at 4^{o}C for 1 hour. The protein: DNA: beads complexes were washed 5 times for 5 minutes each in 150mM NaCl, 10mM HEPES, 0.1% NP-40. The proteins were eluted from the complexes in NuPAGE LDS sample buffer and reducing agent (Thermo Fisher Scientific, Cat. no: NP0008 and NP0009, respectively), and loaded onto precast polyacrylamide gels (Thermo Fisher Scientific, Cat. no: NP0335) for Western blotting.

Name of the probe	**Sequence of the double-stranded DNA probes**
Biotin-rs3095870-C	*BIOT*-GGTCTGATGAAAGCTTGGGGTCTTCTC**C**ATTCCCCAGAACACTGCAGAGAGGAG CCTCTCTGCAGTGTTCTGGGGAATGGAGAAGACCCCAAGCTTTCATCAGACCTC
Biotin-rs3095870-T	*BIOT*-GGTCTGATGAAAGCTTGGGGTCTTCTC**T**ATTCCCCAGAACACTGCAGAGAGGAG CTCCTCTCTGCAGTGTTCTGGGGAATAGAGAAGACCCCAAGCTTTCATCAGACC
Biotin-scramble	*BIOT*-ATTTCACAGTTGTACAGTTGCATGGGTAACAGCTACGAAATCCCGGATAGCTGG TAAAGTGTCAACATGTCAACGTCCCATTGTCCGATGCTTTAGGGCCTATCGACC
Non-Biotin-rs3095870-C	GGTCTGATGAAAGCTTGGGGTCTTCTC**C**ATTCCCCAGAACACTGCAGAGAGGAG CCTCTCTGCAGTGTTCTGGGGAATGGAGAAGACCCCAAGCTTTCATCAGACCTC

Table 2.10 Probes for the pull-down assays.

2.13 Chromatin immunoprecipitation assays (ChIP)

2.13.1 General principle

ChIP assays are a sensitive method that can be used in cultured cells or animal models to identify links between the genome and the proteome by monitoring transcriptional regulation through transcription factor binding on native chromatin of regulatory regions. Whereas EMSAs and pull-down assays look at binding of transcription factors to small linear oligonucleotide probes, ChIP assays capture protein binding on the native chromatin in the cell. The strength of ChIP assays is the ability to capture a snapshot of specific protein: DNA interactions in a system and assess these interactions by PCR. The ChIP assay procedure involves a variety of proteomics and molecular biology methods including crosslinking, cell lysis, DNA shearing, and immunoprecipitation (Figure 2.7).

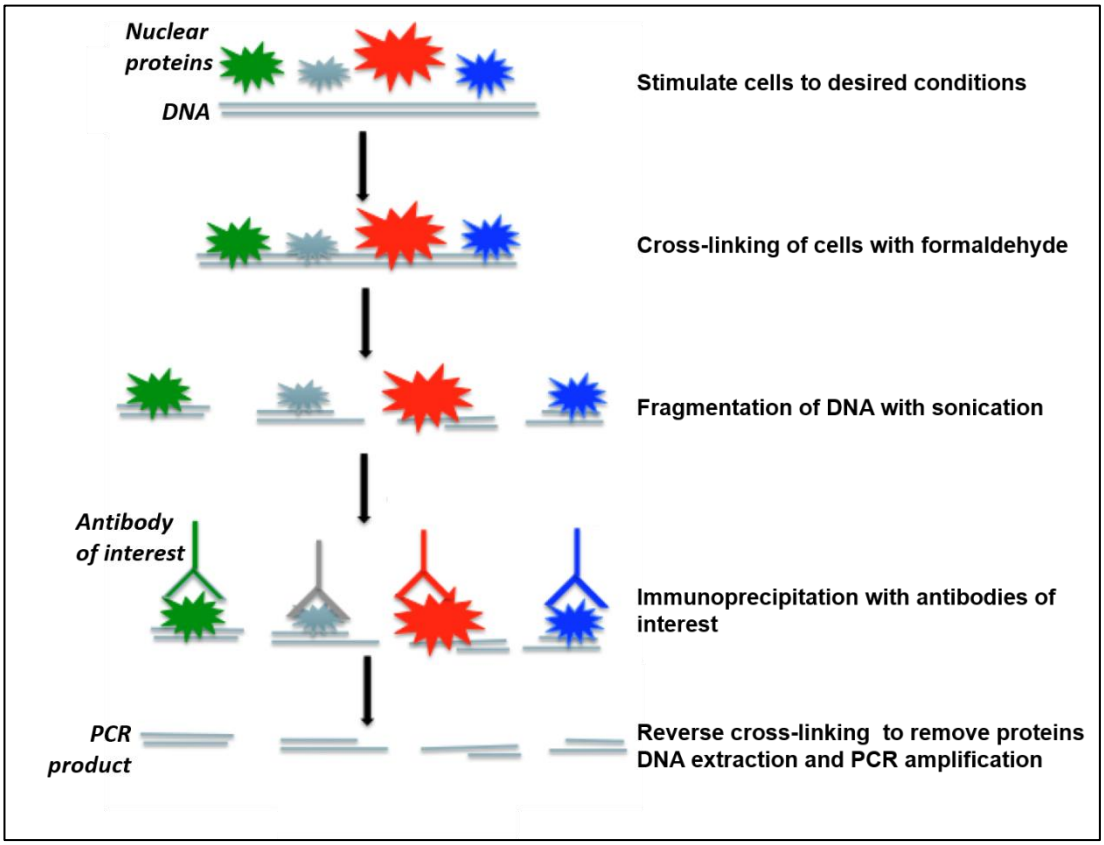


Figure 2.7 Chromatin Immunoprecipitation assay. *Schematic representation of each step of the ChIP protocol.*

2.13.2 Experimental protocol

ImHPASMCs were stimulated with 2ng/ml TGF-β for 16 hours. $6x10^5$-$1x10^6$ cells were used per ChIP assay. Treated cells were cross-linked for 10 minutes with 1% formaldehyde at 37^{o}C. The cells were washed twice with ice-cold PBS supplemented with protease inhibitors and then lysed directly into SDS lysis buffer (1% SDS, 10mM EDTA, 50mM Tris pH 8.1, protease inhibitors). The cell suspension was transferred into 2ml tubes and kept on ice for 15 minutes. The chromatin was sheared to a mean size of 500bp-1Kb by sonication. Sonication settings were empirically determined for each cell line to ensure a range of fragments around 500bp. Cross-linked chromatin was sheared for 15'' at power to generate amplitude of 8 microns. The sonication was repeated 3 times at 30 seconds intervals. The samples were kept on ice during sonication to avoid over-heating and protein degradation. To ensure effective shearing, DNA was recovered by phenol-chlorophorm extraction and run on an agarose gel (Figure 2.8).

Lysed cells were cleared by centrifugation at maximum speed for 10 minutes at 4^{o}C and then immunoprecipitation assays were performed based on a standard protocol (Upstate, Millipore). 2–4µg of antibody (Rabbit IgG, Santa Cruz, sc-2027; TEAD1, Abcam, ab133533; RNA pol II, Santa Cruz, sc-899; phospho-SMAD3, Cell Signalling, 9520; GATA-6, Santa Cruz, sc-9055; C-Jun, Santa Cruz, sc-1694; Mef-2c, Santa Cruz, sc-13268) was added in each reaction and incubated overnight, rotating at 4°C. To reverse the cross-linking, the eluates were incubated overnight at 65°C after the addition of 80mM NaCl. Next day, 4mM EDTA, 16mM Tris-HCl and proteinase K were added to the eluates and incubated at 45oC for one hour. The DNA was phenol: chloroform extracted and ethanol precipitated in the presence of glycogen. The pelleted DNA was resuspended in 30µl water. Standard PCR performed using the Fast-Cycling PCR kit (Qiagen, Cat. no: 203741). DNA of each ChIP (5µl) and 1µl of input were added in a 25µl final volume PCR reaction. The primers that were used for the amplification of the genomic regions around *NKX2-5* transcription start site, the rs3095870 genomic locus and the downstream enhancer are shown in Table 2.11.

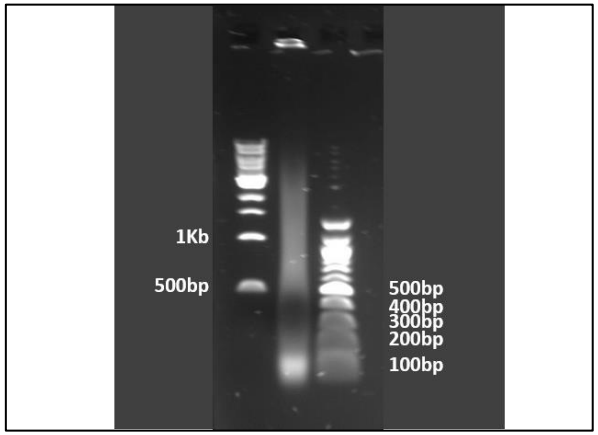


Figure 2.8 Chromatin Shearing.

Region	Name	Forward Primer	Reverse Primer	Annealing Temperature
Promoter	Transcription site (180bp)	GACAGGAGCGATGAGCAGTT	CAGGCTCACATTAGGGAGCA	60^{o}C
	rs3095870 site (180bp)	CCGGGTGGCCTCATTTCTC	ATCCTGTCATCCCCAGCTCT	
Enhancer	Primer Pair 1 (313bp)	GGATGGGACCACGCTACATAC	CTTGTGGCCTCTAAGCCTTG	
	Primer Pair 2 (383bp)	TGACCCAAGGAAACGAAGGG	AACCGGCTCAGAGAAAAGCA	
	Primer Pair 3 (321bp)	GGAGTCGGCAAGGCTTAGAG	GTCCCTTCGTTTCCTTGGGT	
	Primer Pair 4 (265bp)	TTCTCTGAGCCGGTTGAGTT	TGATGGGGCAAGCTGTAGAC	
	Primer Pair 5 (262bp)	CTCCTTGAGTTCTGCCGTCT	ACTGAGGAGGTTACGTGGGT	
	Primer Pair 6 (347bp)	CCGTTCCCGCTTAGAGACTG	CTCCTGTGGGCCTTTTTCAC	

Table 2.11 PCR primers used in the ChIP assays.

2.14 Protein polyacrylamide gel staining

Coomassie Brilliant Blue stain: The polyacrylamide gel was washed 3 times in water for 5 minutes each time, and then stained in Coomassie Brilliant Blue G250 (AMSBio, Cat. no: 17524) for 1 hour with gentle shaking at room temperature. After staining, the gel was washed 3 times in water and de-stained in de-staining solution (30% Methanol, 7% Acetic Acid, ddH_2O), until the background had been cleared.

QC Colloidal Coomassie (Biorad, Cat. no: 1610803) was also used. This staining solution is also a ready-to-use stain, but it does not require alcohols for staining and des-staining steps. In addition, it offers low background and higher sensitivity.

Silver Nitrate Stain: The acrylamide gel was incubated in the fixer solution (40% ethanol, 10% acetic acid, 50% water) for 35 minutes at room temperature. The gel was then washed in water for at least 30 minutes to overnight with frequent changes of water. Overnight washing removes acetic acid, reduces background staining and increases sensitivity. The gel was sensitised in 0.02% sodium thiosulfate for 1 minute only, and washed twice in ddH_2O. A 20 minute incubation of the gel in cold 0.1% silver nitrate solution followed, and 2washes in water. The gel was then transferred in a new 10cm plastic dish and washed in water for 1 minute. The gel was developed in 3% sodium carbonate solution with 0.0175% formaldehyde added just before use. When sufficient stain intensity was achieved (bands start to appear after 2-3 minutes), the gel was washed quickly for 20 seconds in water and the staining is terminated with 5% acetic acid. Stained gels can be stored in the fridge in 1% acetic acid.

2.15 Protein expression analysis

2.15.1 General principle

In molecular and cellular biology, the protein expression studies provide important insight into the state of a cell or its behaviour in response to external stimuli and the biological functions at a given point in time. Therefore, the study of proteins is fundamental to the understanding of the molecular mechanisms that define cell function. There are numerous experimental methods used traditionally by scientists to study proteins including protein detection, isolation and purification, quantification, and characterisation of their structure and function.

Protein electrophoresis and Western blot analysis are widely used to detect the presence or the absence of a protein in a cell or tissue lysate, as well as to identify

differences in the protein expression levels due to a physiological or experimental procedure. There are various methods of protein electrophoresis that differ in the separation matrixes used and the corresponding buffer systems, including polyacrylamide gel electrophoresis (PAGE), isoelectric focusing and 2D electrophoresis. PAGE was used exclusively for the protein analysis in this study. In this technique (410), denatured proteins are separated in an acrylamide gel and migrate through the pores of the gel based on their molecular weight so that smaller proteins travel faster. The proteins move through the gel in response to an electric field that is applied across the buffer chambers and forces the migration. However, in contrast to nucleic acids, proteins are not negatively charged and that interferes with their migration and separation based on the molecular weight. To overcome these limitations, sodium dodecyl sulphate (SDS) detergent was incorporated into the electrophoretic system, which is now widely used and known as SDS-PAGE system. Before electrophoresis, protein samples are mixed with sample buffer (widely known as Laemmli buffer (411)) containing SDS and reducing agents and incubated at high temperatures (95^{o}C for 5 minutes or 70^{o}C for 10 minutes) for complete disruption of molecular interactions. In the presence of sample buffer, the proteins become fully denatured and dissociate from each other. In addition, SDS binds to protein non-covalently, resulting in an overall negative charged protein that migrates in a gel depending purely on its size, enabling molecular weight estimation.

Following SDS-PAGE, proteins are transferred from the acrylamide gel onto a synthetic membrane for further Western blot analysis and immunoblotting (412). The membranes can be made of nitrocellulose, which is the most commonly used, polyvinylidene difluoride (PVDF), activated paper or activated nylon. Electroblotting is the most popular procedure for transferring proteins from a gel to a membrane. The main advantages are the speed and effectiveness of transfer. This process uses an electric current to pull proteins from the gel onto the membrane. It can be achieved by immersion of a gel-membrane sandwich (wet transfer) or by putting the gel-membrane sandwich between absorbent paper that has been soaked in transfer buffer (semidry transfer). The effectiveness of protein transfer depends on the type of gel used, the molecular mass of the protein, and the type of membrane. Some limitations associated with protein transfer include a lower molecular weight limit of ~10kDa, the use of specialised transfer buffers to facilitate transfer of proteins with a high isoelectric point, and problems associated with using a transfer buffer with a lower pH than the isoelectric point.

For protein detection, the protein-bound membranes are immune-decorated with primary antibodies specific to the protein of interest. It is important to prevent non-specific interactions between the primary antibody and the bound proteins. To block nonspecific binding, the membrane is incubated into solutions of blocking reagents such as BSA, non-fat dry milk, or casein. Blocking helps mask any potential nonspecific binding sites on the membrane, thus reducing background ‘‘noise’’, eliminating false positives and providing a clear result.

For the detection of the bound antibodies to the specific protein of interest, various methods are available colorimetric, radioactive, and fluorescent methods. However, chemiluminescent detection is used most often and therefore, will be briefly described here. Enhanced chemiluminescence (ECL) is a sensitive method and can be used for relative quantitation of the protein of interest (413). The primary antibody binds to the protein of interest and the secondary antibody, usually linked to horseradish peroxidase, is used to cleave a chemiluminescent substrate. The reaction product produces luminescence, which is related to the amount of protein. Only a single light detector is required, and the light is detected by photographic film or by a charged couple device camera (more sensitive, greater resolution, and a larger range of exposures than film). It is helpful that many manufacturers produce a variety of ECL-based Western blot detection kits to meet specific needs. Once exposures have been captured, blots can be washed in a buffer and then ‘‘stripped,’’ to remove bound antisera and allow reuse of the blot. Blots can then be stored for future re-probing several more times. However, subsequent re-probing can lead to loss of protein antigens, resulting in a decreased signal (414).

2.15.2 Protein extraction

Two methods of protein extraction were used in these studies. Cytoplasmic and nuclear protein extracts were prepared from cells lysed on ice according to the NE-PER extraction protocol (Thermo Scientific, Cat. no: 78833). This protocol allows a simple, stepwise lysis of cellular and nuclear membranes leading to an efficient isolation of subsequent cytoplasmic and nuclear protein fractions with minimal cross-contamination or interference. Total cell lysates were prepared using ice-cold RIPA extraction buffer (Sigma, Cat. no: R0278). Extraction buffers were supplemented with protease (Complete mini, Roche, Cat. no: 11836170001) and phosphatase inhibitors (Phosphatase Inhibitor Cocktails 2 and 3, Sigma, Cat. no: P5726, R0044). The protein extracts were snap-frozen in liquid nitrogen and stored

at -80^{o}C prior to use. For short-term storage, proteins were denatured and stored at -20^{o}C.

2.15.3 Determination of protein concentration

To determine the total protein concentration in the protein samples, the BCA protein assay (Thermo Fisher Scientific, Cat. no: 23225) was used. The assay utilises the property of the proteins to reduce the charge of Cu^{2+} to Cu^{1+} in an alkaline buffer, and the highly sensitive and selective colorimetric detection of the cuprous cation (Cu^{1+}) by bicinchoninic acid (BCA). When the BCA reagent is added to the protein sample, the cuprous cations are released and react with the bicinchoninic acid creating a dark purple colour. This complex exhibits a strong linear absorbance at 562nm with increasing concentration of protein.

BSA protein standards (Thermo Fisher Scientific, Cat. No: 2320) (125µg/ml – 1mg/ml) were used to create a standard curve for each assay. A blank sample was also included in the assay. After the addition of BCA reagent to the protein samples and the standards, the plate was incubated in the dark for 30 minutes at 37^{o}C. Then the absorbance was measured at 562nm in a Mithras LB 940 Plate Reader. The assay set-up and the creation of a standard curve are described in Figure 2.9.

		BSA Standards, Concentration in µg/µl						
		Blank	0.125	0.25	0.5	0.75	1	Sample
Absorbance at 562nm	Duplicate 1	0.134	0.179	0.26	0.378	0.462	0.571	0.831
	Duplicate 2	0.137	0.193	0.274	0.372	0.476	0.591	0.863
Average		0.1355	0.186	0.267	0.375	0.469	0.581	0.847
Average - Blank		0	0.0505	0.1315	0.2395	0.3335	0.4455	0.7115
Known concentration of Standards		0	0.125	0.25	0.5	0.75	1	1.585689 µg/µl

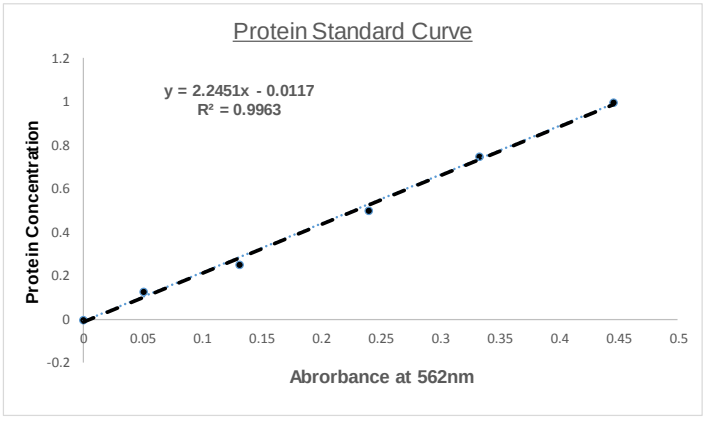


Figure 2.9 Standard curve for protein concentration analysis. *The table shows the absorbance of the blank sample, the BSA standards and the unknown protein sample in duplicates. The average is calculated for all the samples and the raw average value of the blank is subtracted from the rest. Then, a standard curve is plottedbased on the known protein concentrations of the BSA standards, plotting concentration on the X axis, and the absorbance on the Y axis. The unknown protein concentration of the sample is given when the equation is solved for Y.*

2.15.4 SDS-PAGE and Western blotting

After extraction, the proteins were denatured in NuPage LDS Sample buffer and NuPage Reducing agent for 10 minutes at 70^{o}C or 5 minutes at 95^{o}C. The denatured proteins were separated by molecular weight in NuPage Bis-Tris precast gradient (4-12%) polyacrylamide gels (Thermo Fisher Scientific, Cat. no: NP0315) (1.5mm thickness) and run in 1x MOPS SDS Running buffer (Thermo Fisher Scientific, Cat. no: NP0001) for 50 minutes at 185V. Pre-chilled NuPage transfer buffer (Thermo Fisher Scientific, Cat. no: NP0006) was used to transfer the separated proteins to a nitrocellulose membrane (GE Healthcare Life Sciences, Cat. no: 10600048). For transfer of one gel 10% methanol was added to the transfer buffer, if 2 gels were transferred in the same tank the methanol concentration was increased to 20%. Proteins were transferred for 3 hours at 25V or overnight at 12V. Efficient transfer was confirmed using the Ponceau S solution (Sigma, Cat. no: P7170), a reversible stain used to detect proteins on a nitrocellulose or PVDF membrane. Casein blocking buffer (Sigma, Cat. no: C7594) was used to block the protein epitopes. Incubation with the primary antibodies was done overnight on a roller at 4^{o}C, and the HRP-conjugated secondary antibodies diluted in the blocking buffer were added for one hour at room temperature. For the detection of the proteins, the ECL chemiluminescent detection reagent (Amersham, Cat. no: RPN2106) was used.

The primary antibodies used in this study are shown in Table 2.12.

ANTIBODIES	COMPANY	CAT.NUMBER	SPECIES	Concentration (Dilution)
GAPDH	Abcam	ab8245	Mouse	1:50.000
TBP	Abcam	ab51841	Mouse	1:25.000
NKX2-5	Abcam	ab54567	Mouse	1:1000
CTGF	Santa Cruz	sc-14939	Goat	1:2000
Collagen Type 1	Millipore	AB758	Goat	1:2000
AKT	Cell Signalling	9272	Rabbit	1:1000
P-AKT	Cell Signalling	9271	Rabbit	1:1000
p38	Cell Signalling	9212	Rabbit	1:1000
p-p38	Cell Signalling	4511	Rabbit	1:1000
ERK-1,2	Cell Signalling	9102	Rabbit	1:1000
p-ERK-1,2	Cell Signalling	9101	Rabbit	1:1000
ERK-5	Cell Signalling	3372	Rabbit	1:1000
p-ERK-5	Santa Cruz	sc-16564	Goat	1:500
ASK1	Cell Signalling	8662	Rabbit	1:500
p-ASK1	Cell Signalling	3764	Rabbit	1:500
SMAD2/3	Cell Signalling	3102	Rabbit	1:2000
p-SMAD2	Cell Signalling	3101	Rabbit	1:2000
p-SMAD3	Cell Signalling	9520	Rabbit	1:2000
p-SMAD2/3	Cell Signalling	8828	Rabbit	1:2000
VIMENTIN	Cell Signalling	3932	Rabbit	1:1000
N-CADHERIN	Cell Signalling	4061	Rabbit	1:500
YAP1	Cell Signalling	4912	Rabbit	1:2000
p-YAP1	Cell Signalling	4911	Rabbit	1:2000
α-SMA	DAKO	M0851	Mouse	1:2500
TEAD1	Abcam	ab133533	Rabbit	1:2000
TEAD3	Cell Signalling	13224	Rabbit	1:500
p-MKK3/6	Cell Signalling	9231	Rabbit	1:1000

Table 2.12 Antibodies used for Western blotting. *The concentrations that the antibodies were used at are given as the dilution of the primary antibody into the blocking buffer.*

2.15.5 Analysis of relative expression/ densitometry

The relative protein expression was analysed by densitometry, a semi-quantitative measurement of optical density (OD) in light-sensitive materials due to exposure to light. The developed films were scanned and processed in Image J software. It is important to note that overexposed films can provide false data, and therefore exposure must be kept in a linear range in order to be quantitative. A relative value was calculated based on the OD of a given sample compared to the OD of a 'housekeeping' gene, which in most cases was GAPDH. The ratios of the gene-of-interest to GAPDH from the replicate experiments were used to calculate the averages and the standard errors. The results in this thesis represent the relative

protein expression in the given conditions. Statistical analysis was performed in GraphPad Prism 6.

2.16 Gene expression analysis

2.16.1 RNA extraction

Total RNA was extracted from the cells following the instructions of the RNeasy Mini kit extraction protocol (QIAGEN, Cat. no: 74106). This is a column-based system whereby the cells are exposed to a number of reagents, proprietary buffers and centrifugation wash steps, ultimately resulting in RNA binding to a silica membrane from which it can be eluted in RNase-free water.

2.16.2 Measurement of RNA concentration and purity

The quantity and the purity of all RNA samples were measured using the Nanodrop 2000 spectrophotometer. All the RNA samples that were used for downstream applications had a 260/280 ratio of ~2.0 that is considered as “pure”. Samples that did not reach that ratio were not used for RT-qPCR analysis.

2.16.3 Primer design

Primers were designed using the Primer-Blast software available from NCBI. Multiple sets of primers were designed for each gene and tested, and the set that performed best was selected for further analysis. Preferably, primers spanning exon-intron boundaries were selected. The secondary structures including hairpins and homo- and hetero-dimers were assessed with the OligoAnalyzer 3.1 software available from the Integrated DNA Technologies (IDT) who supplied all the oligonucleotides. The GC content and the melting temperature for each primer given the salt conditions of a qPCR reaction were also evaluated. The annealing temperatures were optimised for each primer set separately. For the primer optimisation, a range of primer concentrations were tested with serial dilutions of the RNA. An example of an optimisation is shown in the Figure 2.10. At the end of the RT-qPCR run, the PCR products were electrophoresed in a 1.5% agarose gel to ensure the correct amplicon size and to exclude the possibility of multiple products and excessive primer dimer formation.

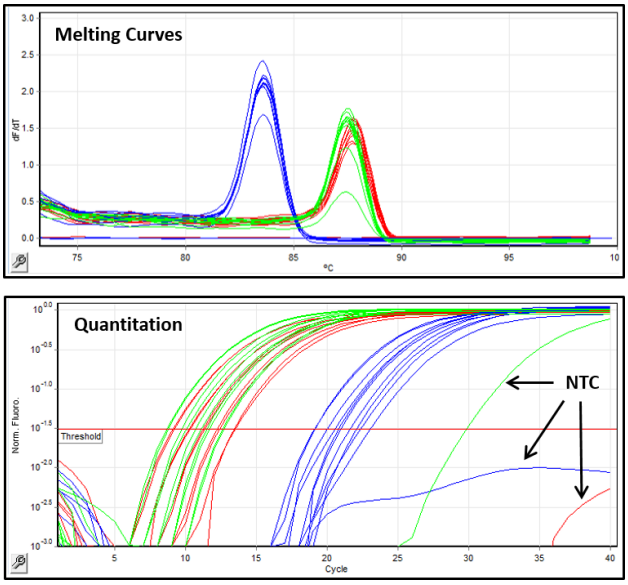


Figure 2.10 Primer optimisation for RT-qPCR. *The figure shows an optimisation for 3 different pairs of primers for the human TEAD1 gene. In the top panel, the melting curves show that a single product is amplified with each primer pair. The lower panel shows the quantitation analysis. Both the green and the red primer pairs were able to identify efficiently the two-fold change in the RNA concentration. The green primer pair was rejected due to non-specific amplification of the NTC sample. The red pair was selected for further analysis.*

2.16.4 Real time-quantitative polymerase chain reaction (RT-qPCR)

RT-qPCR is an adaptation of the standard PCR protocol following the same principle of DNA amplification. RT-qPCR is designed to quantify and determine the initial number of copies of template DNA (the amplification target sequence) with accuracy and high sensitivity over a wide dynamic range. Real-time PCR results can either be qualitative (the presence or absence of a sequence) or quantitative (copy number). Quantitative real-time PCR is thus also known as qPCR analysis.

All qPCR assays rely on the detection of a fluorescent signal which increases as the product amplifies. Fluorescent SYBR green was exclusively used for all qPCR assays in this thesis. SYBR green binds strongly to double-stranded DNA and in the bound state the level of fluorescence increases 100-fold. Initially, fluorescence remains at background levels, and small differences are not detectable even though the amplified product accumulates exponentially after the first few PCR cycles. Eventually, enough amplified product accumulates to yield a detectable fluorescence signal. The cycle number at which this occurs is called the quantification cycle or Cq, and it depends on the amount of the starting material. To analyse the data, the user sets the threshold and the fluorescence intensity is

compared between the samples. To account for differences in the amount of the starting material, fluorescence of the target gene is normalised to that of a 'housekeeping' gene. Housekeeping genes are typically ubiquitously expressed genes that are required for the maintenance of basic cellular functions, and are expressed at relatively constant levels under normal conditions. However, the expression of some housekeeping genes can alter in disease or under certain experimental conditions. In this thesis, I used the human TATA-box binding protein (TBP) that shows the lowest variation in expression among different cell types and tissues (415).

For the qPCR assays, a 1-step method was followed based on the Quantifast SYBR Green RT-PCR protocol (Qiagen, Cat. no: 204154). In this protocol, total RNA, which was the starting material, was transcribed to cDNA before the PCR reaction in the same tube. Total RNA (100ng) was added in 12µl total reaction volume. The qPCR assays were performed in a Corbett Rotor Gene 6000 cycler (Qiagen). The primers used are given in Table 2.13.

Human Gene	Forward Primer	Reverse Primer	Anneal ^{O}C
NKX2-5	GAGCCGAAAAGAAAGAGCTGTG	GGAACCAGATCTTGACCTGCG T	60^{O}C
TBP	AGTGACCCAGCATCACTGTTT	GGCAAACCAGAAACCCTTGC	60^{O}C
TEAD1	CAATGGAGCGACCTTGCCA	GGCCGGGAATGATTCAAACAG	60^{O}C
TEAD3	CTGACACGTACAGCAAACAC	AGCTCCTTCAATCCTCCCTT	60^{O}C
COL1A2	TGCTTGCAGTAACCTTATGCCTA	CAGCAAAGTTCCCACCGAGA	60^{O}C
CTGF	GACCTGGAAGAGAACATTAAGAAGG	TCGGTATGTCTTCATGCTGGTG	60^{O}C
α-SMA	CCGACCGAATGCAGAAGGAG	ACAGAGTATTTGCGCTCCGAA	58^{O}C
SM22	GATTCTGAGCAAGCTGGTGA	TCTGCTTGAAGACCATGGAG	62^{O}C
CD31	ATTGCAGTGGTTATCATCGGAGTG	CTCGTTGTTGGAGTTCAGAAGT GG	62^{O}C
VE-Cadherin	CAG CCC AAA GTG TGT GAG AA	CGG TCA AAC TGC CCA TAC TT	62^{O}C
v-WF	CGGCTTGCACCATTCAGCTA	TGCAGAAGTGAGTATCACAGC CATC	62^{O}C

Table 2.13 qPCR primers for human genes and annealing temperatures.

For the analysis of *NKX2-5* gene expression, numerous sets of primers were tested before one that performed well was identified. In samples where *NKX2-5* expression was low (eg in untreated samples or in normal cells), a non-specific product appeared very early in the amplification. To solve this problem, a specific run file was created that allowed the fluorescence to be read at a time point after the non-specific product was amplified. Thus, only the fluorescence arising from the specific products was analysed.

2.17 Statistical analysis

For the genetics data, statistical analysis was performed in Plink. For the rest of the experimental work, statistical analysis was mainly performed in GraphPad Prism 6 (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001) and occasionally in Microsoft Office Excel. The unpaired student's t-test was exclusively used unless otherwise stated. Student's t-test is most commonly used for testing a hypothesis of a difference of the means between two small sample groups normally distributed (416). Unpaired tests were performed since all the sample groups were independent, and the 2-tailed version was used indicating that the means are not expected to be equal under the null hypothesis. Results represent data from at least 3 independent experiments, unless otherwise stated, and shown as mean + SEM.

CHAPTER 3 - RESULTS: GENETIC ASSOCIATION OF NKX2-5 WITH PULMONARY HYPERTENSION AND SCLERODERMA

3.1 Introduction and study design

A literature review on NKX2-5 reveals dozens of scientific papers highlighting the importance of the gene during mouse embryonic development and its non-redundant role in the heart and vessel formation. However, the majority of published data are focused on the genetic mutations identified along the *NKX2-5* gene and their deleterious effects for the carriers leading to CHD, atrioventricular septal defects, atrial fibrillation, and tetralogy of Fallot. Apart from the mutations leading to malformations of the cardiac system, very little information on genetics of *NKX2-5* is available.

The first aim for this PhD was to explore the hypothesis that *NKX2-5* is genetically associated with vascular pathologies, and in particular with conditions underlined by vascular remodelling such as PH, PAH, atherosclerosis and PAD, etc. Scleroderma is a good model for studying vascular disease due to the extensive fibrosis and vasculopathy. Initial evidence for the hypothesis of association of *NKX2-5* with vascular pathology in scleroderma arose from a case-control association study in SLE patients, where a SNP upstream of *NKX2-5* showed significant association with the disease. SLE and scleroderma are both complex AIDs with vascular complications and shared common pathogenesis.

To investigate the potential association, a candidate gene case-control association study was designed, as described below (Figure 3.1):

1. DNA samples from scleroderma patients and healthy individuals were collected as explained in section 2.1 to form a discovery and a replication cohort.
2. A set of tagging SNPs across the *NKX2-5* genomic locus was selected using different software (section 2.2).
3. The tagging SNPs were genotyped in the discovery cohort and different sets of analyses were performed.

4. The tagging SNPs were genotyped in the replication cohort, and the same sets of analyses were performed.
5. A meta-analysis was performed among the two independent cohorts.

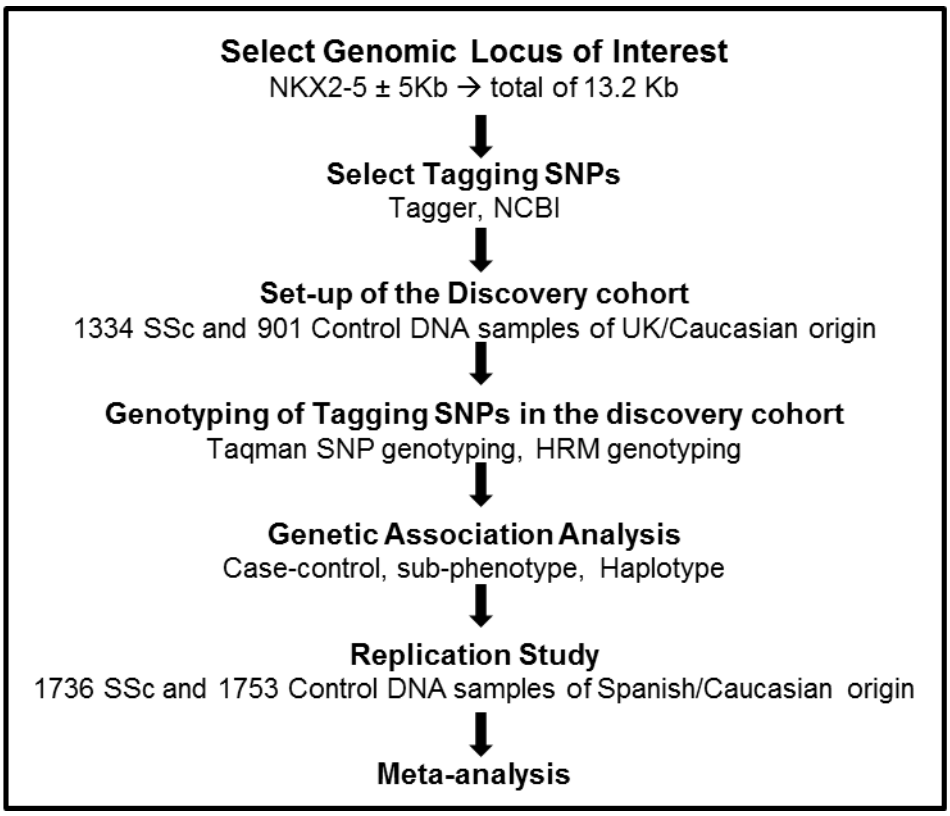


Figure 3.1 Genetic association study design and work-flow. *The NKX2-5 gene is 3.2Kb long and is located in the q arm of chromosome 5. A genomic region 13.2Kb long that centred NKX2-5 leaving 5Kb upstream and downstream of the gene was selected for the study. Six tagging SNPs were selected in the area using different software. The tagging SNPs were genotyped in the discovery and the replication cohorts. The cohorts were independent but shared similar Caucasian origin. DNA samples from 1334 scleroderma patients and 901 healthy individuals were collected in the UK (Centre for Rheumatology, Royal Free NHS Foundation Trust Hospital, London and Institute of Inflammation and Repair, University of Manchester) to form the discovery cohort. DNA samples from 1736 scleroderma patients and 1753 healthy individuals were collected in Spain (Institute de Parasitology and Biomedicine Lopez-Neyra, Granada) to form the replication cohort. Genotype data were used for the overall case-control genetic association analysis and the sub-phenotype analysis in the discovery cohort. The same sets of analyses were repeated in the replication cohort to confirm positive findings. A meta-analysis was performed of both cohorts.*

3.2 Results

3.2.1 Selection and description of tagging SNPs

Genotype data for the HapMap-CEU were used to generate the LD profile of *NKX2-5* region in Haploview (387). As shown in Figure 3.2, a big LD block formed across the region, with 14 SNPs. Tagger software was used to select the tagging SNPs in the region. Two versions of the software were compared: the independent platform through the Broad Institute of Harvard and MIT (386), and the version incorporated in Haploview. The results were similar. A list of 6 tagging SNPs was generated based on their position and annotations regarding their functionality, available through the HaploReg (417) and the ENCODE data at UCSC Genome browser. The selected tagging SNPs are described in detail in Table 3.1.

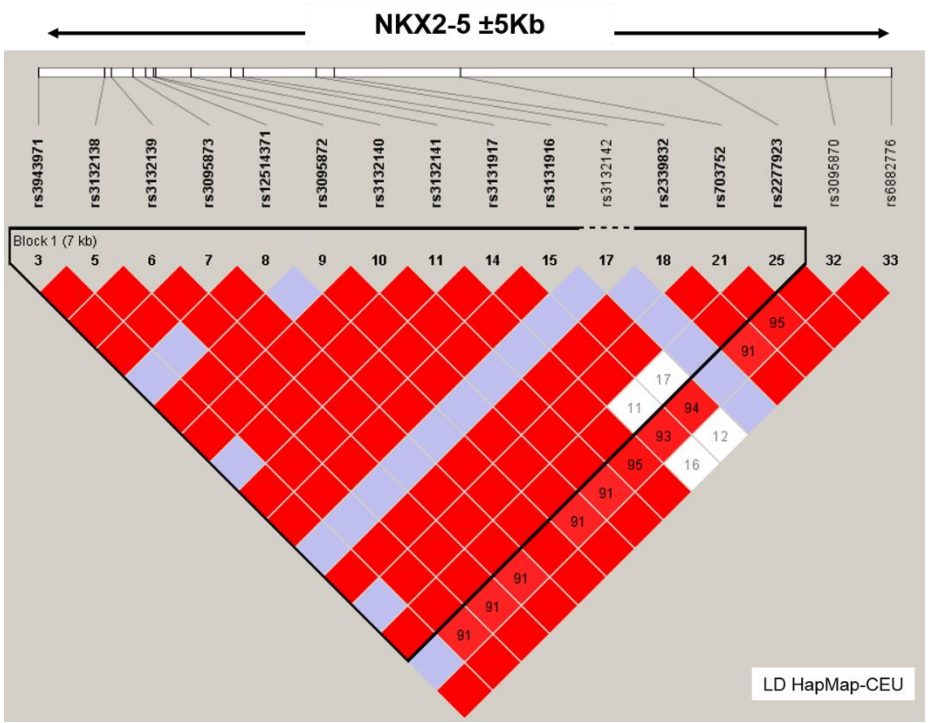


Figure 3.2 LD Block in HapMap-CEU population. *The white horizontal bar illustrates the location of the tagging SNPs in the region on a physical scale. The names of the tagging SNPs are shown above the triangle. The values in the squares represent the pairwise correlation (r^2) between tagging SNPs defined by the upper left and the upper right sides of the square. Shading represents the magnitude and significance of pairwise LD, with a red-to-white gradient reflecting higher to lower LD values. Extended LD covers the region and one haplotype clock is formed, shown in the black triangle. Almost all of the SNPs are in high LD, apart from rs3132142 that is not related to any other SNP. Lilac boxes show pairs of SNPs that are not in LD.*

SNP	Major Allele	Minor Allele	H3K4me3 marks for Promoters	H3K4me1 marks for Enhancers	DNAse Hypersensitivity	Bound Proteins	Changed protein binding motif	Relative posit216ion
rs3095870	C	T	Heart	8 tissues	Heart	SUZ12	NF-kappaB, TEAD1	1.1kb 5' of *NKX2-5*
rs2277923	T	C	8 tissues	15 tissues	14 tissues	SUZ12	TCF12, Myf1	*NKX2-5*, Synonymous
rs703752	C	A	Heart	9 tissues	ESC		Myb	*NKX2-5* 3'UTR
rs31311917	G	T	Heart	7 tissues	Heart		EBF, NRSF, Sin3Ak-20, TCF12	2.1kb 3' of *NKX2-5*
rs12514371	A	G	9 tissues	14 tissues	Heart		AP-1, Myc	3Kb 3' of *NKX2-5*
rs3132139	G	A	4 tissues	12 tissues	Epithelial, Heart, HSC & B-cell	SUZ12	AhR:Arnt, Arnt, Myc	3.4kb 3' of *NKX2-5*

Table 3.1 List of the selected tagging SNPs. *In silico analysis in Haploreg for the selected tagging SNPs along the NKX2-5 genomic locus. The major and minor alleles are shown, and data from the in silico analysis describing histone methylation marks related to promoters and enhancers, as well as DNase hypersensitivity marks and potential binding sites of transcription factors. In the last column, the position of each SNP relative to NKX2-5 gene is given.*

Tagging SNP rs2277923 is located within the first exon, and rs703752 is located at the 3'UTR of *NKX2-5* gene. The remaining 4 SNPs are found in intergenic regions: 1 upstream and 3 downstream of the gene. The minor allele of rs2277923 leads to a synonymous change, and therefore, no change in the protein structure and function is expected. However, the SNP has been reported in few articles as a *NKX2-5* mutation in patients with CHD (418, 419). SNP rs703752 is located in the 3'UTR of *NKX2-5.* The SNP was also reported as a *NKX2-5* mutation in patients with cardiac malformations (419), but no other information is published in literature.

In HaploReg, DNase I hypersensitivity and histone marks specific to promoter or enhancer regions suggest that all the tagging SNPs could be functional. DNase hypersensitive sites are regions of chromatin that are sensitive to DNase I enzyme digestion, leaving the DNA exposed and accessible to transcription factors. In a similar way, enhancer and promoter regions that are devoid of nucleosomes, and histones in the vicinity are tagged with specific marks. It has been shown that H3K4me1 marks are found near enhancer regions, while H3K4me3 marks found near active promoters (420). These marks reveal an open chromatin state, where transcription factors are more able to bind in order to regulate gene expression. The activity of enhancers and sometimes of promoters, is cell type- or tissue-specific depending on the cell requirements. Interestingly, the majority of the functional evidence for the SNPs is derived from heart tissue as expected.

In addition, some of the SNPs are located within consensus protein binding sites that are altered in the presence of the alternative allele. This evidence comes from position weight matrix (PWM)-based probabilities. In brief, the sequence of interest is aligned against other sequences that contain known consensus binding sites for various DNA-binding transcription factors and are thought to be functional. A score is given for each nucleotide based on the similarity with the reference position, and it is independent of other nucleotides. A final score is then calculated by adding the relevant values at each position, which provides an indication of how different or similar is the sequence in question to the reference.

More evidence is based on annotations available through ChIP-seq experiments, where chromatin is immunoprecipitated with different antibodies on a genome-wide scale. The genomic regions are then identified and reported as the genomic signature of the protein used for the immunoprecipitation. SUZ12 (suppressor of zeste-12) is a core component of the polycomb repressive complex 2 (PRC2) that is implicated in transcriptional silencing by generating di- and tri-methylation of lysine

27 on histone H3 (H3K27Me3). In HaploReg, there is evidence that 3 of the NKX2-5 SNPs are located within or near to a SUZ12 binding site.

Overall, the data suggest that all 6 tagging SNPs may display functional potentials, and that are all located in strategic areas in terms of regulation.

3.2.2 Description of the discovery/UK cohort

The Centre for Rheumatology at the Royal Free NHS Foundation Trust Hospital is privileged to retain the biggest cohort of scleroderma patients in the UK, with over 1000 DNA samples deposited to the Biobank from consented patients. Control DNA samples from healthy individuals were collected from various sources (section 2.1). Scleroderma and control samples were genotyped in the laboratory of the Centre for Rheumatology and CTD. An additional cohort of DNA samples of scleroderma patients was included in the study in collaboration with Professors Ariane Herrick and Javier Martin. The samples were originally collected in Manchester under the supervision of Professor Ariane Herrick, and were sent to Granada, Spain as a part of a European scleroderma consortium under the supervision of Professor Javier Martin. The samples were genotyped in the Institute of Parasitology and Biomedicine Lopez-Neyra, Granada, Spain. Genotype data together with matched clinical information were sent to our lab for analysis.

After genotyping and quality control, a total of 1334 cases and 901 control samples were included in the study as the discovery cohort of UK/Caucasian origin (Table 3.2). Assessment of the Hardy-Weinberg equilibrium (HWE) in case-control cohorts has been widely used as a measurement of quality control in order to identify potential genotyping errors (421). All the SNPs were in HWE (Table 3.3).

Sub-Phenotypes	DISCOVERY		REPLICATION
	London	Manchester	
Gender			
Female	771 (85.7%)	354 (81.3%)	1489 (88.8%)
Male	128 (14.3%)	81 (18.6%)	187 (11.1%)
N/A	–	–	60 (3.4%)
Auto-Antibodies			
ATA	203 (22.5%)	50 (11.5%)	351 (22.3%)
ACA	312 (34.7%)	161 (37%)	735 (46.3%)
ARA	172 (19.1%)	N/A	N/A
Other	104 (11.5%)	N/A	N/A
Organ Involvement			
PF	264 (29.3%)	148 (34%)	383 (24.3%)
SSc-PAH	77 (8.5%)	61 (14%)	133 (12.7%)
PH	57 (6.3%)	29 (6%)	230 (22.4%)
RC	78 (8.6%)	N/A	N/A
Disease Subset			
LcSSc	615 (68.4%)	329 (75.6%)	1034 (63.8%)
DcSSc	282 (31.3%)	106 (24.3%)	428 (26.4%)
Overlap	2 (0.2%)	–	23 (1.4%)
Total Cases	**899**	**435**	**1736**
Total Controls	**901**		**1753**
	(UK origin)		(Spanish origin)

Table 3.2 Discovery and replication cohorts. *The patients were categorised in groups based on the presence of auto-antibodies, other pathological complications and the scleroderma subsets. ATA: anti-topoisomerase1; ACA: anti-centromere; ARA: anti-RNA polymerase; PF: Pulmonary fibrosis; PAH: Pulmonary arterial hypertension; PH: Pulmonary Hypertension; RC: Renal crisis; LcSSc: limited scleroderma; DcSSc: diffuse scleroderma. N/A: not available.*

SNP	m	M	Cohort	Genotypes	Observed	Expected	p-value
rs3132139	A	G	UK	95/417/364	0.476	0.4529	0.1364
			Spain	201/816/702	0.4747	0.4575	0.1265
rs12514371	G	A	UK	23/255/598	0.2911	0.2846	0.554
			Spain	46/469/1204	0.2728	0.2731	0.9298
rs3131917	T	G	UK	190/448/248	0.5056	0.4979	0.6856
			Spain	392/847/358	0.5304	0.4998	0.01622
rs703752	A	C	UK	102/416/346	0.4815	0.4601	0.1838
			Spain	154/819/761	0.4723	0.4387	0.00149
rs2277923	C	T	UK	60/388/428	0.4429	0.4118	0.0268
			Spain	157/736/799	0.435	0.428	0.5322
rs3095870	T	C	UK	102/427/338	0.4925	0.463	0.06671
			Spain	142/732/687	0.4689	0.4391	0.00796

Table 3.3 Hardy-Weinberg equilibrium. *The m: minor and M: major alleles, genotype count in the control cohorts, the observed and expected heterozygosity, and the p-value are shown in the table as assessed by Plink. P-value cut-off: 0.001. All the SNPs were in HWE.*

3.2.2.1 LD in the discovery cohort

The genotype data of the discovery cohort were used to examine the LD in Haploview (Figure 3.3). All the SNPs were in strong LD except for SNPs rs3131917 and rs2277923 (r^2=0.11). Two LD blocks were formed: block 1: rs3132139 and rs12514371, and block 2: rs703752 and rs2277923. It is interesting to see that the same LD pattern seen in the HapMap-CEU population (Figure 3.2) is similar here; rs3095870 remains outside the LD block, and also SNPs rs3131917and rs2277923 are not in LD.

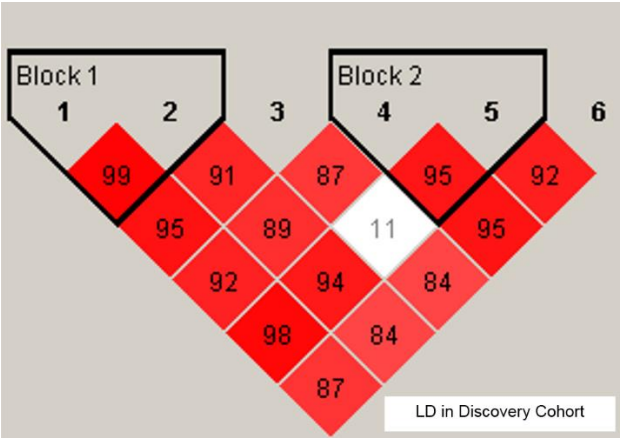


Figure 3.3 LD plot of the discovery cohort. *R^2 values between the pairs of SNPs are indicated in the red squares. Two blocks are formed, each containing two tagging SNPs. The SNPs are: 1: rs3132139, 2: rs12514371, 3: rs3131917, 4: rs703752, 5: rs2277923, 6: rs3095870.*

3.2.2.2 Case-control association study

An overall case-control association study was performed in the discovery cohort. A basic allelic test questioned whether the frequency of the alleles of the SNPs is the same in scleroderma cases and controls. None of the SNPs found to be associated with scleroderma ($p<0.05$) in the allelic test (Table 3.4).

3.2.2.3 Sub-phenotype analysis

Although no significant SNP association was found in the overall case-control association analysis, I further examined whether any of the SNPs was associated with the disease sub-phenotypes: ARA, ACA, ATA, PH, PAH, PF (Table 3.2). The sub-phenotypes based on the organ complications, PH, PAH, and PF, were of particular interest to me. The results of the sub-phenotype association analysis are shown in Tables 3.5 and 3.6.

SNP rs3132139 was found to be associated with PH (p-value= 0.036), and after correcting for multiple testing using permutation analysis, the association was still significant (permutation p-value= 0.039). The odds ratio (OR) was 0.71 and the 95% confidence interval (CI) was 0.57-0.97. To further explore the association, I performed a genetic analysis using a different genetic model. Rs3132139 was significantly associated with PH under the dominant model (p-value= 0.004, permutation p-value= 0.001) (Table 3.5).

SNP	Position	Minor Allele	Major Allele	MAF in Cases	MAF in controls	x^2	P-value	OR (95% CI)	Permutation P-value
rs3132139	173228730	A	G	0.3451	0.3465	0.008434	0.9268	0.994 (0.875-1.129)	0.9351
rs12514371	173229104	G	A	0.1692	0.1718	0.05167	0.8202	0.9814 (0.83-1.15)	0.8342
rs3131917	173230018	T	G	0.4839	0.4673	1.166	0.2801	1.069 (0.94-1.2)	0.2737
rs703752	173232508	A	C	0.3545	0.3588	0.08215	0.7744	0.9816 (0.86-1.11)	0.7782
rs2277923	173235021	C	T	0.2875	0.29	0.0299	0.8627	0.9882 (0.86-1.13)	0.8761
rs3095870	173236451	T	C	0.3514	0.3639	0.6989	0.4031	0.947 (0.83-1.07)	0.3886

Table 3.4 Case-Control association study in the discovery cohort. *The minor and major alleles for each SNP, and their frequencies in the cases of the discovery cohort are shown. The p-values for the x^2 statistics under the basic allelic test and the permuted p-values are given. OR: odds ratio, CI: confidence interval.*

Group	SNP	m/M	Test/Model	MAF in Cases	MAF in Controls	x^2	P-value	OR (95%CI)	Permutation P-value
	rs3132139	A/G	allelic	0.2752	0.3465	4.393	0.03609	0.7163 (0.52-0.97)	0.03996
			dominant	48/61	512/364	8.207	0.004174	NA	0.0119
	rs12514371	C/T	allelic	0.1875	0.1718	0.3195	0.5719	1.112	0.5994
	rs3131917	T/G	allelic	0.4575	0.4673	0.07191	0.7886	0.9616	0.7902
	rs703752	A/C	allelic	0.4095	0.3588	2.081	0.1492	1.239	0.1638
	rs2277923	G/A	allelic	0.2837	0.29	0.0359	0.8497	0.9697	0.8462
PH	rs3095870	A/G	allelic	0.3645	0.3639	0.000284	0.9866	1.003	0.999
	rs3132139	A/G	allelic	0.35	0.3465	0.01251	0.9109	1.016	0.9191
	rs12514371	C/T	allelic	0.152	0.1718	0.61	0.4348	0.8641	0.4266
	rs3131917	T/G	allelic	0.4921	0.4673	0.5445	0.4606	1.104	0.4216
	rs703752	A/C	allelic	0.3555	0.3588	0.01074	0.9175	0.9856	0.9101
	rs2277923	G/A	allelic	0.2823	0.29	0.06263	0.8024	0.963	0.7812
PAH	rs3095870	A/G	allelic	0.3333	0.3639	0.8921	0.3449	0.874	0.3057
	rs3132139	A/G	allelic	0.3367	0.3465	0.2321	0.63	0.9575	0.6244
	rs12514371	C/T	allelic	0.1713	0.1718	0.001035	0.9743	0.9964	0.979
	rs3131917	T/G	allelic	0.4759	0.4673	0.1653	0.6843	1.035	0.6773
	rs703752	A/C	allelic	0.3601	0.3588	0.003694	0.9515	1.005	0.9381
	rs2277923	G/A	allelic	0.2875	0.29	0.01549	0.901	0.9883	0.8951
PF	rs3095870	A/G	allelic	0.3462	0.3639	0.7367	0.3907	0.9254	0.3586

Table 3.5 Sub-phenotype analysis of PH, PAH, PF, RC in the discovery cohort. *The association analysis in each sub-phenotype is presented. The frequencies of the minor (m) and major (M) alleles are shown in cases and controls. For the dominant model, the genotype count is shown. Statistically significant p-values (<0.05) are highlighted.*

Group	SNP	m/M	Test	Freq/Count in Affected	Freq/Count in Unaffected	x^2	P-value	OR (95%CI)	Permutation p-value
ATA	rs3132139	A/G	allelic	0.3252	0.3465	0.7716	0.3797	0.9091	0.3826
	rs12514371	C/T	allelic	0.1833	0.1718	0.3562	0.5506	1.082	0.5554
	rs3131917	T/G	allelic	0.4456	0.4673	0.7343	0.3915	0.9162	0.3816
	rs703752	A/C	allelic	0.3765	0.3588	0.522	0.47	1.079	0.4376
	rs2277923	G/A	allelic	0.2947	0.29	0.0422	0.8372	1.023	0.8482
	rs3095870	A/G	allelic	0.374	0.3639	0.1653	0.6843	1.044	0.6663
ACA	rs3132139	A/G	allelic	0.3557	0.3465	0.2258	0.6347	1.042	0.6284
	rs12514371	C/T	allelic	0.164	0.1718	0.2528	0.6151	0.946	0.6454
	rs3131917	T/G	allelic	0.4922	0.4673	1.494	0.2215	1.105	0.2098
	rs703752	A/C	allelic	0.3473	0.3588	0.3399	0.5599	0.9511	0.5714
	rs2277923	G/A	allelic	0.2777	0.29	0.4363	0.5089	0.9413	0.5335
	rs3095870	A/G	allelic	0.3466	0.3639	0.7608	0.3831	0.9272	0.3806
ARA	rs3132139	A/G	allelic	0.3787	0.3465	1.291	0.2558	1.15	0.2448
	rs12514371	C/T	allelic	0.1391	0.1718	2.19	0.1389	0.7786	0.1309
	rs3131917	T/G	allelic	0.5444	0.4673	6.762	0.00931	1.362 (1.079-1.72)	0.01099
	rs703752	A/C	allelic	0.3363	0.3588	0.6211	0.4306	0.9056	0.4216
	rs2277923	G/A	allelic	0.2725	0.29	0.4197	0.5171	0.917	0.5175
	rs3095870	A/G	allelic	0.3482	0.3639	0.3	0.5839	0.9339	0.5914

Table 3.6 Sub-phenotype analysis of ATA, ACA, ARA in the discovery group. *The association analysis in each sub-phenotype is presented. The frequencies of the minor (m) and major (M) alleles are shown in cases and controls. For the dominant model, the genotype count is shown. Statistically significant p-values (<0.05) are highlighted.*

In the auto-antibody based sub-phenotypes, rs3131917 was significantly associated with the ARA group (p-value= 0.009, permutation p-value= 0.01, OR= 1.36, 95%CI= 1.079-1.72) in the basic allelic test (Table 3.6).

No other SNP showed association with any of the sub-phenotypes.

A subsequent analysis examined whether the association between rs3132139 and PH depends on the presence of scleroderma. To do this, I repeated the association analysis between the PH^+ versus the PH^- scleroderma cases. I hypothesised that if the association found is caused by scleroderma as a common genetic background, the sub-subsequent analysis would be negative. However, as shown in Table 3.7, rs3132139 was still associated with PH in the basic allelic test (p-value= 0.02, permutation p-value= 0.01) and in the model analysis under the dominant model (p-value= 0.005, permutation p-value= 0.01). Next, to further confirm that the association is specific to PH and does not occur due to chance, PH^+ cases were removed from the cohort and the association test was repeated. Indeed, when the PH^+ cases were removed from the analysis, no association was found between the PH^- scleroderma cases compared to controls.

This result verifies that rs3132139 is associated with PH and the association is specific and independent of the presence of scleroderma.

Sub-Phenotype	SNP	Association Test	Allele/Genotype Test	P-value	Permutation P-value
PH	rs3132139	PH+ vs Control	A vs G	0.036	0.039
			GG + AG Vs AA	0.004	0.01
		PH+ vs PH-	A vs G	0.02	0.01
			GG + AG Vs AA	0.005	0.01
		PH- vs Controls	A vs G	0.7341	–
			GG + AG Vs AA	0.726	–

Table 3.7 Subsequent association analysis of rs3132139 in the PH sub-phenotype. *Subsequent sub-phenotype association analysis was performed to further dissect the type of the association or rs3132139 and PH. Statistically significant p-values (<0.05) are highlighted.*

3.2.2.4 Haplotype analysis

I performed a haplotype association analysis in Haploview between the haplotypes occurring in the discovery cohorts (Figure 3.3) and scleroderma, as well as the sub-phenotypes. No haplotype was significantly associated with scleroderma or any other sub-phenotype (Table 3.8). The analysis was repeated in Plink with the same results.

Next, I created an artificial haplotype forcing all 6 tagging SNPs to form a LD block. Surprisingly, this haplotype was significantly associated with scleroderma (p= 4.05×10^{-6}), PH (p-value= 3.32×10^{-12}), and ARA (p-value= 0.0003) (Table 3.8). However, since this haplotype cannot occur physically in the population, the association is just an indication that specific allele combinations might have a synergistic effect on the disease pathogenesis.

Blocks	SNPs in the blocks	Alleles in the Haplotype	Freq. in Cases	Freq. in Controls	x^2	P-value	PH	ARA
Block 1	rs3132139/ rs12514371	GC	0.1668	0.1725	0.1997	0.655	–	–
		AT	0.3484	0.3495	0.0047	0.9451	–	–
		GT	0.4848	0.478	0.161	0.6882	–	–
Block 2	rs703752/ rs2277923	CG	0.2777	0.2845	0.1988	0.6557	–	–
		AA	0.3589	0.3552	0.04961	0.8237	–	–
		CA	0.3634	0.3602	0.03801	0.8454	–	–
	All SNPs	*GTTAAA*	*0.02258*	*0.004235*	*21.24*	*4.05E-06*	*3.32E-12*	*0.0003*

Table 3.8 Haplotype association analysis in the discovery cohort. *Two haplotype blocks were formed in Haploview and were tested for association with scleroderma, PH and ARA. The 6 tagging SNPs were forced together to form a haplotype block (bottom row). Statistically significant p-values (<0.05) are highlighted.*

3.2.3 Description of replication/Spanish cohort

A replication study was designed in collaboration with Professor Javier Martin, Institute of Parasitology and Biomedicine Lopez-Neyra, Granada, Spain. The replication cohort was used to confirm important findings arising from the discovery cohort. DNA samples of scleroderma patients and healthy individuals were collected from hospitals and clinics across Spain and sent for genotyping to the lab at the Institute of Parasitology and Biomedicine Lopez-Neyra. Genotype and clinical data were then sent in our lab for the analysis. After the quality control, 1736 scleroderma cases and 1753 control samples formed the replication cohort and

were included in the study (Table 3.2). No clinical data about RC and ARA were available. All 6 SNPs were in Hardy-Weinberg equilibrium (Table 3.3).

3.2.3.1 LD in the replication cohort

The genotype data were used to construct the LD plot in Haploview (Figure 3.4). The pattern of LD in the replication cohort is similar to that of the discovery cohort, with similar levels of LD (r^2 values) between the pairs of the individual SNPs. SNPs rs3131917 and rs2277923 were unrelated (r^2= 0.19). Two LD blocks were formed again, but different SNPs were included in each block.

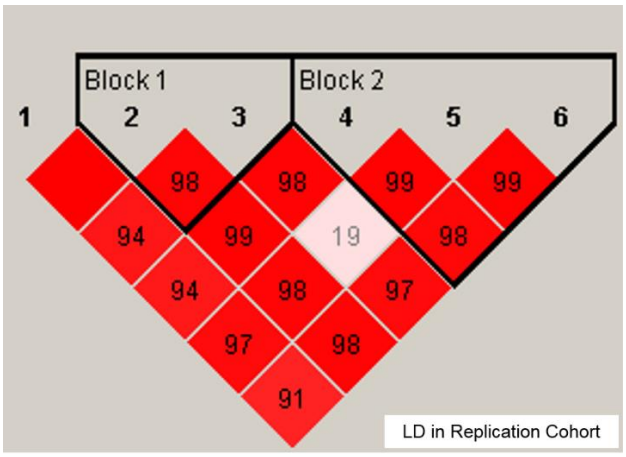


Figure 3.4 LD plot of the replication cohort. *R^2 values between the pairs of SNPs are indicated in the red squares. The SNPs are: 1: rs3132139, 2: rs12514371, 3: rs3131917, 4: rs703752, 5: rs2277923, 6: rs3095870.*

3.2.3.2 Case-control association study

The main aim of a replication study is to confirm positive associations and give strength and robustness to genetic findings. The overall case-control association study in the discovery cohort did not show any significant association. However, since the genotype data of the replication cohort is a valuable source of information, I performed a case-control association study in a similar way.

The results of the study were really intriguing (Table 3.9). Three SNPs showed evidence of association with scleroderma overall. In particular, rs3132139 was associated with scleroderma in the basic allelic test (p-value= 0.038, OR= 1.11, 95%CI= 1.006-1.226), but the association was no longer significant after the permutation analysis (permutation p-value= 0.054). However, in the model analysis, rs3132139 was significantly associated with scleroderma under the recessive model (p-value= 0.005, permutation p-value= 0.01).

Rs3131917 was also found to be significantly associated with scleroderma in a model analysis under the dominant model (p-value=0.004, permutation p-value= 0.008) (Table 3.9).

In the model analysis, rs3095870 also showed evidence of association, with the C allele of the SNP being associated with scleroderma (p-value= 0.03), but the association failed after the permutation analysis (permutation p-value= 0.06) (Table 3.9).

3.2.3.3 Sub-phenotype analysis

A sub-phenotype association analysis was performed in the replication cohort (Table 3.10). SNP rs3132139 was significantly associated with PH (p-value= 0.02203, OR= 1.263, 95% CI= 1.034-1.543, permutation p-value= 0.0205) in the basic allelic test, confirming successfully the association found in the discovery cohort. This result further verifies that rs3132139 is directly associated with PH and the association is replicated in an independent cohort of Caucasian origin.

The sub-phenotype analysis in the replication cohort also revealed that SNPs rs3131917 and rs12514371 are associated with PH (permutation p-value=0.0032 and 0.033, respectively) (Table 3.10). These data suggest a strong overall association between the *NKX2-5* genomic locus and PH.

Furthermore, rs3131917 was also found to be associated with the ATA sub-phenotype (permutation p-value=0.01) (Table 3.10).

SNP	Position	Minor Allele	Major Allele	Test	MAF in Cases	MAF in Controls	x^2	P-value	OR (95%CI)	Permutation P-value
rs3132139	173228730	A	G	allelic	0.3783	0.354	4.285	0.03846	1.11(1.006-1.226)	0.05405
				recessive	254/1453	201/1518	7.554	0.005988	NA	0.01698
rs12514371	173229104	G	A	allelic	0.1562	0.1629	0.5642	0.4526	0.9512	0.6154
rs3131917	173230018	T	G	allelic	0.4863	0.5107	3.758	0.05255	0.9071	0.06383
				dominant	1186/433	1239/358	8.119	0.004379	NA	0.008385
rs703752	173232508	A	C	allelic	0.3115	0.3252	1.47	0.2254	0.9387	0.2
rs2277923	173235021	C	T	allelic	0.3032	0.3102	0.3877	0.5335	0.9675	0.5
rs3095870	173236451	T	C	allelic	0.3101	0.3257	1.813	0.1781	0.9303	0.1667
				dominant	895/813	874/687	4.232	0.03968	NA	0.06429

Table 3.9 Case-control association study in the replication cohort. *The minor and major alleles, and their frequencies in the cases of the replication cohort are shown. The basic allelic test was used for the overall association analysis. Where a marginal or significant association was found, a model analysis was also conducted, and the genetic model with lowest p-value is shown. The x^2 test value, the p-value, the 95% CI of the OR (odds ratio) are provided. Statistically significant p-values (<0.05) are highlighted.*

	SNP	m/M	Test	MAF in Cases	MAF in Controls	x^2	P-value	OR (95%CI)	Permutation P-value
PH	rs3132139	A/G	allelic	0.4093	0.3543	5.244	0.02203	1.263 (1.034-1.543)	0.0205
	rs12514371	C/T	allelic	0.1244	0.1632	4.471	0.03447	0.7289 (0.54-0.978)	0.033
	rs3131917	T/G	allelic	0.4372	0.5106	8.175	0.004247	0.7445 (0.607-0.911)	0.0032
	rs703752	A/C	allelic	0.2917	0.325	2.051	0.1521	0.8553	0.1383
	rs2277923	G/A	allelic	0.2987	0.3103	0.2518	0.6158	0.9466	0.6094
	rs3095870	A/G	allelic	0.292	0.3254	2.02	0.1553	0.855	0.1394
ACA	rs3132139	A/G	allelic	0.387	0.3543	4.689	0.03036	1.151 (1.013-1.3)	0.0399
	rs12514371	C/T	allelic	0.1635	0.1632	0.000609	0.9803	1.002	0.9831
	rs3131917	T/G	allelic	0.4884	0.5106	1.907	0.1674	0.9149	0.1683
			dominant	499/191	1239/358	7.319	0.006823	NA	0.0166
	rs703752	A/C	allelic	0.3051	0.325	1.861	0.1725	0.912	0.1641
	rs2277923	G/A	allelic	0.3011	0.3103	0.4012	0.5265	0.9577	0.5243
	rs3095870	A/G	allelic	0.305	0.3254	1.903	0.1677	0.9096	0.1601

Table 3.10 Sub-phenotype analysis in the replication cohort. *The association analysis in each sub-phenotype is presented. The frequencies of the minor (m) and major (M) alleles are shown in cases and controls. For the dominant model, the genotype count is shown. The x^2 test value, the p-value, the 95% CI of the OR (odds ratio) are provided. Statistically significant p-values (<0.05) are highlighted.*

3.2.3.4 Haplotype analysis

Following the same steps as before, I performed a haplotype association analysis in Haploview between the haplotypes of the replication cohort (Figure 3.4) and the disease overall. No haplotype showed any evidence of association with scleroderma or PH (Table 3.11).

In a similar way, I created an artificial haplotype forcing the 6 tagging SNPs to form a haplotype, which showed evidence of association with scleroderma overall (p-value= 0.01), and also with PH (p-value= 0.007) (Table 3.11).

Blocks	SNPs in the blocks	Alleles in the Haplotype	Freq. in Cases	Freq. in Controls	x^2	P-value	PH
Block 1	rs12514371/ rs3131917	CG	0.1551	0.1617	0.5022	0.4785	–
	rs12514371/ rs3131917	TG	0.3297	0.3478	2.299	0.1294	–
	rs12514371/ rs3131917	TT	0.5152	0.4905	3.814	0.05082	–
Block 2	rs703752/ rs2277923/ rs3095870	AAA	0.3093	0.325	1.921	0.1657	–
	rs703752/ rs2277923/ rs3095870	CGG	0.3053	0.3127	0.4361	0.509	–
	rs703752/ rs2277923/ rs3095870	CAG	0.3854	0.3623	3.859	0.04949	–
	All SNPs	GTTCAG	0.0136	0.007674	5.651	0.0174	0.0072

Table 3.11 Haplotype analysis in the replication cohort. *Two haplotype blocks were formed in Haploview, and tested for association with scleroderma and PH. The 6 tagging SNPs were forced together to form a haplotype (bottom row). Statistically significant p-values (<0.05) are highlighted.*

3.2.4 Meta-analysis

In an attempt to obtain higher statistical power and more robust results, I performed a meta-analysis of the two cohorts. SNP rs3131917 showed evidence of association with scleroderma (p-value= 0.029, OR= 1.089) in the two independent cohorts of Caucasian origin (Table 3.12).

Heterogeneity in a meta-analysis can cause poor results and is often the reason why most associations fail to replicate in independent cohorts. The heterogeneity can be assessed by the Cochran's Q statistic test and the I^2 statistics. I^2 statistics is used to identify the effect of heterogeneity measuring the proportion of total heterogeneity [I^2 = 100% x (Q-df)/Q]. In this study, the p-value for the Cochrane's Q

statistic test >0.1 and I^2 <50%, confirming the absence of heterogeneity in the meta-analysis.

SNP	Minor allele	Major allele	No of studies	P-value	OR	Q	I
rs3132139	A	G	2	0.1164	1.0648	0.1808	44.17
rs12514371	G	A	2	0.4669	0.963	0.7681	0
rs3131917	T	G	2	0.02902	1.089	0.6997	0
rs703752	A	C	2	0.2609	0.9553	0.5917	0
rs2277923	C	T	2	0.5498	0.9752	0.8071	0
rs3095870	T	C	2	0.1162	0.937	0.833	0

Table 3.12 Meta-analysis. *The minor and major alleles are given for each SNP, and the number of individual cohorts, where the meta-analysis is performed (n=2). The p-values and the ORs were calculated based on the allelic test. The p-value (Q) for the Cochran's Q statistic test shows the heterogeneity among the populations. Statistically significant p-values (<0.05) are highlighted.*

3.3 Summary of results

Six tagging SNPs across the *NKX2-5* genomic locus were genotyped in a discovery cohort of scleroderma patients and healthy individuals of UK/Caucasian origin. The tagging SNPs were in strong LD with each other. SNP rs3132139 showed a significant association with PH that was specific to the phenotype and independent of the presence of scleroderma. The association was efficiently replicated in the replication cohort of Spanish/Caucasian origin. The associated SNP is located at the 5' end, downstream of *NKX2-5*. Based on the *in silico* analysis, the region and specifically the SNP might retain functional properties that will be further explored.

While no SNP was associated with scleroderma in the discovery cohort, some positive findings emerged from the replication cohort. In particular, SNPs rs3131917 and rs12514371 were found to be associated with PH in the sub-phenotype analysis. The two SNPs are located near rs3132139 and the positive associations add evidence for the functionality of this genomic region.

A meta-analysis was performed across the discovery and replication cohorts, and SNP rs3131917 showed an overall associated with scleroderma.

In the sub-groups of different auto-antibodies, some positive associations were found in the ACA and ARA sub-groups. In particular, rs3131917 was associated with ACA in the discovery cohort. Unfortunately, no clinical data regarding the presence of ARA antibodies were available for the replication cohort, and the association could not be replicated. In the replication cohort, SNPs rs3131917 and rs3132139 were associated with the ACA sub-group.

The LD followed a similar pattern in both cohorts with similar r^2 values. However, different LD blocks were formed in each cohort. The haplotypes were tested for association with scleroderma and the sub-phenotypes, but no association was found.

3.4 Discussion

The results presented here suggest that the *NKX2-5* gene is a strong candidate gene for the pathogenesis of PH and scleroderma. As previously mentioned, mutations in *NKX2-5* gene cause CHD, and have been extensively studied. Apart from CHD-causing mutations, *NKX2-5* has not previously been genetically associated with any other disease. Following up previous work conducted in the lab, which showed that NKX2-5 is a critical regulator of vascular remodelling in PAH and atherosclerosis (330), I aimed to investigate whether *NKX2-5* is genetically linked with vascular disease.

I designed a candidate gene case-control association study in two independent cohorts of scleroderma patients. Scleroderma is a complex disease with extensive vasculopathy and pulmonary complications. For the purposes of this study, scleroderma patients were categorised in groups based on their clinical findings arising from the RHC, the high resolution computed tomography, and the presence of scleroderma-related auto-antibodies. Patients with a mPAP >25mmHg on RHC in the presence of PF with normal pulmonary arterial wedge pressure <15 mmHg were categorised as PH. Based on the updated clinical classification of pulmonary hypertension, the first group includes connective-tissue associated PAH (180). The pathological features of this category include medial hypertrophy, intimal proliferative and fibrotic changes, adventitial thickening with mild to moderate perivascular inflammatory infiltrates and lymphoid neogenesis, complex lesions (plexiform, dilated lesions) and thrombotic lesions, all of which are inextricably linked to vascular remodelling.

SNP rs3132139 downstream of *NKX2-5* was significantly associated with PH, and the association was efficiently replicated. Further analysis revealed that the association was independent of the presence of scleroderma. This finding supported my initial hypothesis and confirmed that *NKX2-5* genomic locus is associated with PH. Interestingly, the association was not detected with the PAH or PF sub-groups, suggesting that *NKX2-5* is associated with a complex pathological background that involves vascular remodelling and fibrosis. The same SNP, rs3132139, showed a significant association (p=0.006) in a meta-analysis of 46

genome-wide association studies (GWAS) in coronary artery disease (CAD) patients [Data on coronary artery disease/myocardial infarction have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.org (422)]. CAD is an arterial disease, described by plaque formation and constrictive vascular remodelling in the carotid artery, as well as endothelial dysfunction and inflammation (423). This finding further supports our hypothesis for the functional involvement of *NKX2-5* in vascular remodelling.

To date most of our knowledge of the genetics of PH comes from studies on PAH largely focusing on *BMPR2* mutations. More than 300 independent BMPR2 mutations have been described accounting for >75% of familial PAH and 10-25% of sporadic cases of the disease (261). Additional mutations in other genes of the TGF-β pathway have been identified including *SMAD9*, *ACVRL1*, and *ENG* in patients with PAH associated with hereditary haemorrhagic telangiectasia (HHT) (424). Whole-exome sequencing in PAH patients without mutations in *BMPR2* and other TGF-β pathway members, led to the discovery of novel genetic causes including *CAV1* and *KCNK3* (425). *NKX2-5* is a new gene shown to be associated with PH that has already an established role in PAH and vascular remodelling.

In future, it would be interesting to perform a similar genetic study focusing on *NKX2-5* in a cohort of IPAH patients without BMPR2 mutations or other AID. However, getting enough samples to obtain statistical power in a case-control study is challenging. At the moment, a big cohort of DNA samples of PAH patients is being collected in Cambridge under the supervision of Professor Nick Morrell and in collaboration with many scientists and doctors across the country, including the Centre for Rheumatology at Royal Free NHS Foundation Trust Hospital. These DNA samples will be subjected to whole-genome sequencing, and if the data become widely available to scientific community many questions regarding the genetic background of IPAH would be answered.

Due to the strong autoimmune component of scleroderma and the high imbalance of the immune system, the HLA/MHC genes have been prominently associated with scleroderma in the majority of GWAS (290, 292, 296). Outside HLA region, positive associations have been found in genes involved in the pathogenesis of the innate and adaptive immune systems such as type I interferon, TNFα and IL12 pathways, loci associated with T-cell and B-cell immunity, autophagy and inflammation (288, 426). Published data regarding the roles of non-conventional genes in scleroderma are limited. *DNASE1L3, ATG5, PPARγ*, and *CTGF* are few non-HLA and non-

immunity-related genes that have been associated with scleroderma (296, 308, 427), with *CTGF* and *PPARγ* shown to be directly involved in fibrosis and the pathogenesis of scleroderma.

In this study, no positive association was found between any of the SNPs and scleroderma in the discovery cohort. In the replication cohort, however, rs3131917 and rs3132139 were found to be associated with the disease. Case-control association studies can occasionally lack statistical power to detect genetic associations due to small sample size, and sometimes a meta-analysis is performed to identify true associations among different studies, as this often has higher statistical power to reveal true associations. The meta-analysis performed in this study showed that rs3131917 is associated with scleroderma across two independent cohorts of the same ethnic origin.

Several auto-antibodies are present in scleroderma that can be used for diagnosis, sub-grouping of patients and predicting the likelihood of organ complications. For example, the majority of patients with ACA antibodies have LcSSc, and almost 20% of them will develop PAH (428). Also, the frequency of the antibodies can vary among different ethnic populations. For example, although 30% of white scleroderma patients are positive for ACA, the frequency is lower in African-American and Thai scleroderma patients (295). In this study, the percentages of scleroderma patients positive for the ATA antibody were similar (~22%) in the two cohorts. However, the percentage of the ACA-positive scleroderma patients in the replication cohort (46.3%) was much higher compared to the discovery cohort (~36%). Interestingly, a similar increase was also seen in the percentage of SSc-PH; 22.4% in the replication cohort compared to ~6% in the discovery cohort.

Another interesting finding was the marginal association found between rs3095870 and scleroderma in the replication cohort. The same SNP, which is located upstream of *NKX2-5* transcription start site, was associated with SLE in a previous genetic study (331). In particular, a case-control study was performed in two independent Japanese cohorts and showed a functional interaction between rs3095870, upstream of *NKX2-5,* and rs3748079 at *ITPR3* promoter, a locus where NKX2-5 binds. These data provide evidence that rs3095870 might be a functional SNP in SLE and scleroderma, which needs to be explored further. Scleroderma and SLE are both complex AIDs that share some common characteristics such as the increased production of auto-antibodies and the vascular and pulmonary complications. Few years ago, a study reported that the two diseases also share common genetic background including *KIAA0319L, PXK,* and *JAZF1* genes (292).

Taking the above into consideration, as well as the *in silico* analysis, rs3095870 can be considered as a potential regulatory locus for NKX2-5 expression, and more studies will be performed to explore this.

Allele and genotype frequencies and counts were examined in detail and compared between the 2 cohorts, as well as against the reference frequencies obtained from the 100000 Genomes Project. The frequencies were similar between the 2 cohorts and complied with the reference frequencies. The statistical data of the sub-groups were also similar with few changes observed in the ACA group. In addition, LD followed the same pattern with similar r^2 values, however, a different pattern of the LD blocks was observed.

Although the 2 independent cohorts were highly related, differences were seen in the association studies, with more positive findings seen in the replication cohort compared to the discovery cohort. A possible explanation can be that the replication cohort is larger, and therefore it has more statistical power to detect small effects. Another possibility is that the clinical data and classification criteria varied among the different Rheumatology clinics, and consequently the cohorts are more heterogeneous compared to the discovery cohort. In the case of the replication cohort, scleroderma patients were recruited in this study at different hospitals and health centres across Spain, a process that might lack consistency and also affect accuracy.

Scleroderma is a very heterogeneous disease and that can also affect the efficiency of the replication studies. Especially, in the case of the disease-related sub-phenotypes, due to the fact that clinical characteristics of scleroderma patients can change over time during the disease progression. In fact, in many Rheumatology centres, scleroderma patients undergo clinical evaluation every year. In this study as in many other studies, the clinical characteristics of the patients refer to a single time-point that coincides with the study set-up. Given this fact, the results of the genetic study might differ at a later time-point.

To conclude, this is the first time that *NKX2-5*, a gene outside HLA and immunity-related loci, was shown to be associated with scleroderma in a meta-analysis across two independent cohorts of the same ethnic origin. In addition, an association was found between rs3132139, downstream of *NKX2-5* gene, and PH, which was replicated in an independent cohort. The association was independent of the presence of scleroderma. The data propose that *NKX2-5* locus is genetically associated with PH, and further support previous findings showing a critical role of

NKX2-5 in vascular remodelling. The potential functional effect of the associated SNPs on the regulation of NKX2-5 were further explored and the data are presented in the following chapters.

CHAPTER 4 - RESULTS: EXPRESSION OF NKX2-5 IN VASCULAR SMOOTH MUSCLE CELLS

4.1 Introduction

PH is characterised by an increase in peripheral pulmonary vascular resistance and vascular remodelling. Vascular remodelling describes a range of structural and functional changes in the vessel wall that involve cell-cell and cell-matrix interactions. Therefore, different cell types are expected to be implicated in this procedure. Although endothelial dysfunction is lately considered to initiate the disease through the loss of barrier integrity, PASMCs are the important effectors in vascular remodelling. Despite the fact that their precise role is not yet fully understood, published data suggest that cell growth, proliferation, migration and resistance to apoptosis are key characteristics of PASMCs in vascular remodelling [reviewed in (30, 91, 170)].

PASMCs are phenotypically and functionally heterogeneous cells and this has been highlighted in the pulmonary vascular wall. This heterogeneity may be due to developmental factors, such as the embryonic origin of the cells, as well as environmental factors such as increased serum levels, stress etc. The de-differentiation of PASMCs is an example of their plasticity and has been demonstrated in many studies (91, 170, 171). It describes the phenotypic transition of a differentiated contractile cell to a less differentiated embryonic-like synthetic state following injury and stimulation. The two phenotypes are not mutually exclusive, and there is evidence that a synthetic cell can re-acquire the contractile phenotype once injury resolves and the local environment is normalised (30, 33).

Phenotypic differentiation can be monitored by gene and protein markers associated with each phenotype. Although, SMC markers can be selective for SMC in adult vessels, they are also expressed by other cell types such as fibroblasts and myofibroblasts. As a matter of fact, the levels of expression rather than the presence or absence of a protein can better distinguish the different SMC phenotypes. Contractile markers such as SM22, SM-MHC, desmin, calponin, α-SMA, and others are expressed at lower levels in synthetic SMCs, which also produce increased levels of ECM. Therefore, increased expression of collagens, CTGF, MMPs, as well as transcription factors that are abundant in matrix such as KLF4, can confirm the synthetic state.

Heterogeneity has also been observed in the cellular and molecular pathways that regulate vascular remodelling. Many studies have focused on investigating the response of PASMCs to different stimuli such as the secretion of pro-inflammatory molecules and cytokines, other physiological factors and mechanisms, and signalling cascades. The studies show that numerous different pathways can induce similar downstream cellular changes

We have previously shown that NKX2-5 is a major regulator of vascular remodelling, with high levels of expression in synthetic VSMCs (330). When NKX2-5 is knocked down in synthetic VSMCs, the cells revert to a contractile phenotype. Since NKX2-5 is the main focus of this thesis, the aim of this chapter is to address two main areas: a) the gene and protein expression analysis of NKX2-5 in HPASMCs and other VSMC that were used in the experimental work and, b) the cellular and molecular mechanisms that activate NKX2-5 expression in HPASMCs.

Vascular injury, stress and hypoxia result in both immediate and long-term responses mediated by signalling pathways including TGF-β, FGF2, PDGF, Wnt, and the MAPK cascade. Although these pathways are well studied in disease, the trigger that activates NKX2-5 in diseased vessels remains unknown. To explore this, primary and immortalised HPASMCs were stimulated or treated with proteins, cytokines, agonists and antagonists involved in different signalling pathways. The TGF-β superfamily is a key regulator of vascular remodelling (1.2.3.3, Figure 1.4). TGF-β1 signals through a hetero-tetrameric receptor complex composed of two type I and two type II transmembrane receptor subunits, and initiates the downstream signalling through the SMADs, leading to increased ECM production. BMP2 and BMP4 are members of the same superfamily that bind the BMPR2 receptor to initiate downstream signalling. SD208 is a potent inhibitor of TGF-βR1/ALK5 receptor, which recruits and phosphorylates SMAD2/3 and blocks downstream TGF-β signalling.

Apart from the TGF-β superfamily, another focus of this study is also ET-1 (1.1.5.5). ET-1 is a potent mitogen for PASMCs through the activation of phosphorylated-ERK1/2, c-jun and c-fos (429, 430), and it is able to modulate ECM remodelling by stimulating fibroblast-induced collagen synthesis (431). Bosentan is a competitive antagonist of both ETA and ETB receptors and it has been used as a treatment for PAH. More recently, tyrosine kinase receptors PDGFR and FGFR2 have been proposed to be responsible for the increased PASMC growth in PAH (1.2.3.3). FGF2 is mainly produced by endothelial cells and has recently been proposed to contribute to PH through the regulation of EndoMT and TGF-β (114, 132). Imatinib

is another selective inhibitor used to suppress PDGF signalling by inhibiting its receptor PDGFRβ. Furthermore, cells exposed to hypoxia and treated with HIF-1α inhibitor demonstrated a well-established role of hypoxia in driving VSMC proliferation (1.2.5.3.3). The effect of CX4945, a CK2 inhibitor that blocks NKX2-5 phosphorylation and therefore its nuclear translocation and activation, was also evaluated (1.4.5.3). Last but not least, selective inhibitors for members of the MAPK signalling cascade were used to identify whether NKX2-5 is a direct downstream target for any of the kinases.

4.2 Results

4.2.1 NKX2-5 gene and protein analysis

Analysis of NKX2-5 protein and gene expression is complicated by the numerous post-transcriptional and post-translational modifications. The gene spans a 3.2Kb genomic region, and it contains 2 exons that can be transcribed in 4 protein-coding transcripts (Figure 1.8). The longest transcript (1709bp, ENST00000329198.4) is translated to a 324 amino acid protein (CCDS4387, P52952) of 34.9KDa weight.

There are at least 8 commercial polyclonal and monoclonal antibodies available against the human NKX2-5 protein. The specificity of the commercial antibodies is tested in various cells lines (CCRF-CEM, 293T) and those antibodies usually identify an over-expressed protein in transfected cells. Most of the antibodies recognise multiple bands or a triplet between sizes of 30-50KDa.

Since NKX2-5 is a transcription factor, it can only exert its function in the nucleus, where it is able to bind NKE consensus binding elements found in promoters, enhancers or other regulatory regions of downstream target genes and regulate their transcription. Therefore, I have focused on the analysis of the nuclear protein, and the data presented in this thesis refer to the nuclear protein, unless otherwise stated.

To facilitate my work, I isolated cytoplasmic and nuclear protein extracts from primary HPASMCs and used 2 different antibodies to detect the human NKX2-5 protein; a monoclonal Abcam (ab54567) and a polyclonal Santa-Cruz antibody (sc-14033). The distinct cytoplasmic and nuclear fractions were confirmed by the use of antibodies specific to the housekeeping proteins: GAPDH and β-tubulin for the cytoplasmic fraction, and TBP for the nuclear fraction. In the cytoplasmic fraction, the Santa-Cruz polyclonal NKX2-5 antibody detected a triple band around 35KDa (Figure 4.1). In the nuclear fraction, only a single band was detected by the Abcam

monoclonal antibody. The band was higher than the cytoplasmic protein, seen at ~55KDa (Figure 4.1).

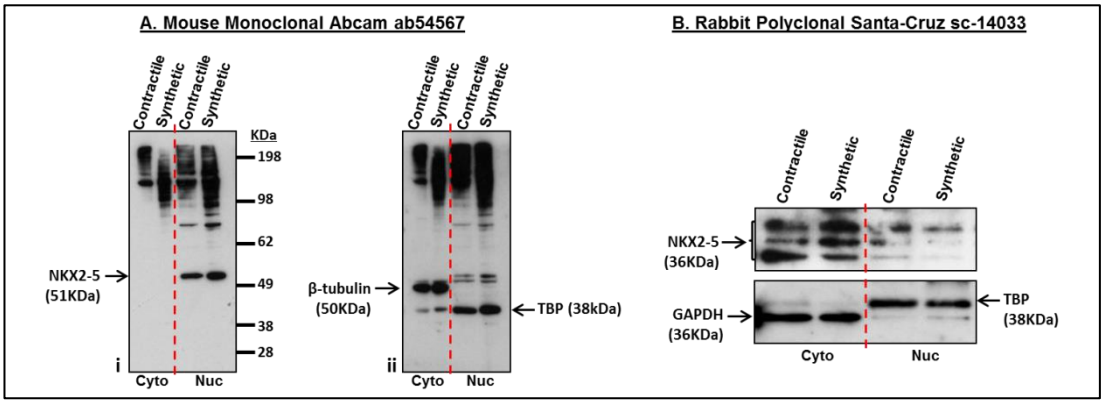


Figure 4.1 Analysis of NKX2-5 human protein. *Cytoplasmic (cyto) and nuclear (nuc) extracts were prepared from contractile and synthetic primary HPASMC and analysed by SDS-PAGE and Western blot to determine the specificity of a mouse monoclonal (left) and a rabbit polyclonal (right) antibodies specific to NKX2-5 protein. Two housekeeping proteins of different sizes were used to determine the effective separation of nuclear and cytoplasmic proteins, and also determine the approximate size of the human NKX2-5 protein: the cytoplasmic β-tubulin at 51KDa and the nuclear TATA-binding protein (TBP) at 38KDa. The Abcam antibody (ab54567) recognises a single nuclear band at approximately 52KDa. The Santa-Cruz antibody (sc-14033) identifies a triplet at around 36KDa in the cytoplasmic fraction. Blot i immunodecorated with NKX2-5 was washed, stripped and re-blotted with β-tubulin and TBP (ii).The same lysates were used to test both antibodies.*

There are at least 5 known sites at which the NKX2-5 protein is modified post-translationally including phosphorylation (Ser164, Thr180, Ser78), di-methylation (Arg225) and sumoylation (K51) (Source: http://www.phosphosite.org). It has been shown that NKX2-5 protein gets phosphorylated at 2 different sites by CK2, promoting its nuclear localisation (364). However, the phosphorylation alone is not enough to explain the differences in the molecular weight of the nuclear protein compared to the cytoplasmic, since each phosphate group is approximately 80Da. In addition, Costa *et al* has shown that transcriptionally-activated NKX2-5 is sumoylated in the nucleus at 2 different sites including K51 (368). SUMO is an 11KDa polypeptide structurally related to ubiquitin that can be covalently conjugated to lysine residues within target proteins at a consensus site.

Taking these into consideration, I concluded that the nuclear protein is phosphorylated and sumoylated and the total extra weight should be roughly 22-23KDa. However, it is possible that additional modifications specific to HPASMCs could occur. Therefore, the 51KDa nuclear protein band detected by the Abcam monoclonal antibody was analysed for all the sub-sequent protein analysis.

The specificity of the results was confirmed by work done by other members of the lab using a NKX2-5 expression vector (unpublished data).

The analysis of gene expression was complicated for similar reasons. Due to the highly repetitive and GC-rich genomic locus, primer design was challenging. In addition, the low levels of *NKX2-5* expression and the absence of expression in normal vessels, made amplification difficult.

4.2.2 Expression of NKX2-5 in HPASMC and other VSMC

4.2.2.1 Primary HPASMC

Primary HPASMC lines (Promocell, Cat. no: C-12521) originating from 6 different human subjects were used in these studies. The company provided information regarding the age and gender, as well as the basic medical history of the subjects. The cells were cultured *in vitro* under conditions that favoured either the contractile or the synthetic state as explained in detail in the section 2.7.

Morphology

The cells were cultured in 5% FCS and photographed at an early passage (passage number 2 or 3). As shown in the pictures below, the elongated spindle-shaped morphology is retained among the different primary cells with some differences in cell size. The size of SMCs can range between 20-500μm, with the largest cells found in the uterus of pregnant women (432). The population doubling time also differed among the primary lines from 30-50 hours, as calculated during culture. In addition, the cells responded differently to increased amount of serum in the synthetic state. The effect ranged from some cells increasing their doubling time immediately, while others took longer to adjust to the new environmental conditions.

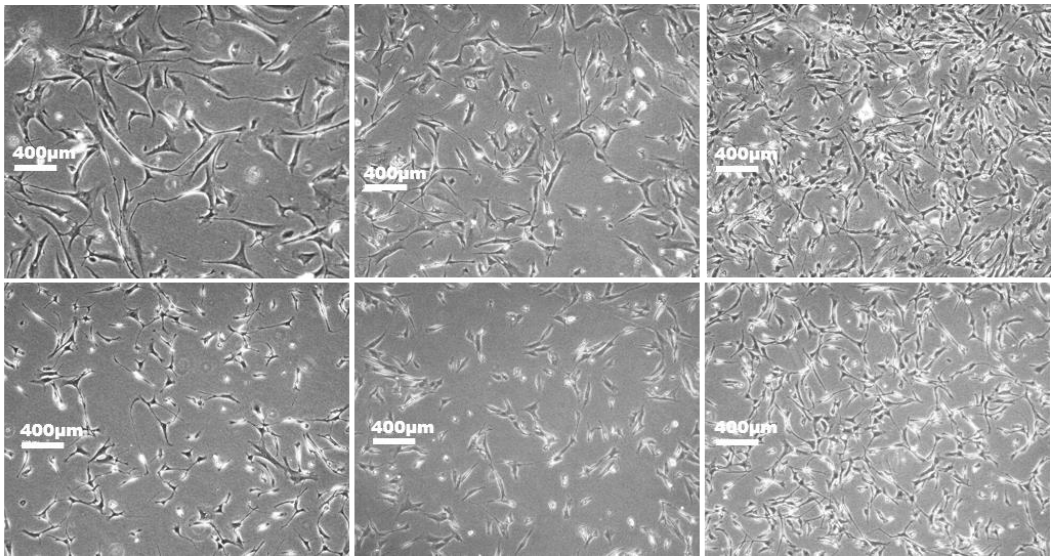


Figure 4.2 Morphology of primary HPASMC. *Primary HPASMC from 6 different human subjects (Promocell, Cat. no: c-12521) were cultured in smooth muscle media containing 5% FCS. The cells were photographed at an early passage to compare their morphology. All the cells present a spindle like shape, but they differ in size, as well as in the doubling time. A bar scale of 400µm is shown for comparison among the cells.*

Expression profile

The expression levels of contractile and synthetic markers during SMC de-differentiation has been extensively studied (1.2.3.2). Increased NKX2-5 expression has been associated with the synthetic state (330). Indeed, protein and gene expression levels were higher in the synthetic compared to contractile HPASMC (Figure 4.3 A, B), but the difference was significant only at the gene level.

The cells were treated with TGF-β and the levels of NKX2-5 were analysed. Since TGF-β has been shown to increase synthesis of fibronectin, collagen, and PAI-1 that are all markers of the synthetic state, NKX2-5 is also expected to increase upon stimulation. Indeed, NKX2-5 expression was induced in response to TGF-β and the effect was greater in contractile compared to synthetic HPASMCs (Figure 4.3 A, B). In the synthetic state, NKX2-5 appears to be activated at maximal levels and therefore the response to TGF-β is not as significant.

To further characterise the primary HPASMCs in both contractile and synthetic phenotypes, I analysed the expression profile in response to a variety of growth factors and cytokines. Contractile and synthetic HPASMCs were cultured accordingly, and after an overnight serum starvation in 0.1% FCS the cells were stimulated with 2ng/ml TGF-β, 50ng/ml FGF2, 100nM ET-1 and 10ng/ml BMP2 for 24 hours. I analysed the protein and gene expression of *NKX2-5* along with the

gene expression of *COL1A2* and *CTGF* as markers of the synthetic phenotype and fibrosis. At the protein level, NKX2-5 was expressed more in synthetic compared to contractile cells (Figure 4.4) and TGF-β stimulated expression similarly as before (Figure 4.3).

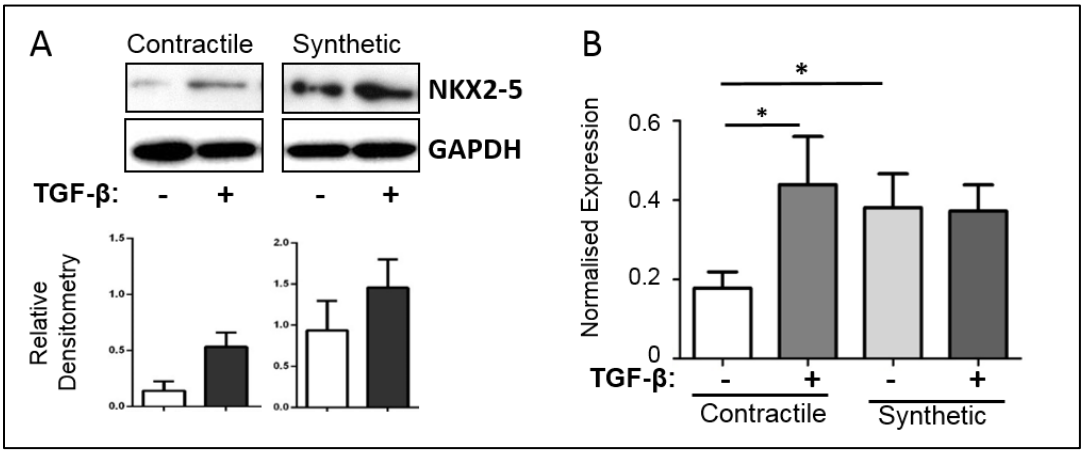


Figure 4.3 NKX2-5 expression in primary HPASMC. *Primary HPASMC were cultured under conditions favouring the contractile or the synthetic state. The cells were serum-starved overnight in 0.1% FCS and then treated with 2ng/ml TGF-β for 24 hours. The cells were lysed and nuclear protein and total RNA were extracted and subjected to SDS-PAGE/Western blot and qPCR analysis respectively. A. NKX2-5 nuclear protein from contractile and synthetic cells stimulated with TGF-β. Protein expression was analysed with densitometry relative to expression of GAPDH (housekeeping protein) in ImageJ, and the results are shown in the bar plot exactly below the western blot, with the bars corresponding to the bands of the blot. B. NKX2-5 gene expression was analysed with qPCR. N=3 (replicate experiments). *p≤0.05.*

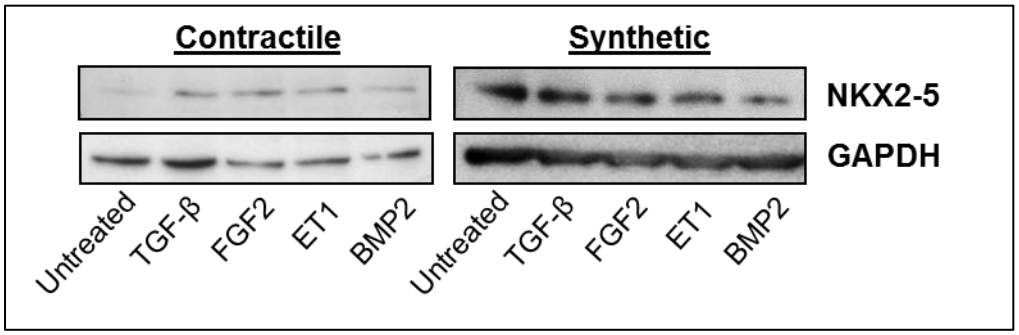


Figure 4.4 NKX2-5 protein expression is response to different stimuli. *Primary HPASMC were cultured under conditions favouring the contractile or the synthetic state. The cells were serum-starved overnight in 0.1% FCS and then treated with the following growth factors for 24 hours: 2ng/ml TGF-β, 50ng/ml FGF2, 100nM ET-1, and 10ng/ml BMP2. After stimulation, nuclear protein was extracted and subjected to SDS-PAGE/Western blot analysis. NKX2-5 protein expression was analysed with densitometry relative to expression of GAPDH (housekeeping protein) in ImageJ. N=3 (replicate experiments).*

As shown before, TGF-β stimulation increased *NKX2-5* gene expression only in the contractile cells, but none of the other mediators affected *NKX2-5* in either state (Figure 4.5). Endogenous gene expression of *COL1A2* and *CTGF* was higher in synthetic than in contractile cells, but upon TGF-β stimulation only *CTGF* was affected in a positive manner. None of the other mediators affected significantly gene expression of *COL1A2* and *CTGF* (Figure 4.5).

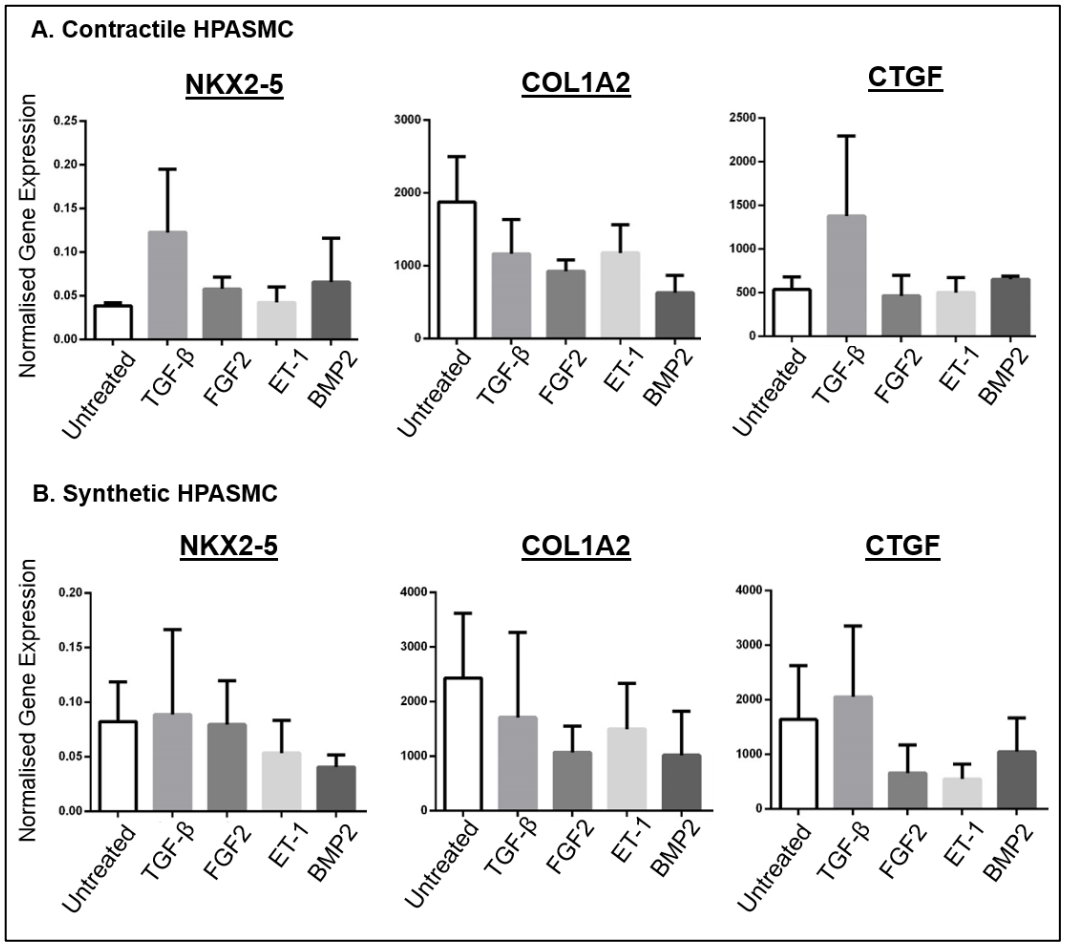


Figure 4.5 Gene expression profile of primary HPASMC. *Primary HPASMC were cultured under conditions favouring the contractile or the synthetic state. The cells were serum-starved overnight in 0.1% FCS and then treated with the following growth factors: 2ng/ml TGF-β, 50ng/ml FGF2, 100nM ET-1, and 10ng/ml BMP2. After 24 hours, total RNA was extracted and subjected to qPCR analysis. Gene expression of NKX2-5, Collagen 1 alpha 2 (COL1A2), and CTGF was analysed against expression of TBP (housekeeping gene) in contractile (A) and in synthetic (B) HPASMC. N=2 (replicate experiments).*

4.2.2.2 Isolated primary HPASMC and VSMC from PAH/PAD/Diabetic patients

VSMCs and lung fibroblasts were isolated from different subjects: a) diseased lung tissue of PAH patients (HPASMCs and fibroblasts) after lung transplantation, b) diseased vessels of diabetic and PAD patients after limb amputation (VSMCs), and c) healthy controls.

Morphology

Isolated HPASMCs exhibited the usual spindle-like morphology. Interestingly, HPASMCs that were isolated from the lung tissue of a young PAH patient (4 year old) were significantly smaller with a very rapid population doubling time (Figure 4.6).

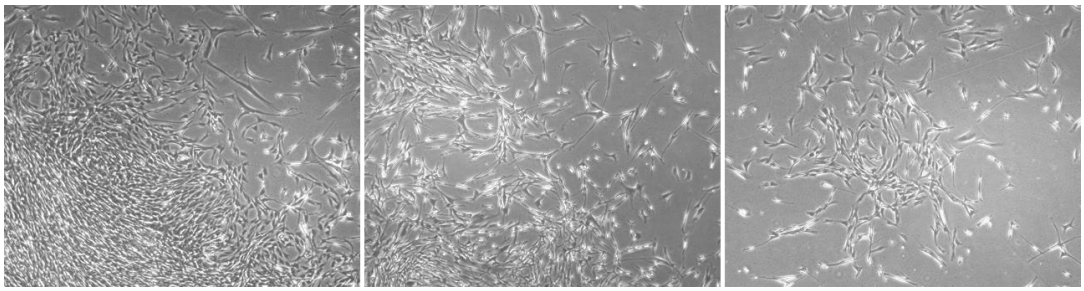


Figure 4.6 Morphology of primary PASMC isolated from a PAH patient. *Primary HPASMC were isolated from a lung transplant of a young PAH patient, and cells were photographed at different time points after isolation. Approximately 2 weeks after the isolation, the cells form small colonies and start to sprout out and proliferate. The isolated cells retain their spindle-like shape, but are relatively small in size compared to cells isolated from an adult. Also, the cells exhibited a very high proliferation rate and short population doubling time.*

Phenotypic characterisation

In order to confirm the efficiency and specificity of my protocol for the isolation of HPASMCs from tissues and vessels, I used immunofluorescence to characterise the phenotype of the isolated cells. VSMCs and adventitial fibroblasts isolated from the pulmonary arteries of PAH and PAD patients and healthy individuals were stained for cell-specific markers such as α-SMA and CD90 (Thy1). α-SMA is a contractile marker of VSMCs, but not specific to VSMC only, since expression has been found in other cell types, such as myofibroblasts. However, the expression of α-SMA is commercially used to identify VSMCs and to determine the cell phenotypic state. CD90 is a surface antigen expressed on fibroblasts and together with vimentin, they both are exclusively used as specific markers for the staining of fibroblasts. Although the specificity of both α-SMA and CD90 is arguable, the two

markers are widely used commercially for the phenotypic characterisation and discrimination among primary cells.

Commercially bought primary HPASMCs that were cultured *in vitro* under the synthetic conditions were used as a control. α-SMA was ubiquitously expressed in all 3 cell types, exhibiting higher levels of expression in PAH PASMCs (Figure 4.7). PAH fibroblasts also expressed very high levels of α-SMA. However, expression of CD90 was specific to fibroblasts, with very little levels of expression in VSMC (≤10%). The findings confirmed the successful isolation of VSMCs from human vessels and tissues.

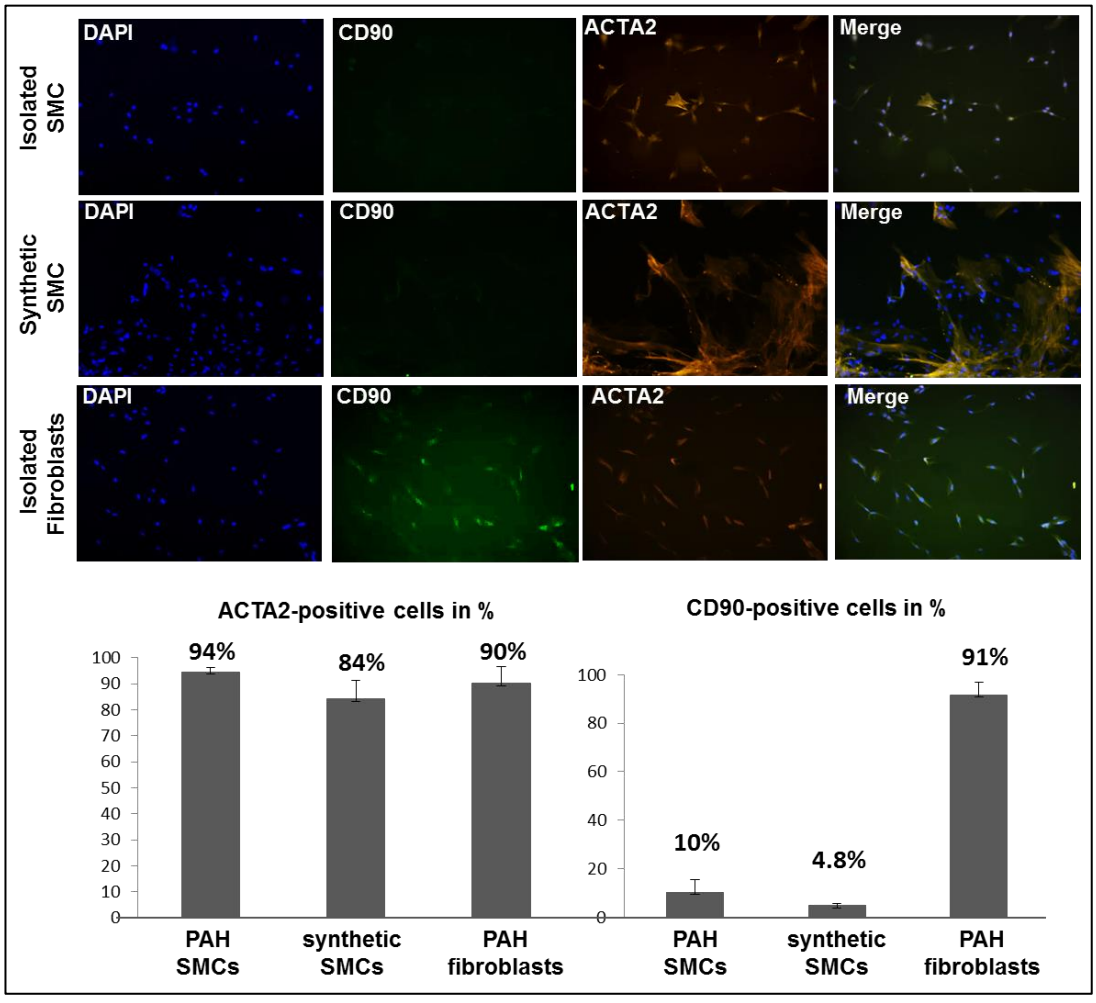


Figure 4.7 Phenotypic characterisation of VSMC and fibroblasts. *Immunofluorescence staining for α-SMA (orange) and CD90 (green) was performed to characterise PASMC and fibroblasts isolated from a single human PAH lung tissue. α-SMA is ubiquitously expressed in PAH and synthetic PASMC as well as in fibroblasts. However, expression of CD90 is limited to fibroblasts. DAPI counterstaining was used to visualise the nuclei (blue). Levels of expression in each cell type are shown as percentages in the bar graphs .The isolated SMCs and fibroblasts were isolated from the same tissue and stained soon after the isolation (P=3). The tissue was obtained during lung transplantation from a 4-year old child with IPAH. Synthetic SMCs commercially bought and cultured under conditions favouring the synthetic phenotype were also stained as a control.*

Expression profile

Protein and gene expression of NKX2-5 in PAH and PAD VSMC was analysed (Figure 4.8). NKX2-5 expression was higher in PAH PASMCs compared to normal PASMCs at both the protein and RNA levels, but the difference was not significant. Inconsistent results were found for NKX2-5 in VSMC from PAD and diabetic patients. The protein expression was increased compared to normal cells, but the qPCR data showed decreased mRNA levels. Increasing the number of patients could provide more information and conclusive data.

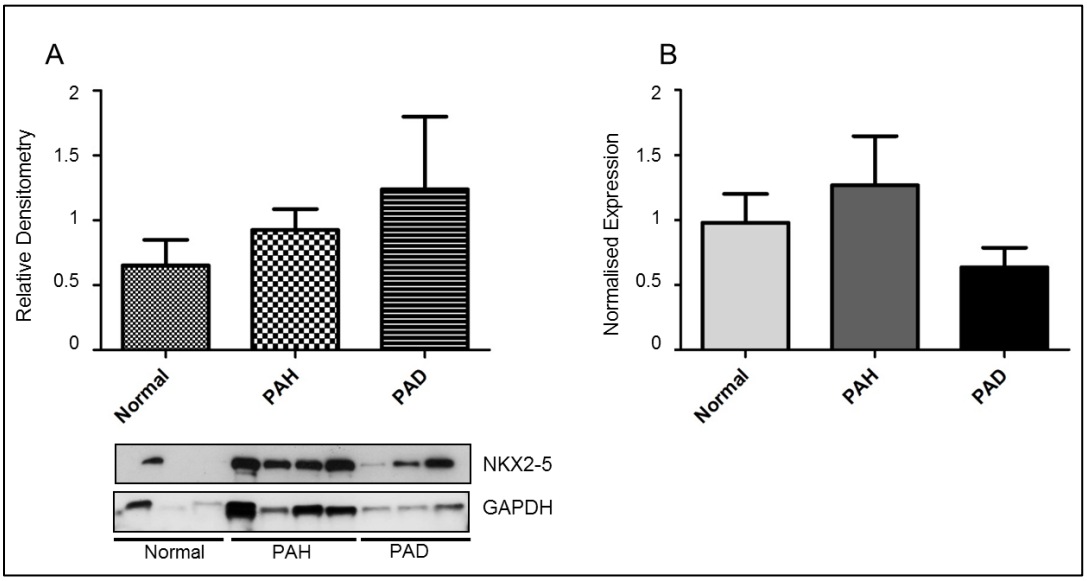


Figure 4.8 NKX2-5 expression in PAH and PAD. *Primary VSMC were isolated from pulmonary arteries of PAH patients (number of patients=4) and healthy donors (number of healthy donors=3), as well as arteries of patients with PAD and diabetes (number of patients=3). Cells were cultured in vitro in 5% FCS. Nuclear protein and total RNA were isolated and were subjected to SDS-PAGE/Western blot and qPCR analysis. A. NKX2-5 protein expression was analysed with densitometry relative to expression of GAPDH (housekeeping gene) in ImageJ (1 experiment performed). B. NKX2-5 gene expression was assessed with qPCR (1 experiment performed).*

4.2.2.3 Immortalised HPASMC (ImHPASMC)

ImHPASMCs were obtained from ABM Good (Cat. no: T0558). The use of these cells facilitated the large-scale experiments conducted in this thesis.

Morphology

ImHPASMCs exhibited similar morphology with the primary cells retaining their spindle-like shape (Figure 4.9). These cells grew faster than primary cells and their growth rate was consistent.

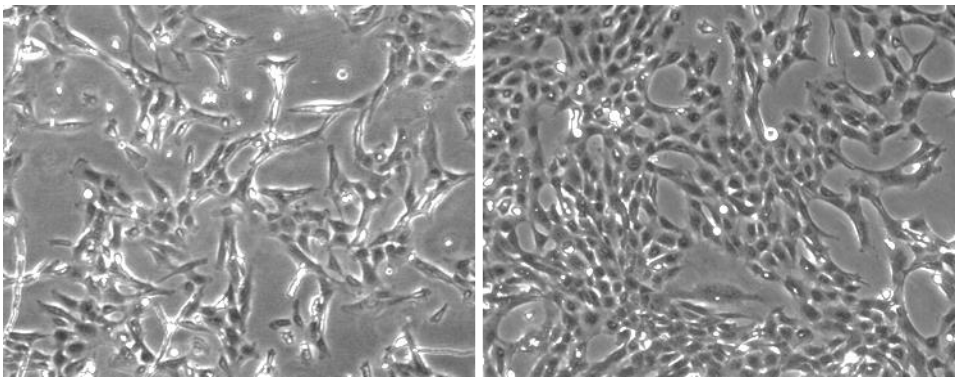


Figure 4.9 Morphology of immortalised HPASMC. *ImHPASMCs were cultured in DMEM supplemented with 5% FCS. The cells were photographed at an early passage to compare their morphology with the primary HPASMC.*

Expression profile

This is the first time that ImHPASMCs have been used in our studies in place of primary cells. Consequently, I aimed to explore the expression profile of the cells and the expression levels of SMC-specific markers in response to different stimuli. ImHPASMCs were serum-starved overnight in 0.1%FCS before stimulation with different mediators for 24 hours. Then, total protein and RNA were isolated in the same way as for primary cells.

Protein levels for NKX2-5, COL1A2, CTGF and the contractile markers SM22, α-SMA and calponin were analysed by Western blotting (Figure 4.10). Increasing amounts of serum (5%-15%) resulted in significant activation of NKX2-5 in a dose-response manner. TGF-β also increased NKX2-5 but did not exhibit as marked an effect as in primary cells. Collagen and CTGF were expressed at low levels, but protein was significantly increased in response to TGF-β. Expression of α-SMA and SM22 was increased upon serum and TGF-β stimulation in a very distinctive way. Expression of calponin was only affected by the serum levels. ET-1, FGF2, BMP2, and BMP4 did not cause any significant differences in protein expression levels. However, a combined stimulation of ImHPASMCs with TGF-β and 10% FCS, significantly increased the protein.

Gene expression for *NKX2-5, COL1A2*, and *CTGF* was analysed by qPCR (Figure 4.11). The pattern of gene expression of *COL1A2* and *CTGF* coincides with protein expression levels. But the gene profile for *NKX2-5* slightly differs from the protein. Gene expression was significantly affected by TGF-β but not serum, suggesting a complex transcriptional regulation.

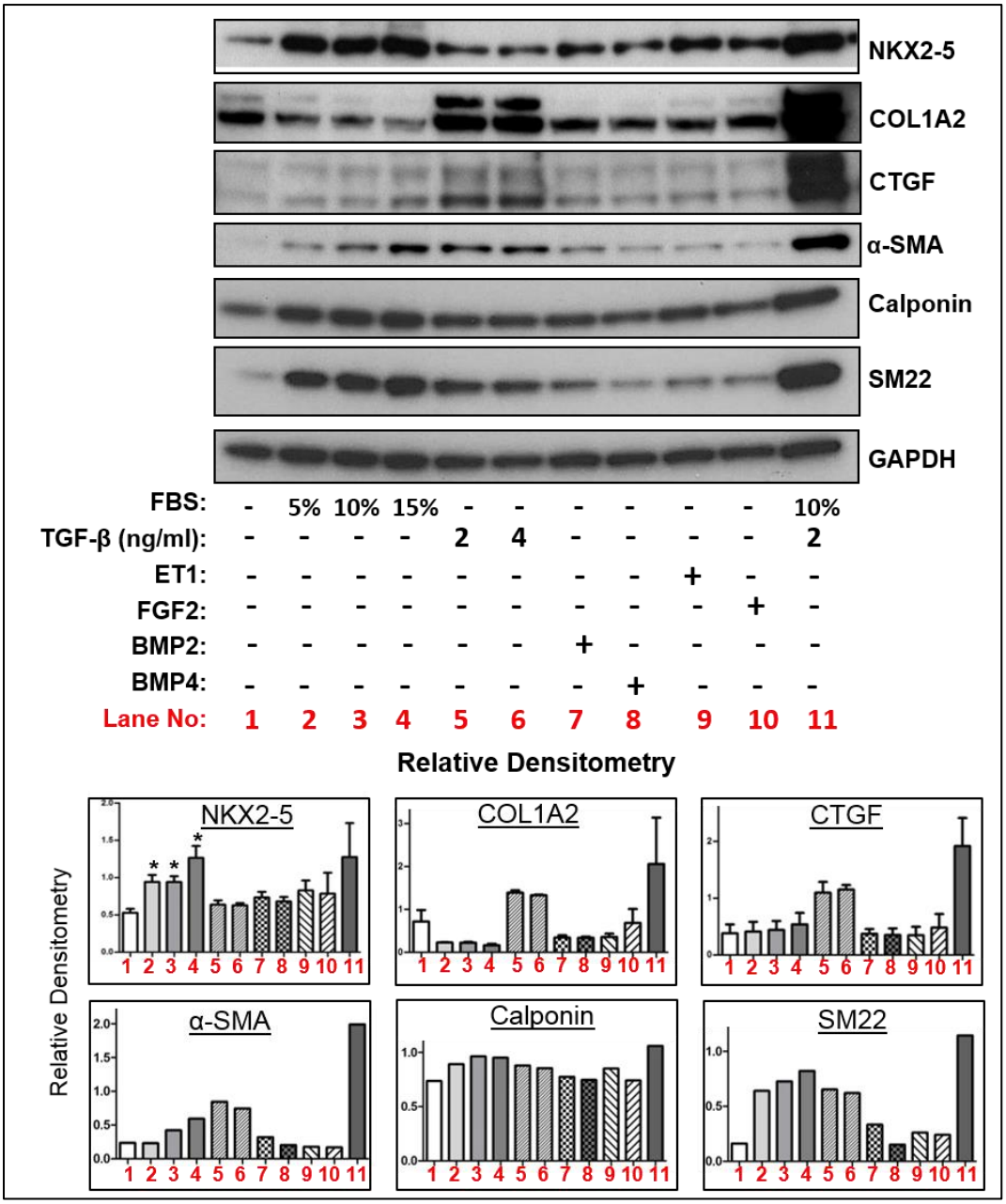


Figure 4.10 Protein expression profile analysis of ImHPASMCs. *ImHPASMCs were serum starved overnight in 0.1% FCS and were then stimulated as follows: 5-15% FCS, 2-4ng/ml TGF-β, 100nM ET-1, 50ng/ml FGF2, 10ng/ml BMP2, and 10ng/ml BMP4. After 24 hours, total protein was extracted and subjected to SDS-PAGE/Western blot analysis. Expression of NKX2-5, collagen 1 alpha 2 (COL1A2), CTGF, α-SMA, Calponin, and transgelin (SM22) was analysed and densitometry relative to GAPDH (housekeeping protein) was assessed in ImageJ. The y axes on the graphs show relative densitometry in arbitrary units. N=3 (replicate experiments). *p≤0.05.*

In addition, the response of gene and protein expression to the higher concentration of TGF-β was an interesting finding. TGF-β (4ng/ml) negatively affected expression of all markers at both protein and gene expression levels (Figures 4.10 and 4.11). High levels of TGF-β are known to cause inhibitory effects on maximally activated cells, probably by activating negative feedback loops that act through the inhibitory partners SMAD6/7 (433).

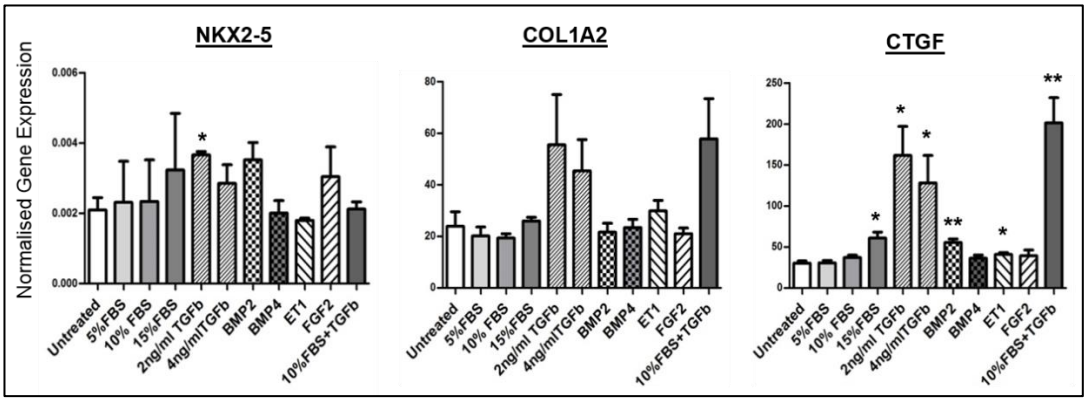


Figure 4.11 Gene expression analysis of ImHPASMCs. *ImHPASMCs were serum starved overnight in 0.1% FCS and then stimulated as follows: 5-15% FCS, 2 or 4ng/ml TGF-β, 100nM ET-1, 50ng/ml FGF2, 10ng/ml BMP2, and 10ng/ml BMP4. After 24 hours, total RNA was extracted and subjected to qPCR analysis. Gene expression of NKX2-5, collagen 1 alpha 2 (COL1A2), and CTGF was analysed. mRNA levels were normalised to TBP expression. N=3 (replicate experiments). *p≤0.05, **p≤0.01.*

Overall, primary and immortalised HPASMCs behave very similarly, although some differences in basal expression of proteins like CTGF and collagen have been observed. Immortalised HPASMC retained a phenotype between the contractile and synthetic state. However, stimulation with TGF-β increased expression of NKX2-5, CTGF and Collagen type I, and increasing amounts of FCS caused dose-dependent increase in expression of NKX2-5 and α-SMA in same way as in primary HPASMC. For the purposes of this study, immortalised and primary HPASMC were treated in similar ways in regards of the presence of FCS in culture media, serum-starvation conditions and TGF-β stimulation.

4.2.3 Cellular and molecular mechanisms that activate NKX2-5 expression in HPASMC

4.2.3.1 Regulation of NKX2-5 through different signalling pathways

In this section, my aim was to dissect the cellular and molecular mechanisms that activate NKX2-5 in HPASMCs. I conducted a thorough literature review to investigate all the mechanisms that are known to be implicated in vascular remodelling and PH. Based on the results, I used a selection of stimuli and inhibitors to treat ImHPASMCs and examine their effect on NKX2-5 expression.

4.2.3.1.1 NKX2-5 is induced by hypoxia

Hypoxia describes a state of lack of oxygen endangering cell function, and it has multiple impacts on the vascular system. Upon exposure to hypoxia, both endothelial cells and VSMC rapidly respond in a manner which involves acute and genomic changes that can lead to either vasodilation or vasoconstriction. To examine whether hypoxia induces NKX2-5 expression, ImHPASMCs were serum-starved overnight in 0.1% FCS, and next day they were treated with 2ng/ml TGF-β for 24 hours. During this incubation period, cells were cultured under normal conditions (normoxia: 21% O_2, 5% $CO_{2,,}$ 37^oC), or exposed to hypoxia. To obtain hypoxic conditions (1% O_2, 5% CO_2, 37^oC), the cells were placed in a custom-made hypoxic chamber (Modular Incubator Chamber, MIC-101; Billups-Rothenberg, Del Mar, CA, USA). The chamber was flushed with gas at 10 l/min for 15 min, then sealed and placed in the incubator for 24 hours.

Hypoxia induced NKX2-5 expression compared to normoxic conditions. TGF-β stimulation caused increased NKX2-5 expression, and the effect was greater in hypoxia (Figure 4.12).

In addition, hypoxia also induced phosphorylated AKT levels in a similar way, while total protein was not significantly altered. Phosphorylated SMAD2 only responded to TGF-β stimulation and remained unaffected by hypoxia (Figure 4.12) The effect of hypoxia on PI3K/AKT on VSMC has been demonstrated previously in many studies and is involved in cell proliferation, migration and apoptosis in vascular remodelling (434, 435). It has also been suggested that activation of P13K and AKT signalling pathways in hypoxia is a mechanism by which cells adapt and survive under conditions of restricted oxygen (436).

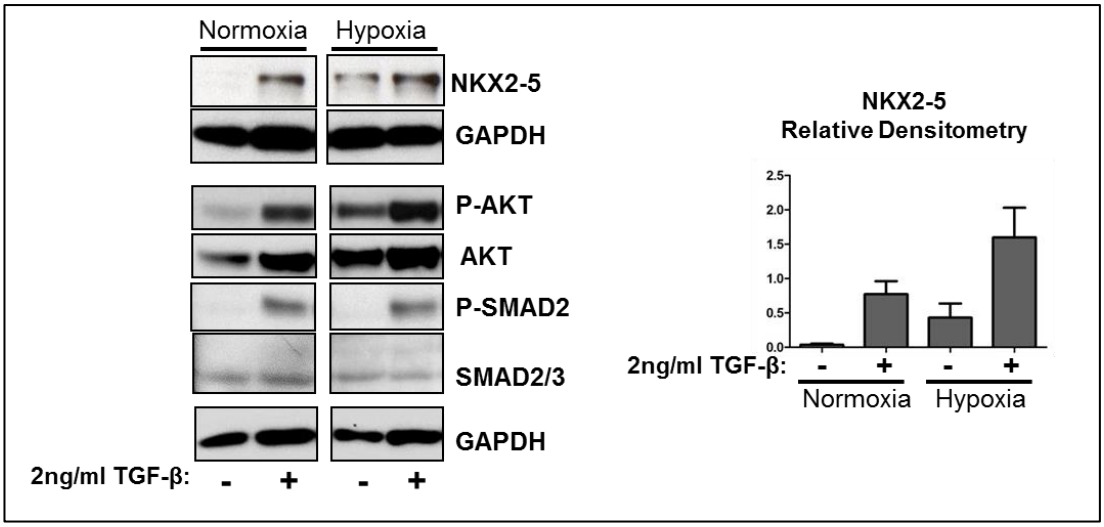


Figure 4.12 Hypoxia activates NKX2-5 expression. *ImHPASMCs were serum starved overnight in 0.1% FCS, and then treated with 2ng/ml TGF-β for 24 hours. During the incubation period, cells were cultured under normal conditions (normoxia: 21% O_2, 5% CO_2, 37^oC) or exposed to hypoxia (1% O_2, 5% CO_2, 37^oC). After the incubation period, cells were lysed and total protein was extracted. A. SDS-PAGE/Western blot for protein levels of NKX2-5, phosphorylated AKT (P-AKT), total AKT, phosphorylated SMAD2 (P-SMAD2) and total SMAD2/3. B. Densitometry analysis of NKX2-5 protein expression relative to GAPDH (housekeeping protein) in ImageJ. N=3 (replicate experiments).*

4.2.3.1.2 Inhibition of TGF-β signalling downregulates NKX2-5 expression

I have already established that TGF-β stimulation activates NKX2-5 expression significantly (Figure 4.3) and it is a critical regulator that promotes HPASMC de-differentiation. Next, I aimed to reverse the effect of stimulation by using the SD208 inhibitor. SD208 is a selective inhibitor of TGF-R1/ALK5receptor developed by Scios, Inc. (Fremont, CA) (Figure 4.13). The compound has been successfully used *in vivo* and *in vitro* to prevent TGF-β–induced ALK5 phosphorylation and subsequent SMAD2 phosphorylation, TGF-β–dependent myofibroblast differentiation, and pulmonary fibrosis (437).

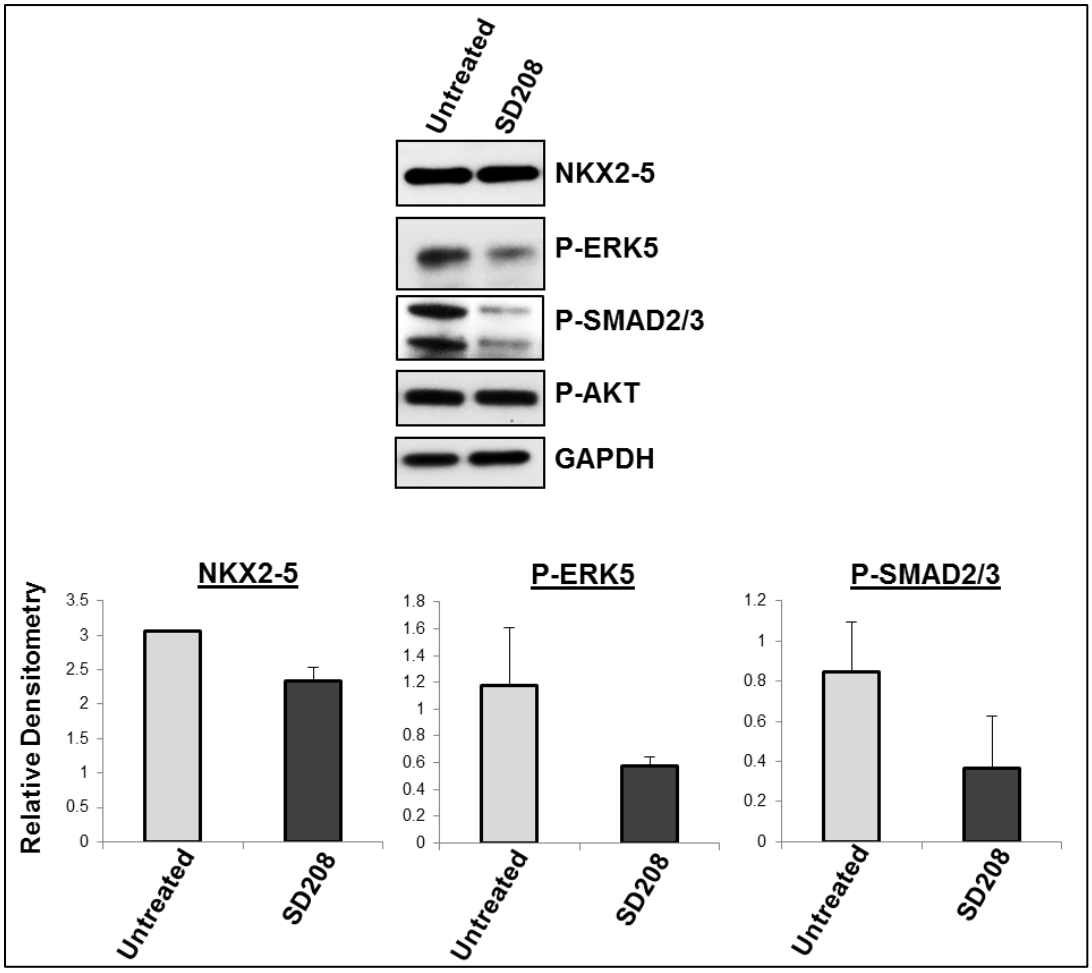


Figure 4.13 Inhibition of TGF-β signalling downregulates NKX2-5. *ImHPASMCs were serum-starved for ~6-7 hours in 0.1% FCS, and then treated with 2ng/ml TGF-β for 24 hours. After 24 hours, the medium was changed to 5% FCS and the cells were treated with 5μM of SD208, an inhibitor of TGF-β signalling pathway, for 24 hours. Cells were lysed, and total protein was extracted and subjected to SDS-PAGE/Western blot analysis. Protein levels of NKX2-5, phosphorylated ERK (P-ERK5), phosphorylated SMAD2/3 (P-SMAD2/3), and phosphorylated AKT (P-AKT) were analysed. Densitometry analysis was conducted in ImageJ relative to GAPDH expression (housekeeping protein). N=3 (replicate experiments).*

ImHPASMCs were serum-starved for ~6-7 hours in 0.1% FCS, and next day they were stimulated with 2ng/ml TGF-β for 24 hours. Following that, the media was changed and the cells were treated with SD208 inhibitor at 5μM final concentration for 24 hours. The concentration was selected based on published data, and levels of phosphorylated-SMAD2/3 proteins were assessed to confirm effective inhibition of the pathway. Indeed, expression of phosphorylated-SMAD2/3 was decreased (Figure 4.13). NKX2-5 protein was also decreased but not significantly, suggesting that other TGF-βRI–independent or TGF-β indirect signalling pathways such as ERK5 might be critical for NKX2-5 activation.

In addition, SD208 treatment decreased phosphorylated ERK5 protein, confirming published data that ERK5 is a mediator of TGF-β1 downstream signalling. Phosphorylated AKT remained unaffected, as expected, since the activation of PI3K/AKT pathway is independent of SMAD2/3 activation (438).

4.2.3.1.3 Signalling pathways that do not affect NKX2-5 expression

The preceding studies showed that TGF-β and hypoxia are potent activators of NKX2-5. Next, a panel of selective inhibitors was used to block a number of pathways that have previously been implicated in PAH and SMC de-differentiation to determine their effect on TGF-β-induced NKX2-5 expression.

5Z-7-Oxoxeaenol (OXO) selectively inhibits TGF-β-activated kinase 1 (TAK1), as well as IL-1-induced activation of NF-κB and JNK/p38 signalling, and the production of inflammatory mediators such as TNFα. BAY 87-2243 is a highly potent and selective inhibitor of hypoxia-induced gene activation that significantly decreases nuclear HIF-1α expression. CX-4945, is a highly specific, ATP-competitive inhibitor of CK2. CX-4945 exerts anti-proliferative effects in haematological tumours by downregulating CK2 expression and suppressing activation of CK2-mediated PI3K/Akt/mTOR signalling pathways. As mentioned earlier, Bosentan is a competitive antagonist of ET-1 receptors and Imatinib (also known as Glivec) blocks the PDGF-Rβ receptor. When the cells were treated with the above inhibitors, the protein expression of NKX2-5 was not affected (Figure 4.14). However, this can only be an indication that NKX2-5 is not regulated through these pathways directly, since only one experiment was performed.

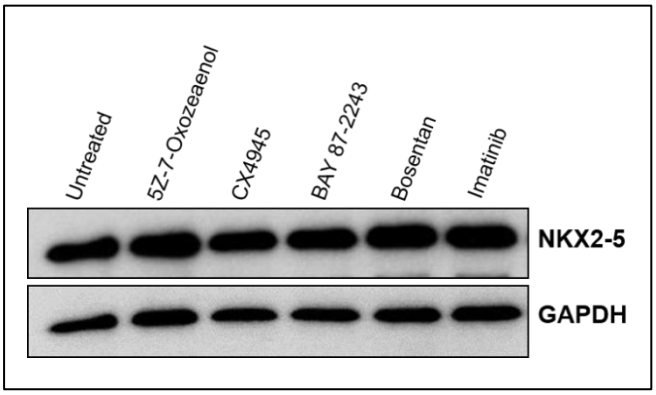


Figure 4.14 Signalling pathways that do not affect NKX2-5 expression. *ImHPASMCs were serum-starved for ~6-7 hours in 0.1% FCS, and then treated with 2ng/ml TGF-β for 24 hours. After 24 hours, the media was changed to 5% FCS and the cells were treated with the following inhibitors for 24 hours: 100nM 5Z-7-Oxozeaenol (inhibitor of TAK1 kinase), 5μM CX4945 (inhibitor of CK2 kinase), 100nM BAY 87-2243 (inhibitor of HIF-1a), 1μM Bosentan (inhibitor of ET-1) and 2μM Imatinib (inhibitor of PDGFR). After treatment, the cells were lysed, and total protein was extracted and subjected to SDS-PAGE/Western blot for the analysis of NKX2-5 protein. N=1 (1 experiment performed).*

4.2.3.1.4 Signalling pathways that affect NKX2-5 expression

Next, I focused on the MAPK cascade in question of the potential effects of PI3K/AKT, JNK, p38, and ERKs on NKX2-5 activation.

Two inhibitors were used for PI3K/AKT signalling: GSK690693 that is a potent and selective pan-AKT kinase inhibitor, and GSK2126453 that is a dual inhibitor selective for both PI3K and mTORC1/2. TC ASK10 was used as a potent ASK1 inhibitor that also blocks downstream JNK1/p38 phosphorylation. For ERK5, ERK5-IN-1 a very specific inhibitor was used that inhibits ERK5 activity and EGFR-induced ERK5 auto-phosphorylation. FR180204 is an inhibitor specific to ERK1 and ERK2 only. SB202190 is a highly selective, potent and cell permeable inhibitor of p38 MAP kinase. BI78D3 was used as a competitive c-Jun N-terminal kinase (JNK) inhibitor, which Inhibits JNK interacting protein 1 (JIP1)-JNK binding and prevents JNK substrate phosphorylation.

ImHPASMCs were serum-starved for ~6-7 hours in 0.1% FCS, before 24 hours of TGF-β stimulation (2ng/ml). After 24 hours, the medium was changed to 5% FCS and the above selective inhibitors were added (Figure 4.15).

Most of the inhibitors down-regulated NKX2-5 protein expression, except for BI78D3 -the JNK inhibitor- which did not affect expression. SD208 showed a similar pattern as before. Unfortunately, GSK690693 and SB202190 inhibitors did not work as expected leaving the phosphorylated AKT and p38 proteins unaffected. This might be explained by the fact that the reagents were old or the concentration used at was not high enough to effectively block the signalling. Similarly, inhibitors for ASK1 and ERK1/2 kinases did not affect NKX2-5 expression (Figure 4.15). However, inhibiting ERK1/2 affected phosphorylated AKT levels negatively, suggesting that AKT activation is also mediated by ERK1/2 (Figure 4.15). On the contrary, use of the ERK5 inhibitor decreased significantly NKX2-5 protein and gene expression, whereas the dual inhibitor of AKT only affected significantly the protein levels of NKX2-5 but not the gene expression.

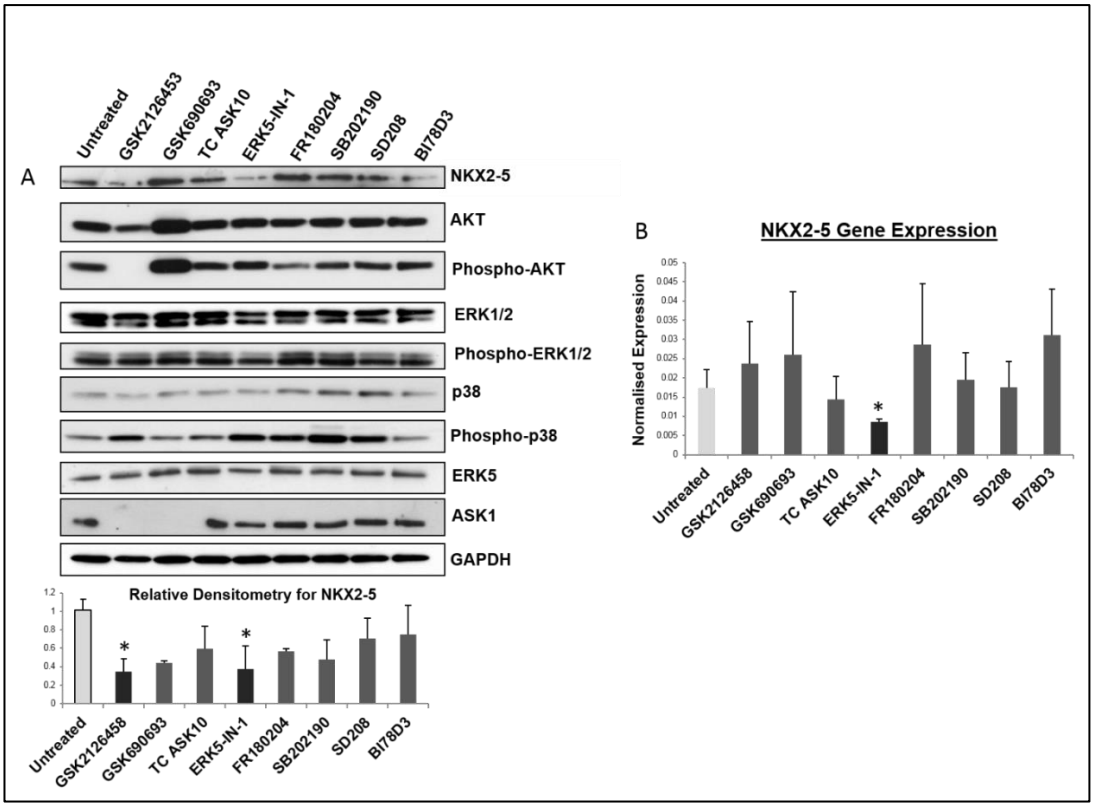


Figure 4.15 Signalling pathways that affect NKX2-5 expression. *ImHPASMCs were serum-starved for ~6-7 hours in 0.1% FCS, and then treated with 2ng/ml TGF-β for 24 hours. After 24 hours, the media was changed to 5% FCS and the cells were treated with the following inhibitors for 24 hours: 1μM GSK2126458 (inhibitor of PI3K kinase and mTORC1/2 receptors), 1μM GSK690693 (pan-AKT kinase inhibitor), 10μM TC ASK10 (inhibitor of ASK1 kinase), 10μM ERK-5-IN-1 (inhibitor of ERK5), 30μM FR180204 (inhibitor of ERK1 and ERK2), 10μM SB202190 (Inhibitor of p38 kinase), 5μM SD208 (Inhibitor of TGF-βRI/ALK5 receptor), 1μM BI78D3 (Inhibitor of JNK). After treatment, the cells were lysed and total protein and RNA were extracted and subjected to SDS-PAGE/Western blot and qPCR analysis. A. Protein expression of NKX2-5, total AKT and phosphorylated-AKT (Phospho-AKT), total ERK1/2 and phosphorylated ERK1/2 (Phospho-ERK1/2), total p38 and phosphorylated p38 (Phospho-p38), total ERK5 and total ASK1 were analysed. Densitometry analysis was performed in ImageJ normalised to GAPDH. B. NKX2-5 gene expression was analysed by qPCR normalised to TBP expression. N=3 (replicate experiments). p*≤0.05*

Then, I focused on the effect of AKT and ERK5 on NKX2-5 activation, since they were the only inhibitors of the panel that decreased NKX2-5 significantly (Figure 4.16). First, I analysed the efficiency of the inhibitors to block their specific signalling cascade. Both AKT and ERK5 inhibitors significantly downregulated phosphorylated AKT and phosphorylated ERK5 protein levels, respectively, and the ratios of phosphorylated to total protein are shown in Figure 4.16 B. NKX2-5 protein was significantly downregulated when both inhibitors were used. However, only the ERK5 inhibitor exhibited a significant effect at the mRNA level (Figure 4.16 C).

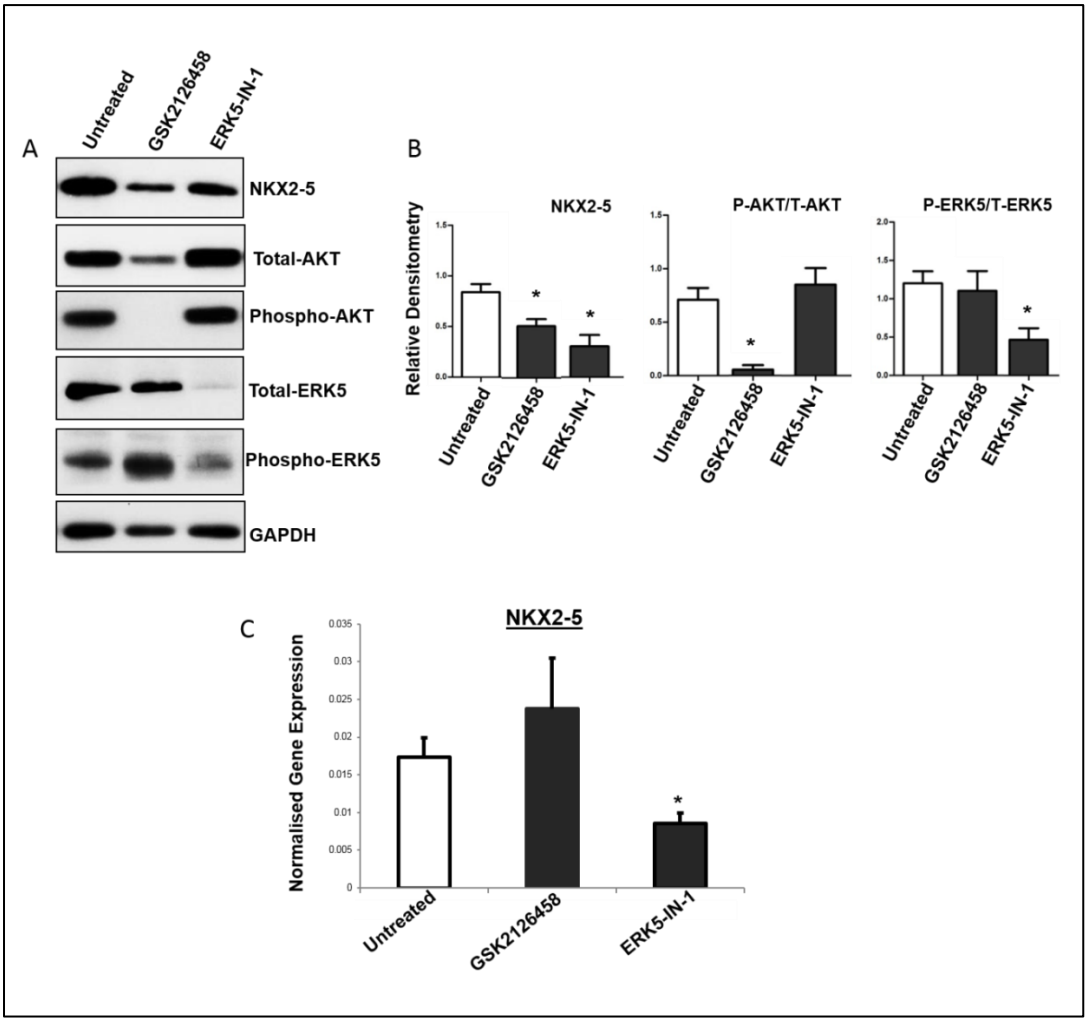


Figure 4.16 NKX2-5 expression is activated through the AKT and ERK5 signalling pathways. *lmHPASMCs were serum-starved for ~6-7 hours in 0.1% FCS, and then treated with 2ng/ml TGF-β for 24 hours. After 24 hours, the media was changed to 5% FCS and the cells were treated with the following inhibitors for 24 hours: 1μM GSK2126458 (inhibitor of PI3K kinase and mTORC1/2 receptors) and 10μM ERK-5-IN-1 (inhibitor of ERK5). After treatment, the cells were lysed and total protein and RNA were extracted and subjected to SDS-PAGE/Western blot and qPCR analysis. A. Protein expression of NKX2-5, total AKT and phosphorylated-AKT (Phospho-AKT), total ERK5 and phosphorylated ERK (Phospho-ERK5) were analysed. B. Densitometry analysis was performed in ImageJ normalised to GAPDH. C. NKX2-5 gene expression was analysed by qPCR normalised to TBP expression. Untreated: Cells treated with 2ng/ml TGF-β after serum starvation, but not treated further with inhibitors. N=3 (replicate experiments). *p≤0.05*

4.2.3.2 Signalling mechanism that regulates NKX2-5 expression

Taking into consideration all the previous findings, I propose a new model of signalling regulation of NKX2-5 activation (Figure 4.17). In brief, hypoxic conditions and TGF-β stimulation activated NKX2-5 expression significantly. When PI3K/AKT and ERK5 pathways were blocked, NKX2-5 was negatively regulated revealing that NKX2-5 activation is mediated directly through PI3K/AKT and ERK5 cascades. Although hypoxia activated directly AKT, TGF-β inhibition by SD208 did not affect the phosphorylated protein. However, it has been shown by others that AKT activation is mediated by BMP-dependent TGF-β signalling. In addition, use of SD208 inhibitor confirmed that ERK5 activation is also mediated by TGF-β.

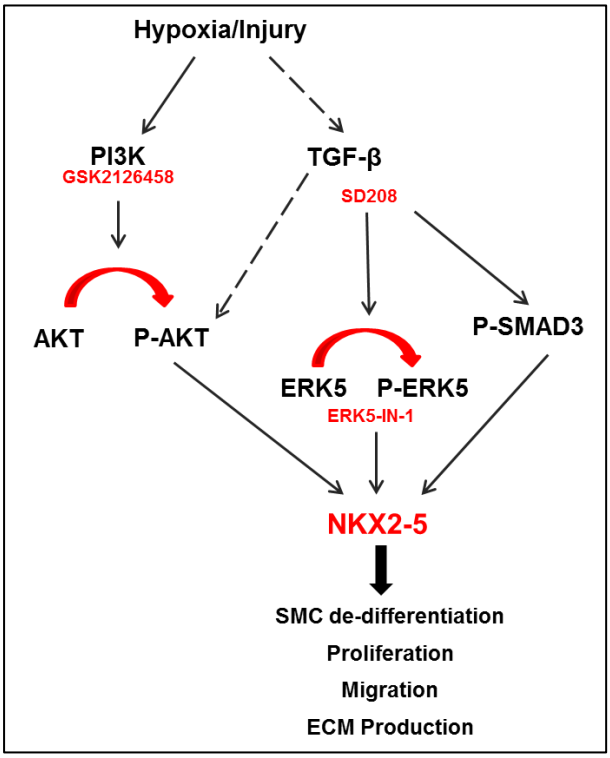


Figure 4.17 Proposed model of signalling mechanism that regulates NKX2-5. *The work conducted in ImHPASMCs using a panel of selective inhibitors for different signalling pathways generated important data regarding the signalling mechanism that upregulates NKX2-5 expression. In detail, hypoxia is able to upregulate a series of events within the cell, and directly leads to NKX2-5 activation. It has been previously shown that TGF-β positively regulates NKX2-5 expression and data presented here provided clear evidence of this mechanism. In addition, the use of selective inhibitors revealed that NKX2-5 is also a downstream target of both the ERK5 and AKT cascades. Broken lines show data published in literature and continuous lines show data produced in this thesis.*

4.3 Summary of results

Commercial primary HPASMC from 6 different human subjects, ImHPASMCs and HPASMCs and VSMC isolated from diseased tissues and vessels were used for the studies described in this chapter. The morphology was similar across the different cells, with small differences occurring in the size and the population doubling rates. Expression levels of NKX2-5 and other gene markers were heterogeneous, a defined characteristic of all SMC that complicates the experimental work resulting in variation and lack of statistical significance (439).

NKX2-5 expression was increased in synthetic compared to contractile HPASMC, and the expression was serum-dependent. Higher levels of NKX2-5 expression were also seen in PAH HPASMCs. However, inconsistency occurred between the NKX2-5 protein and gene expression in different VSMC, which could be due to the low sample number and the different background in disease pathogenesis between PAD and diabetes.

TGF-β is a major stimulator for NKX2-5 expression, with a greater effect in contractile cells compared to synthetic. However, a higher concentration of TGF-β (4ng/ml compared to 2ng/ml that are used) negatively affected not only NKX2-5 expression levels, but also the expression of fibrotic markers such as collagen and CTGF.

Hypoxia (1% O_2) activated NKX2-5 expression; however the use of a HIF-1α inhibitor did not affect NKX2-5 expression. To unravel the molecular and cellular mechanisms of activation of NKX2-5 in HPASMC, I investigated signalling pathways downstream of TGF-β and hypoxia using a panel of selective inhibitors. Inhibitors specific for ET-1, PDGF, and the kinases TAK1, CK2, p38, JNK, ASK1 and ERK1/2 did not affect NKX2-5 expression.

When selective inhibitors were used to block the PI3K/AKT and ERK5 pathways, TGF-β-induced NKX2-5 was significantly decreased indicating that NKX2-5 activation is directly dependent on PI3K/AKT and ERK5. Taking into consideration the findings of this chapter, I proposed for the first time a signalling mechanism of NKX2-5 activation in HPASMC.

4.4 Discussion

Due to the important role of NKX2-5 during embryogenesis, the gene is considered developmental and not many people have studied its post-natal expression. NKX2-5 expression has only been reported in few normal adult human tissues such as the heart, spleen, stomach, and cerebellum (Source: RNAseq, Gene Cards). We have previously shown that NKX2-5 is expressed in diseased vessels, where it controls vascular remodelling, but the regulation of the human gene remains largely unknown. In this chapter, I have provided data showing for the first time a signalling mechanism that activates NKX2-5 expression in adult VSMCs.

NKX2-5 gene and protein expression is under strict control at multiple levels to co-ordinate its important functions, and thus studying its expression has been proved very challenging. To start with, the low levels of expression in human adult tissues and unstimulated cells hinders the quantification of expression and affects the specificity of the chosen experimental method. In particular, to study gene expression, numerous primers had to be designed due to the fact that gene sequence is very repetitive and with high GC content. Low expression led to the amplification of non-specific products, and optimisation of qPCR protocol was required, so that the fluorescence emitted and measured corresponded only to the specific amplified products.

Identification and quantification of protein expression was also difficult. NKX2-5 mRNA and protein are subjected to numerous post-transcriptional and post-translational modifications, most of which affect the localisation of the protein, and therefore the function. NKX2-5 is a transcription factor that binds to specific DNA elements found in downstream target genes to regulate their expression. Consequently, NKX2-5 can only exert its function in the nucleus. Once *NKX2-5* transcriptional activity is required, cytoplasmic protein gets modified and enters the nucleus. Excess cytoplasmic protein is ubiquitinated and degraded. Published data also provide evidence that NKX2-5 can regulate itself through an evolutionary conserved auto-regulation mechanism (440).

Most findings regarding the post-translational modifications have been derived from mouse *Nkx2-5* gene, though a few have also been described in the human protein. Nkx2-5 is negatively regulated through glycosylation (367) and ubiquitination (439). However, another study indicated that small ubiquitin-like modifier (SUMO)-conjugation at a lysine residue (K51R) stabilised and enhanced the transcriptional activity of *Nkx2-5* (366). Nkx2-5 is also modified through phosphorylation,

acetylation and methylation, with phosphorylation been studied extensively. In particular, phosphorylation of the serine 164 (S164) of the homeodomain by CK2 promotes NKX2-5 nuclear translocation, increases DNA binding affinity and transcriptional activity. Taking into consideration the above, as well as that post-translational modifications can also be cell-type specific, NKX2-5 protein can exist in various states of modification that are difficult to discriminate. Since the aim of this thesis was to explore the mechanisms that activate NKX2-5, I focused on studying the nuclear protein expressed in human PASMCs.

The experimental work was mainly conducted in primary HPASMCs. The use of primary cells is challenging. On the one hand, primary cells are most representative of the cells inside human tissues and carry the same genetic background and similar disease phenotypes. However, primary cells cultured *in vitro* loose key functions and characteristics, such as interactions with other cells and with the native matrix.

On the other hand, primary cells originate from different individuals and thus they exhibit high levels of heterogeneity in their morphology, survival, behaviour in culture, responses to stimuli, and gene and protein expression. This is particularly true in the case of VSMCs, which is a highly heterogeneous and plastic cell type not only among individuals but also among different vascular beds within the same organism. This heterogeneity leads to difficulties in the interpretation of data and statistical analysis. High heterogeneity of primary cells causes large variation, and thus low power to detect statistical significant data. This is a challenge that many scientists are called to confront daily and it raises questions such as “do statistical significant data denote significant events in nature?”. Careful experimental design and critical interpretation of experimental results are required to avoid over-interpretation.

Among other disadvantages, primary cells have a finite life-span of less than 10-20 passages *in vitro* depending on the cell type. HPASMCs and other VSMCs are sensitive to different culturing conditions, stimuli and treatments, and often exhibit very inconsistent population doubling-time. For the reasons above, the use of an immortalised cell line that would provide a consistent cellular material, be capable of extended proliferation and possess similar or identical genotype and phenotype as the parental tissue throughout the duration of the project was beneficial. This is the first time that a human immortalised PASMC cell line was used in the field, and consequently assessing the expression profile and phenotypic characteristics was essential. ImHPASMCs are commercially available upon request from ABM Good

and the cells were immortalised from a primary cell line using SV40. Gene and protein expression was assessed in untreated ImHPASMCs as well as under different stimulations and treatments. The expression profile closely resembled the profile of the primary cells.

After establishing the expression profiles of primary and ImHPASMCs, I confirmed that expression of NKX2-5 is strongly associated with the synthetic phenotype. Primary cells were cultured *in vitro* under conditions favouring the contractile or the synthetic state (Section 2.7). NKX2-5 was significantly increased in the synthetic cells (Figure 4.3). TGF-β has previously been described as a potent stimulator of cell proliferation, migration and production of ECM, key characteristics of the synthetic state of VSMCs. When primary and immortalised HPASMCs were stimulated with TGF-β to promote the de-differentiation process, expression of NKX2-5 was increased. However, when the concentration of TGF-β was doubled from 2ng/ml to 4ng/ml, the stimulatory effects were inhibited suggesting a negative feedback loop (Figures 4.10, 4.11). One possible mechanism could involve the inhibitory SMADs, SMAD6 and SMAD7 that are both transcriptionally induced by TGF-β and BMPs (433). Collagen type I and CTGF were also used as markers of the synthetic state, and their expression was enhanced by TGF-β stimulation at both gene and protein levels in ImHPASMCs.

Given that NKX2-5 expression is associated with vascular remodelling in various pathological conditions, I next tested whether NKX2-5 is activated in PAH and in other vascular diseases such as diabetes and PAD. Primary HPASMC were isolated from lung tissue of PAH patients undergoing lung transplantation. The patients suffered from IPAH and were usually very young in age. Control cells were either isolated from normal lung tissues that were rejected as transplants, or bought commercially and cultured under contractile conditions. In addition, VSMCs were isolated from diseased vessels of PAD and diabetic patients who undergone limb amputation due to tissue necrosis in advanced disease stage. Only 3 primary cell lines were generated for the PAD group. NKX2-5 protein expression was increased in both IPAH and PAD, but the difference was not statistically significant (Figure 4.8). Gene expression followed a similar pattern for the IPAH group, although levels in the PAD group were lower compared to IPAH.

Inconsistency between levels of protein and gene expression of NKX2-5 was often observed in experiments undertaken in this project. In this case, the inconsistency could be explained by the different disease pathogenesis and the low sample numbers. In particular, IPAH is a condition with more homogeneous disease

manifestation compared to PAD, which included both diabetes and PAD. Diabetic patients can develop complex vascular complications mainly due to atherosclerosis (441) with a strong inflammatory background and dysregulated lipid metabolism and deposition in the blood vessel wall. Atherosclerosis is also the hallmark of PAD in combination with arteriosclerosis caused by increased fatty deposits on arterial walls leading to limb ischaemia.

Despite the variation in the degree of NKX2-5 activation, my data consistently showed an increase in NKX2-5 in diseased vessels compared to normal. Since a variety of stimuli can initiate vascular remodelling in diseased vessels, the next question was what led to the activation of NKX2-5 in HPASMC. To this end, I examined whether hypoxia affects expression of NKX2-5. Hypoxia is an important regulator of physiological processes, including erythropoiesis, angiogenesis and glycolysis (442). However, hypoxia has also been associated with pathological conditions including, would healing and tissue repair, cancer, inflammation, fibrosis, and PH, where its role has been extensively studied [reviewed in (443-445)]. In particular, hypoxia has been described as a potent mitogen that promotes proliferation of PASMCs leading to vasoconstriction and pulmonary vascular remodelling (443). *In vitro* culture of HPASMCs under hypoxic conditions increased NKX2-5 expression (Figure 4.12), a finding consistent with its role as a regulator of vascular remodelling. Nonetheless, the use of BAY 87-2243, a highly selective and potent inhibitor of HIF-1α, did not affect NKX2-5 expression (Figure 4.14). HIF-1 is a hypoxia-induced transcription factor that mediates hypoxia-dependent transcription, and is thought to regulate PASMC proliferation under hypoxic conditions (446).

The mechanisms by which hypoxia affects PASMC growth are not fully elucidated. Accordingly, this might justify the inconsistent result in the case of BAY inhibitor. In addition, hypoxia-induced pulmonary vascular remodelling is a highly complex process, which may have numerous interactions between the vascular cells including endothelial cells, PASMCs and lung fibroblasts. Therefore, studying the effects of hypoxia in a single cell type excluding cell-cell interactions might hide important piece of information. In this case, the effect of hypoxia on NKX2-5 expression may be mediated via endothelial-derived mechanisms indirectly, and more studies are required to decipher this finding.

Apart from HIF-1α, hypoxia-dependent transcription is also mediated by other growth factors and signalling cascades such as the MAPK cascade. In particular, studies have assessed the effect of hypoxia either directly or through the use of inhibitors in different types of cells such as endothelial cells, macrophages, cancer

cells, and the findings vary significantly [examples of studies with controversial data (447, 448)]. The discrepancies may be caused by the different extents of hypoxia used by investigators. In this study, I evaluated the effect of hypoxia on phosphorylated and active forms of AKT and SMAD2, since the involvement of ERK1/2, PI3/AKT, and TGF-β signalling pathways in hypoxia-mediated cellular responses in PASMCs and endothelial cells is well characterised (449, 450). Indeed, hypoxia induced phosphorylation of AKT and SMAD2 proteins, whereas the total proteins (inactivate forms) remained unaffected (Figure 4.12).

Studies in mice have revealed a vital role for TGF-β signalling through multiple BMP binding sites and GATA transcription factors in Nkx2-5 expression (350). It has also been previously shown that TGF-β stimulates the phenotypic modulation of VSMC in disease (30). In this chapter, I showed that TGF-β efficiently promoted the synthetic state of VSMCs with increased expression of synthetic markers including collagen type I and CTGF (Figure 4.10). Also, TGF-β stimulation activated expression of NKX2-5 at both the protein and gene level, confirming previous findings that NKX2-5 is a regulator of the VSMC phenotypic switch (Figure 4.3). Interestingly, the effect of TGF-β was more prominent in contractile HPASMCs compared to synthetic. In synthetic cells particularly, TGF-β increased protein expression of NKX2-5, but it did not affect the gene expression levels in experimental conditions used in this study. The findings suggest probably that the stimulation is more important for the initiation rather than the continuation of NKX2-5 expression. Similarly, higher concentration of TGF-β does not activate expression, as discussed previously. However, the use of SD208, a selective ALK5 receptor inhibitor, only had a small effect on NKX2-5 protein with a trend to decrease, but did not achieve statistical significance (Figure 4.15). This suggests that more than one TGF-β dependent downstream pathways is implicated in NKX2-5 regulation, and ALK5-dependent pathway is not one of them.

TGF-β is a cytokine with diverse and often contradictory functions. It has been implicated in physiological conditions, such as cell growth, differentiation, and apoptosis, and also in pathological conditions. In the early stages of repair, TGF-β is released from platelets and activated from matrix reservoirs; it then stimulates the chemotaxis of repair cells, modulates immunity and inflammation and induces matrix production. At later stages, it negatively regulates fibrosis through its strong anti-proliferative and apoptotic effects on fibrotic cells. TGF-β has been extensively studied in fibrosis and scleroderma [reviewed in (120, 451)], where it exerts an important profibrotic function inducing mesenchymal fibroblasts to proliferate,

migrate, and synthesise elevated levels of matrix proteins. TGF-β signalling has also been studied in the context of CVD such as hypertension, heart failure, and atherosclerosis. Although initially it was suggested that TGF-β exhibited adverse effects in the cardiovascular system, more recently its role as anti-inflammatory and anti-atherogenic molecule has gained more ground.

The role of TGF-β in VSMC emphasises its distinctive pleiotropic activities. VSMC respond to TGF-β in both Smad-dependent and Smad-independent manners. We and others propose a profibrotic role of TGF-β that promotes the disease associated de-differentiation of VSMCs, characterised by proliferation and migration, and correlates significantly with elevated expression of synthetic markers and ECM in the vasculature (30, 32, 170, 172, 452-456). However, others consider TGF-β as an inhibitory molecule for the proliferation and activation of VSMCs, and that it promotes contractility, the principle function of VSMCs (115, 457, 458). In fact, TGF-β binding and SRF responsive elements are found in the promoters of many contractile genes including SM22, calponin, α-SMA, procollagen I and III (115, 459).

The role of TGF-β has also been studied in other diseases characterised by vascular remodelling. A recent paper by Ha *et al* demonstrated that VSMCa in microvessels of PAD patients express TGF-β, and the expression coincides with expression of collagen and extended fibrosis in the diseased vessels (460). TGF-β expression not only caused the phenotypic de-differentiation of VSMC to the synthetic phenotype, but also stimulated the transition of adjacent fibroblasts to myofibroblasts (460). Another example of vascular disease is atherosclerosis, where the role of TGF-β is described as either atheroprotective or as atherogenic [reviewed in (112, 117)]. A study published few years ago reported that macrophage-specific TGF-β overexpression reduced atherosclerosis in an ApoE deficient mice, and this correlated significantly with fewer macrophages, more VSMCs and more collagen expression pointing to plaque stabilisation (116). However, TGF-β also promotes the synthetic phenotype of VSMCs that contributes significantly to atherosclerosis progression. The role of EndoMT, where TGF-β and other growth factors such as FGF2 are implicated, is gaining increasing attention and provides significant insight in the pathogenesis of atherosclerosis. EndoMT will be discussed in detail in Chapter 6.

The variation in the described functions and roles of TGF-β in disease may have a number of explanations. Firstly, the data depend on the experimental design and conditions followed. Secondly, it is obvious that the growth factor could exert different functions in different cell types, and consequently, the cell-cell and cell-

matrix interactions retain a prominent role that is difficult to mimic in an *in vitro* environment. In vascular diseases in particular, interplay occurs not only between the cells of the vessel wall but also between different processes such as vascular remodelling and inflammation. Thirdly, the tendency to not publish negative data that "do not fit with the hypothesis" or data that are not easily explained by current knowledge is often detrimental in science, since it obscures the real events. Taking into consideration the above, it is obvious that TGF-β orchestrates many processes in the human body with potentially different functions depending on the organism, cell type and environmental stimuli. Further studies are required to unravel the important functions and interactions of TGF-β.

Further in this chapter, I explored whether other signalling pathways could directly affect NKX2-5 expression. The cells were stimulated with TGF-β to maximise NKX2-5 expression, and then treated with a panel of selective inhibitors to evaluate the effect of different signalling pathways. Most of the selected inhibitors targeted pathways that are known to be implicated in vascular diseases such as members of the MAPK cascade or downstream mediators of TGF-β signalling. These pathways involved PI3K/AKT, ERK1/2 and ERK5, ASK1, p38, and JNK kinases. MAP kinases are protein serine/threonine kinases that play an important role in cell differentiation, growth, apoptosis, and the regulation of a variety of transcription factors and gene expression. MAP kinases are activated by phosphorylation on threonine (T) and tyrosine (Y) residues within a T-X-Y phosphorylation motif, where "X" can be glutamine (E), proline (P), or glycine (G) (461).

Most of the inhibitors tended to decrease NKX2-5 protein expression although there was not always a corresponding effect of gene expression (Figure 4.15). However, two selected inhibitors significantly decreased NKX2-5 levels: the PI3K/AKT and ERK5 inhibitors (Figure 4.16). The use of GSK2126458 blocks the activity of PI3K kinase to phosphorylate and activate AKT protein, and the ERK5-IN-1 inhibitor blocks the phosphorylation and activation of ERK5 protein.

The PI3K/AKT cascade has been well studied in endothelial cells and VSMCs in cancer, in angiogenesis, and in cell proliferation and migration is response to apelin, insulin, hypoxia and other stimuli (462-466). Many studies have reported the effect of apelin-a G-coupled receptor important in vascular development, on proliferation of human and rat VSMCs through the PI3K/AKT signalling transduction pathway, with this effect been mediated by phosphorylated ERK1 and ERK2 proteins (463, 467). This pathway that has also been reported to be induced by hypoxia, can serve as a critical therapeutic target in atherosclerosis and other vascular disease (468).

The role of the PI3K/AKT pathway on proliferation of VSMCs has also been examined in PH, where the use of lipocalin-2, a selective up-regulated protein in CHD-PH, significantly promoted HPASMC proliferation by activating PI3K/AKT pathway (469). Another study reported that TGF-β phosphorylates and activates PI3K/AKT with overexpression of Smad3 to significantly augment this effect in rat aortic SMCs leading to cell proliferation and migration (462). Interestingly, the same study implicated phosphorylated p38 as an intermediate factor between AKT and Smad3. The PI3K/AKT pathway has also been shown to be important and a fast acting modulator of contraction of VSMCs, with the use of an AKT-specific inhibitor decreasing vasoconstriction of isolated vessels (470).

ERK5 is a newly identified member of the family of MAP kinase that is activated by ERK kinase 5 (MEK5). Less information is available regarding the role of ERK5 kinase in vascular diseases compared to the rest of the members of the family. A study in rat VSMCs showed that ERK5 activation is implicated in VSMC migration *in vivo* in the context of atherosclerosis (471). In the same study, a role for ERK5 as a survival factor and protector against oxidative stress-induced cell death in VSMCs was also proposed. Another study in human aortic SMCs, reported that ERK5 phosphorylation contributes to MEF-2c activation and subsequent cell hypertrophy induced by angiotensin II, proposing a novel molecular mechanism in vascular remodelling (472). The effect of angiotensin II on activation of ERK5 mediated by MEF-2c was also examined and confirmed in rat aortic SMCs by a different group of investigators, who also established the different phosphorylation profiles and functions between ERK1/2 and ERK5 kinases (473).

Taking into consideration the above, both PI3K/AKT and ERK5 downstream signalling are important regulators of proliferation, migration, hypertrophy and resistance to apoptosis in VSMCs, all of which are characteristics of the synthetic phenotype associated with disease. Blocking both signalling pathways led to decreased NKX2-5 levels, proposing that NKX2-5 is a direct downstream target of AKT and ERK5. As discussed earlier, both kinases can be regulated by TGF-β and TGF-β-dependent mediators, such as Smad3. Based on the evidence accumulated in this chapter along with findings in the literature, I propose a signalling mechanism that regulates the activation of NKX2-5 *in vitro* in HPASMCs (Figure 4.17). The mechanism is driven by hypoxia and TGF-β and mediated by phosphorylation of AKT, ERK5 and SMAD3. Activation of the mechanism may lead to VSMC phenotypic modulation and vascular remodelling.

CHAPTER 5 - RESULTS: TRANSCRIPTIONAL REGULATION OF NKX2-5 EXPRESSION IN VASCULAR SMOOTH MUSCLE CELLS

5.1 Introduction

In Chapter 4, I explored the signalling mechanisms that activate NKX2-5 expression in HPASMCs. The stimuli that activate the gene can vary from growth factors and mitogens, as the products of injury or vascular remodelling, to hypoxia and stress. These stimuli usually initiate downstream signalling mechanisms through GPCR receptors at the nuclear membrane and engage various molecular pathways including TGF-β, PI3K/AKT and ERK5. Although the end product of this cascade is the activation of NKX2-5, how exactly this is accomplished remains unknown. In this chapter, I will address how the *NKX2-5* gene gets activated, and I will focus on the regulation at the transcriptional and post-transcriptional levels.

Because NKX2-5 is not expressed in healthy blood vessels, the regulatory regions responsible for the transcriptional activation in vessels have not been identified. Data from animal studies have shown that the transcriptional regulation of the murine *Nkx2-5* gene is complicated with a number of cis- and trans- acting elements spread over a 23Kb region which regulate expression in tempo-spatial specific ways (474). Despite our knowledge of the mouse *Nkx2-5* gene structure and the high homology between the mouse and the human genes (Figure 1.7), very little is known about the regulation of the human gene.

In this chapter, I will exploit the findings of the genetic study as a guide to potential functional sites. *In silico* analysis has provided evidence that support the potential functionality of the tagging SNPs. The 3 associated SNPs (rs3095870, rs3132139, rs3131917) are located upstream and downstream of the gene in areas that would be key for gene regulation (Figure 5.1). I will address these genomic loci separately and assess their role in the transcriptional regulation of *NKX2-5*. The 3'UTR will also be explored in order to identify other possible post-transcriptional mechanisms.

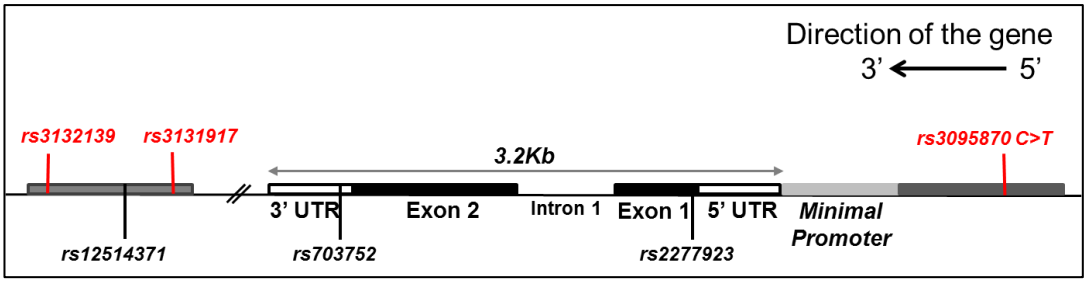


Figure 5.1 Location of tagging SNPs across NKX2-5 genomic locus. *The SNPs shown in red showed association with scleroderma and PH. SNP rs3095870 showed a marginal association with scleroderma in the replication cohort. It is located upstream of the NKX2-5 transcription start site. SNP rs3132139 is associated with PH in the discovery cohort and the association was efficiently replicated. In addition, further analysis showed that the association of rs3132139 with PH is independent of the presence of scleroderma. A meta-analysis among the two independent cohorts showed that rs3131917 is associated with scleroderma. SNPs rs3132139 and rs3131917 are both located downstream of NKX2-5, in a region that was previously identified as a putative enhancer. The rest of the tagging SNPs (rs703752, rs2277923, rs12514371) are shown in black and are located in the first exon of NKX2-5, in the 3'UTR and downstream of the gene respectively.*

At this point, I would like to give some information about the TEAD family of transcription activators, as an important transcriptional regulator of NKX2-5. The TEA/ATTS domain (TEAD) family consists of 4 members TEAD1-4 with distinct and important roles in VSMC differentiation (147, 149, 475). All 4 members recognise and bind specifically at the MCAT consensus element: 5'-TCATTCCT-3'. TEADs have crucial and important roles during embryogenesis. Disruption of TEAD1 during mouse embryonic development prevents heart biogenesis (476), and its expression is required in cardiac muscle (476) (477) (478).

Apart from their essential role in development, TEADs are implicated in various conditions in adulthood related to the cardiovascular system, SMC development and differentiation, skeletal muscle hypertrophy and regeneration. The expression of the TEAD family members is ubiquitous in most adult tissues (479). However, they confer muscle-specific gene transcription in a way that is not yet fully understood. Moreover, each protein exhibits different expression patterns depending on the tissue and developmental stage suggesting that each protein retains a unique function (479). TEAD2 is the first transcription factor expressed immediately after fertilisation, and controls expression of other genes during preimplantation development (479, 480) . TEAD3 is expressed primarily in placenta (481, 482) and in cardiac muscle (483). Also, TEAD3 augments the α_1-adrenergic activation of the skeletal muscle α-actin (484). TEAD4 specifically activates skeletal muscle genes (481) and is involved in skeletal muscle differentiation.

Initial induction of many smooth muscle-specific genes such as myocardin, α-SMA, troponin, and smooth muscle heavy chain is MCAT-dependent (475). Myocardin is expressed in VSMCs and activates muscle-specific genes associated with SMC differentiation (147). Animal studies have revealed a 10Kb distal enhancer that directs the cardiovascular expression *in vivo* and is regulated through the binding of Foxo, E-box, NKX2-5 and TEADs (485). In recent studies, it has been shown that TEAD1 is significantly induced during SMC phenotypic modulation. In particular, TEAD1 represses myocardin-induced activation of smooth muscle-specific genes through disruption of myocardin binding to SRF *in vitro* and *in vivo* (149). Such data highlight the complexity of the function and specificity of the TEAD family members, which can either activate or repress smooth muscle specific genes dependent on the phenotypic state of the cell and the associated co-factors.

The ability of all 4 TEAD family members to bind the canonical consensus MCAT sequence is conserved through evolution. However, they lack a definite transcription activation or repressor domain, and published data suggest that TEAD proteins require co-factors to exert their function (486, 487). Various proteins have been proposed as candidate co-activators, with Yes-associated protein 1 (YAP1) being the best candidate. YAP1 contains an acidic transcriptional activation domain at the carboxyl terminus, but lacks a DNA binding domain (488). All TEAD proteins bind YAP1 specifically and with the same efficiency at their carboxyl-terminal end (489).

YAP1 is the human homolog of Yki and is phosphorylated by the Lats tumour suppressor which is the homolog of Wts in *Drosophila*. They are all components of the Hippo pathway that was first identified in *Drosophila* and is highly conserved through evolution. The role of the Hippo pathway in mammalian cell growth and proliferation, apoptosis, and human cancer is well established (490, 491). Phosphorylated YAP is bound to the phospho-serine/threonine- binding protein 14-3-3 in the cytoplasm. The dephosphorylated protein is the active form found in the nucleus, where it acts as a transcriptional co-activator (489). Zhao *et al* have shown that YAP1 and TEAD1 co-occupy >80% of the promoters that are pulled down by either of them in a ChIP-on-ChIP assay (492). Recently, the Hippo pathway has attracted increasing attention for its role in VSMC phenotypic modulation, with data showing that YAP1 regulates the VSMC phenotype towards the synthetic state (148). It has also been shown that VSMC proliferation is YAP1-dependent, and YAP1 represses expression of CArG-dependent contractile-specific genes such as myocardin, α-SMA, SMMHC and SM22 (148). Taking into consideration these findings, TEAD and YAP proteins form complexes and together they bind MCAT

DNA elements to activate or repress expression of target genes through myocardin/SRF binding, promoting the synthetic state in VSMCs.

5.2 Results

5.2.1 Transcriptional regulation through the NKX2-5 promoter

Eukaryotic promoters can range between 100bp to 1Kb in length. There are usually 3 distinct parts of the promoters: the *core* promoter, which contains the transcription start site (TSS), the RNA polymerase binding site, and binding sites for other transcription factors; the *proximal* promoter with the proximal sequence upstream of the gene that tends to contain primary regulatory elements and it is usually 250bp long; and, the *distal* promoter which compromises the distal sequence upstream of the gene which contains additional regulatory elements with weaker influence.

5.2.1.1 Transcriptional activity of rs3095870 locus

For the study of *NKX2-5* gene, a 578bp long fragment was considered as the minimal promoter of the gene containing the TSS and the RNA polymerase II binding site. As described in section 2.6 the minimal promoter was cloned into the pGL4.10 reporter vector. SNP rs3095870 showed evidence of association with scleroderma but the association was not significant after the permutation analysis (Table 3.9). However, since the *in silico* analysis suggested that this SNP might be functional, I hypothesised that rs3095870 is located within or near a distal promoter element. The genomic locus that contained the SNP was considered as an upstream promoter and was also cloned into the pGL4.10 reporter vector. The altered allele was introduced with site-directed mutagenesis (SDM).

The minimal promoter vector and the upstream/distal promoter vector specific to the C or the T allele of rs3095870 were transfected into primary contractile or synthetic HPASMCs to test the potential of the genomic loci to increase transcriptional activity (Figure 5.2). The rs3095870-C construct significantly increased transcriptional activity compared to the minimal promoter alone in both contractile and synthetic cells. The rs3095870-T construct did not increase transcriptional activity compared to the minimal promoter. Also, the rs3095870-T construct exhibited significantly less transcriptional activity compared to the rs3095870-C.

These data suggest that the genomic locus of rs3095870 retains the ability to increase transcription, and that this property is specific to the major C allele, which provided evidence of association with scleroderma. It is therefore proposed that a

transcriptional mechanism that activates transcription takes place only in the presence of the C allele.

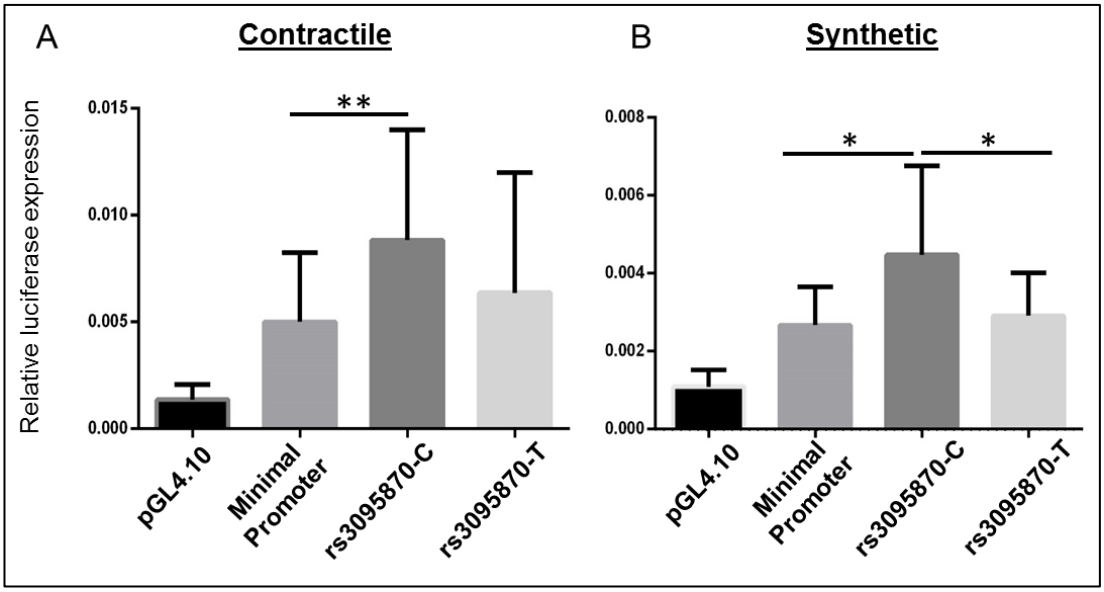


Figure 5.2 Luciferase activity of the minimal and upstream promoter in HPASMC. *Contractile (A) and synthetic (B) primary HPASMCs were transfected with reporter vectors containing the firefly luciferase gene driven by the minimal promoter of NKX2-5 (minimal promoter), the upstream promoter including the rs3095870-C allele (rs3095870-C), and the upstream promoter including the rs3095870-T allele (rs3095870-T), and luciferase assays were performed. Luciferase activity was measured and is presented as the ratio of firefly/renilla expression. The pGL4.10 reporter vector was used as negative control. N=3 (replicate experiments). *p≤0.05, **p≤0.01.*

5.2.1.2 TEAD1 binds at rs3095870 site

An *in silico* analysis focused on the rs3095870 was conducted using TRANSFAC and JASPAR. A panel of different transcription factors were predicted to bind at the region of upstream promoter, but only one protein, TEF-1, could bind at the rs3095870 site (Figure 5.3). In fact, TEF-1 binds only in the presence of the major C allele, whereas when the T allele is present the binding site is abolished.

Taking this into account, the data propose a regulatory mechanism where TEAD1 binds at the upstream promoter only in the presence of the risk/disease-associated rs3095870-C allele and activates or enhances transcription of *NKX2-5,* leading to increased gene expression, which has previously been associated with disease.

In order to confirm the TEAD1 binding at the SNP site, I designed and performed different *in vitro* and *in vivo* binding assays. The design and the details of the experiments are given in sections 2.11 and 2.12.

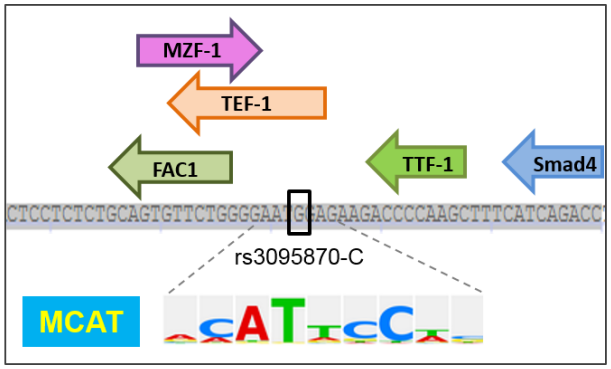


Figure 5.3 TEAD1 binds the rs3095870 site. *An in silico analysis using TRASNFAC and JASPAR revealed that a MCAT consensus sequence coincides with rs3095870. MCAT sites are consensus binding elements for the TEAD family of transcription activators. In the presence of the C allele of rs3095870, TEAD1 or TEF-1 transcription factor binds the consensus element, but when the minor T allele is present the consensus site is abolished.*

5.2.1.2.1 Electrophoretic mobility shift assay (EMSA)

I performed an EMSA to explore whether the C and T alleles of rs3095870 retain different binding affinities for other proteins, indicating that each allele is associated with the binding of different transcription factors.

As it is shown in Figure 5.4A, there are different bands for each allele (red arrows). The two bands are of different sizes which means that different proteins bind each allele confirming different binding properties.

Next, a supershift experiment was performed with an antibody specific to TEAD1. As shown in Figure 5.4B, the original band of rs3095870-C probe was shifted with the TEAD1 antibody, revealing that TEAD1 is indeed a part of this binding complex. In addition, I used an antibody specific to YAP1, which is a partner of TEAD1 that together as a complex regulate expression of downstream targets. The addition of YAP1 antibody resulted in a supershift of the original band, revealing that YAP1 is also member of complex.

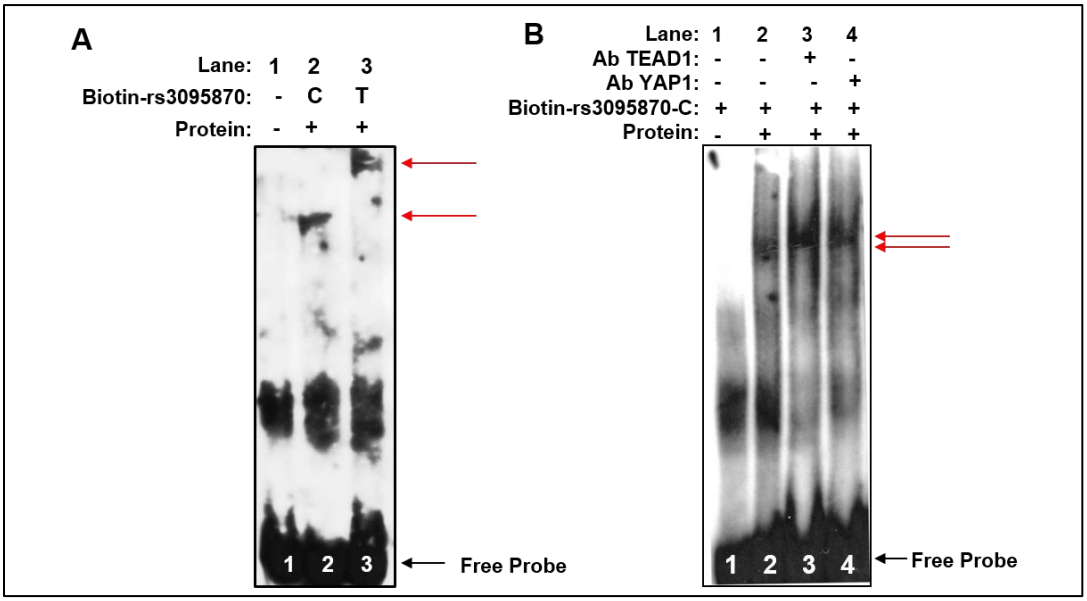


Figure 5.4 Electrophoretic mobility shift assays for rs3095870. *Two biotinylated double-stranded DNA oligonucleotides specific to the rs3095870 SNP were designed containing either the C or the T allele of the SNP. The probes were mixed with the nuclear protein isolated from TGF-β stimulated ImHPASMCs (16 hours of TGF-β stimulation) and the protein/DNA complexes were analysed by EMSA assays. A. The C and T alleles of rs3095870 show different binding affinities, two bands of different sizes are formed (red arrows). B. TEAD1 and YAP1 specific antibodies were used to determine whether either of the proteins are part of the complex. Both antibodies super-shifted the protein/DNA complex created in the presence of the C allele of rs3095870 (red arrows), confirming the presence of TEAD1 and YAP1 proteins in the complex. Free unbound probe is shown by the black arrow at the bottom of the gel.*

5.2.1.2.2 Pull-down assay

To verify the results of the EMSAs, pull-down assays were performed using 2 different biotinylated DNA probes specific to both alleles of rs3095870 and two different control probes. The results confirmed that TEAD1 protein binds specifically the C allele of rs3095870, whereas there is no binding to the T allele probe (Figure 5.5).

Although the biotinylated scrambled probe showed a weak signal, probably due to the strong affinity of biotin and streptavidin, the use of 2 different control probes confirmed that the binding to the C allele is strong and specific.

The binding of TEAD1 to the rs3095870 is very sensitive and numerous experiments were performed to optimise the best binding conditions. The optimisation involved adjusting the amount of detergent in the binding reaction and the salt concentration during washes. After obtaining the best binding conditions, the experiment was performed 3 times to ensure that binding is specific and not an artefact.

This assay is another way to verify the results of the EMSA. Both experiments addressed the same question: whether TEAD1 binds the rs3095870 site and the binding is specific to the C allele or also binds the T allele.

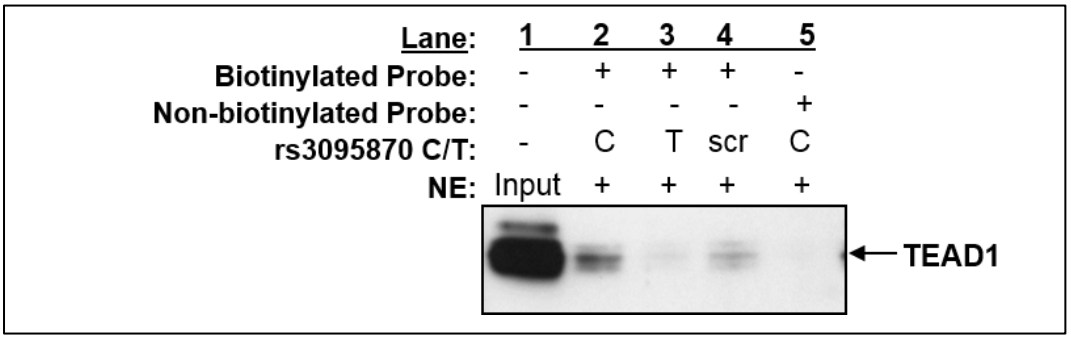


Figure 5.5 Pull-down assay. *Biotinylated probes specific to the C (Lane 2) or the T (Lane 3) allele of rs3095870 were mixed with nuclear protein isolated from TGF-β stimulated ImHPASMCs in order to pull-down proteins that bind specifically the sequence around the SNP. A biotinylated probe with scrambled sequence (scr) (Lane 4) and a non-biotinylated probe specific to the C allele (Lane 5) were used as negative controls. Input: 5% of the total protein used per assay (Lane 1). N=3 replicate experiments were performed, the blot is a representative experiment.*

5.2.1.2.3 Chromatin immunoprecipitation assay (ChIP)

Thirdly, a ChIP assay was performed to monitor *in vivo* the TEAD1 association with the upstream promoter in live ImHPASMCs. The experimental procedure is described in section 2.12. Chromatin immunoprecipitation with the TEAD1 antibody showed that the transcription factor binds the *NKX2-5* upstream promoter (Figure 5.6). TEAD1 binding was also found in the minimal promoter and paired with binding of RNA polymerase II suggesting that TEAD1 participates directly in the transcriptional machinery of *NKX2-5* (Figure 5.6).

In addition, data showed that Smad3 also participates in the transcriptional regulation of *NKX2-5*. Chromatin immunoprecipitation with an antibody specific to phosphorylated-SMAD3 showed binding of the activated form of the protein at the transcriptional start site as well as in the upstream promoter. This finding provides a confirmation that NKX2-5 activation can be mediated by TGF-β as shown in Chapter 4.

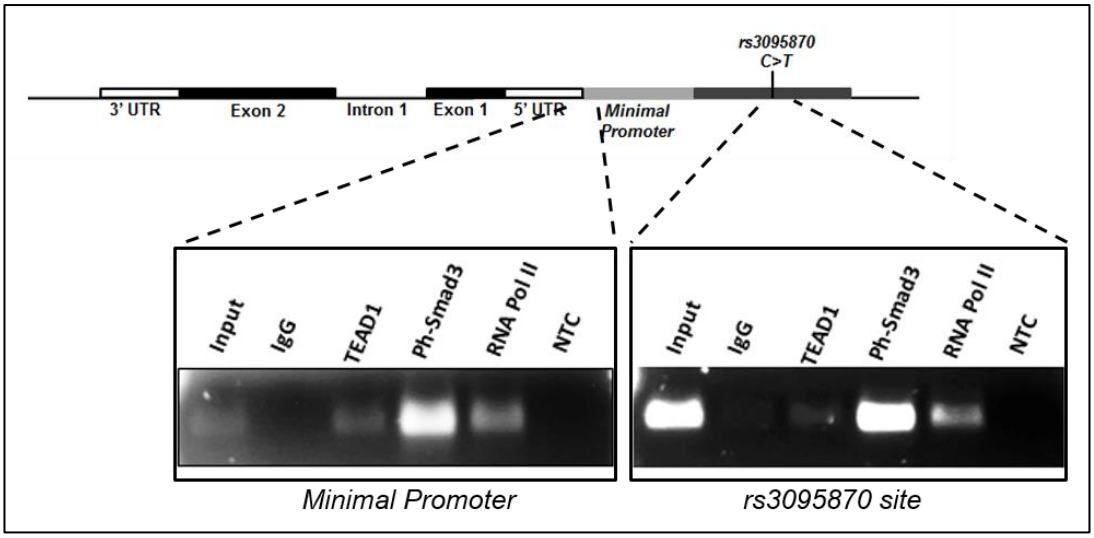


Figure 5.6 Chromatin Immunoprecipitation assay. *ImHPASMCs were treated with TGF-β for 16 hours before the transcription factors were cross-linked to the chromatin. The chromatin was then sheared by sonication into DNA fragments of 500bp-1Kb in length. Chromatin was immunoprecipitated with antibodies specific to TEAD1, phospho-Smad3 and RNA polymerase II. Immunoprecipitation with a non-specific IgG was used as a negative control. Two DNA fragments surrounding the rs3095870 site and the transcription start site of NKX2-5 were amplified by PCR. The PCR products were analysed in an agarose gel and the results showed in the figure above. Input: 10% of initial amount of chromatin used per immunoprecipitation. NTC: no template control PCR reaction.*

5.2.1.3 Transcriptional regulation of NKX2-5 through the TEAD family

Since, TEAD family members can all bind the MCAT consensus element with the same affinity and efficiency, apart from TEAD1, TEAD3 will also be studied in the study.

TEAD3 is generally expressed in cardiac muscle (483, 484), as is NKX2-5. In addition, a recent study showed that TEAD3 is expressed in human aortic SMC and is required for TGF-β signalling (493). Therefore, TEAD3 would be a good candidate effector of NKX2-5 regulation in HPASMC.

5.2.1.3.1 TEAD1 enhances the transcriptional activity of NKX2-5

After confirming the binding of TEAD1 to the rs3095870-C allele and its possible involvement in *NKX2-5* transcriptional regulation, a TEAD1 expression vector was co-transfected into primary HPASMCs together with the minimal and upstream promoter constructs and luciferase assays were performed. TEAD1 co-transfection significantly induced luciferase expression independent of the construct (Figure 5.7), and the same effect was observed in both contractile and synthetic HPASMC.

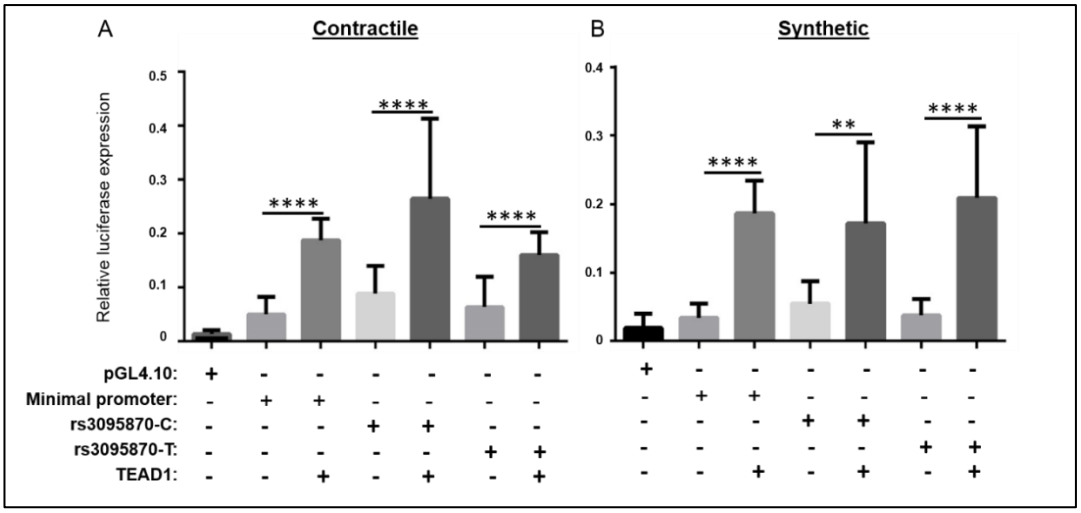


Figure 5.7 TEAD1 enhances transcriptional activity of NKX2-5 upstream promoter. *Contractile (A) and synthetic (B) primary HPASMC were transfected with reporter vectors containing the firefly luciferase gene driven by the minimal promoter of NKX2-5 (minimal promoter), the upstream promoter including the rs3095870-C allele (rs3095870-C), the upstream promoter including the rs3095870-T allele (rs3095870-T). The TEAD1 expression vector was co-transfected with the reporter vectors. Luciferase assays were performed, and luciferase activity was measured and is presented as the ratio of firefly/renilla expression. The pGL4.10 reporter vector was used as negative control. N=3 (replicate experiments). **p≤0.01, ****p≤0.0001.*

5.2.1.3.2 Overexpression of TEAD1 and TEAD3

ImHPASMCs were transfected with TEAD1 and TEAD3 expression vectors to evaluate any potential direct effect of the proteins on NKX2-5 expression. Increasing amounts of DNA ranging from 50ng, 100ng, to 250ng per transfection were used. The cells were transfected efficiently and effectively as shown from the dose-dependent increase in TEAD1 and TEAD3 RNA and protein (Figure 5.8).

Overexpression of TEAD1 or TEAD3 did not significantly affect *NKX2-5* gene (Figure 5.8A) or protein (Figure 5.8B) expression. However, there is an increasing trend in NKX2-5 protein expression upon TEAD3 overexpression. On the contrary, TEAD1 overexpression causes an insignificant decrease in NKX2-5 protein levels.

This experiment, which only performed once, overall led to inconclusive data regarding a potential direct regulatory effect of TEAD1 and TEAD3 on NKX2-5 expression. The data so far indicate that NKX2-5 transcription is under multiple regulatory mechanisms, and more experiments are required to elucidate the transcriptional machinery.

Apart from NKX2-5 expression, the SDS-PAGE/Western blot analysis provided some valuable information about the specificity of the antibodies against TEAD1 and TEAD3. The TEAD3 antibody is not strong enough to detect the endogenous

TEAD3 protein, and also detects a non-specific band exactly above the transfected protein. The TEAD1 antibody detects 3 bands around 49KDa. The middle band corresponds to TEAD1 protein and both the endogenous and transfected proteins are identified very efficiently. The top band of the triplet corresponds to TEAD3 protein, whereas the band at the bottom might be TEAD4 based on the protein homology. All 4 members of the family share great homology, and in particular TEAD1 and TEAD3 are more than 70% similar.

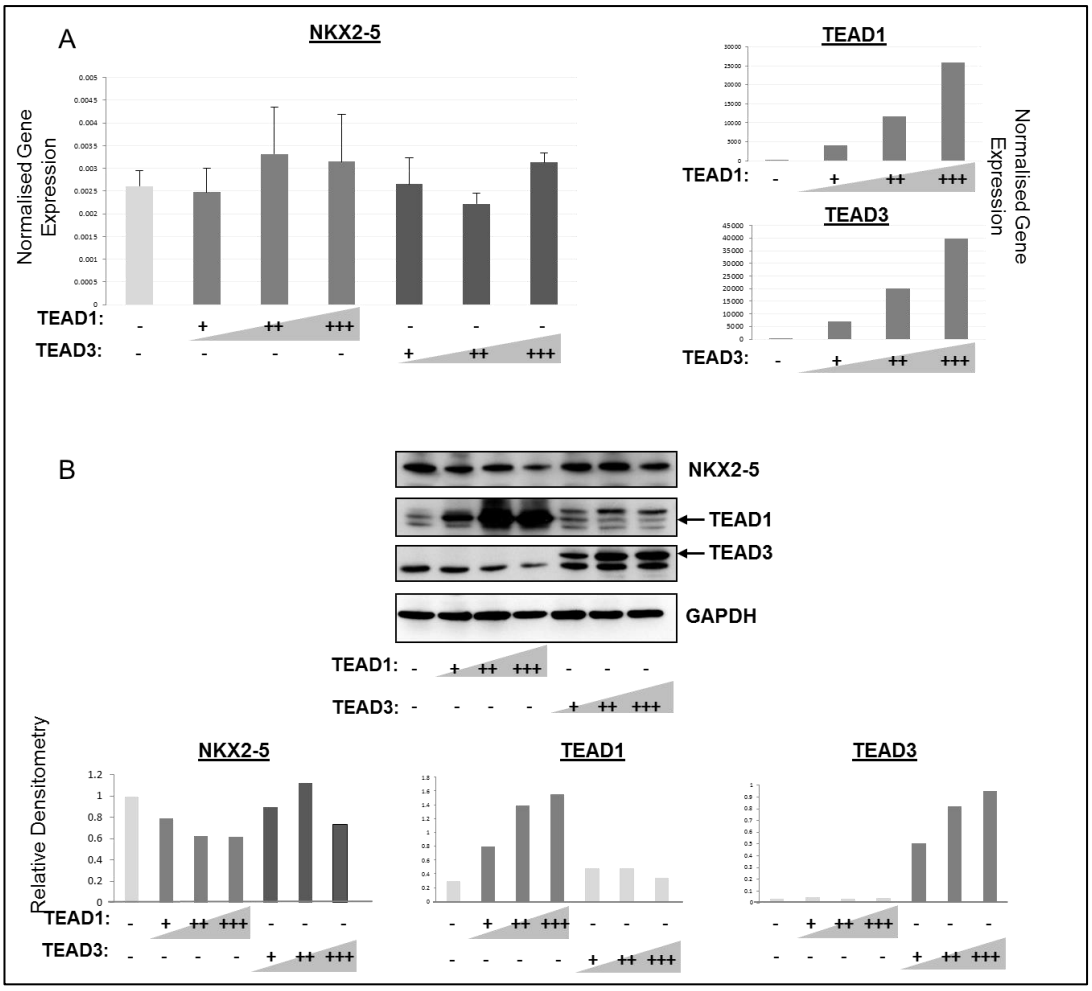


Figure 5.8 Overexpression of TEAD1 and TEAD3. *TEAD1 and TEAD3 expression vectors were transfected into ImHPASMCs and the effect on NKX2-5 expression was assessed. Different amounts of DNA were used for each transfection (+:50ng, ++:100ng, +++:250 ng). The cells were harvested 24 hours after transfection, and total protein and RNA were extracted and subjected to SDS-PAGE/Western blot and qPCR analysis. A. Gene expression of NKX2-5, TEAD1 and TEAD3 was analysed by qPCR. B. Protein expression of NKX2-5, TEAD1 and TEAD3 were analysed and densitometry was performed in K=ImageJ relative to GAPDH expression (housekeeping protein). N=1 (1 experiment performed; the qPCR was performed in triplicate samples).*

5.2.1.3.3 Knock-down of TEAD1 and TEAD3 using siRNA

Since overexpression of TEAD1 and TEAD3 only provided ambiguous data, I next used RNA silencing technology to knock down the proteins with siRNA molecules. The silencing of TEAD1 and TEAD3 proteins was successful (Figure 5.9). At the RNA level, *NKX2-5* was significantly decreased only when TEAD3 protein was knocked down (Figure 5.9A). However, NKX2-5 protein was downregulated when both TEAD1 and TEAD3 were knocked down (Figure 5.9B). Although the decrease in NKX2-5 protein was consistent in all replicate experiments, the difference did not achieve statistical significance.

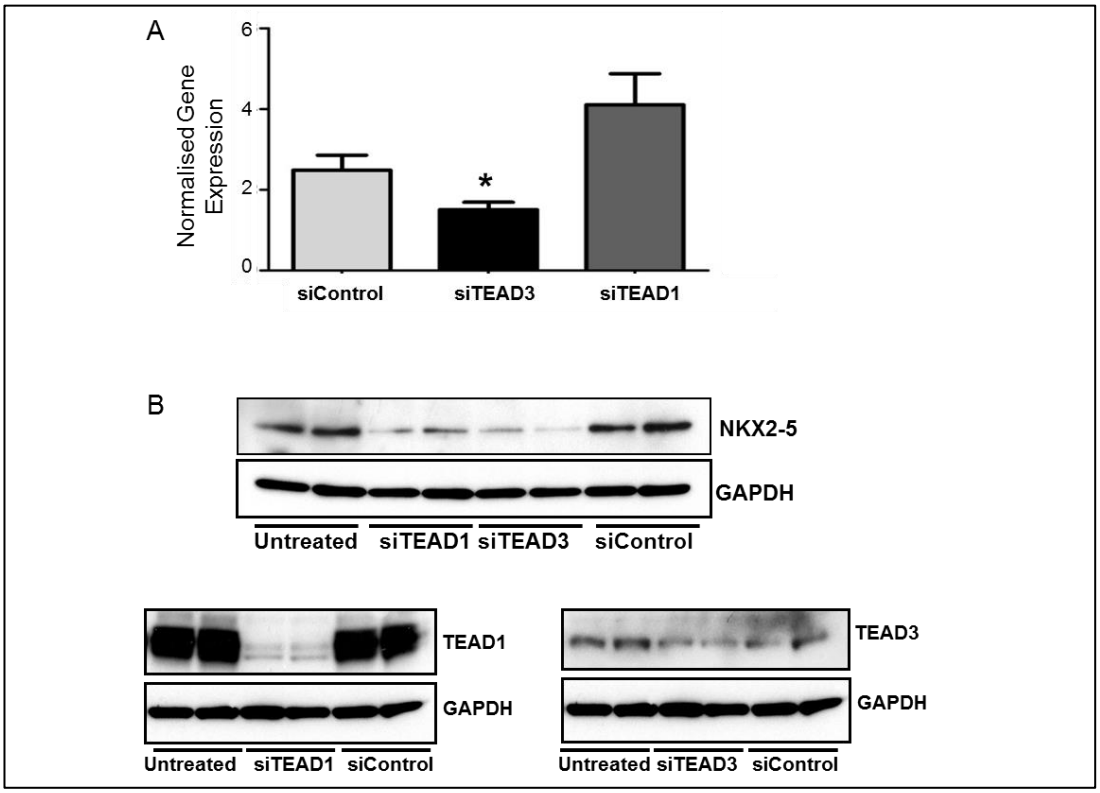


Figure 5.9 Knock-down of TEAD1 and TEAD3. *TEAD1 and TEAD3 were knocked down in ImHPASMCs using specific siRNA oligonucleotides at 100nM final concentration for 72 hours. After treatment total protein and RNA were extracted and subjected to Western blot and qPCR analysis, respectively. A. NKX2-5 gene expression was analysed by qPCR after knock down of TEAD1 and TEAD3. B. NKX2-5 protein expression was analysed after TEAD1/3 knock down. Protein levels of TEAD1 and TEAD3 were also analysed to evaluate the efficiency of the knock down. N=3 (replicate experiments). *p≤0.05.*

5.2.1.3.4 TEAD3 alone is not sufficient to enhance transcriptional activation of NKX2-5

In an attempt to elucidate further the role of TEAD3 on transcriptional regulation of *NKX2-5*, a TEAD3 expression vector was co-transfected together with the constructs of minimal and upstream promoter into ImHPASMCs. Although, there was a trend of increased transcriptional activation in the presence of TEAD3, the effect was not significant for any of the constructs (Figure 5.10). When, the results are compared to the effect of TEAD1 co-transfection on luciferase expression (Figure 5.8), TEAD3 seemed to have no effect overall.

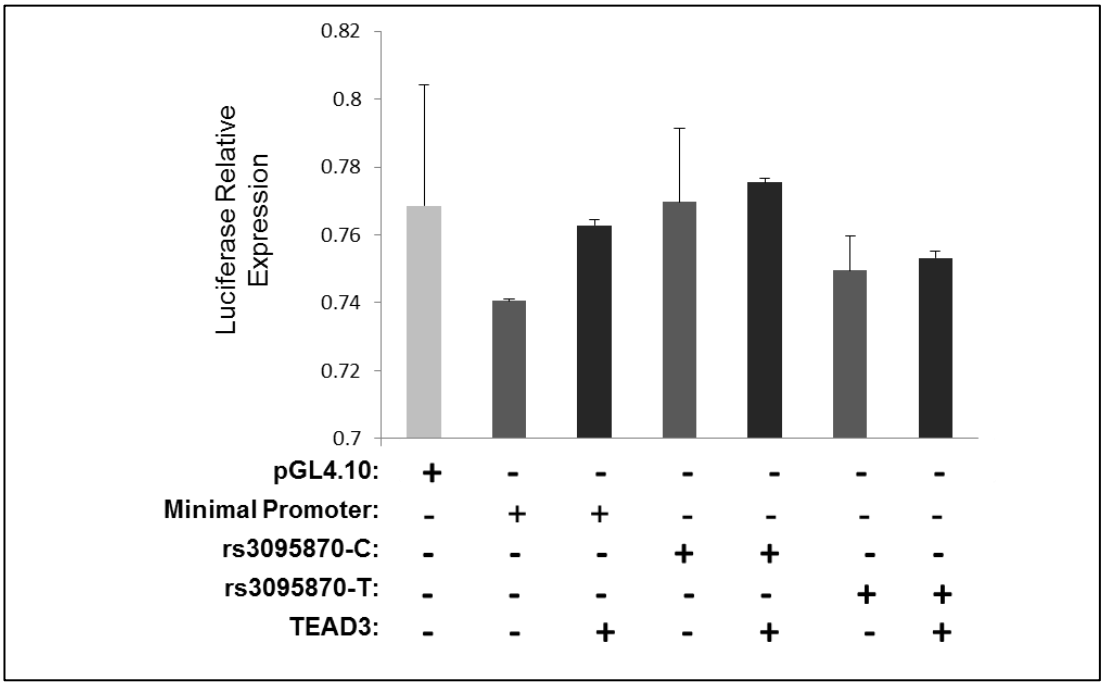


Figure 5.10 TEAD3 is not able to enhance transcriptional activity. *ImHPASMCs were transfected with reporter vectors containing the firefly luciferase gene driven by the minimal promoter of NKX2-5 (minimal promoter), the upstream promoter including the rs3095870-C allele (rs3095870-C), or the upstream promoter including the rs3095870-T allele (rs3095870-T). TEAD3 expression vector was co-transfected with the reporter vectors. Luciferase assays were performed, and luciferase activity was measured and presented as the ratio of firefly/renilla expression. The pGL4.10 reporter vector was used as negative control. N=2 (replicate experiments).*

Taken together, these data suggest that TEAD3 is required for the direct transcriptional activation of *NKX2-5*, but the presence of TEAD1 is needed to mediate and enhance the transcription of the gene. On the other hand, TEAD1 is able to enhance *NKX2-5* transcription at a later time, probably after the initial gene activation. Overall, the data imply a complex transcriptional regulation where TEAD1 and TEAD3 affect NKX2-5 gene and protein levels, but the exact mechanism in not completely clear and requires further investigation.

5.2.1.4 YAP1 is part of the regulatory mechanism of TEAD1/3

Although TEADs have been extensively studied as transcriptional regulators of their downstream targets, the proteins themselves lack a defined transcription activation domain. Ectopic expression of TEAD1 did not enhance TEAD-dependent transcription (486), and overexpression of TEAD1 in cells that do not express TEAD proteins results in repression of transcription (487). These findings suggest that TEAD proteins require co-activators to exert their function. YAP1 (Yes-associated protein 1) has been identified as the best candidate required for TEAD-dependent transcription.

5.2.1.4.1 YAP1 siRNA

To investigate whether YAP1 is the protein through which TEAD1 and TEAD3 regulate NKX2-5, I used RNA silencing to knock down YAP1 protein (Figure 5.11). The experiment showed that *NKX2-5* transcriptional activation depends clearly on YAP1 expression (Figure 5.11). Both NKX2-5 RNA and protein expression levels were significantly decreased when YAP1 was silenced. These data prove unambiguously that TEAD/YAP1 complex is required for *NKX2-5* transcriptional activation.

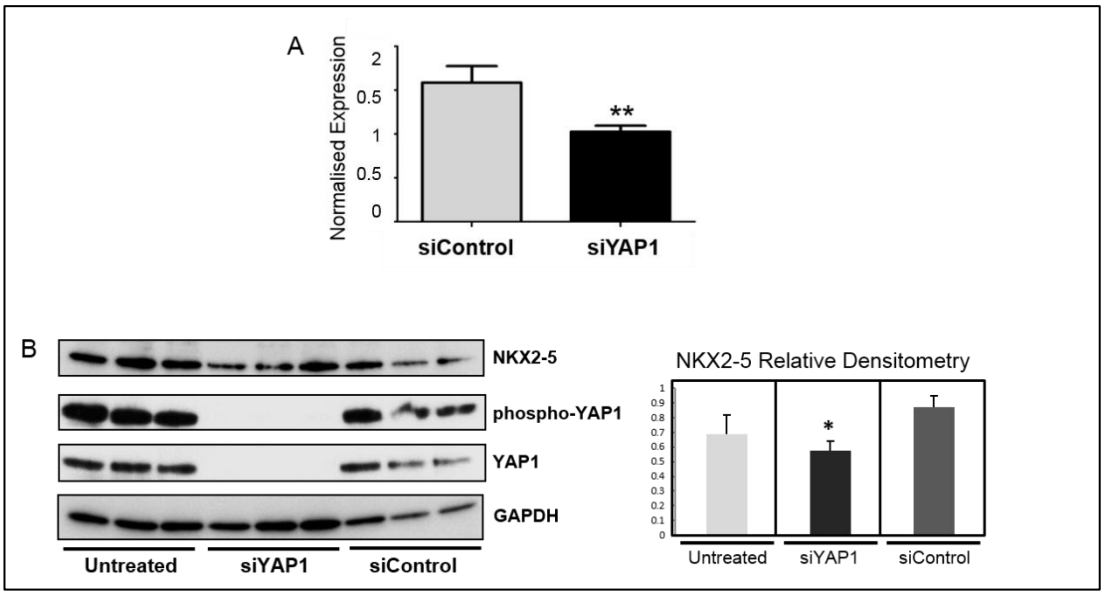


Figure 5.11 Knock-down of YAP1. *YAP1 was knocked down in ImHPASMCs using specific siRNA oligonucleotides at 130nM final concentration for 48 hours. After treatment total protein and RNA were extracted and subjected to SDS-PAGE/Western blot and qPCR analysis. A. NKX2-5 gene expression was analysis by qPCR after knocking down of YAP1 protein. B. Protein expression of NKX2-5, was analysed after YAP1 knock down. Protein levels of phosphorylated YAP1 (Phospho-YAP1) and total YAP1 were also analysed to evaluate the efficiency of the knock down. N=3 (replicate experiments). *p≤0.05, **p≤0.01.*

5.2.1.4.2 Verteporfin blocks the TEAD/YAP complex

After establishing the importance of YAP1 in the transcriptional regulation of NKX2-5, I used verteporfin, a small molecule that inhibits the downstream signalling of TEAD/YAP complexes (494). Verteporfin interferes with the physical interaction of TEAD and YAP proteins and blocks their association. It was developed and has been used as a cancer drug in different malignancies.

Verteporfin is a light-sensitive benzoporphyrin derivative and can be toxic for cells, especially for sensitive primary cells such as HPASMCs. A cell viability assay (MTS) was performed with different concentrations of verteporfin to evaluate its cytotoxic effect in primary HPASMC (Figure 5.12A).

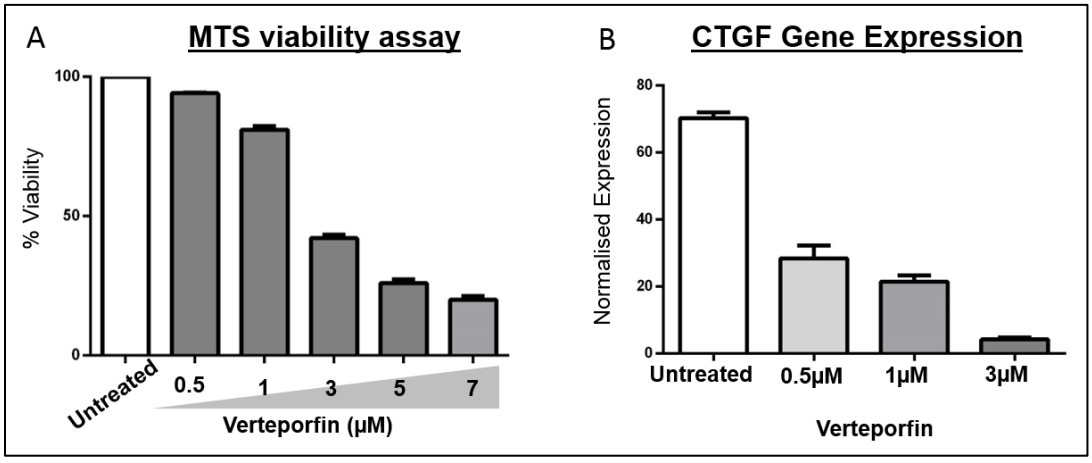


Figure 5.12 Cytotoxicity and effectiveness of Verteporfin. *A. A cell viability (MTS) assay was performed to evaluate the cytotoxic effect of verteporfin on primary HPASMC. Cells were treated with verteporfin: 0.5μM, 1μM, 3μM, 5μM, 7μM for 4 hours, and the absorbance measured at 490nm. B. The less cytotoxic doses were then used to check their potency on the gene expression of CTGF, a known downstream target. Gene expression was assessed by qPCR normalised to TBP expression. N=2.*

Subsequently, to establish how effective the lower doses of 0.5μM, 1μM, and 3μM are, primary HPASMCs were treated with verteporfin in a dose-dependent manner and gene expression levels of *CTGF* were assessed by qPCR (Figure 5.12B). *CTGF* is a known downstream target of TEAD1/YAP1 complex, and verteporfin has been used to inhibit its expression (492). qPCR data showed that increasing concentrations of verteporfin suppressed *CTGF* expression in a dose-dependent manner (Figure 5.12B). Based on these results, the concentration of 1μM verteporfin was selected and used for the experiments.

Primary HPASMC were stimulated with TGF-β, transfected with the TEAD1 expression vector and treated with verteporfin as indicated in Figure 5.13. At the protein level, TGF-β induced NKX2-5 expression in both contractile and synthetic cells, and the effect was maximised by the transfection of TEAD1 (Figure 5.13A). When verteporfin was used, NKX2-5 expression was decreased and reversed the effect of TEAD1, confirming once more the essential role of TEAD1/YAP1 complex on NKX2-5 expression.

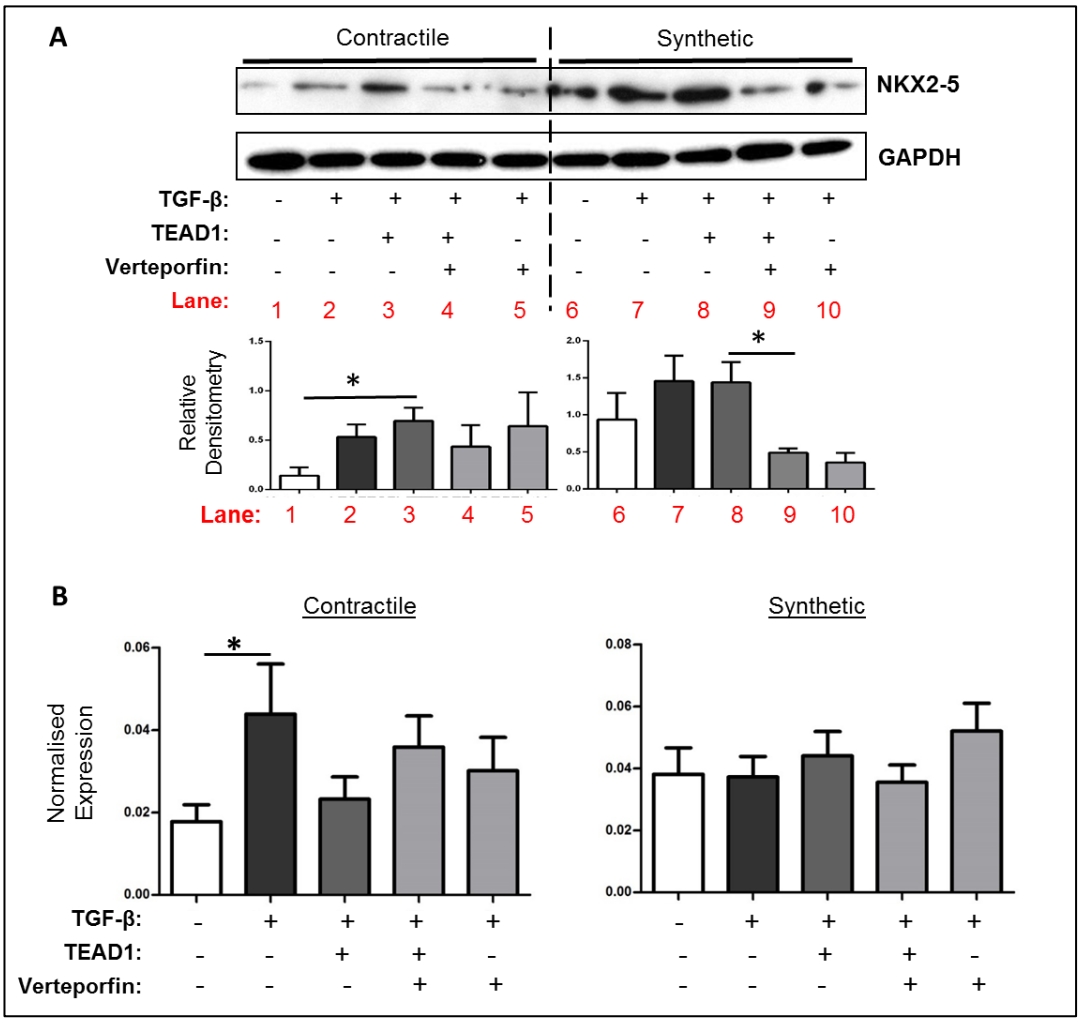


Figure 5.13 Verteporfin interferes with TEAD1/YAP1-mediated NKX2-5 regulation. *Primary HPASMC were cultured under conditions favouring the contractile or the synthetic state. After overnight serum starvation in 0.1% FCS, the cells were either stimulated with 2ng/ml TGF-β, transfected with TEAD1 expression vector or treated with 1μM of verteporfin as indicated. After 24 hours of treatment, total protein and RNA were extracted and subjected to Western blot and qPCR analysis, respectively. A. NKX2-5 protein expression was analysed for the different treatments, and densitometry analysis was performed in ImageJ normalised to GAPDH expression. B. NKX2-5 gene expression was analysed by qPCR normalised to TBP expression. N=3 (replicate experiments). *p≤0.05.*

However, the *NKX2-5* mRNA levels did not correlate with the protein expression. In particular, transfection of TEAD1 did not increase NKX2-5 expression and verteporfin had no effect on expression (Figure 5.13B).

Overall, the data support a role for the TEAD1/YAP1 complex in the regulation of NKX2-5, but the inconsistency between the gene and protein expression needs to be further explored.

5.2.2 Transcriptional regulation through the enhancer

SNPs rs3132139 and rs3131917 were associated with PAH in the discovery cohort and with scleroderma in the meta-analysis, respectively (Chapter 3). The *in silico* analysis using Haploreg, suggested that both SNPs could be functional and they might be near promoter or enhancer regions. The SNPs are located between 2.1-3.4Kb downstream of *NKX2-5*. Tagging SNP rs12514371 is found between them.

A literature review revealed a study by May *et al* (495) where a genome-wide map of predicted cardiac enhancers in the human genome was generated. Briefly, the occupancy profiles of two enhancer-associated co-activator proteins were determined in human foetal and adult heart by ChIP using antibodies specific to p300 and CBP co-activator proteins. The results were categorised by chromosomal region and the "rs" names of the SNPs that reside in each region were also given. On chromosome 5, a genomic region 2.7Kb long downstream of *NKX2-5* is characterised as a putative NKX2-5 enhancer in the heart. The region contains all the 3 SNPs (rs3132139, rs12514371, rs3131917) identified in Chapter 3 as disease-associated polymorphisms.

This enhancer must be activated during heart development when NKX2-5 exerts its unique and non-redundant role. However, I hypothesised that the same mechanism that activates NKX2-5 in embryogenesis, may also activate the gene in diseased vessels in adulthood. Consequently, I considered this region downstream of NKX2-5 as a putative enhancer and further explored this hypothesis.

5.2.2.1 Transcriptional activity of the enhancer

The genomic locus containing the SNPs was cloned into the pGL4.10 reporter vector upstream of the minimal promoter as described in section 2.6. The constructs were transfected into primary HPASMC and luciferase assays were performed.

The construct containing the putative enhancer region increased luciferase activity significantly compared to the minimal and the upstream promoter constructs, confirming that the region is a functional enhancer (Figure 5.14).

As explained in Chapter 2, two constructs were generated based on the associated risk alleles and the haplotypes physically occurring in the populations. The “not-risky” enhancer construct was also able to increase luciferase expression, but not as efficiently as the “risky” construct.

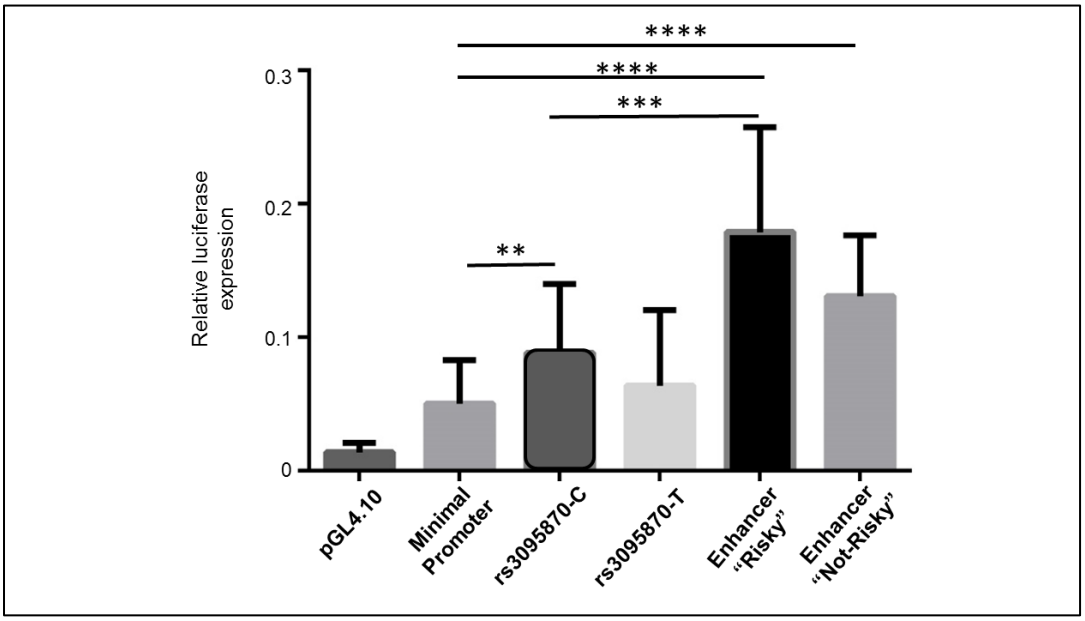


Figure 5.14 The associated SNPs downstream of NKX2-5 are located in a functional enhancer. *Primary HPASMC were transfected with reporter vectors containing the firefly luciferase gene driven by the minimal promoter of NKX2-5 (minimal promoter), the upstream promoter including the rs3095870-C allele (rs3095870-C), the upstream promoter including the rs3095870-T allele (rs3095870-T), the downstream putative enhancer containing the “risky” allele combination associated with disease, and the downstream putative enhancer containing the “not-risky” allele combination. Forty-eight hours after transfection the cells were lysed and luciferase assays were performed. Luciferase activity was measured and presented as the ratio of firefly/renilla expression. The pGL4.10 reporter vector was used as negative control. N=3 (replicate experiments). **p≤0.01, ***p≤0.001, ****p≤0.0001*

5.2.2.2 Transcription factors binding at the enhancer

An *in silico* analysis (TRANSFAC, JASPAR) revealed numerous binding sites for multiple proteins known to regulate transcription through enhancer regions during heart development and SMC differentiation such as GATA6, Mef-2c, c-Jun. Since enhancer effects are usually cell type-specific, I performed ChIP assays in ImHPASMCs stimulated with TGF-β using antibodies specific to GATA6, MEF-2c, c-JUN, Phospho-SMAD3 and RNA polymerase II. The results are presented as fold-enrichment for the binding of each transcription factor in Figure 5.15.

Surprisingly, a significant enrichment was found for MEF-2c, GATA6 and c-JUN at the 5’ end of the enhancer region. Phospho-SMAD3 showed significant binding across the entire region suggesting that TGF-β could also activate the enhancer to

positively regulate NKX2-5 expression. In addition, an interesting finding was the strong enrichment of RNA polymerase II binding towards the first half of the region, where most of the proteins were also bound, suggesting that the enhancer is engaged with the transcriptional machinery. These data signify the downstream genomic locus as a functional enhancer that activates *NKX2-5* transcription in HPASMC through the binding of GATA-6, MEF-2c and c-JUN.

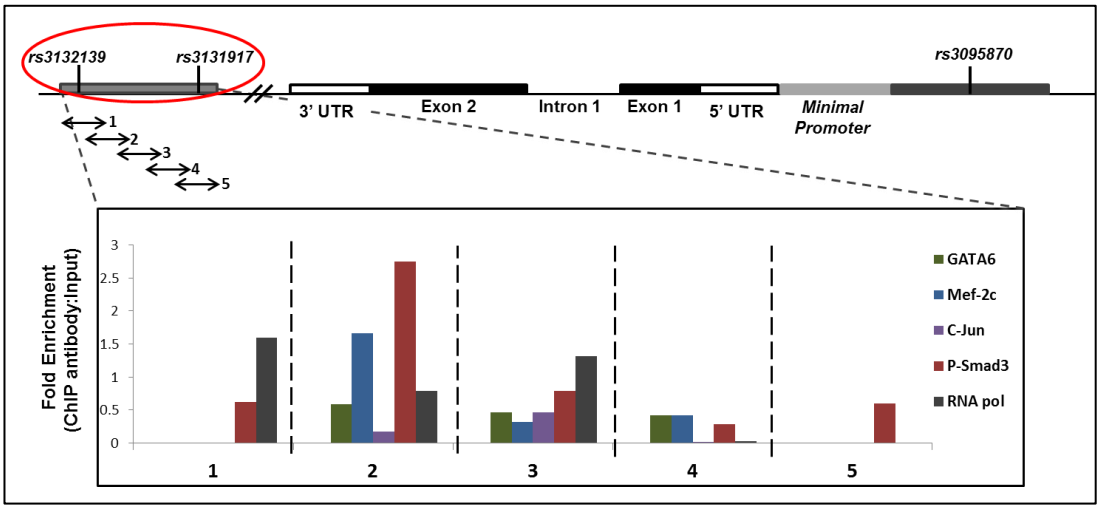


Figure 5.15 Binding of transcription factors in the functional enhancer. *ImHPASMCs were treated with TGF-β for 16 hours before the transcription factors were cross-linked to the chromatin. The chromatin was then sheared by sonication into DNA fragments of 500bp-1Kb in length. Chromatin was immunoprecipitated with antibodies specific to GATA-6, c-JUN, MEF-2c, phospho-SMAD3 and RNA polymerase II. Immunoprecipitation with an IgG antibody was used as a negative control. The enhancer region was divided in 5 DNA fragments (1-5) that were amplified by PCR. The PCR products were run in an agarose gel and the results for the binding of each factor are presented as a ratio of fold-enrichment to the input.*

5.2.2.3 TGF-β is not sufficient to activate the enhancer or the promoter independently

Careful evaluation of the ChIP experiments conducted in the promoter and the enhancer regions of *NKX2-5* revealed that enrichment of phospho-SMAD3 binding was prominent in both cases. These data are in agreement with the findings in Chapter 4 showing that TGF-β activates NKX2-5 expression through phosphorylation of SMAD3.

To examine whether TGF-β induces the transcriptional activity of the constructs, I repeated the luciferase assays in TGF-β-treated and untreated ImHPASMCs (Figure 5.16). The results were inconclusive as TGF-β did not affect the luciferase expression for any of the constructs. Although the experiment was only performed once, the data did not suggest an overall effect of TGF-β.

A possible explanation might be that there is more than one TGF-β-responsive elements at both the upstream promoter and the enhancer that act synergistically to promote the effect of the stimulation. Another explanation is that TGF-β might function through an element found in another region that is not included in the constructs.

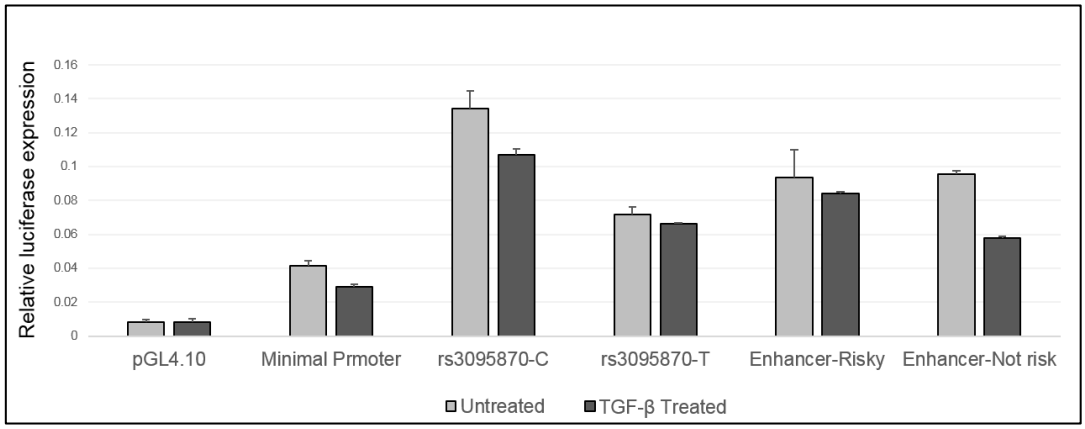


Figure 5.16 TGF-β stimulation did not induce transcriptional activity of the promoter or the enhancer constructs. *Luciferase assays were repeated as for Figure 5.14.24 hours after transfection, the cells were stimulated with 2ng/ml TGF-β for an extra period of 24 hours. The cells were then lysed and luciferase assays were performed. N=1 (one experiment performed in duplicate)*

5.2.3 Post-transcriptional regulation of NX2-5 through the 3'UTR

The tagging SNP rs703752 is located within the 3' untranslated region (3'UTR) of *NKX2-5* gene, 63bp after the last coding exon. The SNP did not show any evidence of association, but the 3'UTR region itself could conceal a powerful mechanism of post-transcriptional regulation. The untranslated regions at the 3' end of the genes are regulatory regions that may influence polyadenylation, translation efficiency, localisation, and stability of the mRNA. They usually contain binding sites for regulatory proteins as well as microRNAs [reviewed in (496)]. In addition, the physical characteristics of the 3'UTR such as the length and secondary structures can contribute to post-transcriptional regulation. These diverse mechanisms of gene regulation confer spatio-temporal specificity to gene transcription.

MicroRNAs are small non-coding single-stranded RNA molecules (~23bp long) encoded by specific genes or introns, they are produced in the nucleus and undergo a maturation process until they are released to the cytoplasm as functional molecules. MicroRNAs bind specific sites within the 3'UTRs and decrease gene

expression of various mRNAs by either inhibiting translation or directly causing degradation of the transcript. Apart from microRNAs, repressor proteins can bind specific silencer regions within the 3'UTRs and inhibit the expression of the mRNAs. In addition, many 3'UTRs contain AU-rich elements (AREs) that are bound by proteins to affect the stability or decay rate of transcripts. Furthermore, some 3'UTRs may contain the sequence “AAUAAA” that causes the synthesis of the poly(A) tail, responsible for mRNA translation, stability, and export [reviewed in (496)].

5.2.3.1 Evaluation of NKX2-5 3’UTR region

To study and explore any functional effect of the 3’UTR on the post-transcriptional regulation of *NKX2-5* gene, the region was cloned into the pmirGLO reporter vector downstream of the firefly luciferase gene as described in section 2.6. The constructs were transfected into primary HPASMC and luciferase assays were performed (Figure 5.17). The 3’UTR construct caused a significant reduction in the luciferase expression in both the contractile and the synthetic cells. This would be explained by the binding of a microRNA on the 3’UTR that prevented the translation of the luciferase gene transcript and decreased expression. The effect was greater in the synthetic cells, and it also was independent of the allele of the SNP.

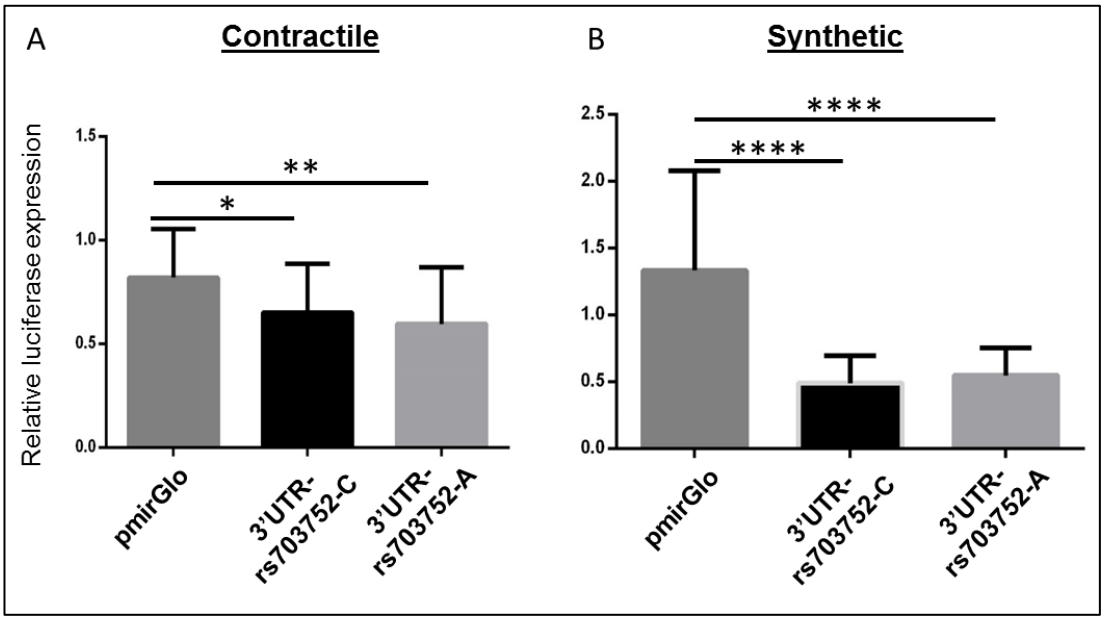


Figure 5.17 Functional properties of NKX2-5 3’UTR. *The 3’UTR of NKX2-5 gene was cloned into the pmirGlO reporter vector downstream of the firefly luciferase gene. The 3’UTR contained rs703752, a tagging SNP. Two versions of the reporter vectors were created: the 3’UTR-rs703752-C, and the 3’UTR-rs703752-A. Primary HPASMC cultured in vitro under the contractile (A) or the synthetic (B) conditions were transfected with the reported vectors. 48 hours after transfection, the cells were lysed and luciferase assays performed. N=3 (replicate experiments). *p≤0.05, **p≤0.01, ****p≤0.0001.*

5.2.3.2 Prediction of microRNA binding at the 3'UTR

To examine the possibility that a microRNA can bind the 3'UTR, an *in silico* analysis was performed using the *miRanda-mirSVR* software (497) (www.microRNA.org). Firstly, a 25bp long region surrounding rs703752 specific to either the A or the C allele was used as the seed. The software uses an algorithm to run the query sequence against all known microRNA sequences available in the database. Based on the results (maximum score and minimum energy) the best candidate microRNAs that bind specifically the A or the C allele of rs703752 were selected and listed in Table 5.1.

	rs703752-A		rs703752-C	
miR name	Score	Energy	Score	Energy
miR-629-3p	158	-21.67	150	-22.29
miR-1260a	140	-12.62	Below threshold	
miR-532-3p	140	-16.34	Below threshold	
miR-4523	145	-22.5	Below threshold	
miR-4713-5p	152	-20.84	Below threshold	
miR-4740-3p	Below threshold		145	-21.08
miR-4279	Below threshold		140	-21.61
miR-4469	Below threshold		143	-27.67
miR-1225-3p	Below threshold		142	-28.57

Table 5.1 microRNAs that bind on the rs703752 site. *A 25bp long region surrounding rs703752 was used as the seed to run a search using the miRanda software. The software uses the seed sequence to identify microRNAs that would bind the sequence based on similarity. The results are described by 2 numbers: a score that shows similarity between the microRNA and the target sequences, and the energy that shows how stable the binding would be, with higher energy scores denoting stronger bindings. The best candidate microRNAs are shown in the table, with those that bind specifically to the rs703752-A allele in red, and those that bind specifically to the rs703752-C allele in blue. miRanda only uses a sequence-based search, without giving any information regarding the expression profile of the candidate microRNAs.*

5.2.3.3 Effect of microRNA binding at the 3'UTR

The first microRNA tested was the hsa-miR-629-3p, which was at the top of the list (Table 5.1), without exhibiting specificity for either allele (A or C). When hsa-miR-

629-3p-mimic was co-transfected together with the 3'UTR constructs, the luciferase expression was significantly increased in both contractile and synthetic cells (Figure 5.18). Transfection of the hsa-miR-629-3p-antagomir brought the expression back down, neutralising the effect. These results suggested that miR-629-3p might act as an antagomir competing with another microRNA that binds the region.

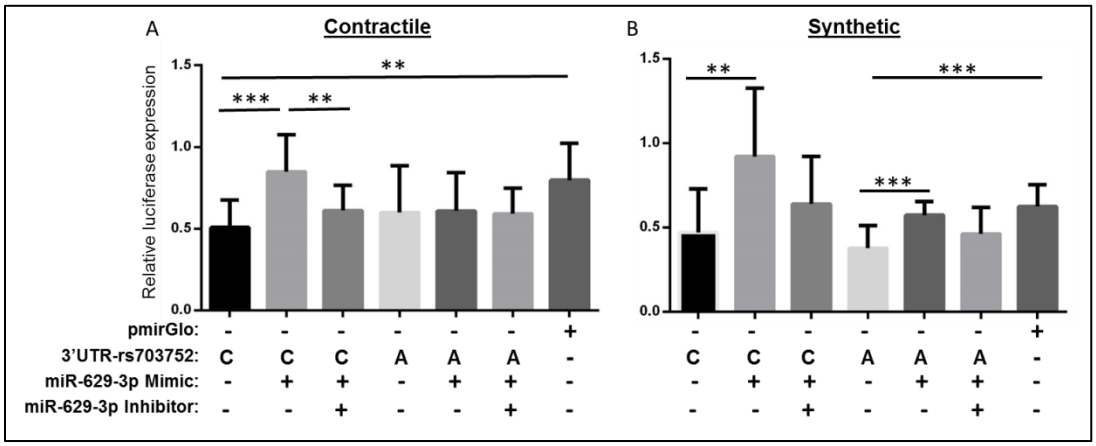


Figure 5.18 Effect of miR-629-3p in primary HPASMC. *The 3'UTR reporter vectors containing either the C or the A allele of rs707352 were transfected into primary HPASMC cultured under contractile (A) or synthetic (B) conditions. The miR-629-3p mimic molecule and inhibitor were co-transfected together with the reporter vectors. The mimic molecule was used at 10nM and the inhibitor at 50nM. Forty-eight hours after transfection the cells were lysed and luciferase assays were performed. N=3 (replicate experiments). **p≤0.01, ***p≤0.001.*

Next, hsa-miR-532-3p, which based on the predictions binds specifically the rs703752-C allele, was co-transfected into primary HPASMC (Figure 5.19). No significant effect was seen in the contractile cells. However, in synthetic cells, transfection of the 3'UTR construct significantly decreased luciferase expression overall, and co-transfection of the inhibitor released luciferase expression in the presence of the A allele. However, as with miR-629-2p, transfection of the mimic molecule did not cause a decrease in luciferase expression suggesting that there is no specific effect.

After failing to identify any functional microRNA that could potentially control NKX2-5 post-transcriptionally using miRanda software, I performed a literature search for candidate microRNAs that a) target NKX2-5 3'UTR, b) are expressed in VSMC, and c) exhibit higher levels of expression in PAH (GSE55427). Based on these criteria, a good candidate microRNA was the hsa-let-7i-3p. Let-7i-3p was co-transfected into

primary HPASMC as before (Figure 5.20). Although there were differences between contractile and synthetic cells, none of the effects was significant.

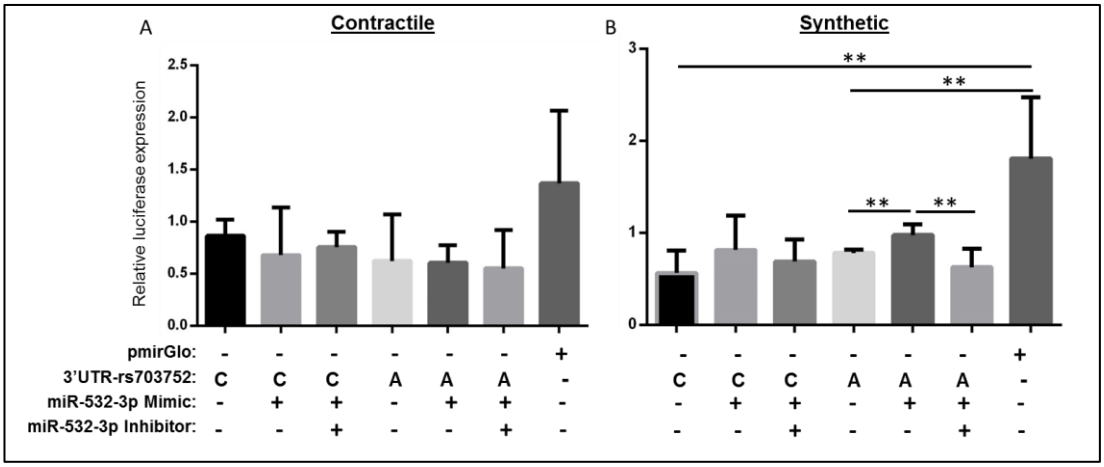


Figure 5.19 Effect of miR-532-3p in primary HPASMC. *The 3'UTR reporter vectors containing either the C or the A allele of rs707352 were transfected into primary HPASMC cultured under contractile (A) or synthetic (B) conditions. The miR-532-3p mimic molecule and inhibitor were co-transfected together with the reporter vectors. Mimic molecule was used at 10nM and the inhibitor at 50nM. Forty-eight hours after transfection the cells were lysed and luciferase assays were performed. N=3 (replicate experiments). **p≤0.01.*

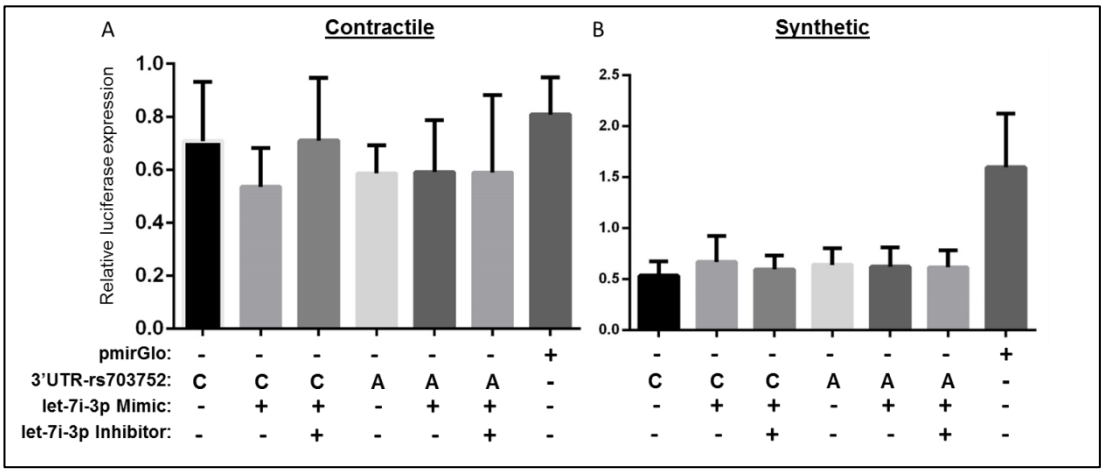


Figure 5.20 Effect of let-7i-3p in primary HPASMC. *The 3'UTR reporter vectors containing either the C or the A allele of rs707352 were transfected into primary HPASMC cultured under contractile (A) or synthetic (B) conditions. The miR-532-3p mimic molecule and inhibitor were co-transfected together with the reporter vectors. The mimic molecule was used at 10nM and the inhibitor at 50nM. Forty-eight hours after transfection the cells were lysed and luciferase assays were performed. N=3 (replicate experiments).*

5.2.4 Methylation studies

Epigenetic mechanisms have been associated with CVD such as atherosclerosis and vascular inflammation. Cytosine methylation and hydroxymethylation, and histone modification are involved in gene regulation in VSMC during adulthood and embryogenesis and some examples are summarised in Webster *et al* (257). DNA methylation has also been associated with cancer. In particular, epigenetic inactivation of genes in cancer cells is based on transcriptional silencing by aberrant CpG methylation of CpG-rich promoter regions (498). Age-related methylation has been studied extensively in prostate cancer and differentially methylated genes can be used as markers for early diagnosis and disease risk assessment (499, 500). In these studies, *NKX2-5* methylation levels were 3-fold higher in prostate than normal tissues, and *NKX2-5* had been proposed as a good marker for distinguishing prostate cancer tissues (500).

The *NKX2-5* genomic locus is rich in CpG islands, with 6 of them spread around the gene (Figure 5.21). Many of the CpG dinucleotides of these islands have been included in methylation arrays, and there is evidence that the locus is unmethylated in aortic SMCs. Methylation is strongly related with disease by preventing transcriptional activation of methylated loci. For example, in a methylation study of patients with Tetralogy of Fallot, *NKX2-5* was found to be methylated, and decreased levels of expression were verified by qPCR (501).

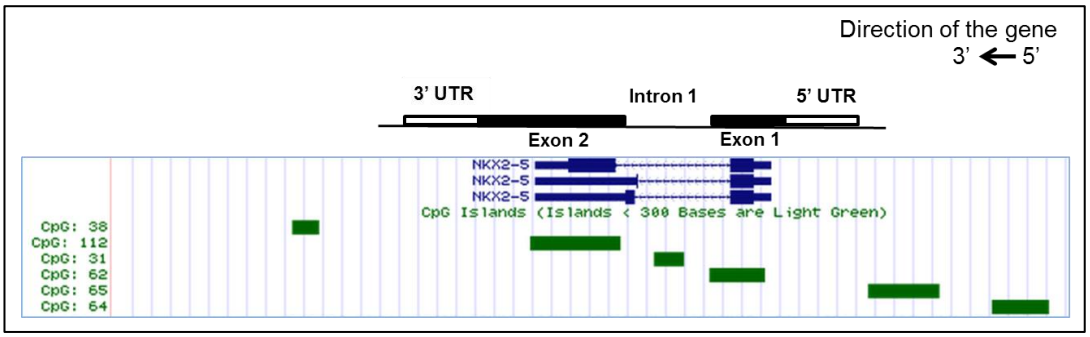


Figure 5.21 CpG islands around NKX2-5 gene. *Six CpG islands are located around the NKX2-5 gene and shown in green (numbers:38, 112, 31, 62, 65, 64). Figure is adapted from the UCSC Genome Browser, GRCh38/hg38 Assembly.*

CpG island 38 is located in the middle of the downstream enhancer. Taking into account the provided data associating NKX2-5 expression with methylation, one of the initial aims of this project was to explore methylation as another epigenetic mechanism of NKX2-5 transcription. This part of the project was not completed due to lack of time, however, it is very important for future studies.

5.3 Summary of results

The focus of the work presented in this chapter was to unravel the mechanism(s) that regulate *NKX2-5* gene expression at the transcriptional, post-transcriptional and epigenetic levels. The findings emerged from the genetic association study were used as tags for potential sites of functional importance, and the associated SNPs were studied thoroughly. SNPs rs3132139 and rs3131917 that were associated with PH and scleroderma, respectively, are located downstream of *NKX2-5* in a region that was described as putative enhancer of *NKX2-5* in the heart tissue. SNP rs3095870 was located upstream of the minimal promoter of *NKX2-5* and could serve as an upstream promoter. Supporting evidence were added through an *in silico* study, which showed that all the SNPs can possibly possess some functional properties.

The genomic region surrounding rs3095870 was considered as a putative upstream promoter and was cloned into a reporter vector to perform luciferase assays. Indeed, the region significantly induced transcriptional activity. A MCAT consensus binding element that is recognised by the TEAD family of transcription factors coincides with rs3095870. Binding assays showed that TEAD1 protein was able to bind specifically only the rs3095870-C allele, and co-transfection of TEAD1 further increased the transcriptional activity of the upstream promoter, suggesting that TEAD1 is a transcriptional enhancer of NKX2-5 through binding at the rs3095870 site.

Since all the TEAD family members recognise and bind the MCAT consensus site with the same efficiency, TEAD1 might not be the only player in the regulation of *NKX2-5*, and TEAD3 protein was also studied. RNA silencing of TEAD1 and TEAD3 proteins using siRNA showed that both proteins are important for the transcriptional regulation of *NKX2-5*. The data also suggested that TEAD3 is required for the initial transcriptional initiation of *NKX2-5* but alone is not able to further enhance transcription, implying that TEAD1 and TEAD3 co-operate to regulate transcription. Silencing of YAP1 expression using siRNA confirmed the EMSA results and showed that YAP1 is a co-factor for TEAD1 and TEAD3, and together they regulate NKX2-5 expression.

Luciferase assays showed that the downstream putative enhancer is a functional enhancer where multiple transcription factors including GATA-6, c-JUN, and MEF-2c bind to further regulate *NKX2*-5 transcription. ChIP assays showed a strong enrichment of phospho-SMAD3 protein validating the results of the previous chapter

that TGF-β activates NKX2-5 expression. However, the exact mechanism requires further investigation. Data from the ChIP experiments also suggested that the enhancer region as well as the upstream promoter are participating directly in the transcriptional machinery of *NKX2-5*, since the binding of the transcription factors coincides with the binding of RNA polymerase II.

Regulation of *NKX2-5* at post-transcriptional and post-translational levels also needs further investigation. In this chapter, the 3'UTR of *NKX2-5* gene was tested for potential binding of microRNAs that could possibly regulate translation. Three candidate microRNAs were tested including miR-629-3p, miR-532-3p and let-7i. None of the candidate microRNAs seemed to regulate transcriptional activity, but the 3'UTR alone showed evidence of functionality, which requires further investigation. Methylation has also been found to regulate NKX2-5 in prostate cancer and in disease related to heart malformations. Taken together, these findings suggest that epigenetic modification could also affect NKX2-5 expression and this could be explored in future.

5.4 Discussion

In the past 10-15 years, scientists have been focused on resolving the genetic code and dissecting the genetic background of simple Mendelian or complex diseases; a period usually referred to as the "genomic era". Tremendous progress has been achieved not only in unravelling genetic information but also giving meaning to it by relating genetics with function. However, the second part has been proved more challenging, as it requires more time, appropriate experimental design and critical thinking. To aid this aim, many powerful databases and browsers have been developed that are used as information depositories, where information inferred by experimental work or based on the literature is stored, and is easily and usually freely available. Examples of such browsers and databases are: NCBI, Ensembl, UniProt, UCSC Genome browser, ENCODE, 1000 Genome Project, Gene Cards, TRANSFAC, Haploreg, JASPAR, GTEx Portal, etc. All of the above engines were used in this study.

SNP rs3095870 is located upstream of the minimal promoter of *NKX2-5* within a region that acts as an upstream promoter and increases transcriptional activity more in the presence of the disease-associated C allele compared to the alternative allele (Figure 5.2). Interestingly, the effect was different between contractile and synthetic HPASMCs, revealing that the two phenotypically different cell types behave differently in regards to gene expression. In general, contractile HPASMCs appear

to be more responsive to stimuli than synthetic cells. An explanation could be that contractile cells retain a stabilised function until a stimulus triggers them to undergo phenotypic modulation, whereas synthetic cells have already changed their expression profile in response to different stimuli. In addition, synthetic cells might exert different responses depending on the stage during the phenotypic modulation. Expression and responsiveness might also depend on the basal levels of expression found in each individual cell, which is also different in between primary cells.

The binding assays performed in TGF-β-treated ImHPASMCs confirmed the binding of TEAD1 transcription factor specifically at the major risk-associated C allele of rs3095870 (Figures 5.4, 5.5, 5.6), which was predicted through the *in silico* analysis (Figure 5.3). The evidence for the binding was strong as it was reproducible. The use of appropriate controls further endorsed the specificity of the assays. Although, ChIP assays are considered technically laborious, in this case performing the pull-down assays proved more challenging for two main reasons: optimisation of a) salt concentration in the solution and b) detergent concentration in the washes to enhance binding, stabilise the complex, and increase specificity without disrupting the protein-DNA association.

Co-transfection of TEAD1 showed that the protein acts as an activator for *NKX2-5* significantly increasing transcriptional activity. The effect was the same in both contractile and synthetic cells, and TEAD1 increased activity in a similar way in the presence of both minimal and upstream promoter independent of the rs3095870 allele. Further studies revealed that TEAD3 is also implicated in the transcriptional regulation of *NKX2-5*, and that YAP1 is required as a co-factor for both TEAD1 and TEAD3. Different methods were used to dissect this information such as gene reporter assays, overexpression of TEAD1 and TEAD3, knock-down of proteins using siRNA oligonucleotides, and inhibitors to block the binding of TEAD/YAP complex to DNA. Although the protein levels were not always consistent with the gene levels, the data collectively point towards a transcriptional regulatory mechanism of the *NKX2-5* gene that involves possibly more than one TEAD/YAP complexes bound at the rs3095870 SNP and elsewhere in the promoter.

The role of TEAD transcription factors have been well studied in VSMC differentiation with findings reporting TEAD-dependent expression of α-SMA. Few recent studies related both TEAD1 and TEAD3 with the disease-associated synthetic phenotype of VSMCs (149, 493). Interestingly, expression of TEAD1 was increased in arterial injury and correlated significantly with VSMC phenotypic switch

(149). Specifically, Liu *et al* showed that TEAD1 competes with myocardin for binding to the SRF elements to block expression of SMC-specific genes and promote the phenotypic switch towards the synthetic state (149). Interaction between other TEAD members and SRF has not been reported as required for promoter activity of other genes in VSMC. However, it has been reported that TEAD factors are able to mediate TGF-β-dependent gene activation (502). Interestingly, a recent study showed that 2 functional SNPs at the 9p21.3 locus, which has constantly been associated with risk of CVD and CAD, disrupt binding of TEAD3 and TEAD4 *in vitro* and *in vivo* in primary HPASMC (493). In addition, it was shown that ablation of TEAD3 binding due to the presence of the risk alleles disrupted SMAD3 binding and downstream TGF-β signalling (493).

Although TEAD proteins recognise and bind specifically to the MCAT DNA elements, they lack a significant activator or repressor function, and therefore require co-factors to regulate transcription of target genes. YAP1, a member of the Hippo pathway, is the most prominent co-factor of TEADs. Under the 'canonical' regulation of Hippo pathway, YAP1 protein is phosphorylated by the Hippo kinases and is retained in the cytoplasm to co-ordinate cytoskeletal functions. However, once the Hippo pathway is dysregulated, YAP1 is dephosphorylated and enters to the nucleus, where it associates with TEADs to regulate transcription. The role of the Hippo pathway and YAP has been extensively studied in various diseases and mainly focused on cancer as a regulator of cell cycle and cell proliferation [reviewed in (503-506)].

The critical role of the YAP1 in cardiac/SMC proliferation during cardiovascular development has been well studied and elucidated. A study in 2012 proposed the implication of YAP1 in VSMC phenotypic modulation by showing that down-regulation of Yap1 promotes VSMC contractile phenotype by up-regulating myocardin and SRF/myocardin-dependent expression of contractile genes (148). This finding was confirmed by other studies (150, 507). It has also been shown that expression of YAP1 is dramatically reduced in the aortic walls of patients with ascending aortic aneurysms, and that YAP1 downregulation in VSMC is associated with ECM disorders in the same context of disease (507). Interestingly, Wang *et al* also showed that YAP1 interacts with NKX2-5 and inhibits binding of NKX2-5 to the 5'-proximal promoter region of myocardin in cardiovascular progenitor cell linage-derived SMCs (150).

Taken together published data and the findings presented in this chapter, there is compelling evidence that TEAD/YAP1 complexes are master regulators of VSMC

phenotypic modulation mediated by NKX2-5. In particular, the finding that YAP1 can physically interact with *NKX2-5* to potentially regulate common downstream target genes suggests another possible way of function for NKX2-5. It is very interesting that most of the studies focus in different CVD, highlighting the importance of this mechanism in vascular remodelling as a general mechanism that underlies CVD. Another intriguing piece of evidence arises from the study by Speight *et al,* in which YAP1 and its counterpart TAZ, are presented as major mechanosensing regulators of organ size, contact inhibition of proliferation, contraction etc (508). This could potentially mean that YAP1 is the link between environmental mechanical inputs and initiation of NKX2-5-dependent transcription to accommodate changes. For instance, in cases of increased stiffness or stretch in the cell-cell environment, YAP1 gets dephosphorylated and activated, enters the nucleus where it associates with TEAD1 and TEAD3 to activate *NKX2-5* transcription, expression of which is further required to promote the synthetic phenotype, proliferation, migration and production of ECM.

The discrepancies seen between the protein and gene expression levels of NKX2-5 upon overexpression or knock-down of TEAD1 and TEAD3 can possibly be explained by the fact that more than one MCAT element is critical for NKX2-5 transcription. Besides the MCAT found on rs3095870, other sites can be located further upstream of *NKX2-5* outside of the genomic region that was cloned into the reporter vectors. This might explain why overexpression of TEAD3 did not affect luciferase expression (Figure 5.10), but protein knock-down decreased significantly gene expression levels of *NKX2-5* (Figure 5.9). Further experiments are required in order to elucidate the exact role of TEAD1 and TEAD3 in transcriptional regulation of *NKX2-5*.

The rest of the associated SNPs are located downstream of *NKX2-5*, in a locus that was previously identified as a putative enhancer of the gene in the heart tissue. This enhancer should be activated when NKX2-5 exerts its unique and non-redundant role during heart development. I hypothesised that the same mechanism that activates the gene in embryogenesis could also activate expression in diseased vessels in adulthood. When the region was cloned into a reporter vector, luciferase activity was significantly increased in the presence of this enhancer compared to the proximal minimal and the upstream promoter (Figure 5.14), confirming that it is a functional enhancer. ChIP assays were conducted to dissect the role of the enhancer, and surprisingly the results showed significant enrichment of transcription factors that are traditionally associated with enhancers such as MEF-2c, GATA6

and c-JUN (Figure 5.15). This result supports early studies showing that MEF-2c/NKX2-5/GATA form a positive regulatory network (347, 350, 351).

Taken together, I propose a new mechanism for the regulation of the *NKX2-5* gene in human HPASMC, as described in detail in Figure 5.22. The mechanism involves an upstream promoter activated through the binding of the TEAD/YAP1 complex, and a downstream enhancer which binds GATA6, MEF-2c and c-JUN. Enrichment of phosphorylated SMAD3 binding further confirms that TGF-β exerts its effect on NKX2-5 regulation through SMAD3 binding on multiple CAGA sites on both the upstream promoter and downstream enhancer. Evidence provided by the ChIP assays shows clearly that both regions interact with RNA polymerase II and are engaged at the transcriptional initiation machinery.

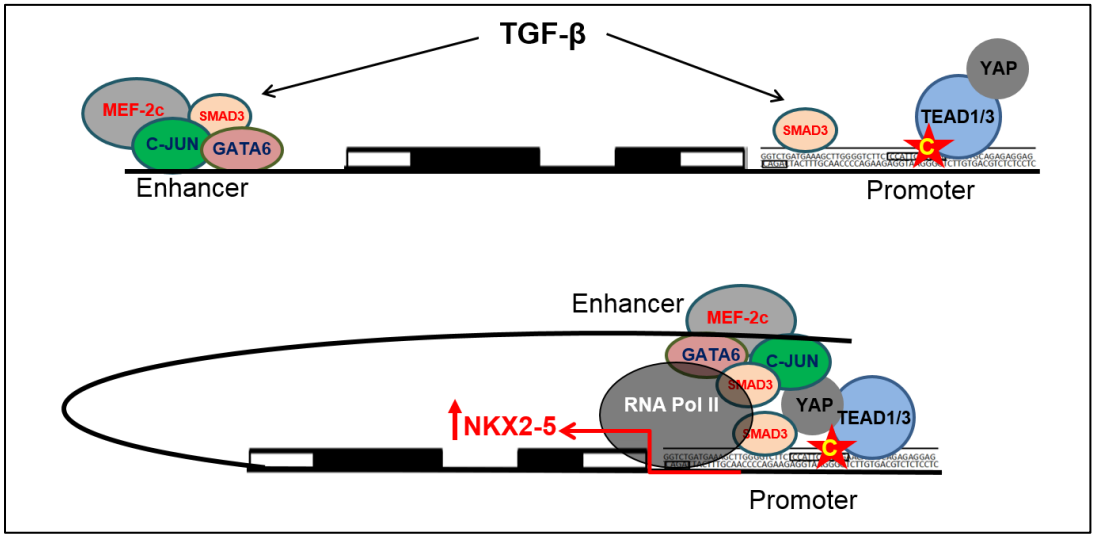


Figure 5.22 Proposed mechanism for the transcriptional regulation of NKX2-5. *Proposed model of transcriptional regulation of NKX2-5. Functional studies revealed an upstream promoter region and a novel downstream functional enhancer that are engaged with the transcriptional initiation machinery of NKX2-5 through the binding of TEAD1/3/YAP1 complex and other transcription activators such as MEF-2c, c-JUN, phospho-SMAD3, GATA6.*

This transcriptional regulatory mechanism may also apply to other conditions characterised by vascular remodelling, including but not limited to PAH, atherosclerosis, CAD, PAD and stroke.

Although the effect of TGF-β on the regulation of NKX2-5 is prominent and possibly signals through the binding of SMAD2/3, the functional CAGA sites along the *NKX2-5* genomic regions were not identified in the course of this study. However, it is absolutely essential to reveal these loci, as this would further elucidate the exact regulatory mechanism. Site-directed mutagenesis to delete individual CAGA sites,

with consequent TGF-β stimulation and performance of gene reporter assays could be one way to evaluate the effect of each individual site. Alternatively, cloning constructs could be made each consisting of a different individual CAGA site to compare the transcriptional activity of each site.

Post-transcriptional and post-translational regulation is very common for most of the genes, although it is less well studied due to each complexity. These mechanisms are usually referred to as epigenetic modifications, and they involve 3 distinct types of regulation: microRNAs, DNA methylation and histone methylation and acetylation. In the past few years, studying epigenetic modifications have become of increasing interest and have shed light on new and important pathogenic mechanisms.

Consequently, my aim was to address whether epigenetic modifications would affect the regulation of NKX2-5. Since one of the tagging SNPs, rs703752, although not associated with disease is located within the 3' UTR, I questioned whether any microRNAs could regulate expression of NKX2-5. The 3'UTR was cloned into the pmirGLO vector, which has been designed specifically for the study of microRNAs. In detail, the 3'UTR is cloned downstream of the firefly luciferase gene and upstream of the renilla luciferase gene, which serves as the internal control. A microRNA targeting the 3'UTR would bind the mRNA of the luciferase gene and block its translation, which would result in decreased firefly luciferase expression, while renilla expression would be unaffected. Cloning of *NKX2-5* 3'UTR decreased luciferase expression in both contractile and synthetic cells, suggesting putative downregulation of expression through a microRNA natively expressed in HPASMC (Figure 5.17). The effect was not dependent on the presence of rs703752, further suggesting that the functional microRNA does not bind the SNP, but binds elsewhere in the 3'UTR. In addition, the effect was more pronounced in the synthetic cells, where NKX2-5 is expressed at higher levels and therefore expected that a microRNA targeting *NKX2-5* mRNA would exert a more dramatic effect when more copies of *NKX2-5* mRNA would be present.

Three different microRNAs were tested in HPASMC for their potential effect on NKX2-5 regulation. MicroRNA mimic molecules and inhibitors were co-transfected with the 3'UTR constructs into HPASMCs. None of the three molecules exhibited a significant functional effect, which would be translated as a decrease in luciferase expression upon transfection of the mimic molecule and release of expression in the presence of the antagomir. The most plausible explanation is that the selection of the putative regulatory microRNAs was not optimal. Most of the available tools for

microRNA selection such as MiRanda (497), microRNA.org (509), miRDB (510) assess target 3'UTRs based on the sequence similarity between microRNA molecules and target sequences, known as the seed. However, microRNAs exhibit a tissue-specific expression and function (511), which is not often taken into account. GEO datasets can be useful as they provide published data on microRNA expression studies conducted in specific human tissues as well as under different treatments.

Although, microRNAs are implicated in the regulation of many genes in the cardiovascular system in adulthood or during development, there is no evidence yet reported that this is another way of regulating *NKX2*-5 gene. However, data presented in this chapter strongly suggest that *NKX2-5* can be regulated through microRNA binding at the 3'UTR, and this requires further investigation.

Another type of epigenetic modification that is known to regulate NKX2-5 expression is DNA methylation. The *in silico* study using the UCSC Genome browser revealed 6 CpG islands neighbouring the gene (Figure 5.21). Studies in prostate cancer report that *NKX2-5* methylation levels were 3-fold higher in cancer tissues compared to normal (500). Other studies reporting regulation of NKX2-5 through methylation have been conducted in the context of heart development and disease (501, 512). However, there is still no reported evidence of gene regulation by DNA methylation in vessels. It is possible that differential methylation of *NKX2-5* could occur in contractile and synthetic phenotypes of VSMCs. In addition, methylation could also be implicated in the normal silencing of the gene in most tissues after birth. Assessment of the methylation status could reveal important information on whether the *NKX2-5* downstream enhancer is active/less methylated in diseased compared to normal vessels. Overall, such a project would determine the role of the epigenetic regulation of NKX2-5 in vascular diseases.

CHAPTER 6 - RESULTS: NKX2-5 EXPRESSION IN ENDOTHELIAL CELLS

6.1 Introduction

Although the role of PASMC in PH is well-established, the importance of endothelial cells (EC) in the pathogenesis of PH has recently gained attention. Many studies currently focus on the endothelial dysfunction as the primary factor of disease initiation. Due to the loss of barrier integrity vasoactive substances, thrombotic mediators and inflammatory cytokines are released from the EC and affect PASMC causing de-differentiation to synthetic phenotype that leads ultimately to vascular remodelling and disease progression. PASMC, EC, myofibroblasts, and undifferentiated cells share equal responsibility in vascular pathologies.

Upon stimulation during vascular remodelling EC can undergo a phenotypic modulation to a mesenchymal phenotype, known as EndoMT. EndoMT is a poorly understood phenomenon that results in normal EC exhibiting phenotype plasticity and acquiring properties of myofibroblasts or mesenchymal cells. The process starts with the loss of cell-to-cell contacts, and loss of endothelial cell-specific markers such as CD31, VE-cadherin, and CD34. The cells progressively express mesenchymal markers such as α-SMA and vimentin and become more migratory.

EndoMT was initially discovered as an essential step in heart and pulmonary artery development. Since, it has been identified as a key modulator in a number of pathological conditions, including cancer, fibrosis, and CVD. More recently, the involvement of EndoMT has been studied and evaluated in PAH and SSc-PAH, neointima formation and atherosclerosis.

Many pathways have been associated with the induction of EndoMT, including autocrine and paracrine mechanisms. Inducing signalling molecules can be produced by tissue injury or immune cells recruited to the sight of injury is response to inflammation (513). The most common cytokines that induce EndoMT are members of TGF-β superfamily (TGF-β1/2, BMP2/4/6/9/10), and other signalling pathways such as Wnt/β-catenin, Notch, and various receptor tyrosine kinases [reviewed in (200)]. All of these pathways induce expression of the transcription factors Snail, Slug, Twist, LEF-1, ZEB1, and ZEB2 that are commonly used as markers of EndoMT (200). These transcription factors play an essential role in downregulating the expression of endothelial markers and proteins maintaining cell–

cell adhesion, such as integrins and focal adhesion kinase, and at the same time they upregulate proteins involved in cell migration and ECM production.

Several microRNAs have been identified as regulators of EndoMT that act through various pathways. miR-9 and miR-21 induce EndoMT through the regulation of TNFα and TGF-β respectively (200). Similarly, miR-31 controls actin remodelling and promotes the secretion of various inflammatory cytokines that induce EndoMT (514). Other positive regulators of EndoMT include bleomycin, parathyroid hormone, cell apoptosis, fluid shear stress, high glucose levels, and hypoxia associated with tissue damage, ischemia and/or inflammation [reviewed in (200)]. On the contrary, molecules that downregulate EndoMT include VEGF-A, miR-15a, miR-23b, miR-126, miR-199a, and miR-155 (200). Interestingly, BMP7 is the only member of the TGF-β superfamily that negatively regulates EndoMT (515).

FGF and TGF-β have been studied extensively as potent regulators of EndoMT. In detail, FGFR1 signalling can inhibit TGF-β induced EndoMT (516). In addition, although FGF2 has been shown to induce EndoMT in some types of endothelial cells (517), it inhibits the transition in others via a miR-20a-mediated inhibition of TGF-β signalling (518). MicroRNA let-7 has also been implicated in regulation of EndoMT. In particular, Chen *et al* has shown that disruption of FGF signalling in the endothelium leads to a dramatic reduction in let-7 levels that, in turn, increases expression of TGF-β ligands and receptors and activation of TGF-β signalling, leading to EndoMT (114). However, in another study, it was shown that AcSDKP, a peptide substrate of angiotensin-converting enzyme ACE, inhibits EndoMT through the upregulation of let-7 and restoration of FGF receptor (519). These conflicting data suggest that more studies are required to elucidate the exact roles of let-7, FGF, and TGF-β signalling pathways.

In this chapter, I will focus on: a) NKX2-5 expression in EC, b) the expression of NKX2-5 during EndoMT, and c) investigating whether the activation signal for NKX2-5 expression in SMCs originates from the EC. Results presented in this chapter are partly generated through collaborations and the people that contributed will be named below. Primary human pulmonary artery EC (HPAEC) and human umbilical vein EC (HUVEC) were used for the experiments, and 2 different protocols were used to induce EndoMT.

6.2 Results

6.2.1 NKX2-5 expression in endothelial cells

Primary HPAEC were cultured in medium containing 2% FCS, treated as indicated in Figure 6.1, and protein extracts were prepared to evaluate the expression profile of the cells by Western blot. FGF2, IL-1β and TNFα are all growth factors and cytokines that are present at elevated levels in vascular and fibrotic conditions, and here are used as stimulators of EndoMT. NKX2-5 is expressed at very low levels in untreated EC. TGF-β and FGF2 slightly induced NKX2-5 expression but this was not statistically significant. On the contrary, increasing serum concentration from 2% to 7% significantly increased NKX2-5 protein expression. Of the rest of the markers tested, CD31 and fibronectin were affected by IL-1β and TNFα, as well as FGF2, respectively, whereas α-SMA and calponin levels did not change. Expression of α-SMA was maintained at a low level.

These data can only be translated as an indication of the effect that these factors had on protein expression, since only one experiment was performed.

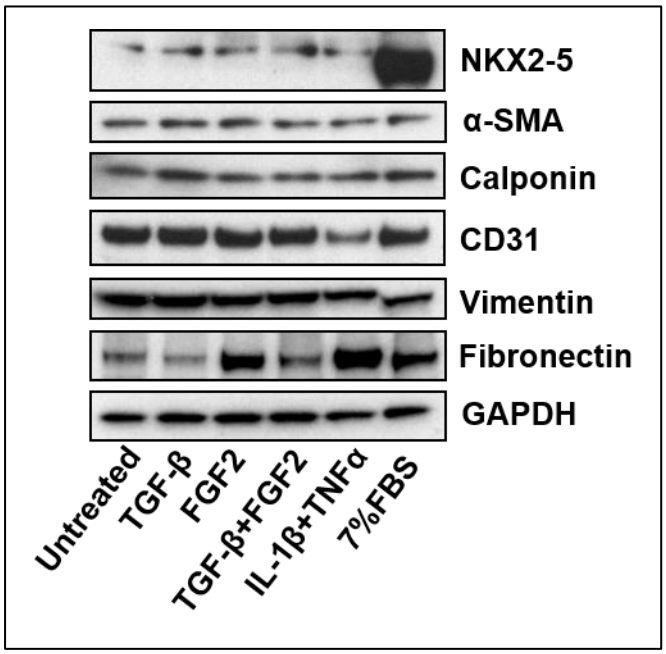


Figure 6.1 Effect of cytokines and growth factors on the expression of endothelial cells. *Primary HPAEC were serum starved overnight in 0.1%FCS and treated with 2ng/ml TGF-β, 50ng/ml FGF2, 10ng/ml TNFα, 10ng/ml IL-1β. After 24 hours, the cells were lysed and total protein was extracted and subjected to Western blot analysis. Protein expression of NKX2-5, α-SMA, calponin, CD31, vimentin and fibronectin was analysed. N=1 (1 replicate experiment performed).*

6.2.2 NKX2-5 expression in EndoMT

Although various studies have reported the involvement of different growth and pro-inflammatory factors in the induction of EndoMT, a special role has been assigned to members of the TGF-β and FGF2 signalling pathways. EndoMT may also occur after exposure to inflammatory molecules such as IL-1β and TNFα, or following AngII receptor type-1 activation. Other studies implicate the MAPK cascade with upregulation of transcription factors such as KLF4 as the initiation step of EndoMT. Progress in the field has been limited by the lack of a successful approach to stimulate EndoMT, the use of endothelial cells obtained from different species and tissues, and indiscriminate use of different isoforms of TGF-β as potent inducers of EndoMT with variable doses and time courses. In this chapter, two different methods of inducing EndoMT were assessed.

6.2.2.1 EndoMT through FRS2α knock-down

A collaboration was established with Professor Michael Simons, Yale School of Medicine, USA. The group has established a model of EndoMT that is based on the silencing of FGF signalling in the EC leading to induced TGF-β signalling and EndoMT (114). FGF signalling, which has recently emerged as a key regulator of the normal vascular state, exerts its function through tyrosine kinase receptors that require the intracellular adaptor FRS2 for the initiation of MAPK signalling. Recent findings also reported that FGF antagonises TGF-β activity in SMC *in vitro*, but the mechanism of this effect and its functional consequences have not been fully investigated (520). In the established model of EndoMT, RNA interference is used (short-hairpin RNA) to knock-down the intracellular adaptor FRS2, silencing the FGF downstream signalling in normal EC. FRS2 knock-down causes EC to change their cell shape and express smooth muscle markers such as calponin, α-SMA, SM22, vimentin, and fibronectin etc. The knock-down also increased TGF-β downstream signalling, with all three receptors being significantly upregulated. The procedure is paired with downregulation of let-7 that is required to maintain low TGF-β levels. The model of EndoMT is described in Figure 6.2.

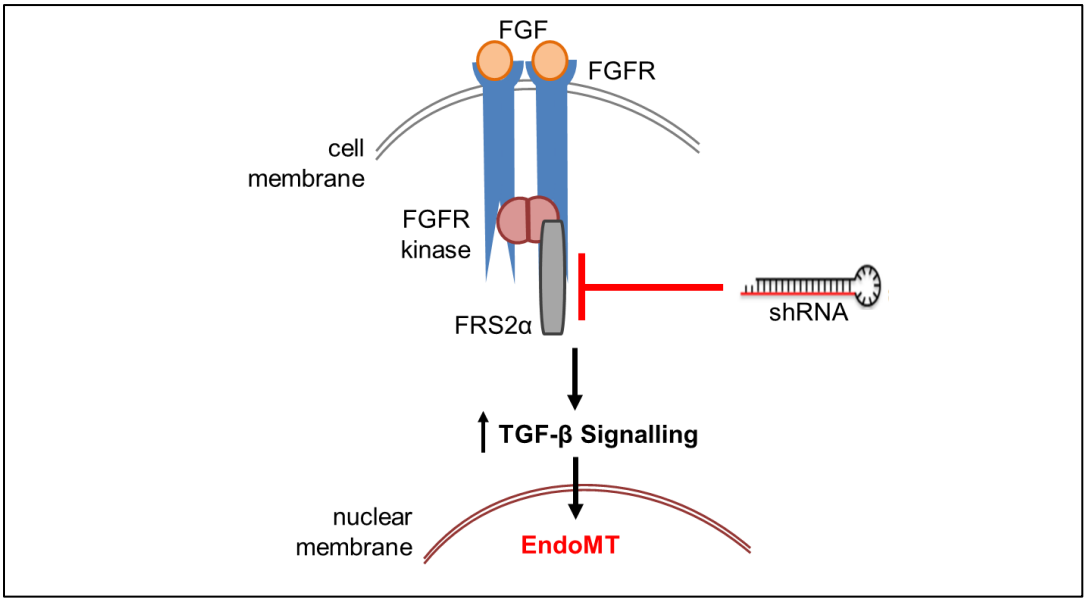


Figure 6.2 Knock-down of FGF signalling in EC leads to EndoMT. *RNA interference is used to knock-down the intracellular adaptor FRS2 and silence FGF downstream signalling. This induces TGF-β signalling and causes EC to undergo EndoMT.*

Both our group and Prof Simons group are working on CVD pathogenesis, with an interest in atherosclerosis, CAD and PAH. Dr Pei-Yu Chen in Prof Simons lab evaluated NKX2-5 expression in the endothelium of CAD patients with different disease severity (Figure 6.3). As disease severity increases from no or mild disease to severe, expression of the endothelial marker CD31 gradually decreased in the endothelium. In parallel, NKX2-5 expression increased significantly ($p<0.001$) following the same trend, with higher levels of expression seen in severe disease.

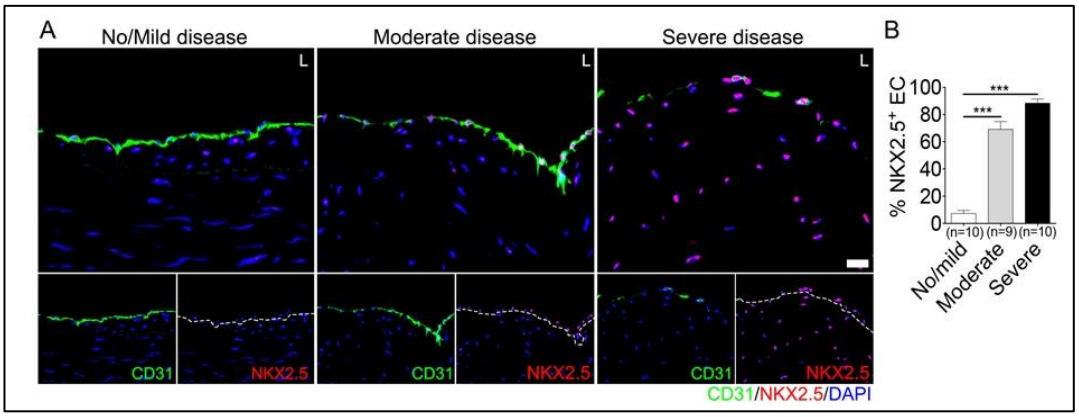


Figure 6.3 NKX2-5 is expressed in the endothelium of CAD patients. *A. Expression of NKX2-5 (red) was analysed by Immunofluorescence in coronary vessels from patients with various extents of CAD. CD31 (green) was used to identify EC, and nuclei were stained with DAPI (blue). B. Percentages of NKX2-5^{+} stained cells in mild, moderate, and severe extent of disease. Number of patient in disease states: No/mild: N=10, Moderate: N=9, Severe: N=10. ***p<0.001.*

Immunofluorescence staining of the coronary specimens provided evidence that NKX2-5 is upregulated in EC and especially in disease when EC undergo EndoMT. These data triggered questions regarding the involvement and role of NKX2-5 in the endothelium. During my PhD I had the opportunity to visit Yale University and work along with Dr Chen for two weeks. In the course of my visit, I managed to validate NKX2-5 expression in RNA and protein extracts prepared by Dr Chen prior to my arrival. The extracts were prepared from 3 independent experiments of EndoMT. In brief, HUVECs, between passages 5-10, were infected with either a control virus or a virus carrying shRNA (adenovirus) specific to FRS2α, and were grown for 4 days before being harvested for RNA and protein. Control and FRS2α knock-down cells were stained for endothelial and mesenchymal markers to determine whether cells undergone EndoMT (Figure 6.4).

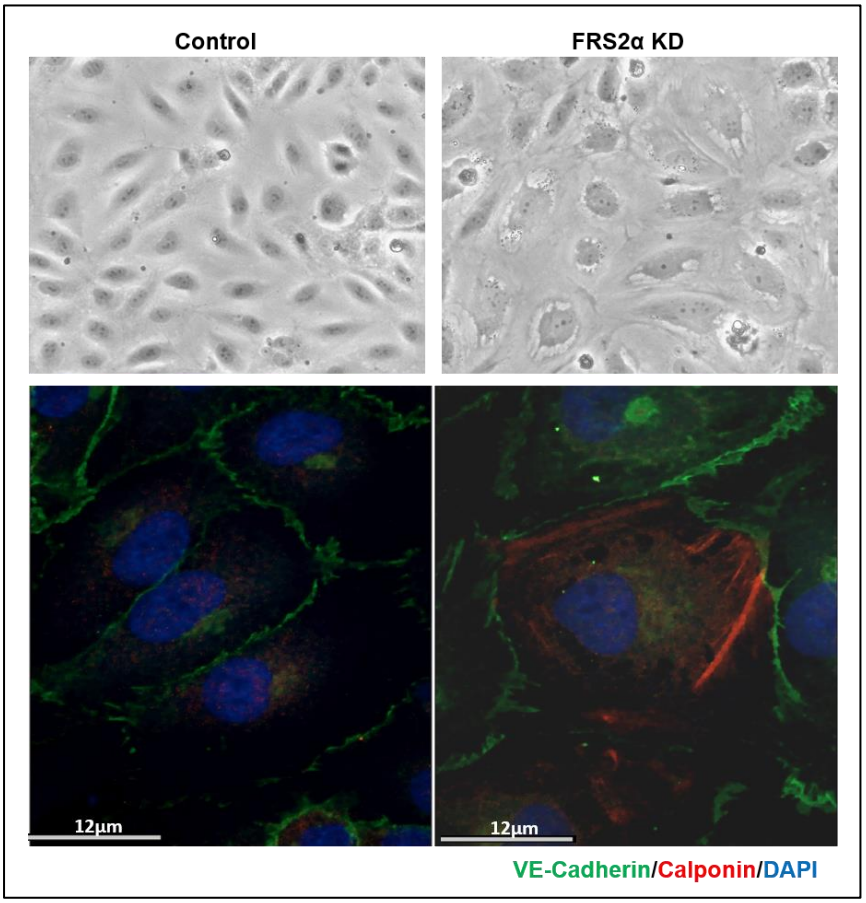


Figure 6.4 Phenotypic changes during FRS2α KD-induced EndoMT. *Primary HUVEC were infected with either a control virus (control) or a virus containing shRNA specific to FRS2α (FRS2α KD). After 4 days, the cells were photographed and stained with immunofluorescence to monitor the changes in morphology and expression of endothelial (VE-Cadherin/green) and mesenchymal markers (calponin/red). DAPI (blue) was used to stain the cell nuclei.*

Four days after treatment with the empty virus, control cells displayed a typical rounded cobblestone morphology with high expression of VE-cadherin and no expression of calponin. FRS2α knock-down resulted in a distinct change in morphology accompanied by induced expression of calponin, a protein not normally expressed in EC, while VE-cadherin expression was maintained.

Next, I used qPCR to evaluate the levels of gene expression of *NKX2-5*, as well as *FGF2* and *TGF-βR1* as genes that are affected directly by the knock-down, together with other mesenchymal and endothelial markers: *VE-CADHERIN, COL1A1, α-SMA,* and *PAI-1*. *NKX2-5* showed a significant increase at gene expression level, and interestingly, the same pattern (of increased expression) was seen in the rest of the genes (Figure 6.5).

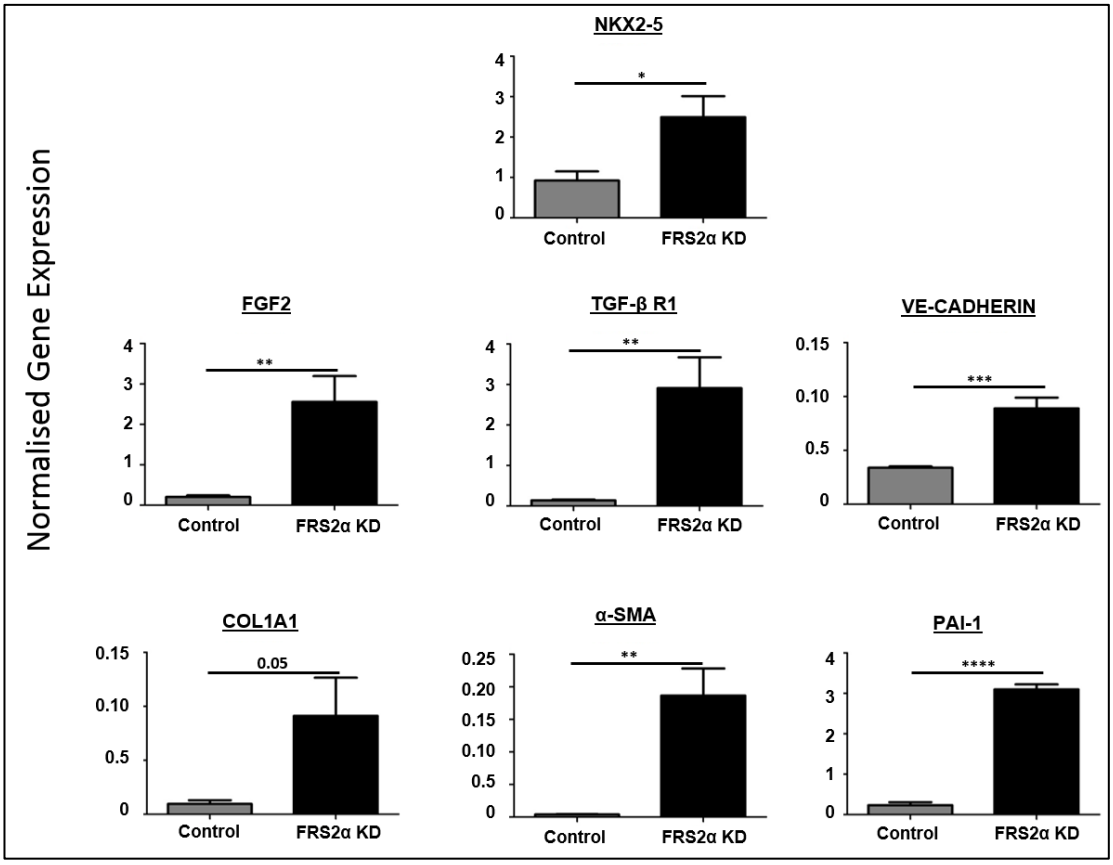


Figure 6.5 Gene expression in HUVEC in FRS2α KD-induced EndoMT. *Primary HUVEC were infected with either a control virus (control) or a virus containing shRNA specific to FRS2α (FRS2α KD). Gene expression of NKX2-5, FGF2, TGF-β R1, VE-cadherin, COL1A1, α-SMA and PAI-1 was assessed by qPCR and normalised to GAPDH expression. N=3 (replicate experiments). *p≤0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001.*

Expression of *TGF-βR1, COL1A1, α-SMA* and *PAI-1* was increased as mesenchymal markers and components of the induced TGF-β signalling pathway. Expression of *VE-cadherin* and *FGF2* were also increased, unexpectedly. Such findings could indicate the ultimate need of the endothelial cell to preserve its phenotype against EndoMT by over-expressing factors critical for the endothelial state.

Protein expression was consistent with gene expression for most of the markers (Figure 6.6). The efficiency of the knock down was confirmed with the complete knock-down of FRS2α expression. Protein levels of the mesenchymal markers collagen, α-SMA, vimentin and PAI-1 were all significantly increased. VE-cadherin protein levels were decreased verifying the loss of the endothelial phenotype. NKX2-5 exhibited higher levels of expression in FRS2α knock-down cells, but the overall difference was not statistically significant.

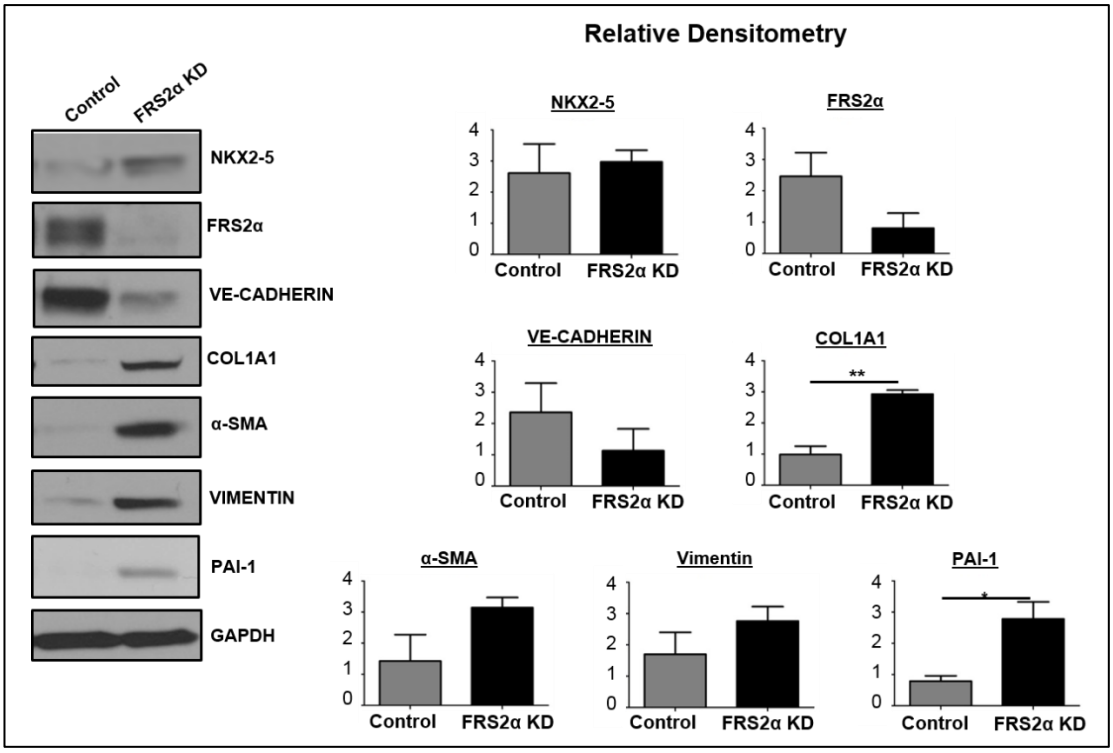


Figure 6.6 Protein expression in HUVEC in FRS2α KD-induced EndoMT. *Primary HUVEC were infected with either a control virus (control) or a virus containing shRNA specific to FRS2α (FRS2α KD). Protein expression of NKX2-5, FRS2α, VE-cadherin, COL1A1, α-SMA, vimentin, and PAI-1 was assessed by Western blot analysis normalised to GAPDH. N=3 (replicate experiments). *p≤0.05, **p≤0.01.*

At this point, it is important to point out that HUVECs are primary cells obtained from the umbilical veins of new-borns. Therefore, the cells preserve an embryonic phenotype where NKX2-5 expression is generally high, and that might explain why

the difference in EndoMT-related genes is not significant. Taken together, the data establish increased NKX2-5 expression in the endothelium of CAD patients, and induced expression during EndoMT

6.2.2.2 Inflammation-induced EndoMT

EndoMT has been linked to fibrosis, where phenotypic differentiation of EC into mesenchymal cells was originally described in experimental wound repair driven by inflammatory stimuli (521). In addition, EC have been found to contribute to the pool of fibroblasts in murine models of cardiac, kidney, and lung fibrosis (16, 17, 522). It has also been proposed that inflammatory cytokines such as TGF-β1, TNFα and IL-1β can induce EndoMT in a similar way to epithelial-to-mesenchymal transition (EMT) based on the mechanistic similarities between the two processes (16). In fact, the inflammation-induced EndoMT mechanism was verified by Rieder *et al* in 2011 (15).

Based on this approach and in collaboration with Dr Robert Good, I assessed NKX2-5 expression in another model of EndoMT. Using the method previously validated and optimised by Dr Good (14), primary human PAEC were treated with a cytokine cocktail of 5ng/ml TGF-β, 5ng/ml TNFα and 0.1ng/ml IL-1β for 5 days to induce EndoMT. The treated and untreated cells were photographed at multiple time-points to monitor the phenotyping changes (Figure 6.7).

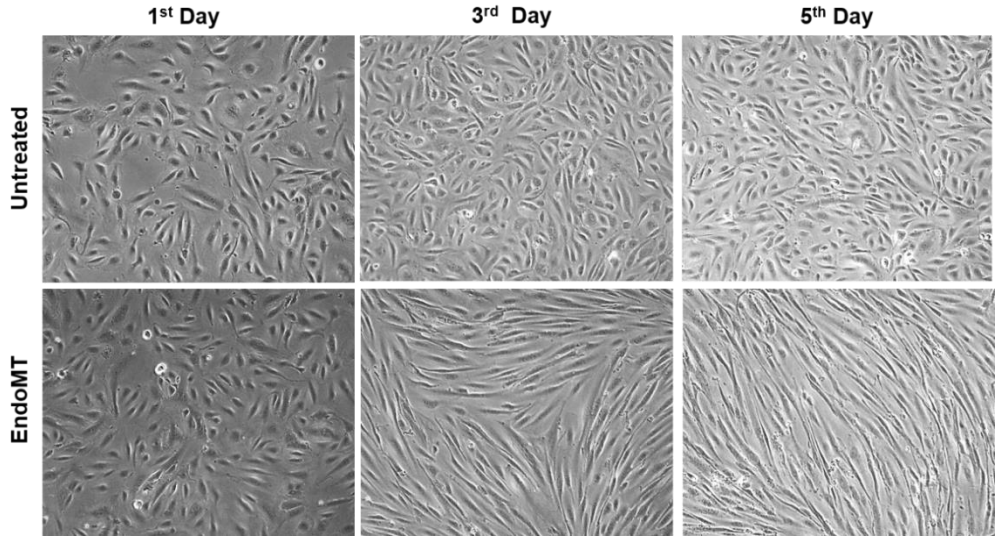


Figure 6.7 Phenotypic differentiation of primary HPAEC treated with a cytokine cocktail to induce EndoMT. *Primary HPAEC were treated with a cytokine cocktail of 5ng/ml TGF-β, 5ng/ml TNFα and 0.1ng/ml IL-1β for 5 days. Untreated and treated cells were photographed on the 1st, 3rd, and 5th day to monitor the phenotypic changes.*

As shown in Figure 6.7, HPAEC treated with the cytokine cocktail presented a change in their morphology by day 3, and their shape was completely changed by the end of the treatment (day 5). In particular, HPAEC lost their typical cobblestone morphology and acquired a more elongated fibroblast-like phenotype. In contrast, untreated cells retained their original morphology.

The change in the morphology was accompanied by corresponding changes in gene expression (Figure 6.8).

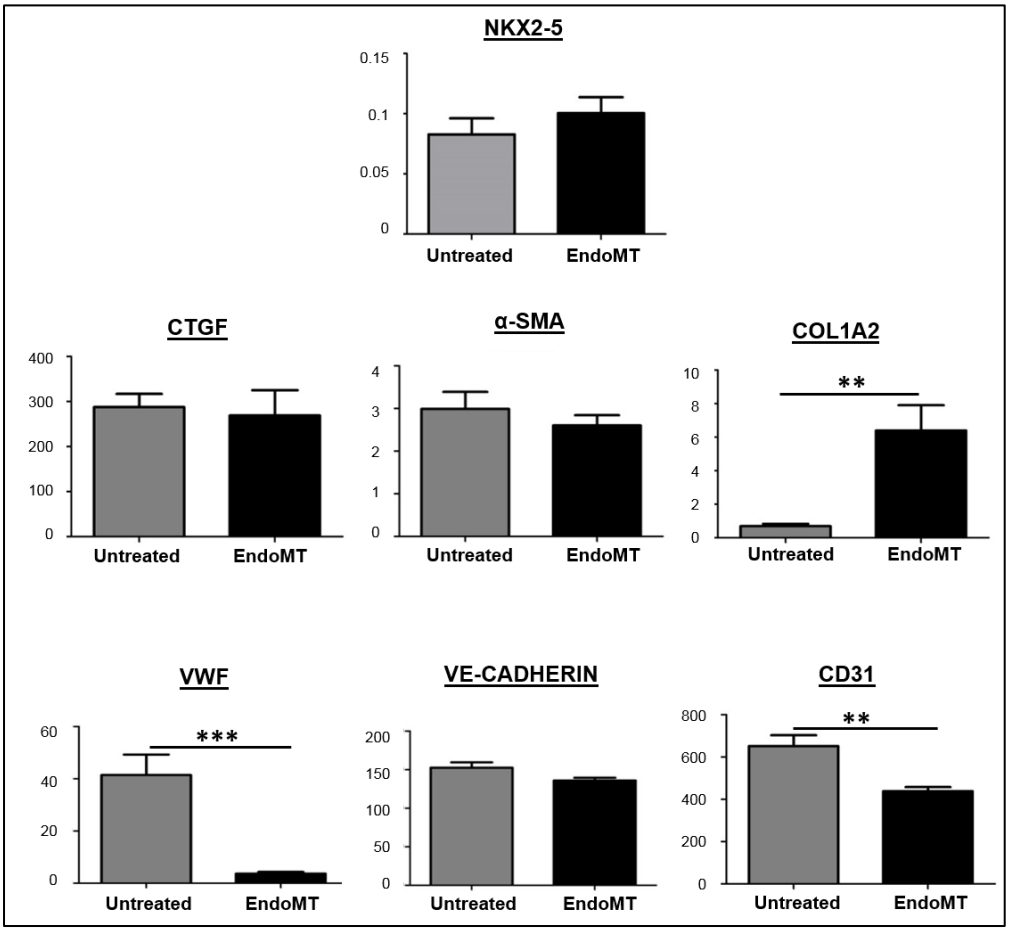


Figure 6.8 Gene Expression of primary HPAEC treated with a cytokine cocktail to induce EndoMT. *Primary HPAEC were treated with a cytokine cocktail of 5ng/ml TGF-β, 5ng/ml TNFα and 0.1ng/ml IL-1β for 5 days. After treatment, the cells were lysed and total RNA was extracted. Gene expression of NKX2-5, CTGF, α-SMA, COL1A2, vWF, VE-Cadherin and CD31 was analysed by qPCR normalised to TBP expression. N=3 (replicate experiments). **p≤0.01, ***p≤0.001.*

Cell lysates of 3 independent replicate experiments of cytokine-induced EndoMT in HPAEC were kindly provided by Dr Good. Total RNA was isolated and subjected to qPCR analysis to validate gene expression of NKX2-5 and other markers. The endothelial markers *CD31* and *von Willebrand* (*vWF*) were significantly downregulated, whereas *VE-cadherin* was only slightly decreased. *COL1A2* gene

expression was significantly induced, but *CTGF* and *α-SMA* showed little change. *NKX2-5* gene expression showed an increasing trend, however this was not statistically significant.

6.3 Summary of results

In this chapter, I focused on NKX2-5 expression in endothelial cells and in cells undergoing EndoMT. Two different methods to induce EndoMT *in vitro* were used in two different cell types. First, the FGF intracellular adaptor FRS2α was knocked down in HUVEC leading to induction of TGF-β signalling and EndoMT. In this model of EndoMT, NKX2-5 expression was upregulated at both the protein and gene levels. NKX2-5 was also found to be expressed in the endothelium of CAD patients and its expression is associated with the severity of the disease.

The second model of EndoMT was inflammation-induced, where HPAEC were stimulated with a cytokine cocktail of TGF-β, TNFα and IL-1β. The cells changed their morphology completely during the course of the treatment, and EndoMT was validated through gene expression profiling. *NKX2-5* gene expression was induced in EndoMT compared to untreated cells, although the difference was not significant.

Overall, these data provide compelling evidence that NKX2-5 is expressed in the endothelium and expression is induced under conditions that result in EndoMT. This new mechanism of EndoMT is poorly understood and requires further investigation.

6.4 Discussion

Combined findings generated during the course of this thesis and previously in the lab have established the expression of NKX2-5 in blood vessels, the role of the gene in vascular remodelling, a signalling mechanism that activates expression in HPASMC, and a transcriptional mechanism that regulates its expression. However, the stimuli that activate NKX2-5 expression in VSMC may originate from different cells or systems, and remain largely unknown. In this chapter, my aim was to explore whether the endothelial cells produce the stimuli that induce NKX2-5 expression in VSMC.

Increasing number of recent findings directly implicate endothelial dysfunction as the trigger of vascular diseases (13-16, 161, 198, 199). Endothelial cells are important regulators of the vascular tone, and they produce vasodilatory or vasoconstrictive factors in response to stimuli circulating in the blood. In disease conditions, the endothelium undergoes functional and structural alterations, thus

losing its protective role and becoming a pro-atherosclerotic structure, a process known as endothelial dysfunction (192, 523). Endothelial dysfunction, detected as a reduced vasodilator response to endothelial stimuli, has been observed to be associated with major cardiovascular risk factors, such as aging, hyperhomocysteinemia, postmenopausal state, smoking, diabetes, hypercholesterolemia, and hypertension (523).

The expression of NKX2-5 as well as other endothelial and mesenchymal markers was examined in HPAEC upon stimulation with growth factors and cytokines known to be implicated and stimulate vascular diseases, including TGF-β, FGF2, IL-1β, and TNFα. FGF2 is a potent mitogen that binds to the FGF receptor family of receptor tyrosine kinases. It is a known angiogenesis-related factor with increased expression in endothelial and epithelial cells, and it has been shown to promote the synthetic phenotype of VSMCs. A recent study showed that FGF2 antagonised and attenuated TGF-β-stimulated differentiation of airway SMC towards a contractile phenotype (118). Migratory effects induced by FGF2 have also been reported in various types of endothelial cells (524).

Expression levels of the endothelial marker CD31 were increased compared to the rest of the markers, and interestingly, upon stimulation with IL-1β and TNF-α, CD31 expression was decreased greatly (Figure 6.1). Expression levels of vimentin, a protein expressed by endothelial and mesenchymal cells, and the mesenchymal markers α-SMA and calponin, which were generally low, did not change in response to stimulation. In contrast, levels of fibronectin were increased when the cells were treated with FGF2 and TNFα. NKX2-5 protein expression in the untreated cells was low compared to all other proteins, but increased upon stimulation with TGF-β and FGF2. However, increasing the serum concentration from 2% to 7% had a dramatic effect on NKX2-5 expression, which increased significantly.

The pattern of NKX2-5 protein expression in EC is similar to that seen in VSMC, suggesting that mechanisms stimulating NKX2-5 activation are common between EC and VSMC. Vascular EC and SMC share a common embryonic origin, as both cell types predominantly derive from the mesoderm lineage, and the primitive streak is the key structural component that discriminates the mesodermal precursors (525). *Nkx2-5*-expressing mesoderm gives rise to the heart muscle and the endothelium of the pharyngeal arch arteries, which contributes to the formation of the great vessels of the heart and the pulmonary arteries (322, 323). *Nkx2-5* was also shown to bind directly the *Er71* gene promoter and activate its expression, which promotes the differentiation of EC and vessel development (526). Early *in*

vivo studies in *Xenopus* and zebrafish embryos showed that FGFs, Wnt, and members of the TGF-β superfamily play important roles in the induction and patterning of mesoderm (525). Therefore, it is possible that specific stimuli that induce post-natal expression of NKX2-5 can occur in adult vascular EC in a similar way to VSMC. Yet, activating stimuli should resemble those occurring in embryonic-like processes such as angiogenesis and differentiation.

Upon these activating stimuli, EC as with many other cell types, can undergo phenotypic modulation to a more mesenchymal phenotype. EndoMT was initially described in embryonic development, where it retains a significant role. In heart development, endocardial cells with a clear endothelial phenotype are able to give rise to mesenchymal heart cushion cells through EndoMT (527, 528). It was also shown that EndoMT is important in aortic and pulmonary artery development and in the development of the normal arterial intima (529). Furthermore, morphological studies in human embryos suggest that EndoMT also occurs during the maturation of vessels, including arteries and veins (530).

Similar to vascular remodelling, EndoMT has also been associated with disease, and it has been studied in the context of various CVD including atherosclerosis, PAH, SSc-PAH etc (13, 14, 17, 158, 198). Environmental stimuli triggering EndoMT might also trigger the phenotypic modulation of VSMC, and thus favour expression of NKX2-5 in the endothelial-derived mesenchymal cells. Immunofluorescent staining of NKX2-5 in human vessels provided evidence supporting this hypothesis (Figure 6.3). NKX2-5 expression was found in the endothelium of CAD patients, and expression showed a significant correlation with the severity of the disease. In contrast, expression of endothelial marker CD31 inversely correlated with the extent of disease, with expression having been ablated in severe CAD (Figure 6.3).

EndoMT has been associated with the progression of atherosclerosis and plaque stability (531). Chen *et al* used an established model of EndoMT by introducing endothelial-specific deletion of Frs2α in ApoE null mice. When double-knockout mice ($Frs2\alpha^{-/-}/ApoE^{-/-}$) were fed a high-fat diet, they developed atherosclerosis at a much earlier time point compared with their $ApoE^{-/-}$ counterparts, demonstrating an 84% increase in total plaque burden (13). In addition, these double knock-out mice exhibited extensive EndoMT, increased deposition of fibronectin, and increased neointima formation. More evidence were provided by Evrard *et al*, who showed that EndoMT-derived fibroblast-like cells are present in intimal plaques throughout atherosclerotic development using an endothelial tracking linage system in ApoE null mice (531). They also showed that this process is driven by TGF-β signalling,

oxidative stress and hypoxia. Both studies provided significant data establishing the contribution of EndoMT in atherosclerosis, and shed light on new mechanisms that drive the progression of CVD.

Another important finding that both studies highlight is the significance of signalling mechanisms that regulate EndoMT and those mainly involve TGF-β and FGF2. In fact, many studies point out that the two growth factors regulate each other (114, 119, 132, 516, 518). In general, the data indicate that FGF signalling is essential to maintain the endothelial phenotype, and once this is blocked TGF-β signalling is upregulated initiating the initiation of the transition towards the mesenchymal phenotype. One way of FGF2 signalling may be suppressed in EC, which has already been described (Figure 6.2), is the knock down of FRS2α that leads to increased TGF-β signalling and TGF-β-dependent transcription of mesenchymal markers through the downregulation of let-7 microRNA (114). In normal conditions, let-7 specifically targets TGF-βR1 mRNA and destroys it preventing TGF-β downstream signalling. miR-20 has also been implicated in EndoMT, since its expression is significantly downregulated upon FGF2 suppression, leading to increased transcription of its downstream targets ALK5 and TGF-βR2 that both promote canonical TGF-β signalling (518).

Interestingly, data presented by the Simons group collectively demonstrate that TGF-β exerts a binary role in endothelial and VSMC cells. On one hand, suppression of FGF signalling and consequent upregulation of TGF-β promotes EndoMT and expression of mesenchymal markers in endothelial-derived cells. On the other hand, FGF suppression in VSMC leads to TGF-β-regulated contractile phenotype with increased contractility and expression of SMC markers (119). This is not consistent with our findings, where TGF-β promotes the phenotypic modulation toward the synthetic phenotype. Furthermore, in a similar study from the same group, it was reported that TGF-β signalling prevents SMC proliferation causing G1/S arrest in FRS2α-knock down aortic SMC (132). However, such a finding might be caused directly due to downregulation of FGF signalling, since it is an important form of paracrine cell-cell communication, rather than TGF-β driven. In addition, expression of proliferative markers was not assessed in any of these studies.

The inflammatory-induced model of EndoMT further supports a role of TGF-β in phenotypic modulation of EC. The cytokine-cocktail induction has previously been described in EMT, a similar process of phenotypic modulation in epithelial cells. In addition, this model also mimics for the inflammatory component of various

conditions such as atherosclerosis and SSc-PAH, without ablating mechanisms that are involved in signal transduction between cells and environment.

Another interesting observation came from a study exploring the effect of EndoMT in hyperplasia and proliferative vascular disease (532), where it was shown that EndoMT is modulated by shear stress in an ERK5-dependent manner. In brief, ERK5 silencing caused spontaneous morphological changes suggestive for EndoMT, which became more apparent with TGF-β stimulation. This study revealed for the first time a crucial role of ERK5 signalling to promote EndoMT, and this effect is enhanced in the presence of TGF-β. This mechanism is consistent with the role of ERK5 as an activator of the signalling cascade for NKX2-5 expression that was discussed earlier (Chapter 4). Interestingly, unpublished work from our lab suggests that the use of ERK5 inhibitor, ERK5-IN-1, blocked EndoMT transition with PAEC retaining their endothelial phenotype after treatment with the cytokine cocktail (TGF-β, TNFα, IL-1β), as confirmed by their morphology and expression profile.

In contrast to VSMC de-differentiation, where cells can acquire either the contractile or the synthetic phenotype dependent on the environmental conditions, there is yet no evidence that EndoMT-derived mesenchymal cells can reverse to their endothelial phenotype. Thus, scientists concentrate on creating drugs that inhibit EndoMT as a prospective therapy. However, this ability of endothelial cells to undergo phenotypic modulation and generate various different types of connective tissue provides a hidden advantage and hope that EndoMT could be used as a potential method for tissue regeneration. Step-wise modulation of endothelial cells through stimulation by key factors could generate osteocyte and chondrocytes that could be used as treatment to osteoporosis, osteonecrosis, and osteoarthritis, respectively. Similarly, generation of myocytes and cardiomyocytes could be beneficial for muscular dystrophy and myocardial infarction. The process may also aid in vascular tissue regeneration, particularly in vasculogenesis through the generation of VSMC and pericytes. Moreover, tissue engineering *ex vivo* could also be used to achieve successfully and efficient replacement or transplantation of degenerated tissues and organs.

Overall, expression of NKX2-5 showed an increasing trend in EndoMT at both the gene and protein levels, however, the difference was statistically significant only at gene expression levels in HUVECs. Nonetheless, HUVECs might not be an ideal and representative cell type to study NKX2-5, since the gene might be expressed due to the embryonic phenotype of the cell. Data presented in this chapter provide evidence of EndoMT being a direct or indirect regulator of NKX2-5. However, this

relationship might be reciprocal since unpublished data generated in our lab showed that NKX2-5 knock down using siRNA blocked the transition of PAEC towards the mesenchymal phenotype. More studies are required to elucidate further the exact role of NKX2-5 in EndoMT.

CHAPTER 7 - CONCLUSION AND FUTURE STUDIES

The main objective of this thesis was to investigate the regulation of *NKX2-5* gene in human adult blood vessels. I have approached the regulation of the gene from different angles and provided compelling evidence at multiple levels. In particular, I have successfully shown that the *NKX2-5* genomic locus is genetically associated with vascular disease (Chapter 3). I have provided strong evidence that a signalling cascade involving hypoxia, TGF-β, AKT and ERK5 is crucial for the activation of NKX2-5 in VSMCs (Chapter 4). In addition, I have identified a transcriptional mechanism by which *NKX2-5* expression is regulated through the interaction of an upstream promoter and a functional downstream enhancer. This mechanism involves the binding of an activator/co-activator complex made of TEAD/YAP1 and the binding of the transcription factors GATA-6, c-JUN, MEF-2c and phosphorylated SMAD3 (Chapter 5). Finally, I have provided evidence that NKX2-5 can be activated in endothelial cells during EndoMT possibly through similar activating mechanisms to those that stimulate expression in VSMCs and involve TGF-β (Chapter 6).

Most studies of NKX2-5 in human report gene mutations that cause or have been associated with different forms of CHD. Until recently, postnatal NKX2-5 expression had only been reported in the heart and few other tissues, excluding blood vessels. This is the first time that expression and regulation of NKX2-5 is studied in human blood vessels, and the findings increase significantly the existing knowledge, most of which comes from earlier animal studies. In addition, this is also the first time that *NKX2-5* has been genetically associated with vascular disease. This finding confirms an earlier study conducted in our lab that showed that conditional knock-out of *NKX2-5* in an animal model of PAH normalised mean pulmonary pressures, and eliminated muscularisation of the vessels and vascular remodelling.

The genetic study was conducted in scleroderma patients and healthy individuals using tagging SNPs across the genomic locus of *NKX2-5*. Scleroderma is an AID with prominent pulmonary complications and vasculopathy, with PAH being the leading cause of death amongst scleroderma patients. The meta-analysis (Chapter 3) showed that *NKX2-5* is genetically associated with scleroderma across two independent cohorts of similar ethnicity (British and Spanish). Scleroderma is a complex disease, and due to the immune system dysfunction, causal genes identified so far by genetic studies are mostly related to immunity, with the exception of *CTGF* gene. The findings of the genetic study contribute new insights in the pathogenesis of scleroderma, and in particular highlight the mechanisms of

extended fibrosis affecting the skin and internal organs. It directly implicates the phenotypic modulation of VSMCs regulated by NKX2-5 in fibrosis, and recognises synthetic VSMCs as an important effector cell type, apart from fibroblasts and myofibroblasts.

NKX2-5 was also found to be associated with PH independent of the presence of scleroderma, and the association was successfully replicated in an independent cohort. This finding further confirms the hypothesis that NKX2-5 is an important regulator of vascular diseases characterised by vascular remodelling. However, to further explore and understand in depth this association, more studies will need to be performed in future. One study could be conducted in other patients with other AIDs such as SLE and RA, that both share common characteristics with scleroderma, and exhibit vascular complications. A study could also be conducted in PH patients that do not have AID and are not diagnosed with BMPR2 mutations. Although this would be the most appropriate group to test the hypothesis, it would be extremely difficult to collect. This issue highlights the importance of collaborations between interdisciplinary scientific groups, and illustrates the need for sharing of knowledge and expertise.

In Chapter 4, I have generated important data that recognise TGF-β as a critical stimulator for NKX2-5 expression. Although this regulatory pathway had been identified before in animal studies, the present study provides the first confirmation of a similar regulation in adult human vessels. I have shown that TGF-β is able to upregulate NKX2-5 and promote the phenotypic modulation towards the synthetic state of VSMCs. The upregulation of NKX2-5 occurs through the phosphorylation of SMAD3, ERK5 and AKT by the respective kinases. This cascade can be activated in response to vascular injury or to other disease-associated stimuli such as hypoxia. However, other stimuli can also trigger the signalling cascade and lead to activation of NKX2-5, including mechanical stretch, stress, metabolic dysfunction and others. These stimuli should be tested in future as alternative mechanisms of NKX2-5 activation.

A panel of selective inhibitors was used with few of them causing a marked decrease at protein and gene expression levels of NKX2-5. We have shown before that conditional knock-out of NKX2-5 ameliorated vascular remodelling and restored its downstream implications. The selective inhibitors that decreased NKX2-5 could be tested in animal models of atherosclerosis or PAH as drugs targeting the activating mechanisms. However, since the targeted cytokines and kinases are

implicated in numerous processes of the human body, a sophisticated and carefully-designed approach would be required.

Findings generated and presented in Chapter 5 may significantly facilitate the designing of strategies to block NKX2-5 activation and expression. In particular, I have identified a transcriptional mechanism that regulates *NKX2-5* and involves two different genomic regions. This transcriptional mechanism can be blocked by disrupting the binding of the TEAD/YAP1 complex on rs3095870 or the binding of other transcription factors (GATA-6, c-JUN, MEF-2c, SMAD3) at the enhancer. Disruption of binding will prevent NKX2-5 expression and therefore suppress pathological vascular remodelling. Although the ChIP assays demonstrated the enrichment of these factors binding to the upstream promoter and the enhancer, more experiments are required in order to identify the exact DNA binding elements. In addition, both TEAD1 and TEAD3 proteins seem to be implicated in the transcriptional regulation, exhibiting different potentials regarding NKX2-5 expression. However, their exact role needs to be further investigated, and additional work is required in order to identify any other functional MCAT consensus elements within the *NKX2-5* promoter that would potentially regulate expression.

Another area that requires investigation is the regulation of NKX2-5 through epigenetic modifications including microRNAs and DNA methylation. There is a body of evidence presented in this thesis and elsewhere suggesting that there is a strong possibility that epigenetic mechanisms are implicated in the regulation of NKX2-5 expression. A careful study design is critical in order to explore post-transcriptional regulation through microRNAs. The study should combine published information regarding the expression profile of microRNAs that were previously implicated in vascular disease and those that would potentially target the *NKX2-5* 3'UTR. Another way to address this question would be to perform a microRNA array in samples that NKX2-5 is overexpressed compared to samples with normal expression. MicroRNAs showing differential expression would be potential candidates. Whole exome sequencing could also be advantageous in this occasion. Exome sequencing detects variants in coding exons, with the capability to expand targeted content to include UTRs and microRNAs for a more comprehensive view of gene regulation.

Existing evidence reports that *NKX2-5* is differentially methylated in cancer. There are six CpG islands adjacent to the *NKX2-5* gene which could either be hypo- or hyper- methylated depending on the cell requirements for NKX2-5 expression (Figure 5.21). Hyper-methylation has been traditionally associated with decreased

expression. Therefore, silencing of postnatal NKX2-5 expression through DNA methylation could be possible and requires further investigation. Methylation levels may be changed in disease allowing expression of NKX2-5 as has been previously reported. Bisulphite conversion and a methylation array could be potentially used to investigate this hypothesis. However, methylation patterns are not ubiquitous and differ between cell types and tissues. Therefore, performing such experiments would require DNA isolated from diseased vessels of patients suffering from vascular diseases such as PAH, atherosclerosis, PAD. Obtaining these kinds of samples is challenging and involves co-operation of patients and clinicians, but may be feasible where there are close collaborations between clinical and basic scientists.

In the final Chapter 6, expression of NKX2-5 was assessed in endothelial cells, where it was shown to be increased during EndoMT, a process of phenotypic modulation whereby endothelial cells acquire a mesenchymal phenotype. EndoMT and VSMC de-differentiation are similar processes that describe the phenotypic switch of two distinct types of vascular cells, which change their morphology and expression profile, and produce and deposit extracellular matrix within the vessel wall. Both processes have been associated with thickening of the blood vessel wall, and formation of plexiform and atherosclerotic lesions. In addition, both vascular endothelial and smooth muscle cells originate from the same embryonic progenitor cell type, where NKX2-5 is expressed during development. It is also apparent that both processes are directly and/or indirectly stimulated by TGF-β, with ERK5 and AKT involved. Taken together, the findings presented in this study propose that similar signalling mechanisms stimulate expression of NKX2-5 during EndoMT and VSMC de-differentiation. However, it is not yet clear if NKX2-5 drives EndoMT upon stimulation by the adjacent activated VSMCs, or it is expressed only by the EndoMT-derived cells, and this question could be better answered using an animal model rather than *in vitro* experiments, due to the high levels of cell-cell and cell-environment interactions present during EndoMT.

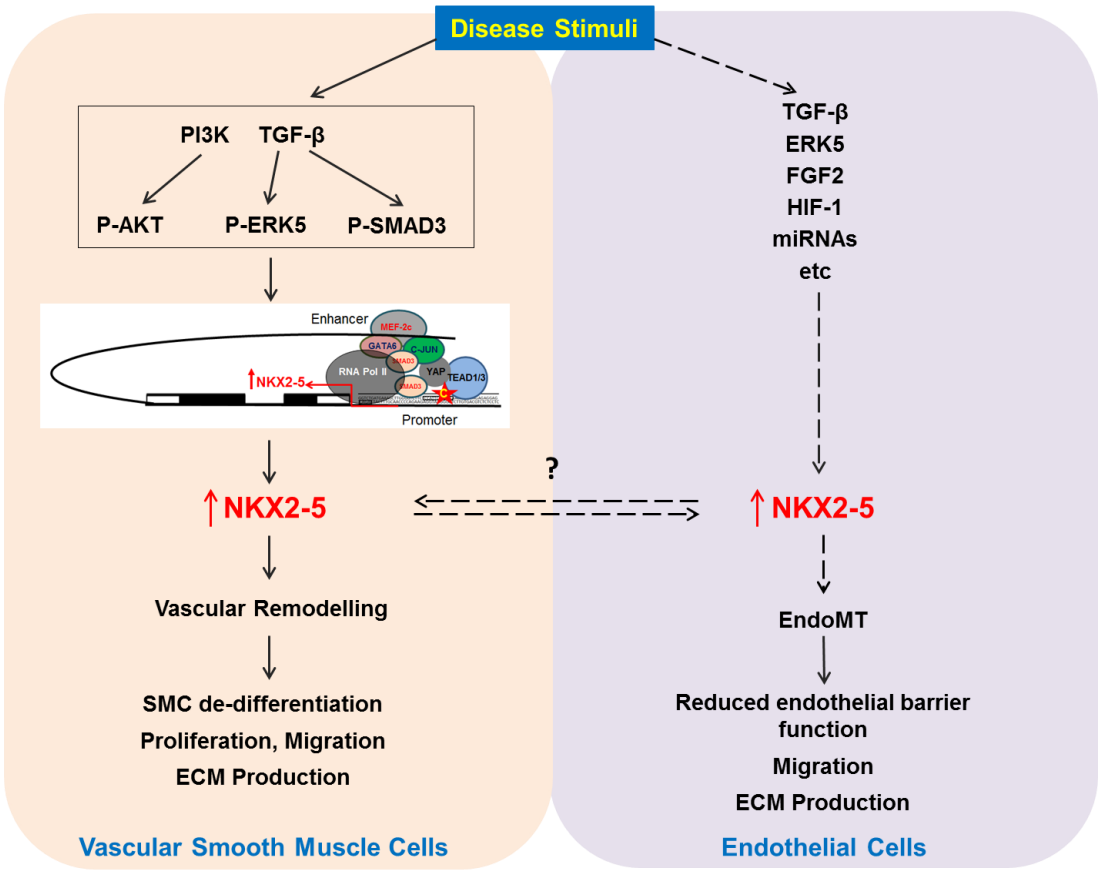


Figure 7.1 The role of NKX2-5 in disease. *Schematic diagram illustrating the role of NKX2-5 and its regulation under disease conditions in vascular smooth muscle and endothelial cells. Continuous lines indicate data generated in this thesis, whereas broken lines indicate evidence found in literature.*

Study Limitations

In general, the main limitations concern the patients cohorts used here and those of future and perspective studies. Scleroderma is a rare condition, the manifestation of which changes overtime. Although clinical criteria for disease diagnosis are well-established and reviewed often, patient phenotyping is critical and often varies between clinical doctors and medical centres. Therefore, important information can be missed or incorrectly annotated. Another issue most people confront in these types of studies concerns the size of the cohorts. For relatively rare diseases such as scleroderma or PAH, generating large cohorts of patients with available clinical information and biological samples such as DNA, serum, tissue biopsies, etc is extremely challenging. It requires collaboration and partnership between clinical doctors from many clinical practices and scientists. Due to these reasons, a replication study in a non-SSc related PAH cohort would be very difficult to co-ordinate.

Another limitation of this study as well as of most of the studies focusing on vascular disease, is the accessibility and availability of clinical samples. In particular, there are limited sources of primary human pulmonary vascular smooth muscle cells, which are usually commercially provided at extreme costs. Other sources could include post-mortem tissues and lung tissues of patients undergoing transplantation, which are both limited sources due to extended protocols and consent processes required. Similarly, lung biopsies or tissue-isolated nucleic acids are also very limited making tissue specific epigenetic studies difficult to conduct.

With reference to the functional work conducted during this thesis, limitations are subject to the lack of general knowledge regarding the biology of the human NKX2-5 gene and protein. As such, TGF-β was used exclusively in this thesis to activate expression of NKX2-5. This effect is easy to control experimentally (ie dose-response and time-course of stimulation). However, TGF-β stimulation is not the only way that NKX2-5 gets activated, and therefore the data presented in this study might be biased towards the TGF-β-dependent mechanisms. Future studies should be conducted to study alternative activation mechanisms, such as hypoxia.

Another drawback of the study is with regards to the microRNA studies presented in this thesis. Currently, there are no known microRNAs to target NKX2-5 mRNA, which could be used as positive controls. Use of a positive control as such could improve the interpretation of the data.

The lack of effective ways to quantify NKX2-5 gene and protein expression is another limitation. Although a panel of monoclonal and polyclonal antibodies are commercially available and tested in these studies, there is no currently available antibody to detect the nuclear NKX2-5 protein with high-specificity and that adversely affects the interpretation of the findings.

In addition, the use of immortalised HPASMC, and the various conditions of serum starvation and TGF-β stimulation are also limitations. This is the first time that an immortalised type of HPASMC is used in vascular studies, and although initial experiments were performed to confirm their suitability in the study, it can be argued that immortalisation may have altered somehow the signalling pathways that activate gene expression. Overall and despite the limitations described above, I have strived to ensure that the experiments were consistently well designed and performed. Successful replication in science is challenging and was accomplished by extensive optimisation ensuring consistency in the materials used and in the experimental protocols.

Future Work

The work conducted during this thesis has been interesting and challenging, and the data presented here contribute significantly to the understanding of the role and regulation of NKX2-5 in human adult vessels. Although there were many significant findings, a number of questions remains unanswered and require further investigation, each of which could generate a new project.

Future work plans can be outlined below in few main points:

- A genetic association study of *NKX2-5* tagging SNPs in:
    - other AIDS, such as SLE and RA
    - PAH patients without AIDs or BMPR2 mutations
- Investigation of other stimuli and environmental insults for potential activation of NKX2-5 expression
- Test of the selective inhibitors that downregulated NKX2-5 expression in animal models of atherosclerosis and/or PAH
- Blocking of NKX2-5 transcriptional activation by inhibiting:
    - The binding of TEAD/YAP activator complex on *NKX2-5* promoter
    - The binding of MEF-2c, GATA6, c-JUN on the *NKX2-5* downstream enhancer by deleting the consensus DNA elements
    - TGF-β stimulation mediated by the binding of phosphorylated SMAD3 on NKX2-5 regulatory regions
- Further investigation of other potential TGF-β responsive elements
- Investigation of the presence of other MCAT elements within the *NKX2-5* proximal or upstream promoter
- The effect of microRNAs on the post-transcriptional regulation of *NKX2-5*
- The effect of methylation on the epigenetic regulation of *NKX2-5* in diseased vessels compared to controls
- Investigation of the role of NKX2-5 during EndoMT in animal models

Overall, this project provides important insights for the regulatory and activating mechanisms of NKX2-5 expression in diseased blood vessels. These mechanisms may be similar in other diseases characterised by vascular remodelling. Data presented here could ultimately be used to develop strategies to prevent pathological NKX2-5 expression, and therefore lead to the design of targeted therapies for CVD.

APPENDIX

Websites, Genome Browsers, Tools

https://www.ncbi.nlm.nih.gov/pubmed

http://www.uniprot.org/

http://www.ensembl.org/index.html

The International HapMap Project (NCBI retired HapMap Resource, June 16, 2016)

http://www.internationalgenome.org/home

https://genome.ucsc.edu/

http://www.genecards.org/

http://pngu.mgh.harvard.edu/~purcell/plink/

https://www.broadinstitute.org/haploview/haploview

http://archive.broadinstitute.org/mpg/tagger/

http://imagej.net/Welcome

http://www.graphpad.com/scientific-software/prism/

http://bioinfo.ut.ee/primer3-0.4.0/primer3/

http://archive.broadinstitute.org/mammals/haploreg/haploreg.php

http://www.microrna.org/microrna/home.do

http://www.gtexportal.org/home/

http://www.genecodes.com/sequencher

http://biologylabs.utah.edu/jorgensen/wayned/ape/

REFERENCES

1. Klabunde RE. Cardiovascular Physiology Concepts: Lippincott Williams & Wilkins; 2005.
2. Ogawa T, de Bold AJ. The heart as an endocrine organ. Endocrine Connections. 2014;3(2):R31-44.
3. Song W, Wang H, Wu Q. Atrial Natriuretic Peptide in Cardiovascular Biology and Disease (NPPA). Gene. 2015;569(1):1-6.
4. Saladin K. Anatomy and Physiology : The Unity of Form and Function 7th. McGraw-Hill Education. 2014;7th Edition.
5. Sandoo A, van Zanten J, Metsios GS, Carroll D, Kitas GD. The Endothelium and Its Role in Regulating Vascular Tone. The Open Cardiovascular Medicine Journal. 2010;4:302-12.
6. Rumbaut RE TP. Platelet-Vessel Wall Interactions in Hemostasis and Thrombosis. San Rafael (CA): Morgan & Claypool Life Sciences; 2010. 2010.
7. Haurani MJ, Pagano PJ. Adventitial fibroblast reactive oxygen species as autacrine and paracrine mediators of remodeling: bellwether for vascular disease? Cardiovascular research. 2007;75(4):679-89.
8. Galkina E, Kadl A, Sanders J, Varughese D, Sarembock IJ, Ley K. Lymphocyte recruitment into the aortic wall before and during development of atherosclerosis is partially L-selectin dependent. The Journal of experimental medicine. 2006;203(5):1273-82.
9. Tieu BC, Lee C, Sun H, Lejeune W, Recinos A, 3rd, Ju X, et al. An adventitial IL-6/MCP1 amplification loop accelerates macrophage-mediated vascular inflammation leading to aortic dissection in mice. The Journal of clinical investigation. 2009;119(12):3637-51.
10. Swedenborg J, Mayranpaa MI, Kovanen PT. Mast cells: important players in the orchestrated pathogenesis of abdominal aortic aneurysms. Arteriosclerosis, thrombosis, and vascular biology. 2011;31(4):734-40.
11. Yuan SY RR. Regulation of Endothelial Barrier Function. San Rafael (CA): Morgan & Claypool Life Sciences; 2010. 2010.
12. Michiels C. Endothelial cell functions. Journal of cellular physiology. 2003;196(3):430-43.
13. Chen PY, Qin L, Baeyens N, Li G, Afolabi T, Budatha M, et al. Endothelial-to-mesenchymal transition drives atherosclerosis progression. The Journal of clinical investigation. 2015;125(12):4514-28.
14. Good RB, Gilbane AJ, Trinder SL, Denton CP, Coghlan G, Abraham DJ, et al. Endothelial to Mesenchymal Transition Contributes to Endothelial Dysfunction in Pulmonary Arterial Hypertension. The American journal of pathology. 2015;185(7):1850-8.
15. Rieder F, Kessler SP, West GA, Bhilocha S, de la Motte C, Sadler TM, et al. Inflammation-induced endothelial-to-mesenchymal transition: a novel mechanism of intestinal fibrosis. The American journal of pathology. 2011;179(5):2660-73.
16. Zeisberg EM, Tarnavski O, Zeisberg M, Dorfman AL, McMullen JR, Gustafsson E, et al. Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. Nature medicine. 2007;13(8):952-61.
17. Hashimoto N, Phan SH, Imaizumi K, Matsuo M, Nakashima H, Kawabe T, et al. Endothelial-mesenchymal transition in bleomycin-induced pulmonary fibrosis. American journal of respiratory cell and molecular biology. 2010;43(2):161-72.
18. Brozovich F, Nicholson C, Degen C, Gao YZ, Aggarwal M, Morgan K. Mechanisms of Vascular Smooth Muscle Contraction and the Basis for

Pharmacologic Treatment of Smooth Muscle Disorders. Pharmacological Reviews. 2016;68(2):476-532.
19. Wynne BM, Chiao CW, Webb RC. Vascular Smooth Muscle Cell Signaling Mechanisms for Contraction to Angiotensin II and Endothelin-1. Journal of the American Society of Hypertension : JASH. 2009;3(2):84-95.
20. Komari SO, Headley PC, Klausner AP, Ratz PH, Speich JE. Evidence for a common mechanism for spontaneous rhythmic contraction and myogenic contraction induced by quick stretch in detrusor smooth muscle. Physiological reports. 2013;1(6):e00168.
21. Mosse PR, Campbell GR, Wang ZL, Campbell JH. Smooth muscle phenotypic expression in human carotid arteries. I. Comparison of cells from diffuse intimal thickenings adjacent to atheromatous plaques with those of the media. Laboratory investigation; a journal of technical methods and pathology. 1985;53(5):556-62.
22. Mosse PR, Campbell GR, Campbell JH. Smooth muscle phenotypic expression in human carotid arteries. II. Atherosclerosis-free diffuse intimal thickenings compared with the media. Arteriosclerosis (Dallas, Tex). 1986;6(6):664-9.
23. Campbell JH, Kocher O, Skalli O, Gabbiani G, Campbell GR. Cytodifferentiation and expression of alpha-smooth muscle actin mRNA and protein during primary culture of aortic smooth muscle cells. Correlation with cell density and proliferative state. Arteriosclerosis (Dallas, Tex). 1989;9(5):633-43.
24. Hayward IP, Bridle KR, Campbell GR, Underwood PA, Campbell JH. Effect of extracellular matrix proteins on vascular smooth muscle cell phenotype. Cell biology international. 1995;19(10):839-46.
25. Byers HR, White GE, Fujiwara K. Organization and function of stress fibers in cells in vitro and in situ. A review. Cell and muscle motility. 1984;5:83-137.
26. Pellegrin S, Mellor H. Actin stress fibres. Journal of Cell Science. 2007;120(20):3491-9.
27. Campbell GR, Campbell JH, Manderson JA, Horrigan S, Rennick RE. Arterial smooth muscle. A multifunctional mesenchymal cell. Archives of pathology & laboratory medicine. 1988;112(10):977-86.
28. Ang AH, Tachas G, Campbell JH, Bateman JF, Campbell GR. Collagen synthesis by cultured rabbit aortic smooth-muscle cells. Alteration with phenotype. The Biochemical journal. 1990;265(2):461-9.
29. Merrilees MJ, Campbell JH, Spanidis E, Campbell GR. Glycosaminoglycan synthesis by smooth muscle cells of differing phenotype and their response to endothelial cell conditioned medium. Atherosclerosis. 1990;81(3):245-54.
30. Owens GK, Kumar MS, Wamhoff BR. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. Physiological reviews. 2004;84(3):767-801.
31. Shimizu RT, Blank RS, Jervis R, Lawrenz-Smith SC, Owens GK. The smooth muscle alpha-actin gene promoter is differentially regulated in smooth muscle versus non-smooth muscle cells. The Journal of biological chemistry. 1995;270(13):7631-43.
32. Owens GK. Regulation of differentiation of vascular smooth muscle cells. Physiological reviews. 1995;75(3):487-517.
33. Alexander MR, Owens GK. Epigenetic control of smooth muscle cell differentiation and phenotypic switching in vascular development and disease. Annual review of physiology. 2012;74:13-40.
34. Orlandi A, Ehrlich HP, Ropraz P, Spagnoli LG, Gabbiani G. Rat aortic smooth muscle cells isolated from different layers and at different times after endothelial denudation show distinct biological features in vitro. Arteriosclerosis and thrombosis : a journal of vascular biology / American Heart Association. 1994;14(6):982-9.

35. Orlandi A, Ferlosio A, Gabbiani G, Spagnoli LG, Ehrlich PH. Phenotypic heterogeneity influences the behavior of rat aortic smooth muscle cells in collagen lattice. Experimental cell research. 2005;311(2):317-27.
36. Majesky MW. Developmental basis of vascular smooth muscle diversity. Arteriosclerosis, thrombosis, and vascular biology. 2007;27(6):1248-58.
37. Manner J, Perez-Pomares JM, Macias D, Munoz-Chapuli R. The origin, formation and developmental significance of the epicardium: a review. Cells, tissues, organs. 2001;169(2):89-103.
38. Bergers G, Song S. The role of pericytes in blood-vessel formation and maintenance. Neuro-Oncology. 2005;7(4):452-64.
39. Cheung C, Bernardo AS, Pedersen RA, Sinha S. Directed differentiation of embryonic origin-specific vascular smooth muscle subtypes from human pluripotent stem cells. Nature protocols. 2014;9(4):929-38.
40. Bargehr J, Low L, Cheung C, Bernard WG, Iyer D, Bennett MR, et al. Embryological Origin of Human Smooth Muscle Cells Influences Their Ability to Support Endothelial Network Formation. Stem cells translational medicine. 2016;5(7):946-59.
41. Lyons I, Parsons LM, Hartley L, Li R, Andrews JE, Robb L, et al. Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes & development. 1995;9(13):1654-66.
42. Tanaka M, Kasahara H, Bartunkova S, Schinke M, Komuro I, Inagaki H, et al. Vertebrate homologs of tinman and bagpipe: roles of the homeobox genes in cardiovascular development. Developmental genetics. 1998;22(3):239-49.
43. Kasahara H, Bartunkova S, Schinke M, Tanaka M, Izumo S. Cardiac and extracardiac expression of Csx/Nkx2.5 homeodomain protein. Circulation research. 1998;82(9):936-46.
44. Prall OW, Menon MK, Solloway MJ, Watanabe Y, Zaffran S, Bajolle F, et al. An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. Cell. 2007;128(5):947-59.
45. Dorn T, Goedel A, Lam JT, Haas J, Tian Q, Herrmann F, et al. Direct nkx2-5 transcriptional repression of isl1 controls cardiomyocyte subtype identity. Stem cells (Dayton, Ohio). 2015;33(4):1113-29.
46. Zhang L, Nomura-Kitabayashi A, Sultana N, Cai W, Cai X, Moon AM, et al. Mesodermal Nkx2.5 is necessary and sufficient for early second heart field development. Developmental biology. 2014;390(1):68-79.
47. Tanaka M, Chen Z, Bartunkova S, Yamasaki N, Izumo S. The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development. Development (Cambridge, England). 1999;126(6):1269-80.
48. Yamagishi H, Yamagishi C, Nakagawa O, Harvey RP, Olson EN, Srivastava D. The combinatorial activities of Nkx2.5 and dHAND are essential for cardiac ventricle formation. Developmental biology. 2001;239(2):190-203.
49. Pashmforoush M, Lu JT, Chen H, Amand TS, Kondo R, Pradervand S, et al. Nkx2-5 pathways and congenital heart disease; loss of ventricular myocyte lineage specification leads to progressive cardiomyopathy and complete heart block. Cell. 2004;117(3):373-86.
50. Abdul Samad F, Suliman BA, Basha SH, Manivasagam T, Essa MM. A Comprehensive In Silico Analysis on the Structural and Functional Impact of SNPs in the Congenital Heart Defects Associated with NKX2-5 Gene-A Molecular Dynamic Simulation Approach. 2016;11(5):e0153999.
51. Edelberg JM, Christie PD, Rosenberg RD. Regulation of vascular bed-specific prothrombotic potential. Circulation research. 2001;89(2):117-24.
52. VanderLaan PA, Reardon CA, Getz GS. Site specificity of atherosclerosis: site-selective responses to atherosclerotic modulators. Arteriosclerosis, thrombosis, and vascular biology. 2004;24(1):12-22.

53. Berk BC. Vascular smooth muscle growth: autocrine growth mechanisms. Physiological reviews. 2001;81(3):999-1030.
54. 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. Circulation research. 1994;75(4):669-81.
55. Frid MG, Aldashev AA, Dempsey EC, Stenmark KR. Smooth muscle cells isolated from discrete compartments of the mature vascular media exhibit unique phenotypes and distinct growth capabilities. Circulation research. 1997;81(6):940-52.
56. Yang X, Long L, Southwood M, Rudarakanchana N, Upton PD, Jeffery TK, et al. Dysfunctional Smad signaling contributes to abnormal smooth muscle cell proliferation in familial pulmonary arterial hypertension. Circulation research. 2005;96(10):1053-63.
57. Kumar A, Raju S, Das A, Mehta AC. Vessels of the Central Airways: A Bronchoscopic Perspective. Chest. 2016;149(3):869-81.
58. Tuder RM, Abman SH, Braun T, Capron F, Stevens T, Thistlethwaite PA, et al. Development and pathology of pulmonary hypertension. Journal of the American College of Cardiology. 2009;54(1 Suppl):S3-9.
59. Badesch DB, Champion HC, Sanchez MA, Hoeper MM, Loyd JE, Manes A, et al. Diagnosis and assessment of pulmonary arterial hypertension. Journal of the American College of Cardiology. 2009;54(1 Suppl):S55-66.
60. Lammers S, Scott D, Hunter K, Tan W, Shandas R, Stenmark KR. Mechanics and function of the pulmonary vasculature: implications for pulmonary vascular disease and right ventricular function. Comprehensive Physiology. 2012;2(1):295-319.
61. Boron WF BE. Medical Physiology. Philadelphia, Pennsylvania 686-711. Updated edition ed. 2005.
62. Ogut O, Brozovich FV. Regulation of force in vascular smooth muscle. Journal of molecular and cellular cardiology. 2003;35(4):347-55.
63. Williams DA, Becker PL, Fay FS. Regional changes in calcium underlying contraction of single smooth muscle cells. Science (New York, NY). 1987;235(4796):1644-8.
64. Rybalkin SD, Yan C, Bornfeldt KE, Beavo JA. Cyclic GMP phosphodiesterases and regulation of smooth muscle function. Circulation research. 2003;93(4):280-91.
65. Sim JY. Nitric oxide and pulmonary hypertension. Korean Journal of Anesthesiology. 2010;58(1):4-14.
66. Vane JR, Anggard EE, Botting RM. Regulatory functions of the vascular endothelium. The New England journal of medicine. 1990;323(1):27-36.
67. Mitchell JA, Ahmetaj-Shala B, Kirkby NS, Wright WR, Mackenzie LS, Reed DM, et al. Role of prostacyclin in pulmonary hypertension. Global Cardiology Science & Practice. 2014;2014(4):382-93.
68. Ozkor MA, Quyyumi AA. Endothelium-derived hyperpolarizing factor and vascular function. Cardiology research and practice. 2011;2011:156146.
69. Archer SL, Gragasin FS, Wu X, Wang S, McMurtry S, Kim DH, et al. Endothelium-derived hyperpolarizing factor in human internal mammary artery is 11,12-epoxyeicosatrienoic acid and causes relaxation by activating smooth muscle BK(Ca) channels. Circulation. 2003;107(5):769-76.
70. Luksha L, Agewall S, Kublickiene K. Endothelium-derived hyperpolarizing factor in vascular physiology and cardiovascular disease. Atherosclerosis. 2009;202(2):330-44.
71. Shimokawa H, Morikawa K. Hydrogen peroxide is an endothelium-derived hyperpolarizing factor in animals and humans. Journal of molecular and cellular cardiology. 2005;39(5):725-32.

72. You J, Golding EM, Bryan RM. Arachidonic acid metabolites, hydrogen peroxide, and EDHF in cerebral arteries. American Journal of Physiology - Heart and Circulatory Physiology. 2005;289(3):H1077-H83.
73. Ergul A. Endothelin-1 and endothelin receptor antagonists as potential cardiovascular therapeutic agents. Pharmacotherapy. 2002;22(1):54-65.
74. Galie N, Manes A, Branzi A. The endothelin system in pulmonary arterial hypertension. Cardiovascular research. 2004;61(2):227-37.
75. The 10 leading causes of death in the world, 2000 and 2012. Published by WHO, 2014. http://www.who.int/mediacentre/factsheets/fs310/en/,
76. Foundation BH. Cardiovascular Disease Statistics 2015. https://wwwbhforguk/publications/statistics/cvd-stats-2015. 2015.
77. Lloyd-Jones DM, Nam BH, D'Agostino RB, Sr., Levy D, Murabito JM, Wang TJ, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. Jama. 2004;291(18):2204-11.
78. Middlemass JB, Yazdani MF, Kai J, Standen PJ, Qureshi N. Introducing genetic testing for cardiovascular disease in primary care: a qualitative study. The British Journal of General Practice. 2014;64(622):e282-9.
79. Khan TA, Shah T, Prieto D, Zhang W, Price J, Fowkes GR, et al. Apolipoprotein E genotype, cardiovascular biomarkers and risk of stroke: systematic review and meta-analysis of 14,015 stroke cases and pooled analysis of primary biomarker data from up to 60,883 individuals. International journal of epidemiology. 2013;42(2):475-92.
80. Chaturvedi N. Ethnic differences in cardiovascular disease. Heart. 2003;89(6):681-6.
81. Dimsdale JE. Psychological Stress and Cardiovascular Disease. Journal of the American College of Cardiology. 2008;51(13):1237-46.
82. Jones CA, Perera A, Chow M, Ho I, Nguyen J, Davachi S. Cardiovascular Disease Risk Among the Poor and Homeless – What We Know So Far. Current Cardiology Reviews. 2009;5(1):69-77.
83. Psaltis PJ, Simari RD. Vascular wall progenitor cells in health and disease. Circulation research. 2015;116(8):1392-412.
84. Mitchell A, Fujisawa T, Newby D, Mills N, Cruden NL. Vascular injury and repair: a potential target for cell therapies. Future cardiology. 2015;11(1):45-60.
85. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. Circulation. 2016;133(4):e38-360.
86. Renna NF, de Las Heras N, Miatello RM. Pathophysiology of vascular remodeling in hypertension. International journal of hypertension. 2013;2013:808353.
87. Schoenhagen P, Ziada KM, Vince DG, Nissen SE, Tuzcu EM. Arterial remodeling and coronary artery disease: the concept of "dilated" versus "obstructive" coronary atherosclerosis. Journal of the American College of Cardiology. 2001;38(2):297-306.
88. Campbell JH, Campbell GR. Smooth muscle phenotypic modulation--a personal experience. Arteriosclerosis, thrombosis, and vascular biology. 2012;32(8):1784-9.
89. Manderson JA, Mosse PR, Safstrom JA, Young SB, Campbell GR. Balloon catheter injury to rabbit carotid artery. I. Changes in smooth muscle phenotype. Arteriosclerosis (Dallas, Tex). 1989;9(3):289-98.
90. Berk BC, Vekshtein V, Gordon HM, Tsuda T. Angiotensin II-stimulated protein synthesis in cultured vascular smooth muscle cells. Hypertension. 1989;13(4):305-14.

91. Tajsic T, Morrell NW. Smooth muscle cell hypertrophy, proliferation, migration and apoptosis in pulmonary hypertension. Comprehensive Physiology. 2011;1(1):295-317.
92. Differentiation. CGTCAMAneSMSACPiDa.
93. Gerthoffer WT. Mechanisms of vascular smooth muscle cell migration. Circulation research. 2007;100(5):607-21.
94. Burg ED, Remillard CV, Yuan JX. Potassium channels in the regulation of pulmonary artery smooth muscle cell proliferation and apoptosis: pharmacotherapeutic implications. British journal of pharmacology. 2008;153 Suppl 1:S99-s111.
95. Small JV. Structure-function relationships in smooth muscle: the missing links. BioEssays : news and reviews in molecular, cellular and developmental biology. 1995;17(9):785-92.
96. Rolfe BE, Muddiman JD, Smith NJ, Campbell GR, Campbell JH. ICAM-1 expression by vascular smooth muscle cells is phenotype-dependent. Atherosclerosis. 2000;149(1):99-110.
97. Sekiguchi K, Kurabayashi M, Oyama Y, Aihara Y, Tanaka T, Sakamoto H, et al. Homeobox protein Hex induces SMemb/nonmuscle myosin heavy chain-B gene expression through the cAMP-responsive element. Circulation research. 2001;88(1):52-8.
98. Eberhard A, Kahlert S, Goede V, Hemmerlein B, Plate KH, Augustin HG. Heterogeneity of angiogenesis and blood vessel maturation in human tumors: implications for antiangiogenic tumor therapies. Cancer research. 2000;60(5):1388-93.
99. Morikawa S, Baluk P, Kaidoh T, Haskell A, Jain RK, McDonald DM. Abnormalities in pericytes on blood vessels and endothelial sprouts in tumors. The American journal of pathology. 2002;160(3):985-1000.
100. Rzucidlo EM, Martin KA, Powell RJ. Regulation of vascular smooth muscle cell differentiation. Journal of vascular surgery. 2007;45 Suppl A:A25-32.
101. Perros F, Montani D, Dorfmuller P, Durand-Gasselin I, Tcherakian C, Le Pavec J, et al. Platelet-derived growth factor expression and function in idiopathic pulmonary arterial hypertension. American journal of respiratory and critical care medicine. 2008;178(1):81-8.
102. Holycross BJ, Blank RS, Thompson MM, Peach MJ, Owens GK. Platelet-derived growth factor-BB-induced suppression of smooth muscle cell differentiation. Circulation research. 1992;71(6):1525-32.
103. Ten Freyhaus H, Berghausen EM, Janssen W, Leuchs M, Zierden M, Murmann K, et al. Genetic Ablation of PDGF-Dependent Signaling Pathways Abolishes Vascular Remodeling and Experimental Pulmonary Hypertension. Arteriosclerosis, thrombosis, and vascular biology. 2015;35(5):1236-45.
104. ten Freyhaus H, Dumitrescu D, Berghausen E, Vantler M, Caglayan E, Rosenkranz S. Imatinib mesylate for the treatment of pulmonary arterial hypertension. Expert opinion on investigational drugs. 2012;21(1):119-34.
105. Merklinger SL, Jones PL, Martinez EC, Rabinovitch M. Epidermal growth factor receptor blockade mediates smooth muscle cell apoptosis and improves survival in rats with pulmonary hypertension. Circulation. 2005;112(3):423-31.
106. Tamaoka M, Hassan M, McGovern T, Ramos-Barbon D, Jo T, Yoshizawa Y, et al. The epidermal growth factor receptor mediates allergic airway remodelling in the rat. The European respiratory journal. 2008;32(5):1213-23.
107. Jackson CL, Reidy MA. Basic fibroblast growth factor: its role in the control of smooth muscle cell migration. The American journal of pathology. 1993;143(4):1024-31.
108. Tu L. Autocrine fibroblast growth factor-2 signaling contributes to altered endothelial phenotype in pulmonary hypertension. 2011;45(2):311-22.

109. Evans JDW, Girerd B, Montani D, Wang XJ, Galiè N, Austin ED, et al. BMPR2 mutations and survival in pulmonary arterial hypertension: an individual participant data meta-analysis. The Lancet Respiratory Medicine. 2016;4(2):129-37.
110. Gordon KJ, Blobe GC. Role of transforming growth factor-beta superfamily signaling pathways in human disease. Biochimica et biophysica acta. 2008;1782(4):197-228.
111. Mack CP. Signaling mechanisms that regulate smooth muscle cell differentiation. Arteriosclerosis, thrombosis, and vascular biology. 2011;31(7):1495-505.
112. Grainger DJ. TGF-β and atherosclerosis in man. Cardiovascular research. 2007;74(2):213-22.
113. Abdel-Latif A, Zuba-Surma EK, Case J, Tiwari S, Hunt G, Ranjan S, et al. TGF-beta1 enhances cardiomyogenic differentiation of skeletal muscle-derived adult primitive cells. Basic research in cardiology. 2008;103(6):514-24.
114. Chen PY, Qin L, Barnes C, Charisse K, Yi T, Zhang X, et al. FGF regulates TGF-beta signaling and endothelial-to-mesenchymal transition via control of let-7 miRNA expression. Cell reports. 2012;2(6):1684-96.
115. Guo X, Chen SY. Transforming growth factor-β and smooth muscle differentiation. World Journal of Biological Chemistry. 2012;3(3):41-52.
116. Reifenberg K, Cheng F, Orning C, Crain J, Küpper I, Wiese E, et al. Overexpression of TGF-ß1 in Macrophages Reduces and Stabilizes Atherosclerotic Plaques in ApoE-Deficient Mice. PloS one. 2012;7(7):e40990.
117. Toma I, McCaffrey TA. Transforming growth factor-β and atherosclerosis: interwoven atherogenic and atheroprotective aspects. Cell and tissue research. 2012;347(1):155-75.
118. Schuliga M, Javeed A, Harris T, Xia Y, Qin C, Wang Z, et al. Transforming growth factor-beta-induced differentiation of airway smooth muscle cells is inhibited by fibroblast growth factor-2. American journal of respiratory cell and molecular biology. 2013;48(3):346-53.
119. Chen PY, Qin L, Li G, Tellides G, Simons M. Fibroblast growth factor (FGF) signaling regulates transforming growth factor beta (TGFbeta)-dependent smooth muscle cell phenotype modulation. Scientific reports. 2016;6:33407.
120. Meng X-m, Nikolic-Paterson DJ, Lan HY. TGF-[beta]: the master regulator of fibrosis. Nat Rev Nephrol. 2016;12(6):325-38.
121. Qiu P, Feng XH, Li L. Interaction of Smad3 and SRF-associated complex mediates TGF-beta1 signals to regulate SM22 transcription during myofibroblast differentiation. Journal of molecular and cellular cardiology. 2003;35(12):1407-20.
122. Qiu P, Ritchie RP, Fu Z, Cao D, Cumming J, Miano JM, et al. Myocardin enhances Smad3-mediated transforming growth factor-beta1 signaling in a CArG box-independent manner: Smad-binding element is an important cis element for SM22alpha transcription in vivo. Circulation research. 2005;97(10):983-91.
123. Nabel EG, Shum L, Pompili VJ, Yang ZY, San H, Shu HB, et al. Direct transfer of transforming growth factor beta 1 gene into arteries stimulates fibrocellular hyperplasia. Proceedings of the National Academy of Sciences of the United States of America. 1993;90(22):10759-63.
124. Li W, Li Q, Jiao Y, Qin L, Ali R, Zhou J, et al. Tgfbr2 disruption in postnatal smooth muscle impairs aortic wall homeostasis. The Journal of clinical investigation. 2014;124(2):755-67.
125. Lutgens E, Gijbels M, Smook M, Heeringa P, Gotwals P, Koteliansky VE, et al. Transforming growth factor-beta mediates balance between inflammation and fibrosis during plaque progression. Arteriosclerosis, thrombosis, and vascular biology. 2002;22(6):975-82.
126. Huang JS, Wang YH, Ling TY, Chuang SS, Johnson FE, Huang SS. Synthetic TGF-beta antagonist accelerates wound healing and reduces scarring.

FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2002;16(10):1269-70.
127. Suwanabol PA, Kent KC, Liu B. TGF-β and Restenosis Revisited: A Smad Link. The Journal of surgical research. 2011;167(2):287-97.
128. Morrell NW, Yang X, Upton PD, Jourdan KB, Morgan N, Sheares KK, et al. Altered growth responses of pulmonary artery smooth muscle cells from patients with primary pulmonary hypertension to transforming growth factor-beta(1) and bone morphogenetic proteins. Circulation. 2001;104(7):790-5.
129. Feinberg MW, Watanabe M, Lebedeva MA, Depina AS, Hanai J, Mammoto T, et al. Transforming growth factor-beta1 inhibition of vascular smooth muscle cell activation is mediated via Smad3. The Journal of biological chemistry. 2004;279(16):16388-93.
130. Cooley BC, Nevado J, Mellad J, Yang D, St Hilaire C, Negro A, et al. TGF-beta signaling mediates endothelial-to-mesenchymal transition (EndMT) during vein graft remodeling. Science translational medicine. 2014;6(227):227ra34.
131. Jiang Z, Tao M, Omalley KA, Wang D, Ozaki CK, Berceli SA. Established neointimal hyperplasia in vein grafts expands via TGF-beta-mediated progressive fibrosis. American journal of physiology Heart and circulatory physiology. 2009;297(4):H1200-7.
132. Chen PY, Qin L, Li G, Tellides G, Simons M. Smooth muscle FGF/TGFbeta cross talk regulates atherosclerosis progression. 2016;8(7):712-28.
133. Kawai-Kowase K, Sato H, Oyama Y, Kanai H, Sato M, Doi H, et al. Basic fibroblast growth factor antagonizes transforming growth factor-beta1-induced smooth muscle gene expression through extracellular signal-regulated kinase 1/2 signaling pathway activation. Arteriosclerosis, thrombosis, and vascular biology. 2004;24(8):1384-90.
134. Denton CP, Abraham DJ. Transforming growth factor-beta and connective tissue growth factor: key cytokines in scleroderma pathogenesis. Current opinion in rheumatology. 2001;13(6):505-11.
135. Denton CP, Khan K, Hoyles RK, Shiwen X, Leoni P, Chen Y, et al. Inducible lineage-specific deletion of TbetaRII in fibroblasts defines a pivotal regulatory role during adult skin wound healing. The Journal of investigative dermatology. 2009;129(1):194-204.
136. Derrett-Smith EC, Dooley A, Gilbane AJ, Trinder SL, Khan K, Baliga R, et al. Endothelial injury in a transforming growth factor beta-dependent mouse model of scleroderma induces pulmonary arterial hypertension. Arthritis and rheumatism. 2013;65(11):2928-39.
137. Ponticos M, Harvey C, Ikeda T, Abraham D, Bou-Gharios G. JunB mediates enhancer/promoter activity of COL1A2 following TGF-beta induction. Nucleic acids research. 2009;37(16):5378-89.
138. Min LJ, Mogi M, Tsukuda K, Jing F, Ohshima K, Nakaoka H, et al. Direct stimulation of angiotensin II type 2 receptor initiated after stroke ameliorates ischemic brain damage. American journal of hypertension. 2014;27(8):1036-44.
139. Brown MJ. Therapeutic potential of vaccines in the management of hypertension. Drugs. 2008;68(18):2557-60.
140. Mehta PK, Griendling KK. Angiotensin II cell signaling: physiological and pathological effects in the cardiovascular system. American journal of physiology Cell physiology. 2007;292(1):C82-97.
141. Wang M, Tang YB, Ma MM, Chen JH, Hu CP, Zhao SP, et al. TRPC3 channel confers cerebrovascular remodelling during hypertension via transactivation of EGF receptor signalling. Cardiovascular research. 2016;109(1):34-43.
142. Noseda M, Fu Y, Niessen K, Wong F, Chang L, McLean G, et al. Smooth Muscle alpha-actin is a direct target of Notch/CSL. Circulation research. 2006;98(12):1468-70.

143. Tang Y, Urs S, Liaw L. Hairy-related transcription factors inhibit Notch-induced smooth muscle alpha-actin expression by interfering with Notch intracellular domain/CBF-1 complex interaction with the CBF-1-binding site. Circulation research. 2008;102(6):661-8.
144. Karimi Galougahi K, Ashley EA, Ali ZA. Redox regulation of vascular remodeling. Cellular and molecular life sciences : CMLS. 2016;73(2):349-63.
145. Miano JM. Serum response factor: toggling between disparate programs of gene expression. Journal of molecular and cellular cardiology. 2003;35(6):577-93.
146. Miano JM, Long X, Fujiwara K. Serum response factor: master regulator of the actin cytoskeleton and contractile apparatus. American journal of physiology Cell physiology. 2007;292(1):C70-81.
147. Yoshida T, Sinha S, Dandre F, Wamhoff BR, Hoofnagle MH, Kremer BE, et al. Myocardin is a key regulator of CArG-dependent transcription of multiple smooth muscle marker genes. Circulation research. 2003;92(8):856-64.
148. Xie C, Guo Y, Zhu T, Zhang J, Ma PX, Chen YE. Yap1 protein regulates vascular smooth muscle cell phenotypic switch by interaction with myocardin. The Journal of biological chemistry. 2012;287(18):14598-605.
149. Liu F, Wang X, Hu G, Wang Y, Zhou J. The transcription factor TEAD1 represses smooth muscle-specific gene expression by abolishing myocardin function. The Journal of biological chemistry. 2014;289(6):3308-16.
150. Wang L, Qiu P, Jiao J, Hirai H, Xiong W, Zhang J, et al. Yes-Associated Protein Inhibits Transcription of Myocardin and Attenuates Differentiation of Vascular Smooth Muscle Cell from Cardiovascular Progenitor Cell Lineage. Stem cells (Dayton, Ohio). 2016.
151. Miano JM. Myocardin in biology and disease. Journal of biomedical research. 2015;29(1):3-19.
152. Guignabert C, Tu L, Girerd B, Ricard N, Huertas A, Montani D, et al. New molecular targets of pulmonary vascular remodeling in pulmonary arterial hypertension: importance of endothelial communication. Chest. 2015;147(2):529-37.
153. Vanhoutte PM, Shimokawa H, Tang EH, Feletou M. Endothelial dysfunction and vascular disease. Acta physiologica (Oxford, England). 2009;196(2):193-222.
154. Vascular Inflammation and Endothelial Dysfunction in Experimental Hypertension. International journal of hypertension. 2011;2011.
155. Medici D, Shore EM, Lounev VY, Kaplan FS, Kalluri R, Olsen BR. Conversion of vascular endothelial cells into multipotent stem-like cells. Nature medicine. 2010;16(12):1400-6.
156. Cavarretta E, Latronico MV, Condorelli G. Endothelial-to-mesenchymal transition and microRNA-21: the game is on again. Arteriosclerosis, thrombosis, and vascular biology. 2012;32(2):165-6.
157. Gonzalez DM, Medici D. Signaling mechanisms of the epithelial-mesenchymal transition. Sci Signal. 2014;7(344):re8.
158. Piera-Velazquez S, Mendoza FA, Jimenez SA. Endothelial to Mesenchymal Transition (EndoMT) in the Pathogenesis of Human Fibrotic Diseases. J Clin Med. 2016;5(4).
159. Coll-Bonfill N, Musri MM, Ivo V, Barbera JA, Tura-Ceide O. Transdifferentiation of endothelial cells to smooth muscle cells play an important role in vascular remodelling. Am J Stem Cells. 2015;4(1):13-21.
160. Moore-Morris T, Guimaraes-Camboa N, Banerjee I, Zambon AC, Kisseleva T, Velayoudon A, et al. Resident fibroblast lineages mediate pressure overload-induced cardiac fibrosis. The Journal of clinical investigation. 2014;124(7):2921-34.
161. Welch-Reardon KM, Wu N, Hughes CC. A role for partial endothelial-mesenchymal transitions in angiogenesis? Arteriosclerosis, thrombosis, and vascular biology. 2015;35(2):303-8.

162. Pasterkamp G, den Ruijter HM, Libby P. Temporal shifts in clinical presentation and underlying mechanisms of atherosclerotic disease. Nat Rev Cardiol. 2016.
163. Hamburg NM, Balady GJ. Exercise rehabilitation in peripheral artery disease: functional impact and mechanisms of benefits. Circulation. 2011;123(1):87-97.
164. Pant R, Marok R, Klein LW. Pathophysiology of coronary vascular remodeling: relationship with traditional risk factors for coronary artery disease. Cardiology in review. 2014;22(1):13-6.
165. Nicolosi PA, Tombetti E, Maugeri N, Rovere-Querini P, Brunelli S, Manfredi AA. Vascular Remodelling and Mesenchymal Transition in Systemic Sclerosis. Stem Cells Int. 2016;2016.
166. Mensah KA, Yadav R, Trow TK, Brunet CM, Fares WH. Lupus-Associated Pulmonary Arterial Hypertension: Variable Course and Importance of Prompt Recognition. Case reports in medicine. 2015;2015:328435.
167. Sandoo A, Veldhuijzen van Zanten JJ, Metsios GS, Carroll D, Kitas GD. Vascular function and morphology in rheumatoid arthritis: a systematic review. Rheumatology (Oxford, England). 2011;50(11):2125-39.
168. Kolodgie FD, Nakazawa G, Sangiorgi G, Ladich E, Burke AP, Virmani R. Pathology of Atherosclerosis and Stenting. Neuroimaging clinics of North America 2007;17(3):285-vii. doi:10.1016/j.nic.2007.03.006.
169. Dalager S, Paaske WP, Kristensen IB, Laurberg JM, Falk E. Artery-related differences in atherosclerosis expression: implications for atherogenesis and dynamics in intima-media thickness. Stroke; a journal of cerebral circulation. 2007;38(10):2698-705.
170. Bennett MR, Sinha S, Owens GK. Vascular Smooth Muscle Cells in Atherosclerosis. Circulation research. 2016;118(4):692-702.
171. Ackers-Johnson M, Talasila A, Sage AP, Long X, Bot I, Morrell NW, et al. Myocardin regulates vascular smooth muscle cell inflammatory activation and disease. Arteriosclerosis, thrombosis, and vascular biology. 2015;35(4):817-28.
172. Shankman LS, Gomez D, Cherepanova OA, Salmon M, Alencar GF, Haskins RM, et al. KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. Nature medicine. 2015;21(6):628-37.
173. Varnava AM, Mills PG, Davies MJ. Relationship between coronary artery remodeling and plaque vulnerability. Circulation. 2002;105(8):939-43.
174. Silver AE, Vita JA. Shear-stress-mediated arterial remodeling in atherosclerosis: too much of a good thing? Circulation. 2006;113(24):2787-9.
175. Peach G, Griffin M, Jones KG, Thompson MM, Hinchliffe RJ. Diagnosis and management of peripheral arterial disease. BMJ. 2012;345:e5208.
176. Ouriel K. Peripheral arterial disease. Lancet. 2001;358(9289):1257-64.
177. Hirsch AT, Criqui MH, Treat-Jacobson D, Regensteiner JG, Creager MA, Olin JW, et al. Peripheral arterial disease detection, awareness, and treatment in primary care. Jama. 2001;286(11):1317-24.
178. Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Haase N, et al. Heart disease and stroke statistics--2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation. 2008;117(4):e25-146.
179. Belch JJ, Topol EJ, Agnelli G, Bertrand M, Califf RM, Clement DL, et al. Critical issues in peripheral arterial disease detection and management: a call to action. Arch Intern Med. 2003;163(8):884-92.
180. Simonneau G, Gatzoulis MA, Adatia I, Celermajer D, Denton C, Ghofrani A, et al. Updated clinical classification of pulmonary hypertension. Journal of the American College of Cardiology. 2013;62(25 Suppl):D34-41.

181. "2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS)." Nazzareno Galie, Marc Humbert, Jean-Luc Vachiery, Simon Gibbs, Irene Lang, Adam Torbicki, Gerald Simonneau, Andrew Peacock, Anton Vonk Noordegraaf, Maurice Beghetti, Ardeschir Ghofrani, Miguel Angel Gomez Sanchez, Georg Hansmann, Walter Klepetko, Patrizio Lancellotti, Marco Matucci, Theresa McDonagh, Luc A. Pierard, Pedro T. Trindade, Maurizio Zompatori and Marius Hoeper. Eur Respir J 2015; 46: 903-975. The European respiratory journal. 2015;46(6):1855-6.
182. Humbert M, Sitbon O, Chaouat A, Bertocchi M, Habib G, Gressin V, et al. Pulmonary arterial hypertension in France: results from a national registry. American journal of respiratory and critical care medicine. 2006;173(9):1023-30.
183. Harvey A, Montezano AC, Lopes RA, Rios F, Touyz RM. Vascular Fibrosis in Aging and Hypertension: Molecular Mechanisms and Clinical Implications. The Canadian journal of cardiology. 2016;32(5):659-68.
184. Badesch DB, Raskob GE, Elliott CG, Krichman AM, Farber HW, Frost AE, et al. Pulmonary arterial hypertension: baseline characteristics from the REVEAL Registry. Chest. 2010;137(2):376-87.
185. Rich S, Dantzker DR, Ayres SM, Bergofsky EH, Brundage BH, Detre KM, et al. Primary pulmonary hypertension. A national prospective study. Ann Intern Med. 1987;107(2):216-23.
186. Callan P, Clark AL. Right heart catheterisation: indications and interpretation. Heart. 2016;102(2):147-57.
187. Shi Y, O'Brien JE, Fard A, Mannion JD, Wang D, Zalewski A. Adventitial myofibroblasts contribute to neointimal formation in injured porcine coronary arteries. Circulation. 1996;94(7):1655-64.
188. Stevens T, Phan S, Frid MG, Alvarez D, Herzog E, Stenmark KR. Lung vascular cell heterogeneity: endothelium, smooth muscle, and fibroblasts. Proc Am Thorac Soc. 2008;5(7):783-91.
189. Demerouti EA, Manginas AN, Athanassopoulos GD, Karatasakis GT. Complications leading to sudden cardiac death in pulmonary arterial hypertension. Respiratory care. 2013;58(7):1246-54.
190. Chin KM, Rubin LJ. Pulmonary arterial hypertension. Journal of the American College of Cardiology. 2008;51(16):1527-38.
191. de Jesus Perez VA. Molecular pathogenesis and current pathology of pulmonary hypertension. Heart Fail Rev. 2016;21(3):239-57.
192. Rodrigues SF, Granger DN. Blood cells and endothelial barrier function. Tissue barriers. 2015;3(1-2):e978720.
193. Morrell NW, Adnot S, Archer SL, Dupuis J, Jones PL, MacLean MR, et al. Cellular and molecular basis of pulmonary arterial hypertension. Journal of the American College of Cardiology. 2009;54(1 Suppl):S20-31.
194. Humbert M, Montani D, Perros F, Dorfmuller P, Adnot S, Eddahibi S. Endothelial cell dysfunction and cross talk between endothelium and smooth muscle cells in pulmonary arterial hypertension. Vascul Pharmacol. 2008;49(4-6):113-8.
195. Duluc L, Wojciak-Stothard B. Rho GTPases in the regulation of pulmonary vascular barrier function. Cell and tissue research. 2014;355(3):675-85.
196. Storck EM, Wojciak-Stothard B. Rho GTPases in pulmonary vascular dysfunction. Vascul Pharmacol. 2013;58(3):202-10.
197. Wojciak-Stothard B. New drug targets for pulmonary hypertension: Rho GTPases in pulmonary vascular remodelling. Postgraduate medical journal. 2008;84(993):348-53.
198. Hopper RK, Moonen JR, Diebold I, Cao A, Rhodes CJ, Tojais NF, et al. In Pulmonary Arterial Hypertension, Reduced BMPR2 Promotes Endothelial-to-

Mesenchymal Transition via HMGA1 and Its Target Slug. Circulation. 2016;133(18):1783-94.
199. Ranchoux B, Antigny F, Rucker-Martin C, Hautefort A, Pechoux C, Bogaard HJ, et al. Endothelial-to-mesenchymal transition in pulmonary hypertension. Circulation. 2015;131(11):1006-18.
200. Medici D. Endothelial-Mesenchymal Transition in Regenerative Medicine. Stem Cells Int. 2016;2016:6962801.
201. Stenmark KR, Frid M, Perros F. Endothelial-to-Mesenchymal Transition: An Evolving Paradigm and a Promising Therapeutic Target in PAH. Circulation. 2016;133(18):1734-7.
202. Austin ED, Cogan JD, West JD, Hedges LK, Hamid R, Dawson EP, et al. Alterations in oestrogen metabolism: implications for higher penetrance of familial pulmonary arterial hypertension in females. The European respiratory journal. 2009;34(5):1093-9.
203. Austin ED, Loyd JE, Phillips JA, 3rd. Genetics of pulmonary arterial hypertension. Seminars in respiratory and critical care medicine. 2009;30(4):386-98.
204. Frump AL, Goss KN, Vayl A, Albrecht M, Fisher A, Tursunova R, et al. Estradiol improves right ventricular function in rats with severe angioproliferative pulmonary hypertension: effects of endogenous and exogenous sex hormones. Am J Physiol Lung Cell Mol Physiol. 2015;308(9):L873-90.
205. Stenmark KR, Fagan KA, Frid MG. Hypoxia-induced pulmonary vascular remodeling: cellular and molecular mechanisms. Circulation research. 2006;99(7):675-91.
206. Ma H, Xu D, Wu Y, Ma Y, Li Z. To decipher the hypoxic pulmonary hypertension: Vascular heterogeneity and the hypothesis of hypoxic responsive threshold. Journal of Medical Hypotheses and Ideas. 2015;9(1):29-37.
207. Lanner MC, Raper M, Pratt WM, Rhoades RA. Heterotrimeric G proteins and the platelet-derived growth factor receptor-beta contribute to hypoxic proliferation of smooth muscle cells. American journal of respiratory cell and molecular biology. 2005;33(4):412-9.
208. Moudgil R, Michelakis ED, Archer SL. Hypoxic pulmonary vasoconstriction. J Appl Physiol (1985). 2005;98(1):390-403.
209. Ball MK, Waypa GB, Mungai PT, Nielsen JM, Czech L, Dudley VJ, et al. Regulation of hypoxia-induced pulmonary hypertension by vascular smooth muscle hypoxia-inducible factor-1alpha. American journal of respiratory and critical care medicine. 2014;189(3):314-24.
210. Carmeliet P, Dor Y, Herbert JM, Fukumura D, Brusselmans K, Dewerchin M, et al. Role of HIF-1alpha in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. Nature. 1998;394(6692):485-90.
211. Schultz K, Fanburg BL, Beasley D. Hypoxia and hypoxia-inducible factor-1alpha promote growth factor-induced proliferation of human vascular smooth muscle cells. American journal of physiology Heart and circulatory physiology. 2006;290(6):H2528-34.
212. Li X, He Y, Xu Y, Huang X, Liu J, Xie M, et al. KLF5 mediates vascular remodeling via HIF-1alpha in hypoxic pulmonary hypertension. Am J Physiol Lung Cell Mol Physiol. 2016;310(4):L299-310.
213. Courboulin A, Tremblay VL, Barrier M, Meloche J, Jacob MH, Chapolard M, et al. Kruppel-like factor 5 contributes to pulmonary artery smooth muscle proliferation and resistance to apoptosis in human pulmonary arterial hypertension. Respir Res. 2011;12:128.
214. Fijalkowska I, Xu W, Comhair SA, Janocha AJ, Mavrakis LA, Krishnamachary B, et al. Hypoxia inducible-factor1alpha regulates the metabolic shift of pulmonary hypertensive endothelial cells. The American journal of pathology. 2010;176(3):1130-8.

215. Devlin AM, Clark JS, Reid JL, Dominiczak AF. DNA synthesis and apoptosis in smooth muscle cells from a model of genetic hypertension. Hypertension. 2000;36(1):110-5.
216. Gobe G, Browning J, Howard T, Hogg N, Winterford C, Cross R. Apoptosis occurs in endothelial cells during hypertension-induced microvascular rarefaction. Journal of structural biology. 1997;118(1):63-72.
217. Pollman MJ, Yamada T, Horiuchi M, Gibbons GH. Vasoactive substances regulate vascular smooth muscle cell apoptosis. Countervailing influences of nitric oxide and angiotensin II. Circulation research. 1996;79(4):748-56.
218. Jais X, Launay D, Yaici A, Le Pavec J, Tcherakian C, Sitbon O, et al. Immunosuppressive therapy in lupus- and mixed connective tissue disease-associated pulmonary arterial hypertension: a retrospective analysis of twenty-three cases. Arthritis and rheumatism. 2008;58(2):521-31.
219. Lepetit H, Eddahibi S, Fadel E, Frisdal E, Munaut C, Noel A, et al. Smooth muscle cell matrix metalloproteinases in idiopathic pulmonary arterial hypertension. The European respiratory journal. 2005;25(5):834-42.
220. Montani D, Seferian A, Savale L, Simonneau G, Humbert M. Drug-induced pulmonary arterial hypertension: a recent outbreak. Eur Respir Rev. 2013;22(129):244-50.
221. Dempsie Y, Morecroft I, Welsh DJ, MacRitchie NA, Herold N, Loughlin L, et al. Converging evidence in support of the serotonin hypothesis of dexfenfluramine-induced pulmonary hypertension with novel transgenic mice. Circulation. 2008;117(22):2928-37.
222. Galie N, Manes A, Branzi A. Prostanoids for pulmonary arterial hypertension. Am J Respir Med. 2003;2(2):123-37.
223. Galie N, Rubin L, Hoeper M, Jansa P, Al-Hiti H, Meyer G, et al. Treatment of patients with mildly symptomatic pulmonary arterial hypertension with bosentan (EARLY study): a double-blind, randomised controlled trial. Lancet. 2008;371(9630):2093-100.
224. Ghofrani HA, Osterloh IH, Grimminger F. Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond. Nature reviews Drug discovery. 2006;5(8):689-702.
225. Bhatia S, Frantz RP, Severson CJ, Durst LA, McGoon MD. Immediate and long-term hemodynamic and clinical effects of sildenafil in patients with pulmonary arterial hypertension receiving vasodilator therapy. Mayo Clinic proceedings. 2003;78(10):1207-13.
226. Ahuja J, Arora D, Kanne JP, Henry TS, Godwin JD. Imaging of Pulmonary Manifestations of Connective Tissue Diseases. Radiol Clin North Am. 2016;54(6):1015-31.
227. Ling Y, Johnson MK, Kiely DG, Condliffe R, Elliot CA, Gibbs JS, et al. Changing demographics, epidemiology, and survival of incident pulmonary arterial hypertension: results from the pulmonary hypertension registry of the United Kingdom and Ireland. American journal of respiratory and critical care medicine. 2012;186(8):790-6.
228. Kherbeck N, Tamby MC, Bussone G, Dib H, Perros F, Humbert M, et al. The role of inflammation and autoimmunity in the pathophysiology of pulmonary arterial hypertension. Clin Rev Allergy Immunol. 2013;44(1):31-8.
229. Elhai M, Meune C, Avouac J, Kahan A, Allanore Y. Trends in mortality in patients with systemic sclerosis over 40 years: a systematic review and meta-analysis of cohort studies. Rheumatology (Oxford, England). 2012;51(6):1017-26.
230. Chifflot H, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. Seminars in arthritis and rheumatism. 2008;37(4):223-35.
231. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of

Rheumatology/European League against Rheumatism collaborative initiative. Arthritis and rheumatism. 2013;65(11):2737-47.
232. Jablonska S, Blaszczyk M. Scleroderma overlap syndromes. Advances in experimental medicine and biology. 1999;455:85-92.
233. Bhansing KJ, Lammens M, Knaapen HK, van Riel PL, van Engelen BG, Vonk MC. Scleroderma-polymyositis overlap syndrome versus idiopathic polymyositis and systemic sclerosis: a descriptive study on clinical features and myopathology. Arthritis research & therapy. 2014;16(3):R111.
234. Chung L, Utz PJ. Antibodies in scleroderma: direct pathogenicity and phenotypic associations. Current rheumatology reports. 2004;6(2):156-63.
235. Valerio CJ, Schreiber BE, Handler CE, Denton CP, Coghlan JG. Borderline mean pulmonary artery pressure in patients with systemic sclerosis: transpulmonary gradient predicts risk of developing pulmonary hypertension. Arthritis and rheumatism. 2013;65(4):1074-84.
236. Coghlan JG, Denton CP, Grunig E, Bonderman D, Distler O, Khanna D, et al. Evidence-based detection of pulmonary arterial hypertension in systemic sclerosis: the DETECT study. Annals of the rheumatic diseases. 2014;73(7):1340-9.
237. Cappelli S, Bellando Randone S, Camiciottoli G, De Paulis A, Guiducci S, Matucci-Cerinic M. Interstitial lung disease in systemic sclerosis: where do we stand? Eur Respir Rev. 2015;24(137):411-9.
238. Fisher MR, Forfia PR, Chamera E, Housten-Harris T, Champion HC, Girgis RE, et al. Accuracy of Doppler echocardiography in the hemodynamic assessment of pulmonary hypertension. American journal of respiratory and critical care medicine. 2009;179(7):615-21.
239. Condliffe R, Howard LS. Connective tissue disease-associated pulmonary arterial hypertension. F1000Prime Rep. 2015;7:06.
240. Lefevre G, Dauchet L, Hachulla E, Montani D, Sobanski V, Lambert M, et al. Survival and prognostic factors in systemic sclerosis-associated pulmonary hypertension: a systematic review and meta-analysis. Arthritis and rheumatism. 2013;65(9):2412-23.
241. Overbeek MJ, Mouchaers KT, Niessen HM, Hadi AM, Kupreishvili K, Boonstra A, et al. Characteristics of interstitial fibrosis and inflammatory cell infiltration in right ventricles of systemic sclerosis-associated pulmonary arterial hypertension. Int J Rheumatol. 2010;2010.
242. Mathai SC, Hassoun PM. Pulmonary arterial hypertension in connective tissue diseases. Heart Fail Clin. 2012;8(3):413-25.
243. Thakkar V, Lau EM. Connective tissue disease-related pulmonary arterial hypertension. Best Pract Res Clin Rheumatol. 2016;30(1):22-38.
244. Cambien F, Tiret L. Genetics of cardiovascular diseases: from single mutations to the whole genome. Circulation. 2007;116(15):1714-24.
245. Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. Hum Genet. 2011;129(4):351-70.
246. Roberts R. A genetic basis for coronary artery disease. Trends Cardiovasc Med. 2015;25(3):171-8.
247. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.
248. Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet. 2003;4(8):587-97.
249. Patnala R, Clements J, Batra J. Candidate gene association studies: a comprehensive guide to useful in silico tools. BMC Genet. 2013;14:39.
250. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science (New York, NY). 2005;308(5720):385-9.

251. Elder SJ, Lichtenstein AH, Pittas AG, Roberts SB, Fuss PJ, Greenberg AS, et al. Genetic and environmental influences on factors associated with cardiovascular disease and the metabolic syndrome. J Lipid Res. 2009;50(9):1917-26.
252. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(30):10604-9.
253. Voelter-Mahlknecht S. Epigenetic associations in relation to cardiovascular prevention and therapeutics. Clin Epigenetics. 2016;8:4.
254. Schiano C, Vietri MT, Grimaldi V, Picascia A, De Pascale MR, Napoli C. Epigenetic-related therapeutic challenges in cardiovascular disease. Trends in Pharmacological Sciences.36(4):226-35.
255. Turgeon PJ, Sukumar AN, Marsden PA. Epigenetics of Cardiovascular Disease - A New "Beat" in Coronary Artery Disease. Med Epigenet. 2014;2(1):37-52.
256. Abi Khalil C. The emerging role of epigenetics in cardiovascular disease. Ther Adv Chronic Dis. 2014;5(4):178-87.
257. Webster AL, Yan MS, Marsden PA. Epigenetics and cardiovascular disease. The Canadian journal of cardiology. 2013;29(1):46-57.
258. Kim J, Kim JY, Song KS, Lee YH, Seo JS, Jelinek J, et al. Epigenetic changes in estrogen receptor beta gene in atherosclerotic cardiovascular tissues and in-vitro vascular senescence. Biochimica et biophysica acta. 2007;1772(1):72-80.
259. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19(1):92-105.
260. Exiqon. What are lncRNAs?
261. Machado RD, Eickelberg O, Elliott CG, Geraci MW, Hanaoka M, Loyd JE, et al. Genetics and genomics of pulmonary arterial hypertension. Journal of the American College of Cardiology. 2009;54(1 Suppl):S32-42.
262. Nichols WC, Koller DL, Slovis B, Foroud T, Terry VH, Arnold ND, et al. Localization of the gene for familial primary pulmonary hypertension to chromosome 2q31-32. Nature genetics. 1997;15(3):277-80.
263. Morse JH, Jones AC, Barst RJ, Hodge SE, Wilhelmsen KC, Nygaard TG. Mapping of familial primary pulmonary hypertension locus (PPH1) to chromosome 2q31-q32. Circulation. 1997;95(12):2603-6.
264. Deng Z, Morse JH, Slager SL, Cuervo N, Moore KJ, Venetos G, et al. Familial primary pulmonary hypertension (gene PPH1) is caused by mutations in the bone morphogenetic protein receptor-II gene. American journal of human genetics. 2000;67(3):737-44.
265. Larkin EK, Newman JH, Austin ED, Hemnes AR, Wheeler L, Robbins IM, et al. Longitudinal analysis casts doubt on the presence of genetic anticipation in heritable pulmonary arterial hypertension. American journal of respiratory and critical care medicine. 2012;186(9):892-6.
266. White K, Dempsie Y, Nilsen M, Wright AF, Loughlin L, MacLean MR. The serotonin transporter, gender, and 17beta oestradiol in the development of pulmonary arterial hypertension. Cardiovascular research. 2011;90(2):373-82.
267. Machado RD, Southgate L, Eichstaedt CA, Aldred MA, Austin ED, Best DH, et al. Pulmonary Arterial Hypertension: A Current Perspective on Established and Emerging Molecular Genetic Defects. Hum Mutat. 2015;36(12):1113-27.
268. Antigny F, Hautefort A, Meloche J, Belacel-Ouari M, Manoury B, Rucker-Martin C, et al. Potassium Channel Subfamily K Member 3 (KCNK3) Contributes to the Development of Pulmonary Arterial Hypertension. Circulation. 2016;133(14):1371-85.

269. Best DH, Sumner KL, Austin ED, Chung WK, Brown LM, Borczuk AC, et al. EIF2AK4 mutations in pulmonary capillary hemangiomatosis. Chest. 2014;145(2):231-6.
270. Eyries M, Montani D, Girerd B, Perret C, Leroy A, Lonjou C, et al. EIF2AK4 mutations cause pulmonary veno-occlusive disease, a recessive form of pulmonary hypertension. Nature genetics. 2014;46(1):65-9.
271. Germain M, Eyries M, Montani D, Poirier O, Girerd B, Dorfmuller P, et al. Genome-wide association analysis identifies a susceptibility locus for pulmonary arterial hypertension. Nature genetics. 2013;45(5):518-21.
272. Wang G, Knight L, Ji R, Lawrence P, Kanaan U, Li L, et al. Early onset severe pulmonary arterial hypertension with 'two-hit' digenic mutations in both BMPR2 and KCNA5 genes. Int J Cardiol. 2014;177(3):e167-9.
273. Huston JH, Ryan JJ. The emerging role of epigenetics in pulmonary arterial hypertension: an important avenue for clinical trials (2015 Grover Conference Series). Pulm Circ. 2016;6(3):274-84.
274. Kim JD, Lee A, Choi J, Park Y, Kang H, Chang W, et al. Epigenetic modulation as a therapeutic approach for pulmonary arterial hypertension. Exp Mol Med. 2015;47:e175.
275. Parikh VN, Jin RC, Rabello S, Gulbahce N, White K, Hale A, et al. MicroRNA-21 integrates pathogenic signaling to control pulmonary hypertension: results of a network bioinformatics approach. Circulation. 2012;125(12):1520-32.
276. Iannone L, Zhao L, Dubois O, Duluc L, Rhodes Christopher J, Wharton J, et al. miR-21/DDAH1 pathway regulates pulmonary vascular responses to hypoxia. Biochemical Journal. 2014;462(1):103-12.
277. Wang D, Zhang H, Li M, Frid MG, Flockton AR, McKeon BA, et al. MicroRNA-124 controls the proliferative, migratory, and inflammatory phenotype of pulmonary vascular fibroblasts. Circulation research. 2014;114(1):67-78.
278. Pullamsetti SS, Doebele C, Fischer A, Savai R, Kojonazarov B, Dahal BK, et al. Inhibition of microRNA-17 improves lung and heart function in experimental pulmonary hypertension. American journal of respiratory and critical care medicine. 2012;185(4):409-19.
279. Caruso P, Dempsie Y, Stevens HC, McDonald RA, Long L, Lu R, et al. A role for miR-145 in pulmonary arterial hypertension: evidence from mouse models and patient samples. Circulation research. 2012;111(3):290-300.
280. Courboulin A, Paulin R, Giguere NJ, Saksouk N, Perreault T, Meloche J, et al. Role for miR-204 in human pulmonary arterial hypertension. The Journal of experimental medicine. 2011;208(3):535-48.
281. Bonnet S, Michelakis ED, Porter CJ, Andrade-Navarro MA, Thebaud B, Bonnet S, et al. An abnormal mitochondrial-hypoxia inducible factor-1alpha-Kv channel pathway disrupts oxygen sensing and triggers pulmonary arterial hypertension in fawn hooded rats: similarities to human pulmonary arterial hypertension. Circulation. 2006;113(22):2630-41.
282. McKinsey TA. Therapeutic potential for HDAC inhibitors in the heart. Annual review of pharmacology and toxicology. 2012;52:303-19.
283. Stern EP, Denton CP. The Pathogenesis of Systemic Sclerosis. Rheum Dis Clin North Am. 2015;41(3):367-82.
284. Arnett FC, Cho M, Chatterjee S, Aguilar MB, Reveille JD, Mayes MD. Familial occurrence frequencies and relative risks for systemic sclerosis (scleroderma) in three United States cohorts. Arthritis and rheumatism. 2001;44(6):1359-62.
285. Feghali-Bostwick C, Medsger TA, Jr., Wright TM. Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies. Arthritis and rheumatism. 2003;48(7):1956-63.
286. Allanore Y, Dieude P, Boileau C. Updating the genetics of systemic sclerosis. Current opinion in rheumatology. 2010;22(6):665-70.

287. Salazar G, Mayes MD. Genetics, Epigenetics, and Genomics of Systemic Sclerosis. Rheum Dis Clin North Am. 2015;41(3):345-66.
288. Bossini-Castillo L, Lopez-Isac E, Mayes MD, Martin J. Genetics of systemic sclerosis. Seminars in immunopathology. 2015;37(5):443-51.
289. Zhou X, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, et al. HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. Arthritis and rheumatism. 2009;60(12):3807-14.
290. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nature genetics. 2010;42(5):426-9.
291. Dieude P, Boileau C, Guedj M, Avouac J, Ruiz B, Hachulla E, et al. Independent replication establishes the CD247 gene as a genetic systemic sclerosis susceptibility factor. Annals of the rheumatic diseases. 2011;70(9):1695-6.
292. Martin JE, Assassi S, Diaz-Gallo LM, Broen JC, Simeon CP, Castellvi I, et al. A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. Human molecular genetics. 2013;22(19):4021-9.
293. Lopez-Isac E, Bossini-Castillo L, Simeon CP, Egurbide MV, Alegre-Sancho JJ, Callejas JL, et al. A genome-wide association study follow-up suggests a possible role for PPARG in systemic sclerosis susceptibility. Arthritis research & therapy. 2014;16(1):R6.
294. Muller-Hilke B. HLA class II and autoimmunity: epitope selection vs differential expression. Acta histochemica. 2009;111(4):379-81.
295. Reveille JD, Fischbach M, McNearney T, Friedman AW, Aguilar MB, Lisse J, et al. Systemic sclerosis in 3 US ethnic groups: a comparison of clinical, sociodemographic, serologic, and immunogenetic determinants. Seminars in arthritis and rheumatism. 2001;30(5):332-46.
296. Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. American journal of human genetics. 2014;94(1):47-61.
297. Dieude P, Dawidowicz K, Guedj M, Legrain Y, Wipff J, Hachulla E, et al. Phenotype-haplotype correlation of IRF5 in systemic sclerosis: role of 2 haplotypes in disease severity. The Journal of rheumatology. 2010;37(5):987-92.
298. Wu SP, Leng L, Feng Z, Liu N, Zhao H, McDonald C, et al. Macrophage migration inhibitory factor promoter polymorphisms and the clinical expression of scleroderma. Arthritis and rheumatism. 2006;54(11):3661-9.
299. Bossini-Castillo L, Simeon CP, Beretta L, Vonk MC, Callejas-Rubio JL, Espinosa G, et al. Confirmation of association of the macrophage migration inhibitory factor gene with systemic sclerosis in a large European population. Rheumatology (Oxford, England). 2011;50(11):1976-81.
300. Fagerholm SC, Varis M, Stefanidakis M, Hilden TJ, Gahmberg CG. alpha-Chain phosphorylation of the human leukocyte CD11b/CD18 (Mac-1) integrin is pivotal for integrin activation to bind ICAMs and leukocyte extravasation. Blood. 2006;108(10):3379-86.
301. Carmona FD, Simeon CP, Beretta L, Carreira P, Vonk MC, Rios-Fernandez R, et al. Association of a non-synonymous functional variant of the ITGAM gene with systemic sclerosis. Annals of the rheumatic diseases. 2011;70(11):2050-2.
302. Rueda B, Broen J, Simeon C, Hesselstrand R, Diaz B, Suarez H, et al. The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. Human molecular genetics. 2009;18(11):2071-7.
303. Yokoyama K, Su Ih IH, Tezuka T, Yasuda T, Mikoshiba K, Tarakhovsky A, et al. BANK regulates BCR-induced calcium mobilization by promoting tyrosine phosphorylation of IP(3) receptor. The EMBO journal. 2002;21(1-2):83-92.

304. Dieude P. Rheumatic diseases: environment and genetics. Joint, bone, spine : revue du rhumatisme. 2009;76(6):602-7.
305. Gourh P, Agarwal SK, Martin E, Divecha D, Rueda B, Bunting H, et al. Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. Journal of autoimmunity. 2010;34(2):155-62.
306. Coustet B, Dieude P, Guedj M, Bouaziz M, Avouac J, Ruiz B, et al. C8orf13-BLK is a genetic risk locus for systemic sclerosis and has additive effects with BANK1: results from a large french cohort and meta-analysis. Arthritis and rheumatism. 2011;63(7):2091-6.
307. Lee YH, Choi SJ, Ji JD, Song GG. The association between the PTPN22 C1858T polymorphism and systemic sclerosis: a meta-analysis. Molecular biology reports. 2012;39(3):3103-8.
308. Fonseca C, Lindahl GE, Ponticos M, Sestini P, Renzoni EA, Holmes AM, et al. A polymorphism in the CTGF promoter region associated with systemic sclerosis. The New England journal of medicine. 2007;357(12):1210-20.
309. Zhang X, Nie S, Si X, Luo Y, Tang W. Association between the CTGF -945C/G polymorphism and systemic sclerosis: a meta-analysis. Gene. 2012;509(1):1-6.
310. Kawaguchi Y, Ota Y, Kawamoto M, Ito I, Tsuchiya N, Sugiura T, et al. Association study of a polymorphism of the CTGF gene and susceptibility to systemic sclerosis in the Japanese population. Annals of the rheumatic diseases. 2009;68(12):1921-4.
311. Mak AC, Tang PL, Cleveland C, Smith MH, Kari Connolly M, Katsumoto TR, et al. Brief Report: Whole-Exome Sequencing for Identification of Potential Causal Variants for Diffuse Cutaneous Systemic Sclerosis. Arthritis & rheumatology (Hoboken, NJ). 2016;68(9):2257-62.
312. Wei P, Yang Y, Guo X, Hei N, Lai S, Assassi S, et al. Identification of an Association of TNFAIP3 Polymorphisms With Matrix Metalloproteinase Expression in Fibroblasts in an Integrative Study of Systemic Sclerosis-Associated Genetic and Environmental Factors. Arthritis & rheumatology (Hoboken, NJ). 2016;68(3):749-60.
313. Makino T, Jinnin M. Genetic and epigenetic abnormalities in systemic sclerosis. J Dermatol. 2016;43(1):10-8.
314. Asano Y, Bujor AM, Trojanowska M. The impact of Fli1 deficiency on the pathogenesis of systemic sclerosis. J Dermatol Sci. 2010;59(3):153-62.
315. Lian X, Xiao R, Hu X, Kanekura T, Jiang H, Li Y, et al. DNA demethylation of CD40l in CD4+ T cells from women with systemic sclerosis: a possible explanation for female susceptibility. Arthritis and rheumatism. 2012;64(7):2338-45.
316. Harvey RP. NK-2 homeobox genes and heart development. Developmental biology. 1996;178(2):203-16.
317. Evans SM. Vertebrate tinman homologues and cardiac differentiation. Seminars in cell & developmental biology. 1999;10(1):73-83.
318. Chen CY, Schwartz RJ. Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, nkx-2.5. The Journal of biological chemistry. 1995;270(26):15628-33.
319. Sepulveda JL, Belaguli N, Nigam V, Chen CY, Nemer M, Schwartz RJ. GATA-4 and Nkx-2.5 coactivate Nkx-2 DNA binding targets: role for regulating early cardiac gene expression. Molecular and cellular biology. 1998;18(6):3405-15.
320. Pradhan L, Genis C, Scone P, Weinberg EO, Kasahara H, Nam HJ. Crystal structure of the human NKX2.5 homeodomain in complex with DNA target. Biochemistry. 2012;51(32):6312-9.
321. Bodmer R. The gene tinman is required for specification of the heart and visceral muscles in Drosophila. Development (Cambridge, England). 1993;118(3):719-29.

322. Paffett-Lugassy N1 SR, Nevis KR, Guner-Ataman B, O'Loughlin E, Jahangiri L, Harvey RP, Burns CG, Burns CE. Heart field origin of great vessel precursors relies on nkx2.5-mediated vasculogenesis. Nat Cell Biol. 2013 Nov;15(11):1362-9.
323. Nagelberg D, Wang J, Su R, Torres-Vazquez J, Targoff KL, Poss KD, et al. Origin, Specification, and Plasticity of the Great Vessels of the Heart. Current biology : CB. 2015;25(16):2099-110.
324. Kasahara H, Wakimoto H, Liu M, Maguire CT, Converso KL, Shioi T, et al. Progressive atrioventricular conduction defects and heart failure in mice expressing a mutant Csx/Nkx2.5 homeoprotein. The Journal of clinical investigation. 2001;108(2):189-201.
325. Lee YS, Wysocki A, Warburton D, Tuan TL. Wound Healing in Development. Birth defects research Part C, Embryo today : reviews. 2012;96(3):213-22.
326. Gallo V, Deneen B. Glial Development: The Crossroads of Regeneration and Repair in the CNS. Neuron. 2014;83(2):283-308.
327. Ponticos M, Abraham D, Alexakis C, Lu QL, Black C, Partridge T, et al. Col1a2 enhancer regulates collagen activity during development and in adult tissue repair. Matrix biology : journal of the International Society for Matrix Biology. 2004;22(8):619-28.
328. Ponticos M, Partridge T, Black CM, Abraham DJ, Bou-Gharios G. Regulation of collagen type I in vascular smooth muscle cells by competition between Nkx2.5 and deltaEF1/ZEB1. Molecular and cellular biology. 2004;24(14):6151-61.
329. Ponticos M, Smith BD. Extracellular matrix synthesis in vascular disease: hypertension, and atherosclerosis. Journal of biomedical research. 2014;28(1):25-39.
330. Ponticos M DA, Santos J, Baliga R, Trinder S, Papaioannou I, Shiwen X, Hobbs A, Jackson C, Denton C, Abraham D. NKX2-5 is a critical regulator of vessel remodelling in vascular disease. Ready for Submission to Nature Communications. 2017.
331. Oishi T, Iida A, Otsubo S, Kamatani Y, Usami M, Takei T, et al. A functional SNP in the NKX2.5-binding site of ITPR3 promoter is associated with susceptibility to systemic lupus erythematosus in Japanese population. Journal of human genetics. 2008;53(2):151-62.
332. Hoffman JI, Kaplan S. The incidence of congenital heart disease. Journal of the American College of Cardiology. 2002;39(12):1890-900.
333. Wang J, Liu XY, Yang YQ. Novel NKX2-5 mutations responsible for congenital heart disease. Genetics and molecular research : GMR. 2011;10(4):2905-15.
334. Su W, Zhu P, Wang R, Wu Q, Wang M, Zhang X, et al. Congenital heart diseases and their association with the variant distribution features on susceptibility genes. Clinical genetics. 2016.
335. Tong YF. Mutations of NKX2.5 and GATA4 genes in the development of congenital heart disease. Gene. 2016;588(1):86-94.
336. Xie X, Shi X, Xun X, Rao L. Associations of NKX2-5 Genetic Polymorphisms with the Risk of Congenital Heart Disease: A Meta-analysis. Pediatric cardiology. 2016;37(5):953-61.
337. Harris BS, Spruill L, Edmonson AM, Rackley MS, Benson DW, O'Brien TX, et al. Differentiation of cardiac Purkinje fibers requires precise spatiotemporal regulation of Nkx2-5 expression. Developmental dynamics : an official publication of the American Association of Anatomists. 2006;235(1):38-49.
338. Thomas PS, Kasahara H, Edmonson AM, Izumo S, Yacoub MH, Barton PJ, et al. Elevated expression of Nkx-2.5 in developing myocardial conduction cells. The Anatomical record. 2001;263(3):307-13.
339. Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, et al. Congenital heart disease caused by mutations in the transcription factor NKX2-5. Science (New York, NY). 1998;281(5373):108-11.

340. Xu H, Morishima M, Wylie JN, Schwartz RJ, Bruneau BG, Lindsay EA, et al. Tbx1 has a dual role in the morphogenesis of the cardiac outflow tract. Development (Cambridge, England). 2004;131(13):3217-27.
341. Mommersteeg MT, Brown NA, Prall OW, de Gier-de Vries C, Harvey RP, Moorman AF, et al. Pitx2c and Nkx2-5 are required for the formation and identity of the pulmonary myocardium. Circulation research. 2007;101(9):902-9.
342. Saadane N, Alpert L, Chalifour LE. Expression of immediate early genes, GATA-4, and Nkx-2.5 in adrenergic-induced cardiac hypertrophy and during regression in adult mice. British journal of pharmacology. 1999;127(5):1165-76.
343. Takimoto E, Mizuno T, Terasaki F, Shimoyama M, Honda H, Shiojima I, et al. Up-regulation of natriuretic peptides in the ventricle of Csx/Nkx2-5 transgenic mice. Biochemical and biophysical research communications. 2000;270(3):1074-9.
344. Toko H, Zhu W, Takimoto E, Shiojima I, Hiroi Y, Zou Y, et al. Csx/Nkx2-5 is required for homeostasis and survival of cardiac myocytes in the adult heart. The Journal of biological chemistry. 2002;277(27):24735-43.
345. Schwartz RJO, E. N. Building the heart piece by piece: modularity of cis-elements regulating Nkx2-5 transcription. Development (Cambridge, England). 1999;126(19):4187-92.
346. Chi X, Chatterjee PK, Wilson W, 3rd, Zhang SX, Demayo FJ, Schwartz RJ. Complex cardiac Nkx2-5 gene expression activated by noggin-sensitive enhancers followed by chamber-specific modules. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(38):13490-5.
347. Lien CL, Wu C, Mercer B, Webb R, Richardson JA, Olson EN. Control of early cardiac-specific transcription of Nkx2-5 by a GATA-dependent enhancer. Development (Cambridge, England). 1999;126(1):75-84.
348. Searcy RD, Vincent EB, Liberatore CM, Yutzey KE. A GATA-dependent nkx-2.5 regulatory element activates early cardiac gene expression in transgenic mice. Development (Cambridge, England). 1998;125(22):4461-70.
349. Lien CL, McAnally J, Richardson JA, Olson EN. Cardiac-specific activity of an Nkx2-5 enhancer requires an evolutionarily conserved Smad binding site. Developmental biology. 2002;244(2):257-66.
350. Brown CO, 3rd, Chi X, Garcia-Gras E, Shirai M, Feng XH, Schwartz RJ. The cardiac determination factor, Nkx2-5, is activated by mutual cofactors GATA-4 and Smad1/4 via a novel upstream enhancer. The Journal of biological chemistry. 2004;279(11):10659-69.
351. Skerjanc IS, Petropoulos H, Ridgeway AG, Wilton S. Myocyte enhancer factor 2C and Nkx2-5 up-regulate each other's expression and initiate cardiomyogenesis in P19 cells. The Journal of biological chemistry. 1998;273(52):34904-10.
352. Nagao K, Taniyama Y, Kietzmann T, Doi T, Komuro I, Morishita R. HIF-1alpha signaling upstream of NKX2.5 is required for cardiac development in Xenopus. The Journal of biological chemistry. 2008;283(17):11841-9.
353. Schultheiss TM, Burch JB, Lassar AB. A role for bone morphogenetic proteins in the induction of cardiac myogenesis. Genes & development. 1997;11(4):451-62.
354. Liberatore CM, Searcy-Schrick RD, Vincent EB, Yutzey KE. Nkx-2.5 gene induction in mice is mediated by a Smad consensus regulatory region. Developmental biology. 2002;244(2):243-56.
355. Lim JY, Kim WH, Kim J, Park SI. Involvement of TGF-beta1 signaling in cardiomyocyte differentiation from P19CL6 cells. Molecules and cells. 2007;24(3):431-6.
356. Cohen ED, Tian Y, Morrisey EE. Wnt signaling: an essential regulator of cardiovascular differentiation, morphogenesis and progenitor self-renewal. Development (Cambridge, England). 2008;135(5):789-98.

357. Clevers H, Nusse R. Wnt/beta-catenin signaling and disease. Cell. 2012;149(6):1192-205.
358. Marvin MJ, Di Rocco G, Gardiner A, Bush SM, Lassar AB. Inhibition of Wnt activity induces heart formation from posterior mesoderm. Genes & development. 2001;15(3):316-27.
359. Pal R, Khanna A. Role of smad- and wnt-dependent pathways in embryonic cardiac development. Stem cells and development. 2006;15(1):29-39.
360. Cambier L, Plate M, Sucov HM, Pashmforoush M. Nkx2-5 regulates cardiac growth through modulation of Wnt signaling by R-spondin3. Development (Cambridge, England). 2014;141(15):2959-71.
361. Patel R, Kos L. Endothelin-1 and Neuregulin-1 convert embryonic cardiomyocytes into cells of the conduction system in the mouse. Developmental dynamics : an official publication of the American Association of Anatomists. 2005;233(1):20-8.
362. Wang Z, Xu G, Wu Y, Guan Y, Cui L, Lei X, et al. Neuregulin-1 enhances differentiation of cardiomyocytes from embryonic stem cells. Medical & biological engineering & computing. 2009;47(1):41-8.
363. Funke-Kaiser H, Lemmer J, Langsdorff CV, Thomas A, Kovacevic SD, Strasdat M, et al. Endothelin-converting enzyme-1 (ECE-1) is a downstream target of the homeobox transcription factor Nkx2-5. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2003;17(11):1487-9.
364. Kasahara H, Izumo S. Identification of the in vivo casein kinase II phosphorylation site within the homeodomain of the cardiac tisue-specifying homeobox gene product Csx/Nkx2.5. Molecular and cellular biology. 1999;19(1):526-36.
365. Ponticos, M; Baliga, R; Holmes, AM; Dooley, A; Hobbs, A; Denton, CP; Abraham, DA; Tissue specific conditional deletion of Nkx2-5 inhibits vascular remodelling in the chronic hypoxia-induced mouse model of pulmonary arterial hypertetion. In: (Proceedings) Keystone symposia: Pulmonary Vascular Disease and Right Ventricular Dysfunction: Current Concepts and Future Therapies. 2012
366. Wang J, Zhang H, Iyer D, Feng XH, Schwartz RJ. Regulation of cardiac specific nkx2.5 gene activity by small ubiquitin-like modifier. The Journal of biological chemistry. 2008;283(34):23235-43.
367. Kim HS, Woo JS, Joo HJ, Moon WK. Cardiac transcription factor Nkx2.5 is downregulated under excessive O-GlcNAcylation condition. PloS one. 2012;7(6):e38053.
368. Costa MW, Lee S, Furtado MB, Xin L, Sparrow DB, Martinez CG, et al. Complex SUMO-1 regulation of cardiac transcription factor Nkx2-5. PloS one. 2011;6(9):e24812.
369. Tang X, Ma H, Han L, Zheng W, Lu YB, Chen XF, et al. SIRT1 deacetylates the cardiac transcription factor Nkx2.5 and inhibits its transcriptional activity. Scientific reports. 2016;6.
370. Ponticos M, Jackson C, Bou-Gharios G, Abraham D. The role of the homeobox transcription factor NKX2.5 in the regulation of collagen in vessels and in the atherosclerotic lesion. Vascular Pharmacology. 2006;45(3):e45-e6.
371. Ponticos M. The role of the Homedomian family member Nkx2-5 in the cardiovascular system. C. Handler, D. Abraham, M. Dashwood, & G. Coghlan (Eds.), Springer UK. 2010:Chapter 7;120-4.
372. Riazi AM, Takeuchi JK, Hornberger LK, Zaidi SH, Amini F, Coles J, et al. NKX2-5 regulates the expression of beta-catenin and GATA4 in ventricular myocytes. PloS one. 2009;4(5):e5698.
373. Kwon C, Arnold J, Hsiao EC, Taketo MM, Conklin BR, Srivastava D. Canonical Wnt signaling is a positive regulator of mammalian cardiac progenitors. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(26):10894-9.

374. Ueyama T, Kasahara H, Ishiwata T, Nie Q, Izumo S. Myocardin Expression Is Regulated by Nkx2.5, and Its Function Is Required for Cardiomyogenesis. Molecular and cellular biology. 2003;23(24):9222-32.
375. Moskowitz IP, Kim JB, Moore ML, Wolf CM, Peterson MA, Shendure J, et al. A molecular pathway including Id2, Tbx5, and Nkx2-5 required for cardiac conduction system development. Cell. 2007;129(7):1365-76.
376. Ganga M, Espinoza HM, Cox CJ, Morton L, Hjalt TA, Lee Y, et al. PITX2 isoform-specific regulation of atrial natriuretic factor expression: synergism and repression with Nkx2.5. The Journal of biological chemistry. 2003;278(25):22437-45.
377. Biben C, Weber R, Kesteven S, Stanley E, McDonald L, Elliott DA, et al. Cardiac septal and valvular dysmorphogenesis in mice heterozygous for mutations in the homeobox gene Nkx2-5. Circulation research. 2000;87(10):888-95.
378. Chen F, Kook H, Milewski R, Gitler AD, Lu MM, Li J, et al. Hop is an unusual homeobox gene that modulates cardiac development. Cell. 2002;110(6):713-23.
379. Dupays L, Jarry-Guichard T, Mazurais D, Calmels T, Izumo S, Gros D, et al. Dysregulation of connexins and inactivation of NFATc1 in the cardiovascular system of Nkx2-5 null mutants. Journal of molecular and cellular cardiology. 2005;38(5):787-98.
380. Kasahara H, Ueyama T, Wakimoto H, Liu MK, Maguire CT, Converso KL, et al. Nkx2.5 homeoprotein regulates expression of gap junction protein connexin 43 and sarcomere organization in postnatal cardiomyocytes. Journal of molecular and cellular cardiology. 2003;35(3):243-56.
381. Chen CY, Schwartz RJ. Competition between negative acting YY1 versus positive acting serum response factor and tinman homologue Nkx-2.5 regulates cardiac alpha-actin promoter activity. Molecular endocrinology (Baltimore, Md). 1997;11(6):812-22.
382. Ikle JM, Tavares AL, King M, Ding H, Colombo S, Firulli BA, et al. Nkx2.5 regulates endothelin converting enzyme-1 during pharyngeal arch patterning. 2017;55(3).
383. Meng ZZ, Liu W, Xia Y, Yin HM, Zhang CY, Su D, et al. The pro-inflammatory signalling regulator Stat4 promotes vasculogenesis of great vessels derived from endothelial precursors. Nat Commun. 2017;8:14640.
384. Zhou JZ, Jiang JX. Gap junction and hemichannel-independent actions of connexins on cell and tissue functions – An update. FEBS letters. 2014;588(8):1186-92.
385. Oyamada M, Takebe K, Oyamada Y. Regulation of connexin expression by transcription factors and epigenetic mechanisms. Biochimica et biophysica acta. 2013;1828(1):118-33.
386. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nature genetics. 2005;37(11):1217-23.
387. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics (Oxford, England). 2005;21(2):263-5.
388. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007;81(3):559-75.
389. Maniatis N. Linkage Disequilibrium Maps and Disease-Association Mapping. In: Collins AR, editor. Linkage Disequilibrium and Association Mapping: Analysis and Applications. Totowa, NJ: Humana Press; 2007. p. 109-21.
390. Liu N, Zhang K, Zhao H. Haplotype-association analysis. Advances in genetics. 2008;60:335-405.
391. Ziegler A KI. A statistical approach to genetic epidemiology: concepts and applications, 2nd edn. Wiley-VCH, Weinheim 2010.

392. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. American journal of human genetics. 2005;76(6):967-86.
393. Ziegler A, Van Steen K, Wellek S. Investigating Hardy–Weinberg equilibrium in case–control or cohort studies or meta-analysis. Breast Cancer Research and Treatment. 2011;128(1):197-201.
394. Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. The American Statistician. 2016;70(2):129-33.
395. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. Nat Rev Genet. 2014;15(5):335-46.
396. Polychronakos C. Clinical and Investigative Medicine Medecine Clinique et Experimentale.
397. Gao X, Becker LC, Becker DM, Starmer JD, Province MA. Avoiding the high Bonferroni penalty in genome-wide association studies. Genetic epidemiology. 2010;34(1):100-5.
398. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics. 1994;138(3):963-71.
399. Stewart SA, Weinberg RA. Senescence: does it all happen at the ends? Oncogene. 2002;21(4):627-30.
400. Jha KK, Banga S, Palejwala V, Ozer HL. SV40-Mediated immortalization. Experimental cell research. 1998;245(1):1-7.
401. Lundberg AS, Hahn WC, Gupta P, Weinberg RA. Genes involved in senescence and immortalization. Current opinion in cell biology. 2000;12(6):705-9.
402. Hemmer JD, Dean D, Vertegel A, Langan E, 3rd, LaBerge M. Effects of serum deprivation on the mechanical properties of adherent vascular smooth muscle cells. Proceedings of the Institution of Mechanical Engineers Part H, Journal of engineering in medicine. 2008;222(5):761-72.
403. Horita H, Wysoczynski CL, Walker LA, Moulton KS, Li M, Ostriker A, et al. Nuclear PTEN functions as an essential regulator of SRF-dependent transcription to control smooth muscle differentiation. Nat Commun. 2016;7:10830.
404. Rosner M, Schipany K, Hengstschlager M. Merging high-quality biochemical fractionation with a refined flow cytometry approach to monitor nucleocytoplasmic protein expression throughout the unperturbed mammalian cell cycle. Nature protocols. 2013;8(3):602-26.
405. Mehrhof FB, Schmidt-Ullrich R, Dietz R, Scheidereit C. Regulation of vascular smooth muscle cell proliferation: role of NF-kappaB revisited. Circulation research. 2005;96(9):958-64.
406. Zimmermann O, Zwaka TP, Marx N, Torzewski M, Bucher A, Guilliard P, et al. Serum starvation and growth factor receptor expression in vascular smooth muscle cells. Journal of vascular research. 2006;43(2):157-65.
407. Alaseem AM, Madiraju P, Aldebeyan SA, Noorwali H, Antoniou J, Mwale F. Naproxen Induces Type X Collagen Expression in Human Bone-Marrow-Derived Mesenchymal Stem Cells Through the Upregulation of 5-Lipoxygenase. Tissue Engineering Part A. 2015;21(1-2):234-45.
408. Menon MB, Kotlyarov A, Gaestel M. SB202190-Induced Cell Type-Specific Vacuole Formation and Defective Autophagy Do Not Depend on p38 MAP Kinase Inhibition. PloS one. 2011;6(8):e23054.
409. Schöckel L, Glasauer A, Basit F, Bitschar K, Truong H, Erdmann G, et al. Targeting mitochondrial complex I using BAY 87-2243 reduces melanoma tumor growth. Cancer & Metabolism. 2015;3(1):11.
410. Biorad. <A guide to Polyacrylamide Gel Electrophoresis and Detection.pdf>.
411. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227(5259):680-5.
412. Jensen EC. The basics of western blotting. Anatomical record (Hoboken, NJ : 2007). 2012;295(3):369-71.

413. MacPhee DJ. Methodological considerations for improving Western blot analysis. Journal of pharmacological and toxicological methods. 2010;61(2):171-7.
414. Alegria-Schaffer A, Lodge A, Vattem K. Performing and optimizing Western blots with an emphasis on chemiluminescent detection. Methods in enzymology. 2009;463:573-99.
415. Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. Guideline to reference gene selection for quantitative real-time PCR. Biochemical and biophysical research communications. 2004;313(4):856-62.
416. Fay DS, Gerow K. A biologist's guide to statistical thinking and analysis. WormBook : the online review of C elegans biology. 2013:1-54.
417. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research. 2012;40(Database issue):D930-4.
418. Draus JM, Jr., Hauck MA, Goetsch M, Austin EH, 3rd, Tomita-Mitchell A, Mitchell ME. Investigation of somatic NKX2-5 mutations in congenital heart disease. Journal of medical genetics. 2009;46(2):115-22.
419. Reamon-Buettner SM, Hecker H, Spanel-Borowski K, Craatz S, Kuenzel E, Borlak J. Novel NKX2-5 mutations in diseased heart tissues of patients with cardiac malformations. The American journal of pathology. 2004;164(6):2117-25.
420. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014;15(4):272-86.
421. Gomes I, Collins A, Lonjou C, Thomas NS, Wilkinson J, Watson M, et al. Hardy-Weinberg quality control. Annals of human genetics. 1999;63(Pt 6):535-8.
422. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nature genetics. 2015;47(10):1121-30.
423. Heusch G, Libby P, Gersh B, Yellon D, Bohm M, Lopaschuk G, et al. Cardiovascular remodelling in coronary artery disease and heart failure. Lancet. 2014;383(9932):1933-43.
424. Harrison RE, Flanagan JA, Sankelo M, Abdalla SA, Rowell J, Machado RD, et al. Molecular and functional analysis identifies ALK-1 as the predominant cause of pulmonary hypertension related to hereditary haemorrhagic telangiectasia. Journal of medical genetics. 2003;40(12):865-71.
425. Austin ED, Hamid R, Hemnes AR, Loyd JE, Blackwell T, Yu C, et al. BMPR2 expression is suppressed by signaling through the estrogen receptor. Biology of sex differences. 2012;3(1):6.
426. Bossini-Castillo L, Lopez-Isac E, Martin J. Immunogenetics of systemic sclerosis: Defining heritability, functional variants and shared-autoimmunity pathways. Journal of autoimmunity. 2015;64:53-65.
427. Marangoni RG, Korman BD, Allanore Y, Dieude P, Armstrong LL, Rzhetskaya M, et al. A candidate gene study reveals association between a variant of the Peroxisome Proliferator-Activated Receptor Gamma (PPAR-gamma) gene and systemic sclerosis. Arthritis research & therapy. 2015;17:128.
428. Mehra S, Walker J, Patterson K, Fritzler MJ. Autoantibodies in systemic sclerosis. Autoimmunity Reviews. 2013;12(3):340-54.
429. Komuro I, Kurihara H, Sugiyama T, Yoshizumi M, Takaku F, Yazaki Y. Endothelin stimulates c-fos and c-myc expression and proliferation of vascular smooth muscle cells. FEBS letters. 1988;238(2):249-52.
430. Taylor DS, Cheng X, Pawlowski JE, Wallace AR, Ferrer P, Molloy CJ. Epiregulin is a potent vascular smooth muscle cell-derived mitogen induced by angiotensin II, endothelin-1, and thrombin. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(4):1633-8.
431. Horstmeyer A, Licht C, Scherr G, Eckes B, Krieg T. Signalling and regulation of collagen I synthesis by ET-1 and TGF-beta1. The FEBS journal. 2005;272(24):6297-309.

432. Lowe JS, Anderson PG. Stevens & Lowe's Human Histology: With STUDENT CONSULT Online Access: Elsevier Health Sciences UK; 2014.
433. Nicklas D, Saiz L. Characterization of Negative Feedback Network Motifs in the TGF-β Signaling Pathway. PloS one. 2013;8(12):e83531.
434. Garat CV, Crossno JT, Jr., Sullivan TM, Reusch JE, Klemm DJ. Inhibition of phosphatidylinositol 3-kinase/Akt signaling attenuates hypoxia-induced pulmonary artery remodeling and suppresses CREB depletion in arterial smooth muscle cells. Journal of cardiovascular pharmacology. 2013;62(6):539-48.
435. Zhang H, Gong Y, Wang Z, Jiang L, Chen R, Fan X, et al. Apelin inhibits the proliferation and migration of rat PASMCs via the activation of PI3K/Akt/mTOR signal and the inhibition of autophagy under hypoxia. Journal of cellular and molecular medicine. 2014;18(3):542-53.
436. Beitner-Johnson D, Rust RT, Hsieh TC, Millhorn DE. Hypoxia activates Akt and induces phosphorylation of GSK-3 in PC12 cells. Cellular signalling. 2001;13(1):23-7.
437. Kapoun AM, Gaspar NJ, Wang Y, Damm D, Liu YW, O'Young G, et al. Transforming growth factor-beta receptor type 1 (TGFbetaRI) kinase activity but not p38 activation is required for TGFbetaRI-induced myofibroblast differentiation and profibrotic gene expression. Molecular pharmacology. 2006;70(2):518-31.
438. Wilkes MC, Mitchell H, Penheiter SG, Dore JJ, Suzuki K, Edens M, et al. Transforming growth factor-beta activation of phosphatidylinositol 3-kinase is independent of Smad2 and Smad3 and regulates fibroblast responses via p21-activated kinase-2. Cancer research. 2005;65(22):10431-40.
439. Jeong HS, Jung ES, Sim YJ, Kim SJ, Jang JW, Hong KS, et al. Fbxo25 controls Tbx5 and Nkx2-5 transcriptional activity to regulate cardiomyocyte development. Biochimica et biophysica acta. 2015;1849(6):709-21.
440. Clark CD, Zhang B, Lee B, Evans SI, Lassar AB, Lee KH. Evolutionary conservation of Nkx2.5 autoregulation in the second heart field. Developmental biology. 2013;374(1):198-209.
441. Paneni F, Beckman JA, Creager MA, Cosentino F. Diabetes and vascular disease: pathophysiology, clinical consequences, and medical therapy: part I. European heart journal. 2013;34(31):2436-43.
442. Michiels C. Physiological and Pathological Responses to Hypoxia. The American journal of pathology. 2004;164(6):1875-82.
443. Welsh DJ, Peacock AJ. Cellular responses to hypoxia in the pulmonary circulation. High altitude medicine & biology. 2013;14(2):111-6.
444. Balamurugan K. HIF-1 at the crossroads of hypoxia, inflammation, and cancer. International journal of cancer. 2016;138(5):1058-66.
445. Hong WX, Hu MS, Esquivel M, Liang GY, Rennert RC, McArdle A, et al. The Role of Hypoxia-Inducible Factor in Wound Healing. Advances in Wound Care. 2014;3(5):390-9.
446. Veith C, Schermuly RT, Brandes RP, Weissmann N. Molecular mechanisms of hypoxia-inducible factor-induced pulmonary arterial smooth muscle cell alterations in pulmonary hypertension. The Journal of Physiology. 2016;594(5):1167-77.
447. Olszewska-Pazdrak B, Hein TW, Olszewska P, Carney DH. Chronic hypoxia attenuates VEGF signaling and angiogenic responses by downregulation of KDR in human endothelial cells. American Journal of Physiology - Cell Physiology. 2009;296(5):C1162-C70.
448. Deguchi J-o, Yamazaki H, Aikawa E, Aikawa M. Chronic Hypoxia Activates the Akt and β-Catenin Pathways in Human Macrophages. Arteriosclerosis, thrombosis, and vascular biology. 2009;29(10):1664-70.
449. Li GW, Xing WJ, Bai SZ, Hao JH, Guo J, Li HZ, et al. The calcium-sensing receptor mediates hypoxia-induced proliferation of rat pulmonary artery smooth

muscle cells through MEK1/ERK1,2 and PI3K pathways. Basic & clinical pharmacology & toxicology. 2011;108(3):185-93.
450. Zhang H, Akman HO, Smith ELP, Zhao J, Murphy-Ullrich JE, Batuman OA. Cellular response to hypoxia involves signaling via Smad proteins. Blood. 2003;101(6):2253-60.
451. Lafyatis R. Transforming growth factor beta--at the centre of systemic sclerosis. Nature reviews Rheumatology. 2014;10(12):706-19.
452. Gomez D, Owens GK. Smooth muscle cell phenotypic switching in atherosclerosis. Cardiovascular research. 2012;95(2):156-64.
453. Lee RM, Owens GK, Scott-Burden T, Head RJ, Mulvany MJ, Schiffrin EL. Pathophysiology of smooth muscle in hypertension. Canadian journal of physiology and pharmacology. 1995;73(5):574-84.
454. Sinha S, Hoofnagle MH, Kingston PA, McCanna ME, Owens GK. Transforming growth factor-beta1 signaling contributes to development of smooth muscle cells from embryonic stem cells. American journal of physiology Cell physiology. 2004;287(6):C1560-8.
455. Tabas I, Garcia-Cardena G, Owens GK. Recent insights into the cellular biology of atherosclerosis. The Journal of cell biology. 2015;209(1):13-22.
456. Zhu SB, Zhu J, Zhou ZZ, Xi EP, Wang RP, Zhang Y. TGF-beta1 induces human aortic vascular smooth muscle cell phenotype switch through PI3K/AKT/ID2 signaling. American journal of translational research. 2015;7(12):2764-74.
457. Li HX, Han M, Bernier M, Zheng B, Sun SG, Su M, et al. Kruppel-like factor 4 promotes differentiation by transforming growth factor-beta receptor-mediated Smad and p38 MAPK signaling in vascular smooth muscle cells. The Journal of biological chemistry. 2010;285(23):17846-56.
458. Goumans M-J, Liu Z, ten Dijke P. TGF-[beta] signaling in vascular biology and dysfunction. Cell Res. 2009;19(1):116-27.
459. Deaton RA, Su C, Valencia TG, Grant SR. Transforming growth factor-beta1-induced expression of smooth muscle marker genes involves activation of PKN and p38 MAPK. The Journal of biological chemistry. 2005;280(35):31172-81.
460. Ha DM, Carpenter LC, Koutakis P, Swanson SA, Zhu Z, Hanna M, et al. Transforming growth factor-beta 1 produced by vascular smooth muscle cells predicts fibrosis in the gastrocnemius of patients with peripheral artery disease. Journal of translational medicine. 2016;14:39.
461. Honda T, Obara Y, Yamauchi A, Couvillon AD, Mason JJ, Ishii K, et al. Phosphorylation of ERK5 on Thr732 Is Associated with ERK5 Nuclear Localization and ERK5-Dependent Transcription. PloS one. 2015;10(2):e0117914.
462. Suwanabol PA, Seedial SM, Zhang F, Shi X, Si Y, Liu B, et al. TGF-β and Smad3 modulate PI3K/Akt signaling pathway in vascular smooth muscle cells. American journal of physiology Heart and circulatory physiology. 2012;302(11):H2211-9.
463. Liu C, Su T, Li F, Li L, Qin X, Pan W, et al. PI3K/Akt signaling transduction pathway is involved in rat vascular smooth muscle cell proliferation induced by apelin-13. Acta biochimica et biophysica Sinica. 2010;42(6):396-402.
464. Karar J, Maity A. PI3K/AKT/mTOR Pathway in Angiogenesis. Frontiers in Molecular Neuroscience. 2011;4.
465. Tsuji-Tamura K, Ogawa M. Inhibition of the PI3K–Akt and mTORC1 signaling pathways promotes the elongation of vascular endothelial cells. Journal of Cell Science. 2016;129(6):1165-78.
466. Polivka J, Jr., Janku F. Molecular targets for cancer therapy in the PI3K/AKT/mTOR pathway. Pharmacology & therapeutics. 2014;142(2):164-75.
467. Li F, Li L, Qin X, Pan W, Feng F, Chen F, et al. Apelin-induced vascular smooth muscle cell proliferation: the regulation of cyclin D1. Frontiers in bioscience : a journal and virtual library. 2008;13:3786-92.

468. Lv D, Li H, Chen L. Apelin and APJ, a novel critical factor and therapeutic target for atherosclerosis. Acta biochimica et biophysica Sinica. 2013;45(7):527-33.
469. Wang G, Ma N, Meng L, Wei Y, Gui J. Activation of the phosphatidylinositol 3-kinase/Akt pathway is involved in lipocalin-2-promoted human pulmonary artery smooth muscle cell proliferation. Molecular and cellular biochemistry. 2015;410(1-2):207-13.
470. Sinagra T, Tamburella A, Urso V, Siarkos I, Drago F, Bucolo C, et al. Reversible inhibition of vasoconstriction by thiazolidinediones related to PI3K/Akt inhibition in vascular smooth muscle cells. Biochemical pharmacology. 2013;85(4):551-9.
471. Yoshizumi M, Kyotani Y, Zhao J, Nagayama K, Ito S, Tsuji Y, et al. Role of big mitogen-activated protein kinase 1 (BMK1) / extracellular signal-regulated kinase 5 (ERK5) in the pathogenesis and progression of atherosclerosis. Journal of pharmacological sciences. 2012;120(4):259-63.
472. Zhao Z, Geng J, Ge Z, Wang W, Zhang Y, Kang W. Activation of ERK5 in angiotensin II-induced hypertrophy of human aortic smooth muscle cells. Molecular and cellular biochemistry. 2009;322(1):171-8.
473. Sharma G, Goalstone ML. Regulation of ERK5 by insulin and angiotensin-II in vascular smooth muscle cells. Biochemical and biophysical research communications. 2007;354(4):1078-83.
474. Reecy JM, Li X, Yamada M, DeMayo FJ, Newman CS, Harvey RP, et al. Identification of upstream regulatory regions in the heart-expressed homeobox gene Nkx2-5. Development (Cambridge, England). 1999;126(4):839-49.
475. Gan Q, Yoshida T, Li J, Owens GK. Smooth muscle cells and myofibroblasts use distinct transcriptional mechanisms for smooth muscle alpha-actin expression. Circulation research. 2007;101(9):883-92.
476. Chen Z, Friedrich GA, Soriano P. Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. Genes & development. 1994;8(19):2293-301.
477. Gupta MP, Amin CS, Gupta M, Hay N, Zak R. Transcription enhancer factor 1 interacts with a basic helix-loop-helix zipper protein, Max, for positive regulation of cardiac alpha-myosin heavy-chain gene expression. Molecular and cellular biology. 1997;17(7):3924-36.
478. Butler AJ, Ordahl CP. Poly(ADP-ribose) polymerase binds with transcription enhancer factor 1 to MCAT1 elements to regulate muscle-specific transcription. Molecular and cellular biology. 1999;19(1):296-306.
479. Kaneko KJ, DePamphilis ML. Regulation of gene expression at the beginning of mammalian development and the TEAD family of transcription factors. Developmental genetics. 1998;22(1):43-55.
480. Wang Q, Latham KE. Translation of maternal messenger ribonucleic acids encoding transcription factors during genome activation in early mouse embryos. Biology of reproduction. 2000;62(4):969-78.
481. Jacquemin P, Hwang JJ, Martial JA, Dolle P, Davidson I. A novel family of developmentally regulated mammalian transcription factors containing the TEA/ATTS DNA binding domain. The Journal of biological chemistry. 1996;271(36):21775-85.
482. Jiang SW, Wu K, Eberhardt NL. Human placental TEF-5 transactivates the human chorionic somatomammotropin gene enhancer. Molecular endocrinology (Baltimore, Md). 1999;13(6):879-89.
483. Azakie A, Larkin SB, Farrance IK, Grenningloh G, Ordahl CP. DTEF-1, a novel member of the transcription enhancer factor-1 (TEF-1) multigene family. The Journal of biological chemistry. 1996;271(14):8260-5.
484. Maeda T, Mazzulli JR, Farrance IK, Stewart AF. Mouse DTEF-1 (ETFR-1, TEF-5) is a transcriptional activator in alpha 1-adrenergic agonist-stimulated cardiac myocytes. The Journal of biological chemistry. 2002;277(27):24346-52.

485. Creemers EE, Sutherland LB, McAnally J, Richardson JA, Olson EN. Myocardin is a direct transcriptional target of Mef2, Tead and Foxo proteins during cardiovascular development. Development (Cambridge, England). 2006;133(21):4245-56.
486. Xiao JH, Davidson I, Matthes H, Garnier JM, Chambon P. Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. Cell. 1991;65(4):551-68.
487. Jiang SW, Eberhardt NL. TEF-1 transrepression in BeWo cells is mediated through interactions with the TATA-binding protein, TBP. The Journal of biological chemistry. 1996;271(16):9510-8.
488. Yagi R, Chen LF, Shigesada K, Murakami Y, Ito Y. A WW domain-containing yes-associated protein (YAP) is a novel transcriptional co-activator. The EMBO journal. 1999;18(9):2551-62.
489. Vassilev A, Kaneko KJ, Shu H, Zhao Y, DePamphilis ML. TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. Genes & development. 2001;15(10):1229-41.
490. Hao Y, Chun A, Cheung K, Rashidi B, Yang X. Tumor suppressor LATS1 is a negative regulator of oncogene YAP. The Journal of biological chemistry. 2008;283(9):5496-509.
491. Zhao B, Wei X, Li W, Udan RS, Yang Q, Kim J, et al. Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. Genes & development. 2007;21(21):2747-61.
492. Zhao B, Ye X, Yu J, Li L, Li W, Li S, et al. TEAD mediates YAP-dependent gene induction and growth control. Genes & development. 2008;22(14):1962-71.
493. Almontashiri NA, Antoine D, Zhou X, Vilmundarson RO, Zhang SX, Hao KN, et al. 9p21.3 Coronary Artery Disease Risk Variants Disrupt TEAD Transcription Factor-Dependent Transforming Growth Factor beta Regulation of p16 Expression in Human Aortic Smooth Muscle Cells. Circulation. 2015;132(21):1969-78.
494. Liu-Chittenden Y, Huang B, Shim JS, Chen Q, Lee SJ, Anders RA, et al. Genetic and pharmacological disruption of the TEAD-YAP complex suppresses the oncogenic activity of YAP. Genes & development. 2012;26(12):1300-5.
495. May D, Blow MJ, Kaplan T, McCulley DJ, Jensen BC, Akiyama JA, et al. Large-scale discovery of enhancers from human heart tissue. Nature genetics. 2012;44(1):89-93.
496. Barrett LW, Fletcher S, Wilton SD. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cellular and molecular life sciences : CMLS. 2012;69(21):3613-34.
497. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. PLoS biology. 2004;2(11):e363.
498. Feinberg AP, Tycko B. The history of cancer epigenetics. Nature reviews Cancer. 2004;4(2):143-53.
499. Kwabi-Addo B, Wang S, Chung W, Jelinek J, Patierno SR, Wang BD, et al. Identification of differentially methylated genes in normal prostate tissues from African American and Caucasian men. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010;16(14):3539-47.
500. Kwabi-Addo B, Chung W, Shen L, Ittmann M, Wheeler T, Jelinek J, et al. Age-related DNA methylation changes in normal human prostate tissues. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13(13):3796-802.
501. Sheng W, Qian Y, Wang H, Ma X, Zhang P, Diao L, et al. DNA methylation status of NKX2-5, GATA4 and HAND1 in patients with tetralogy of fallot. BMC medical genomics. 2013;6:46.
502. Fujii M, Toyoda T, Nakanishi H, Yatabe Y, Sato A, Matsudaira Y, et al. TGF-β synergizes with defects in the Hippo pathway to stimulate human malignant mesothelioma growth. The Journal of experimental medicine. 2012;209(3):479-94.

503. Harvey KF, Zhang X, Thomas DM. The Hippo pathway and human cancer. Nature reviews Cancer. 2013;13(4):246-57.
504. Piersma B, Bank RA, Boersema M. Signaling in Fibrosis: TGF-beta, WNT, and YAP/TAZ Converge. Frontiers in medicine. 2015;2:59.
505. Saito A, Nagase T. Hippo and TGF-beta interplay in the lung field. Am J Physiol Lung Cell Mol Physiol. 2015;309(8):L756-67.
506. Zhou Q, Li L, Zhao B, Guan KL. The hippo pathway in heart development, regeneration, and diseases. Circulation research. 2015;116(8):1431-47.
507. Li H, Jiang W, Ren W, Guo D, Guo J, Wang X, et al. Downregulation of the Yes-Associated Protein Is Associated with Extracellular Matrix Disorders in Ascending Aortic Aneurysms. Stem Cells Int. 2016;2016:6786184.
508. Speight P, Kofler M, Szaszi K, Kapus A. Context-dependent switch in chemo/mechanotransduction via multilevel crosstalk among cytoskeleton-regulated MRTF and TAZ and TGFbeta-regulated Smad3. Nat Commun. 2016;7:11642.
509. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome biology. 2010;11(8):R90.
510. Wang X. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. Bioinformatics (Oxford, England). 2016;32(9):1316-22.
511. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, et al. Distribution of miRNA expression across human tissues. Nucleic acids research. 2016.
512. Pasini A, Bonafe F, Govoni M, Guarnieri C, Morselli PG, Sharma HS, et al. Epigenetic signature of early cardiac regulatory genes in native human adipose-derived stem cells. Cell biochemistry and biophysics. 2013;67(2):255-62.
513. Mahler GJ, Farrar EJ, Butcher JT. Inflammatory cytokines promote mesenchymal transformation in embryonic and adult valve endothelial cells. Arteriosclerosis, thrombosis, and vascular biology. 2013;33(1):121-30.
514. Katsura A, Suzuki HI, Ueno T, Mihira H, Yamazaki T, Yasuda T, et al. MicroRNA-31 is a positive modulator of endothelial-mesenchymal transition and associated secretory phenotype induced by TGF-beta. Genes Cells. 2016;21(1):99-116.
515. Xu X, Friehs I, Zhong Hu T, Melnychenko I, Tampe B, Alnour F, et al. Endocardial fibroelastosis is caused by aberrant endothelial to mesenchymal transition. Circulation research. 2015;116(5):857-66.
516. Chen PY, Qin L, Tellides G, Simons M. Fibroblast growth factor receptor 1 is a key inhibitor of TGFbeta signaling in the endothelium. Sci Signal. 2014;7(344):ra90.
517. Lee JG, Ko MK, Kay EP. Endothelial mesenchymal transformation mediated by IL-1beta-induced FGF-2 in corneal endothelial cells. Exp Eye Res. 2012;95(1):35-9.
518. Correia AC, Moonen JR, Brinker MG, Krenning G. FGF2 inhibits endothelial-mesenchymal transition through microRNA-20a-mediated repression of canonical TGF-beta signaling. J Cell Sci. 2016;129(3):569-79.
519. Nagai T, Kanasaki M, Srivastava SP, Nakamura Y, Ishigaki Y, Kitada M, et al. N-acetyl-seryl-aspartyl-lysyl-proline inhibits diabetes-associated kidney fibrosis and endothelial-mesenchymal transition. Biomed Res Int. 2014;2014:696475.
520. Papetti M, Shujath J, Riley KN, Herman IM. FGF-2 antagonizes the TGF-beta1-mediated induction of pericyte alpha-smooth muscle actin expression: a role for myf-5 and Smad-mediated signaling pathways. Investigative ophthalmology & visual science. 2003;44(11):4994-5005.
521. Romero LI, Zhang DN, Herron GS, Karasek MA. Interleukin-1 induces major phenotypic changes in human skin microvascular endothelial cells. Journal of cellular physiology. 1997;173(1):84-92.

522. Zeisberg M, Hanai J, Sugimoto H, Mammoto T, Charytan D, Strutz F, et al. BMP-7 counteracts TGF-beta1-induced epithelial-to-mesenchymal transition and reverses chronic renal injury. Nature medicine. 2003;9(7):964-8.
523. Versari D, Daghini E, Virdis A, Ghiadoni L, Taddei S. Endothelial Dysfunction as a Target for Prevention of Cardiovascular Disease. Diabetes Care. 2009;32(Suppl 2):S314-21.
524. Barkefors I, Le Jan S, Jakobsson L, Hejll E, Carlson G, Johansson H, et al. Endothelial Cell Migration in Stable Gradients of Vascular Endothelial Growth Factor A and Fibroblast Growth Factor 2: Effects on chemotaxis and chemokinesis. Journal of Biological Chemistry. 2008;283(20):13905-12.
525. Sinha S, Iyer D, Granata A. Embryonic origins of human vascular smooth muscle cells: implications for in vitro modeling and clinical application. Cellular and molecular life sciences : CMLS. 2014;71(12):2271-88.
526. Park C, Kim TM, Malik AB. Transcriptional Regulation oa Endothelial Cell And Vascular Development. Circulation research. 2013;112(10):1380-400.
527. Armstrong EJ, Bischoff J. Heart valve development: endothelial cell signaling and differentiation. Circulation research. 2004;95(5):459-70.
528. Liebner S, Cattelino A, Gallini R, Rudini N, Iurlaro M, Piccolo S, et al. Beta-catenin is required for endothelial-mesenchymal transformation during heart cushion development in the mouse. The Journal of cell biology. 2004;166(3):359-67.
529. Arciniegas E, Neves CY, Carrillo LM, Zambrano EA, Ramirez R. Endothelial-mesenchymal transition occurs during embryonic pulmonary artery development. Endothelium. 2005;12(4):193-200.
530. Hall SM, Hislop AA, Haworth SG. Origin, differentiation, and maturation of human pulmonary veins. American journal of respiratory cell and molecular biology. 2002;26(3):333-40.
531. Evrard SM, Lecce L, Michelis KC, Nomura-Kitabayashi A, Pandey G, Purushothaman KR, et al. Endothelial to mesenchymal transition is common in atherosclerotic lesions and is associated with plaque instability. Nat Commun. 2016;7:11853.
532. Moonen JR, Lee ES, Schmidt M, Maleszewska M, Koerts JA, Brouwer LA, et al. Endothelial-to-mesenchymal transition contributes to fibro-proliferative vascular disease and is modulated by fluid shear stress. Cardiovascular research. 2015;108(3):377-86.