

National Heart and Lung Institute
Imperial College London
London
W12 0NN

Investigating the Biology of *SOX17* in Pulmonary Arterial Hypertension

Thesis submitted for the degree of

Doctor of Philosophy

By

Rachel Walters

2022

Statement of Originality

I declare that this thesis was conducted and written by myself, and all work included is my own unless declared otherwise.

Rachel Walters

Copyright Declaration

‘The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non-Commercial 4.0 International License (CC BY-NC).

Under this license, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the license terms clear to others by naming the license and linking to the license text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this license or permitted under UK Copyright Law.’

Acknowledgements

The research documented in this thesis and my experience throughout my PhD would not have been possible if not for the continued help and dedication of my supervisor Chris Rhodes. Thank you for the support and guidance throughout my time at Imperial, especially through the working-at-home days of the pandemic! Thank you for always getting my writing back to me as soon as possible and for creating a fun and exciting research environment.

Thank you to my supervisor Martin Wilkins for always asking what the bigger picture is and for reading through and giving me feedback on this thesis. I am also very grateful to Anna Randi and Jorge Ferrer for their helpful comments and suggestions during my early and late stage assessments. Also, thank you to Wayne Mitchell for taking me under his wing, building my confidence in teaching and helping me successfully apply for my AFHEA. Teaching was one of my favorite experiences during my PhD!

To Eleni, my lab-wife, there are not enough words to express my gratitude for your help and support for the last three and a half years. This PhD would have been an entirely different, and much less enjoyable experience without you.

Thank you to Pablo, Yukyee, Liliana, Marili, Alex, Adam, Beata, Lan, John, Anna, Jurjan, and all past and present members of the PH groups at the NHLI, Imperial. Thank you for sharing both your equipment and time throughout my project, for the confidence boosts after every presentation and the fabulous lunch breaks!

I would also like to thank my dad, my mum, George and Rosie for their support throughout this PhD and my entire education. To my dad, thank you for the endless supply of whisky and always wanting to "talk science". Be it discussions on geology, chemistry, physics or even biology, thank you for encouraging me from an early age to stay curious. To my mum, thank you for being an incredible role model and for encouraging me to continue with my a-level in history, without that writing this would have been much harder.

Finally, thank you to my partner and best friend Wesley. Thank you for your never-ending encouragement, and for always being there for me, especially whilst I was learning to code. Your support has meant more to me than you could possibly realise and these last three and a half years would not have been possible without you.

Abstract

Background: Pulmonary arterial hypertension (PAH) is a rare disease characterised by remodelling of the pulmonary arteries and increased vascular resistance. A genome wide association study (GWAS) of idiopathic/heritable PAH established novel genetic associations. This GWAS identified a novel genomic locus upstream of the *SOX17* gene promoter containing two independent signals inside a putative enhancer region. Sox17 is an important transcription factor in human pulmonary artery endothelial cells (hPAECs) and rare pathogenic mutations in *SOX17* have previously been linked to PAH.

Hypothesis and aims: There is allele-specific transcription factor (TF) binding at the enhancer region upstream of *SOX17* and PAH-associated variation in this region affects *SOX17* expression, hPAEC function and PAH development. This thesis aims to define the TFs binding at the *SOX17* locus and downstream pathways and functional changes regulated by Sox17. Also, to explore how *SOX17* expression and function can be rescued by novel compounds.

Methods and Results: Electromobility shift assay (EMSA) have been used to identify TFs which differentially bind to the enhancer locus present upstream of *SOX17*. These assays have demonstrated differential binding of hPAECs nuclear proteins to the risk and non-risk alleles from a *SOX17* GWAS signal. ROR-alpha, has been identified through in silico analysis and antibody supershift assays to be potentially responsible for this binding. siRNA and CRISPR-Inhibition have been used to manipulate *SOX17* and the PAH associated loci to determine downstream genes and pathways affected by *SOX17* manipulation. Inhibition of a GWAS signal upstream of *SOX17* decreases *SOX17* expression. Manipulation of *SOX17* expression affects PAH-relevant pathways and functions in hPAECs. *In-silico* analysis has been used to predict compounds which rescue *SOX17* dysfunction. Multiple candidate compounds have been tested and show a reversal of gene changes induced by *SOX17* in hPAECs.

Conclusions: The PAH GWAS signal is now confirmed as an active enhancer site that regulates *SOX17*, a TF that regulates PAH-relevant pathways and functions in hPAECs. Novel compounds may be able to rescue these functions through restoration of downstream *SOX17* targets, and thus be candidates for the treatment of PAH.

Abstract Figure

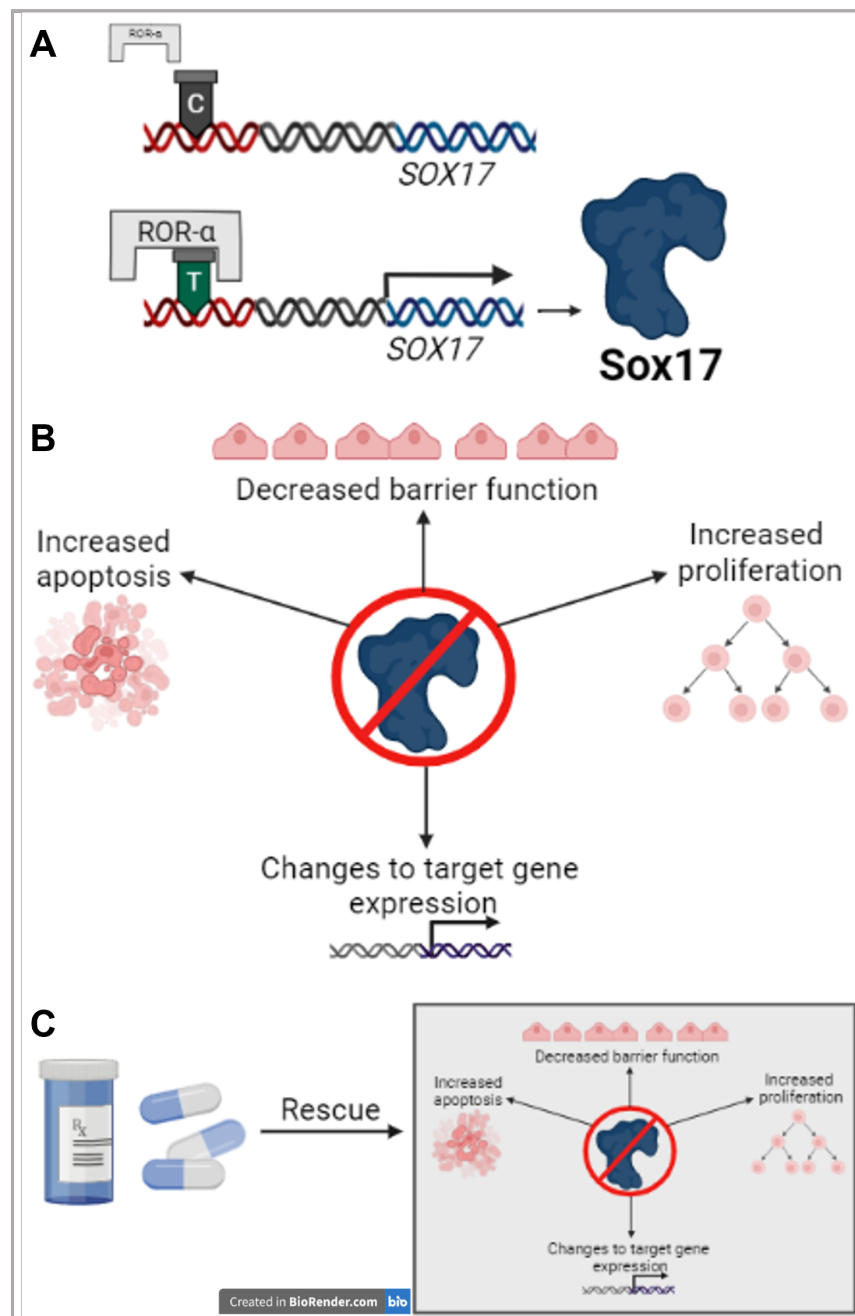


Figure 0.1

Thesis Abstract Figure. Representation of the work included in this thesis. **A)** The non-risk allele (T) present at *SOX17*-signal 2 (rs765727) upstream of *SOX17* causes additional binding of the TF ROR-alpha in hPAECs. This locus is located in an enhancer region which affects the expression of *SOX17*. **B)** Knockdown of *SOX17* expression leads to changes in hPAEC proliferation, apoptosis, viability and barrier function and affects the expression of *CDH5*, *NOTCH4*, *ADAMTS12*, *MRP17*, *LAMB3* and *COL4a6* in hPAECs. **C)** Compounds can cause the opposite effects of genetic changes which occur due to *SOX17* dysfunction in hPAECs. This thesis does not give evidence for the functional changes which occur due to rescue of *SOX17*-dysfunction with candidate compounds.

Contents

Figures	VI
---------	----

Tables	VIII
--------	------

Abbreviations	X
---------------	---

1 Introduction	1
----------------	---

1.1 Pulmonary Hypertension	1
--------------------------------------	---

1.1.1 Pulmonary Vasculature	1
---------------------------------------	---

1.1.2 Pulmonary Hypertension	2
--	---

1.2 Pulmonary Arterial Hypertension	3
---	---

1.2.1 Epidemiology	3
------------------------------	---

1.2.2 Pathophysiology	4
---------------------------------	---

1.2.3 Endothelial Cells	4
-----------------------------------	---

1.2.4 Medial and Adventitial Cells	6
--	---

1.2.5 Inflammatory Cells	7
------------------------------------	---

1.3 Current Treatments for PAH	8
--	---

1.3.1 Calcium Channel Blockers	8
--	---

1.3.2 The Prostacyclin Pathway	9
--	---

1.3.3 The Endothelin Pathway	9
--	---

1.3.4 The Nitric Oxide Pathway	10
--	----

1.4 Genetics & Genomics of PAH	11
--	----

1.4.1 TGF β Superfamily & BMPR2 Signaling	14
---	----

1.4.2 Other Genes Associated with hPAH	16
--	----

1.4.3 Common Variants associated with PAH	18
---	----

1.4.4 Epigenetic Regulation and PAH	19
---	----

1.5 <i>SOX17</i>	20
----------------------------	----

1.5.1 Members of the SOXF Family	20
--	----

1.5.2 <i>SOX17</i> Location, Expression & Function	21
--	----

1.5.3 Sox17 Related Signalling Pathways	22
---	----

1.5.4 Sox17 & the TGF β Superfamily	28
---	----

1.5.5	Common Variants Affecting SRY box transcription factor 17 (<i>SOX17</i>)	28
1.6	Summary	32
1.7	Hypotheses	32
1.8	Aims	33
2	Methods	34
2.1	Declaration of Responsibilities	34
2.2	Materials	34
2.3	Cell Culture	34
2.4	Primer & Probe Design	34
2.5	Electromobility Shift Assay	36
2.5.1	In-Silico Analysis of TF binding at <i>SOX17</i> -signal 2 . . .	36
2.5.2	Electromobility Shift Assay Protocol	36
2.5.3	Supershift EMSA assay	37
2.5.4	EMSA and Supershift Imaging	37
2.6	CRISPR-Inhibition	37
2.6.1	Inhibition of putative upstream enhancer of <i>SOX17</i> . . .	37
2.6.2	Relative Expression of <i>SOX17</i>	38
2.7	siRNA	38
2.7.1	siRNA of SOXF Factors	38
2.7.2	Relative Expression of SOXF Factors	39
2.7.3	Protein Quantification by Immunoblotting	39
2.8	Functional Assays	40
2.8.1	MTT Assay	40
2.8.2	Caspase 3/7 Assay	41
2.8.3	CellTitre-Glo Assay	41
2.8.4	Boyden Chamber Permeability Assay	41
2.9	RNA Sequencing	41
2.9.1	Gene Ontology Analysis	42
2.10	The Connectivity Map: prediction of rescue compounds from RNA signatures	42

2.10.1 RNA-Sequencing of Genetically Modified hPAECs	42
2.10.2 In-Silico Analysis of Genetically Modified <i>SOX17</i> RNA-Sequencing Data	43
2.10.3 Using The Connectivity Map to Predict Compounds to Reverse <i>SOX17 Dysfunction</i>	43
2.10.4 Cell Culture	44
2.10.5 Drug Exposure	44
2.10.6 Analysis of Candidate Gene Expression	44
2.11 Graphing & Statistical Analysis	45
3 Defining Upstream Regulators of <i>SOX17</i> & Effects of Common Variation at the <i>SOX17</i> Locus in PAH	46
3.1 Introduction	46
3.1.1 Hypotheses	50
3.1.2 Aims	50
3.2 Methods	51
3.2.1 Optimisation of EMSA protocol	51
3.2.2 Imaging of EMSA, competition EMSA and supershift assays	52
3.2.3 Creation of sgRNA for CRISP-INH of rs765727	52
3.2.4 Statistical analysis	53
3.3 Results	53
3.3.1 Differential binding of hPAEC nuclear protein to the non-risk allele	53
3.3.2 In-silico analysis suggests multiple TFs which could preferentially bind to the non-risk allele	55
3.3.3 Competition EMSA show loss of shift with ROR α TF probes	56
3.3.4 Supershift EMSA suggest ROR- α is responsible for the differential binding at rs765727	58
3.3.5 CRISPR-INH of the region surrounding rs765727 affects <i>SOX17</i> expression	60

3.3.6	RNAseq of hPAEC targeted by CRISPR-INH	60
3.4	Discussion	65
3.5	Conclusion	69
4	Functional Analysis of Sox17 Manipulation in Pulmonary	
	Vascular Cells	70
4.1	Introduction	70
4.1.1	Hypothesis	71
4.1.2	Aims	72
4.2	Methods	73
4.2.1	RNAseq and principal component analysis	73
4.2.2	Selection of <i>SOX17</i> -related genes	73
4.2.3	Statistical Analysis	73
4.3	Results	75
4.3.1	<i>SOX17</i> -siRNA successfully knocks down <i>SOX17</i> at the mRNA and protein level	75
4.3.2	Knockdown of <i>SOX17</i> leads to downstream genetic and molecular pathway changes	76
4.3.3	Knockdown of <i>SOX17</i> influences downstream targets in hPAECs	80
4.3.4	Knockdown of <i>SOX17</i> influences hPAEC functions . . .	83
4.3.5	Comparison of differentially expressed genes following CRISPR- INH of <i>SOX17</i> -signal 1 and 2 and <i>SOX17</i> -siRNA	84
4.3.6	Knockdown of individual SOXF TFs by small interfering RNA (siRNA) has an effect on other SOXF TFs in hPAECs	87
4.4	Discussion	93
4.5	Conclusion	97
5	Repurposing of Drug Compounds to Rescue Loss of SOX17	
	Function in PAH	98
5.1	Introduction	98
5.1.1	Hypothesis	100

5.1.2	Aims	101
5.2	Methods	102
5.2.1	Connectivity Map Query and Analysis	102
5.2.2	Statistical Analysis	102
5.3	Results	103
5.3.1	Analysis of differentially expressed genes following <i>SOX17</i> promoter and signal 1 manipulation	103
5.3.2	Comparison of connectivity map perturbagen signatures with RNAseq analysis of <i>SOX17</i> inhibition	105
5.3.3	Exposure of hPAEC to CMap Compounds	106
5.4	Discussion	112
5.5	Conclusion	116
6	Conclusions & Future Work	117
6.1	Defining <i>SOX17</i> -signal 2 & regulators of <i>SOX17</i>	117
6.2	The role of <i>SOX17</i> in hPAECs and PAH	119
6.3	The role of <i>SOX17</i> in intracranial aneurysms and cancer	120
6.4	<i>SOX17</i> and therapies in PAH	121
6.5	Summary	125
	References	127
7	Supplemental Work	166
7.1	Supplemental Tables	168
7.2	Supplemental Figures	177

Figures

Figure 0.1:	Thesis Abstract Figure	5
Figure 1.1:	Structure of pulmonary arteries and pathophysiology of PAH.	5
Figure 1.2:	Main pathways targeted with current PAH therapeutics.	11
Figure 1.3:	The transforming growth factor beta superfamily.	16
Figure 1.4:	The SOXF transcription factor family.	21
Figure 1.5:	<i>SOX17</i> related signalling pathways.	27
Figure 1.6:	<i>SOX17</i> genetic variations associated with PAH.	29
Figure 1.7:	Common <i>SOX17</i> genetic variation associated with PAH.	31
Figure 3.1:	Electromobility shift and supershift assay.	48
Figure 3.2:	Optimisation of EMSA protocol.	52
Figure 3.3:	Differential binding of proteins to rs765727	54
Figure 3.4:	Relative expression of transcription factors of interest in hPAECs	56
Figure 3.5:	Differential binding of competition TF consensus binding sequence probes to rs765727	57
Figure 3.6:	Supershift EMSA	59
Figure 3.7:	Knockdown of <i>SOX17</i> through CRISPR-Inhibition of <i>SOX17</i> -signal 2	60
Figure 3.8:	Differentially expressed genes following CRISPR-inhibition of <i>SOX17</i> -signal 2 in hPAECs	62
Figure 3.9:	Over-representation analysis for enriched pathways, functions and molecular processes following CRISPR-inhibition of <i>SOX17</i> -Signal 2	65
Figure 4.1:	Knockdown of <i>SOX17</i> following siRNA.	75
Figure 4.2:	Differentially expressed genes following siRNA- <i>SOX17</i> in hPAECs	79

Figure 4.3:	Over-representation analysis for enriched pathways and functions following siRNA-<i>SOX17</i> . . .	80
Figure 4.4:	Validation of gene ontology target genes	81
Figure 4.5:	Validation of potential <i>SOX17</i>-associated genes	82
Figure 4.6:	Effect of <i>SOX17</i>-siRNA on hPAEC function . .	84
Figure 4.7:	Comparison of DEG following CRISPR-INH of <i>SOX17</i>-signal 1 and 2 and <i>SOX17</i>-siRNA . . .	85
Figure 4.8:	Knockdown of SOXF factors.	88
Figure 4.9:	Comparison of knockdown of SOXF factors. . .	90
Figure 5.1:	Connectivity map workflow.	100
Figure 5.2:	Differentially expressed genes following CRISPR-activation of <i>SOX17</i> promoter in hPAECs . . .	103
Figure 5.3:	Differentially expressed genes following CRISPR-inhibition of <i>SOX17</i> promoter in hPAECs . . .	104
Figure 5.4:	Differentially expressed genes following CRISPR-inhibition of <i>SOX17</i>-Signal 1 in hPAECs	104
Figure 5.5:	Heatmaps of clustered regularly interspaced short palindromic repeats (CRISPR) fold change and CMap perturbagen Z-scores	106
Figure 5.6:	Differentially expressed genes following Sirolimus exposure	108
Figure 5.7:	Differentially expressed genes following Aminopurvalanol-a exposure	109
Figure 5.8:	Differentially expressed genes following Aminopurvalanol-a exposure	110
Figure 5.9:	Differentially expressed genes following YK-4279 exposure	111
Figure 7.1:	Knockdown of <i>SOX7</i> following siRNA.	177
Figure 7.2:	Knockdown of <i>SOX18</i> following siRNA.	178
Figure 7.3:	Full images of western blots shown in this thesis	179

Tables

Table 1.1:	Clinical grouping of PH	2
Table 1.2:	Genes which are affected by rare and common deleterious genetic variants and have been associated with PAH.	13
Table 1.3:	Common variants associated with PAH.	32
Table 2.1:	EMSA biotin (-bio) labelled and unlabeled probe sequences.	35
Table 2.2:	Transcription factor competition EMSA probes.	35
Table 2.3:	CRISPRi guide RNA sequences	35
Table 2.4:	qPCR gene primers	36
Table 2.5:	Antibodies	40
Table 2.6:	CRISPR guides for <i>SOX17</i>-signal 1 and <i>SOX17</i> promoter	43
Table 3.1:	Binding reactions for EMSA optimisation.	51
Table 3.2:	Top 50 significantly differentially expressed genes following <i>SOX17</i>-signal 2 inhibition	64
Table 4.1:	Top 50 significantly differentially expressed genes following <i>SOX17</i>-siRNA.	78
Table 4.2:	GO analysis and <i>SOX17</i>-associated gene changes in RNAseq and qPCR analysis.	83
Table 4.3:	Differentially expressed genes in common between siRNA-<i>SOX17</i> and <i>SOX17</i>-signal 2 hPAECs	86
Table 4.4:	Differentially expressed genes in common between siRNA conditions	93
Table 5.1:	Top compounds which reached a tau score cutoff of +/-90 in CMap queries.	105
Table 7.1:	Materials and reagents.	171
Table 7.2:	CMap <i>SOX17</i> promoter repression gene lists	172
Table 7.3:	CMap <i>SOX17</i> promoter activation gene lists	173
Table 7.4:	CMap <i>SOX17</i>-signal 1 repression gene lists	174

Table 7.5: Gene, drug, condition and primer information for	
connectivity map qPCR	175

Abbreviations

A. acid	arachidonic acid	CHD	congenital heart diseases
<i>ACT</i>β	actin beta	ChIP	chromatin immunoprecipitation
<i>ADAM10</i>	A disintegrin and metalloproteinase domain-containing protein 10	Chr	chromosome
ANOVA	analysis of variance	CMap	connectivity map
BFP	blue fluorescent protein	CRISPR	clustered regularly interspaced short palindromic repeats
BMP	bone morphogenetic protein	CRISPR–ACT	CRISPR activation
BMPR	bone morphogenetic protein receptor	CRISPR–INH	CRISPR inhibition
<i>BMPR2</i>	bone morphogenetic protein receptor 2	dCas9	dead Cas9
BOEC	blood outgrowth endothelial cell	DEG	differentially expressed gene
bp	base pair	DLL	delta-like protein
cAMP	cyclic adenosine monophosphate	DMSO	dimethyl sulfoxide
CCB	calcium channel blockers	DNA	deoxyribonucleic acid
CDK	cyclin dependent kinase	EC	endothelial cell
cDNA	complimentary DNA	ECM	extracellular matrix
cGMP	cyclic guanosine monophosphate	EGM2	endothelial growth medium 2
		EMSA	electromobility shift assay

EndoMT	endothelial-to-mesenchymal transitioning	hPAEC	human pulmonary artery endothelial cell
EGR1	early growth response protein 1	hPAH	heritable PAH
ET-1	endothelin-1	HRE	hypoxia response elements
ETA	endothelin receptor A	HUVEC	human umbilical vein endothelial cell
ETB	endothelin receptor B	IL8	interleukin 8
FBS	fetal bovine serum	iPAH	idiopathic PAH
FDR	false discovery rate	JAG	jagged
FZD	frizzled receptors	KRAB	Kruppel associated box
GFP	green fluorescent protein	L-a	L-arginine
GMP	guanosine monophosphate	LPR	low-density lipoprotein receptor-related protein co-receptor
GO	gene ontology	LPS	lipopolysaccharide
GSK3β	glycogen synthase kinase 3 β	mPAP	mean pulmonary artery pressure
GTP	guanosine triphosphate	mRNA	messenger ribonucleic acid
GWAS	genome wide association study	MRPL15	mitochondrial ribosomal protein L15
HEK293FT	human embryonic kidney 293FT	mTOR	mammalian target of rapamycin
HIF	hypoxia-inducible factor	mTORC1	mTOR complex 1
HMG	high mobility group	NO	nitric oxide

NICD	notch intracellular domain	qPCR	quantitative-PCR
NR2F2	nuclear receptor subfamily 2 group F member 2	RISC	RNA induced silencing complex
PAEC	pulmonary artery endothelial cell	RNAi	RNA interference
PAH	pulmonary arterial hypertension	RNAseq	RNA sequencing
PASMC	pulmonary artery smooth muscle cell	ROR-<i>alpha</i>	RAR-related orphan receptor alpha
PCR	polymerase chain reaction	RTPCR	reverse transcription-PCR
PDE5	phosphodiesterase type 5	SMAD	mothers against decapentaplegic homologue
PDGF	platelet-derived growth factor	sGC	soluble guanylate cyclase
PGI2	prostacyclin	siRNA	small interfering RNA
PH	pulmonary hypertension	SMC	smooth muscle cell
PHD2	prolyl hydroxylase	SNP	single nucleotide polymorphism
PBS	phosphate buffered saline	SOX7	SRY box transcription factor 7
pos	position	SOX17	SRY box transcription factor 17
PVR	pulmonary vascular resistance	SOX18	SRY box transcription factor 18
PVOD	Pulmonary Veno-occlusive Disease	TAD	topologically associated domain

TEF–LEF	T-cell factor/lymphoid enhancer factor	VE–Cadherin	vascular endothelial–cadherin
TF	transcription factor	VEGF	vascular endothelial growth factor
TGFβ	transforming growth factor β	VEGFR2	vascular endothelial growth factor receptor 2
TGFBR2	transforming growth factor β receptor 2	VR–PAH	PAH vasoresponders
TMEM68	transmembrane protein 68	Wnt–PCP	Wnt/planar cell polarity

1 Introduction

1.1 Pulmonary Hypertension

1.1.1 Pulmonary Vasculature

The pulmonary vasculature is responsible for the flow of deoxygenated blood from the right ventricle of the heart, through the lungs to be oxygenated and back to the left heart for systemic circulation. It begins with the pulmonary artery which is around 30 mm in diameter [Townesley, 2012] and splits into the right and left side to deliver blood to each lung. The pulmonary arteries become smaller, becoming the pulmonary arterioles eventually reaching the capillaries where the alveoli are located and where gas exchange occurs. These capillaries branch off the main artery at an oblique angle following airway branching and become smaller as they reach the periphery of the lung; ending in around 10 to 15 μm in diameter [Townesley, 2012]. The systolic pressure of the pulmonary artery is 14 mmHg (± 3) (millimetres of mercury), an order of magnitude lower than that of systemic arteries [Leslie and Wick, 2005]. The pulmonary circulatory system is under low pressure when compared to the systemic circulatory system which requires a high pressure to deliver oxygenated blood to the body's extremities. The pulmonary circulation does not require this high pressure as it only delivers blood a short distance from the heart to the lungs, and back to the heart.

All arteries have a similar basic structure (Figure 1.1A). The tunica intima is the innermost layer and consists of a single layer of pulmonary artery endothelial cells (PAECs) that are in contact with the blood. The PAECs act as a barrier between the blood and the outer vessel wall. They are also important signalling mediators, reacting to small changes in shear stress and pressure which can occur at branching points or as the vessel diameter becomes smaller. Differentiation of PAECs is under tight regulation by extrinsic and intrinsic factors to ensure the correct lineage is followed to create either venous, arterial or lymphatic endothelial cells (ECs). For example, Notch is an essential factor for the differentiation into PAECs and it is believed to prevent venous EC fate determination [Villa et al., 2001]. More detail on this topic is given in section 1.5.1.

The intima is separated from the second layer, the tunica media, by the elastic lamina which provides structure to the arterial wall and prevents collapse. The tunica media is characterized by pulmonary artery smooth muscle cells (PASMCs) surrounded by another layer of the elastic lamina. The PASMCs allow the artery wall to contract and relax, allowing the maintenance of pulsatile flow. The outermost layer is the tunica adventitia and consists of collagen fibres, fibroblasts, and elastic lamina. It is the strongest layer of the artery wall and prevents it from over-expanding. Although all arteries have this basic structure, different types of arteries exist with different roles depending on factors such

as pressure and shear stress. The largest arteries contain more elastic layers interspersed throughout the PASMCs in the medial layer. This gives them more rigidity and strength to withstand high blood flow from the heart to the smaller arteries. Transitional arteries contain features of both larger elastic arteries and smaller arteries. The smaller arteries can be categorised into muscular and non-muscular which have higher or lower levels of PASMCs respectively. Capillaries are the smallest non-muscular arteries and consist of ECs and an elastic lamina only. This results in a short distance between blood and air, maximizing the efficiency of gas exchange [Leslie and Wick, 2005].

There are diseases which affect the arteries of the pulmonary circulation including pulmonary hypertension (PH), which this thesis investigates.

1.1.2 Pulmonary Hypertension

PHs is categorised into five main groups (Table 1.1) depending on the physiology and underlying cause of the disease. These multiple forms have different drives and diseases associated with them. These classifications are subject to updates by the European Society of Cardiology and the European Respiratory Society and are based on the most up-to-date research [Galiè et al., 2019]. This thesis focuses on clinical group 1 or pulmonary arterial hypertension (PAH).

Clinical Group	Name	Definition
1	PAH Idiopathic PAH Heritable PAH PAH associated with other Diseases Drug-induced PAH	Pre-Capillary PH
2	PH due to Left Heart Disease	Post-Capillary PH
3	PH due to lung disease	Pre-Capillary PH
4	Chronic Thromboembolic PH	Pre-Capillary PH
5	PH due to miscellaneous mechanisms	Pre-Capillary PH Post-Capillary PH

Table 1.1: Clinical grouping of PH. PH, pulmonary hypertension. PAH, pulmonary arterial hypertension. Taken from <https://phassociation.org/>

PAH is characterised by remodelling of the structures of the pulmonary vasculature which causes an increase in blood pressure in the pulmonary arteries. The healthy pulmonary vasculature is a low-pressure system, and the arteries are adapted to this low-pressure environment. The increased pulmonary vascular resistance in PAH triggers remodelling

of all structures of the pulmonary circulatory system, increasing the pressure and right ventricular afterload and eventually causing death through right heart hypertrophy and failure. The remodelling process (Figure 1.1) includes stiffening of larger arteries and muscularization and obliteration of smaller (non-muscular) arteries, fibrosis, and the development of plexiform lesions; network-like, complex vascular formations which occur during the final stages of the remodelling process.

In PH the resting mean pulmonary artery pressure (mPAP) is ≥ 25 mmHg, whereas the average in healthy individuals is 14.0 ± 3.3 mmHg [Kovacs et al., 2009]. However, a mPAP of 19–24 mmHg is classified as borderline PH and has a higher mortality risk than an mPAP of under 19 mmHg, highlighting the large spectrum of this disease [Lau et al., 2015]. In the most recent PH international symposium, it was suggested that the resting mPAP used for PAH diagnosis should be lowered to <20 mmHg. However, it was decided that this is too low to be used independently and will often capture patients who do not have a vascular disease but do have a higher cardiac output. Therefore it was proposed that pulmonary vascular resistance (PVR) ≥ 3 wood units be included in diagnosing PAH [Galiè et al., 2019] as this will minimise false positive diagnoses.

1.2 Pulmonary Arterial Hypertension

1.2.1 Epidemiology

PAH corresponds to clinical group 1 PH (Table 1.1) and is a rare disease with an incidence of between 2.1 and 7.4 cases per million population per year [Humbert et al., 2006, Peacock et al., 2007, Hoeper et al., 2016a]. With no intervention, the median survival is 2.8 years [D’Alonzo et al., 1991] and with modern therapeutic intervention, the rate of mortality in the first year is around 15% [Thenappan et al., 2007]. Patients suffer significant morbidities, with reduced exercise capacity and quality of life. The mean age of onset is between 54 and 68 years [Benza et al., 2010, Hoeper et al., 2013]. There is a higher percentage of women who suffer from PAH with between 57.0% and 79.5% of patients being female [Humbert et al., 2006, Badesch et al., 2010, Ling et al., 2012]. However, the disease prognosis is worse in men [Ventetuolo et al., 2014]. It is unclear why this paradox occurs, however there is research which suggests sex hormones such as estrogen play a contrasting role in PH development and progression, leading to a higher risk in females, but these patients surviving longer [White, 2016].

There are four sub-groups of PAH (Table 1.1), idiopathic PAH (iPAH), heritable PAH (hPAH), PAH associated with other diseases (such as congenital heart disease) or viruses (such as HIV) and drug-induced PAH. iPAH refers to PAH where there is no family history or obvious underlying cause and is the most common form of PAH [Galiè et al.,

2015a]. Research often focuses on the idiopathic form which is less confounded by known co-morbidities.

1.2.2 Pathophysiology

PAH categories may differ in their underlying causes but they are all characterised by a resting mPAP of equal to or greater than 25 mmHg and pulmonary arterial wedge pressure less than 15mmHg (and PVR > 3 Woods units) [Galiè et al., 2015a]. Increased pulmonary vascular resistance in PAH is driven by structural changes and a pre-capillary pathology involving vasoconstriction, inflammation and proliferative and destructive remodelling of the artery wall. This results in the occlusion of small arteries [Schermuly et al., 2011, Lan et al., 2018] (Figure 1.1B). It is not clear how each of these factors contributes to the onset and progression of the disease.

1.2.3 Endothelial Cells

It is known that PAH affects all layers of the pulmonary artery wall. In healthy pulmonary arteries, the ECs form a semi-permeable barrier between the blood and the outer layers of the artery. They perform important functions in response to external stimuli such as a change in shear stress or oxygen levels and release growth, angiogenic, coagulation and other factors into their environment. The tight regulation of these pathways is essential for maintaining barrier function and for vascular homeostasis. The initial trigger of PAH is thought to be PAEC dysfunction and injury and in severe disease, the intima has been found to increase in thickness by up to 3-fold [Stacher et al., 2012]. Although the exact cause of PAEC injury in iPAH is unknown it is thought to be a combination of genetic factors, hypoxia, change in shear stress, inflammation, or toxins. It is known it leads to the release of pro-proliferative factors such as VEGF, a reduction in vasodilators such as nitric oxide and an increase in vasoconstrictors such as ET-1. This leads to vascular remodelling of the pulmonary artery wall through PAEC proliferation, loss of barrier integrity, permeability changes, dysregulated angiogenesis, invasion of the intimal layers by cells exhibiting fibroblast markers and endothelial-to-mesenchymal transitioning (for full review see: [Vaillancourt et al., 2015]).

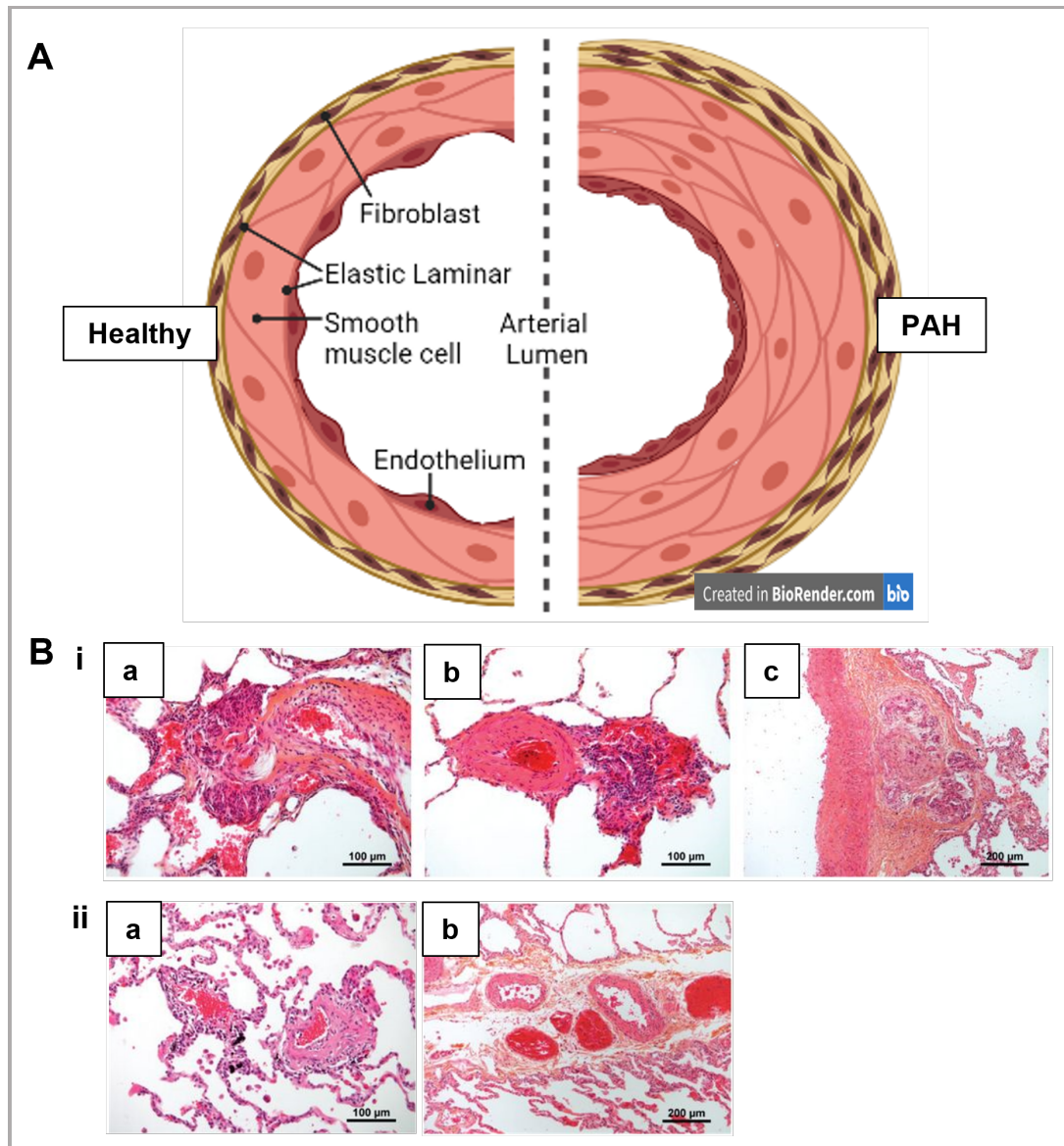


Figure 1.1

Structure of pulmonary arteries and pathophysiology of PAH. A) Structural changes of the pulmonary artery during PAH. Endothelial cells line the tunica intima, smooth muscle cells line the tunica media, and the tunica adventitia is made up of connective tissue and fibroblasts. The elastic lamina sits between the ECs and the SMCs and between the SMCs and the adventitia. Increasing severity of PAH is characterised by endothelial cell and smooth muscle cell proliferation and decreased apoptosis and fibroblast proliferation. As PAH progresses, the arterial wall becomes thicker and stiffer, leading to increased vascular resistance and heart failure. **B) Representative images of vascular lesion.** Images taken from: [Humbert et al., 2019]. **i.** Images of plexiform lesions taken from PAH patient lungs. **ii.** Images of representative vascular lesions taken from PAH patient microvasculature. **a)** Image of microvessel showing inflammation and remodelling. **b)** Image of microvessel showing muscular hyperplasia and remodelling in PAH.

One form of dysfunction of PAECs is the production of a cell population which have a hyper-proliferative/anti-apoptotic state [Jonigk et al., 2011]. In PAH, this state appears to be the result of aberrant growth factor release by the PAECs themselves, an initial

stage of apoptosis and the subsequent selection of anti-apoptotic PAECs [Campbell et al., 2001, Zhao et al., 2003]. Studies have shown that a higher rate of proliferation and increased resistance to apoptosis is found in hPAEC derived from iPAH patient lung [Masri et al., 2007]. At later stages of the disease, complex plexiform lesions can appear which cause further obstruction of the lumen. Plexiform lesions have a complex and dynamic structure and are a result of disorganized angiogenesis of PAECs into vessels containing apoptosis-resistant myofibroblasts and matrix proteins [Cool et al., 1999, Yi et al., 2000]. The microenvironment of plexiform lesions is affected by this remodelling process. Upregulation of hypoxia and shear stress-induced angiogenic mediators has been found in plexiform lesions and they are a possible source of these factors in the blood. It is thought that they contribute to increased angiogenesis in other areas of the pulmonary system in PAH [Jonigk et al., 2011].

1.2.4 Medial and Adventitial Cells

In PAH, in the medial layer, PASMCs are characterised by a hyper- proliferative, anti-apoptotic state which results in increased thickness of the medial layer termed medial hypertrophy. Muscularization of distal, small non-muscular arteries also occurs through the hyperproliferation and migration of PASMCs and differentiation of PAECs into a smooth muscle-like state [Jones et al., 1999]. At the border between muscularized and non-muscularized arteries, PASMC progenitors have been identified which are thought to be “primed” for distal expansion and muscularization of smaller arteries under specific conditions [Sheikh et al., 2015]. In severe PAH, the differentiation and migration of PASMCs can contribute to the formation of an additional artery wall layer between the intima and the internal elastic lamina called the neointima which can form occlusive lesions in severe disease.

The neointima consists of extracellular matrix and differentiated cells with PASMC markers and/or mesenchymal markers. However, there is debate about the origin of these differentiated cells. One possible origin is the aforementioned differentiated PASMCs. It has been shown in a house dust mite mouse model of PAH that a small number of PASMCs breach the internal elastic lamina and begin the formation of the neointima through a notch-dependent process [Steffes et al., 2020]. Other studies suggest another origin for the neointima to be PAECs, which differentiate through endothelial-to-mesenchymal transitioning (EndoMT) [Ranchoux et al., 2015, Good et al., 2015]. EndoMT is regulated by the transcription factors snail, slug and Twist-1 which cause the downregulation of PAEC-specific genes and junction proteins such as vascular endothelial–cadherin (VE–Cadherin) and p120-Catenin. These proteins are essential for the maintenance of the semi-permeable barrier through the formation of tight junctions between PAECs. The loss of cell-cell junction proteins causes the PAECs to detach from the monolayer

and migrate into the medial layer of the artery. EndoMT, therefore, results in a decrease in PAEC barrier function and affects the semi-permeable nature of the intimal layer of pulmonary arteries [Good et al., 2015]. Interestingly, hPAEC cells derived from iPAH patient lungs migrate more than control hPAEC from control lungs in a cell migration assay [Masri et al., 2007]. This process is important during the early development of pulmonary arteries [Arciniegas et al., 2005] but has been implicated in PAH [Ranchoux et al., 2015]. Lineage tracing studies using monocrotaline mouse models of PAH have shown that PAECs can contribute to the creation of the neointima via EndoMT [Qiao et al., 2014]. It is not clear why transcription factors involved in EndoMT become more active in PAECs in PAH, but one such factor, early growth response protein 1 (EGR1) is increased in PAH patient PAECs and experimental-flow PAECs [Dickinson et al., 2011]. EGR1 is known to induce the expression of slug, a known EndoMT mediator [Cheng et al., 2013]. Other genetic factors, such as members of the transforming growth factor β (TGF β) family, could play a role (discussed in section 1.4.1) [Ranchoux et al., 2015]. Other work suggests EndoMT occurs after exposure to inflammatory molecules, hemodynamic forces and cyclic strain [Krenning et al., 2016, Suzuki et al., 2018]. The formation of the neointima, whether caused by PAECs, PASMCs or a combination of these and other cell types, represents an important stage in PAH development and highlights the severity of the remodelling which the pulmonary arteries undergo.

The tunica adventitia is primarily made up of fibroblasts and the elastic laminar and collagen which they produce. Studies suggest that adventitia act as the primary “injury-sensing” tissue and, in response to cellular injury, fibroblasts deposit more extracellular matrix components and release growth factors that affect the other cell types of the artery wall (for full review see: [Stenmark et al., 2006]).

1.2.5 Inflammatory Cells

PAECs, PASMCs and fibroblasts, whilst important in the origin and development of PAH, are not the only cell types to consider in this disease; the inflammatory system is also extremely influential. For example, both T- and B-lymphocyte cells are found in plexiform lesions, although it is unclear whether they are a causative agent of these novel formations or migrate here once the lesion is formed [Tuder et al., 1994]. In disease-associated PAH, cellular injury attracts inflammatory mediators which are a possible source of aberrant growth factors which could cause the pro-proliferative, anti-apoptotic environment which is prevalent in the pulmonary arteries during PAH [Cool et al., 1997]. Inflammatory markers found in the blood may also be useful for the prognosis of patients. The levels of circulating white blood cells in patients have been explored as a potential biomarker for PAH as the neutrophil/lymphocyte ratio is a potential prognostic marker for disease severity [Harbaum et al., 2017].

The interactions between the cells of the pulmonary artery wall, the inflammatory system and the circulating blood have been well studied and are known to be important in the onset and progression of PAH (for full review see: [Nogueira-Ferreira et al., 2014]). Figure 1.2 highlights three pathways that affect multiple cell types of the artery wall and are targeted by current therapies. To fully elucidate the causes of PAH and to produce more effective therapies, the interplay between these cell types must be appreciated.

1.3 Current Treatments for PAH

Current treatments for PAH mainly target pathways that affect pulmonary vasoconstriction and pulmonary vascular cell proliferation and are often used in combination [Hoeper et al., 2016b, Sommer et al., 2020]. There are three main pathways targeted, prostacyclin (PGI₂), endothelin-1 (ET-1) and nitric oxide (NO). Vasodilator mediators such as PGI₂ and NO are known to be reduced in PAH and vasoconstrictors (ET-1) are upregulated [Christman et al., 1992, Giaid et al., 1993, Giaid and Saleh, 1995]. Although these therapies can improve outcomes for patients [Sitbon et al., 2019], often they do not reverse pulmonary artery remodelling and so the patient remains unwell. Furthermore even with modern therapeutic intervention, the rate of mortality in the first year is still around 15% [Thenappan et al., 2007]. Additionally, many of the current therapeutics also cause unacceptable side effects which frequently lead to patients switching therapies. These issues highlight the need for a personalised approach to PAH treatment in the future.

1.3.1 Calcium Channel Blockers

Calcium channel blockers (CCB) disrupt the flow of calcium into vascular cells leading to the relaxation of PSMCs and vasodilation. They are not thought to affect the proliferation of vascular cell types. Nifedipine was shown to have a short- and long-term effect on PAH patients cardiac output and ventricular ejection fraction as well as improving exercise capacity [Rubin et al., 1983]. It is primarily given to patients with a relatively slower heart rate [BJC, 2020]. Diltiazem has been shown to reduce pulmonary arterial pressure in PAH [Islam et al., 1999] and is given to patients with a relatively faster heart rate [BJC, 2020]. Although current therapeutics are only partially beneficial to most, there is a small group of patients who have a significant drop in pulmonary artery pressure and have extended long-term survival after treatment with calcium channel blockers, these are known as PAH vasoresponders (VR-PAH) [Sitbon et al., 2005]. VR-PAH account for <10% of the iPAH patient population [Sitbon et al., 2005]. To be classified as VR-PAH, a patient must have a decrease in mPAP (>10mmHg, or a reduction to less than 40mmHg) and an increase in cardiac output after a vasodilator challenge with inhaled nitric oxide, iloprost, intravenous epoprostenol or adenosine [Sitbon et al., 1998,

Jing et al., 2009, Maron et al., 2012, Oliveira et al., 2010]. This is performed at the start of the treatment journey. If this challenge is positive CCB such as diltiazem or nifedipine can be given long-term. Monitoring of these patients is essential, as only 50% maintain a long-term response to these therapies [Chin and Rubin, 2008].

1.3.2 The Prostacyclin Pathway

PGI₂ is produced in PAECs of the artery through the conversion of arachidonic acid by prostacyclin synthase. Prostacyclin receptors are located on PASMCs of the tunica media and when they are bound by PGI₂ change downstream cyclic adenosine monophosphate (cAMP) levels which causes relaxation of the smooth muscle and vasodilation [Clapp et al., 2002] (Figure 1.2). In PAH, the levels of circulating PGI₂ are decreased [Christman et al., 1992] and drugs targeting this pathway attempt to make up for the loss of PGI₂. However, PGI₂ is not stable and so cannot be used as a therapeutic, therefore, analogues such as iloprost, treprostinil and epoprostenol have been created and are collectively known as prostanoids. They bind to PGI₂ receptors on PASMCs and trigger the production of cAMP. All three have shown an increase in positive outcomes for patients [Barst et al., 1996, Olschewski et al., 2002, Jing et al., 2013]. Epoprostenol is a therapy of choice for many PAH patients presenting with extreme symptoms and was the first targeted therapy to be approved. However, administration of epoprostenol requires continuous infusion via a catheter, causing issues for patient comfort and increased infections due to the method of administrations. New versions of epoprostenol have reduced the frequency of pump changes, thus reducing the risk of infection and improving patient comfort [Sitbon et al., 2014]. Selexipag is a new treatment that binds selectively to PGI₂ I receptors only and acts as an agonist. A phase 3 randomized, double-blind, placebo-controlled trial, found the risk of death or complication to be significantly lower with selexipag than a placebo [Sitbon et al., 2015].

1.3.3 The Endothelin Pathway

ET-1 is a potent vasoconstrictor and has two receptor subtypes, endothelin receptor A (ETA) and endothelin receptor B (ETB) (Figure 1.2). Both receptors are found on PASMCs, although a low level of ETB is also found in PAECs and have their effects on different pathways. Through these two receptors, the ET-1 pathway causes vessel narrowing through the constriction of PASMCs and proliferation of PASMCs and PAECs. The ET-1 pathway is upregulated in PAH [Giaid et al., 1993] and drugs targeting this pathway attempt to antagonise ETA and ETB and block this pathway (Figure 1.2). Three ETA and ETB antagonists have been licensed and are now used in PAH therapy, bosentan, ambrisentan and macitentan. Bosentan and macitentan are dual receptor antagonists and bind to both ETA and ETB. Macitentan is a modified version of bosentan with increased

receptor affinity. Ambrisentan is a selective antagonist and binds to only ETA. Having drugs that target the different ET-1 receptors is important to allow the selective targeting of different downstream effects of the ET-1 pathway. All of these therapies have been shown to improve exercise capacity and hemodynamics of patients in controlled trials [Rubin et al., 2002, Humbert et al., 2007, Iglarz et al., 2008, Galiè et al., 2008, Ben-Yehuda et al., 2012]. Sitaxentan is another drug of this class, but it is no longer prescribed due to increased liver toxicity and death in patients [Lavelle et al., 2009].

1.3.4 The Nitric Oxide Pathway

NO is produced from L-arginine via nitric oxide synthase enzymes. NO activates soluble guanylate cyclase (sGC) which catalyses the formation of cyclic guanosine monophosphate (cGMP) and causes smooth muscle relaxation. Phosphodiesterase type 5 (PDE5) is responsible for the degradation of cGMP (Figure 1.2). The NO pathway is known to be reduced in PAH [Giaid and Saleh, 1995] and PDE5 is highly expressed in the lungs of PAH patients [Galiè et al., 2005]. There are two classes of therapeutics which target this pathway, PDE5 inhibitors and sGC stimulators [Stasch et al., 2011]. Riociguat is an sGC stimulant and stabilizes the NO-sGC complex. This class of drug can also stimulate sGC in a NO-independent manner to ensure the formation of cGMP in PASMCs [Stasch et al., 2001]. In a placebo-controlled phase 3 trial, riociguat was found to improve 6-minute walk distance and pulmonary vascular resistance [Ghofrani et al., 2013]. Sildenafil and tadalafil are PDE5 inhibitors and prevent cGMP hydrolysis, thus allowing for it to have its downstream effects on PASMCs. Both therapeutics have been found to improve the 6-minute walk distance of patients in a dose-dependent manner [Galiè et al., 2005, 2009]. The PATENT-PLUS study investigated whether there is a therapeutic synergy between sGC stimulant (riociguat) and PDE5 inhibitor (sildenafil). However, no clinical benefit of using both therapies was found, and some patients discontinued the treatment due to higher occurrence of systemic hypotension [Galiè et al., 2015b].

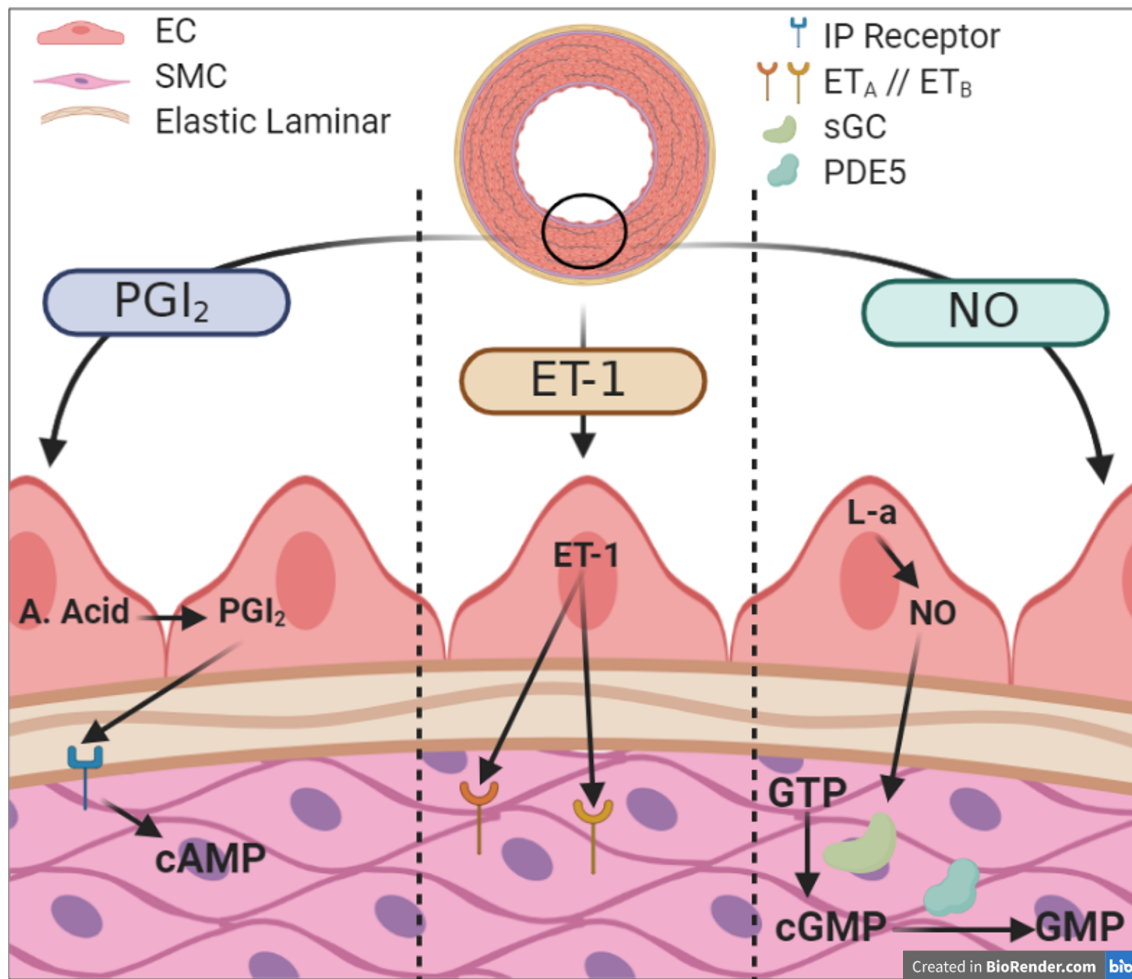


Figure 1.2

Main pathways targeted with current PAH therapeutics. Adapted from [Humbert et al., 2014]. endothelial cell (EC). smooth muscle cell (SMC). **A) Prostacyclin Pathway.** PGI₂ is produced in ECs of the artery through the conversion of arachidonic acid by prostacyclin synthase. Prostacyclin receptors are located on SMCs and, when they are bound by PGI₂, change downstream cAMP levels. Prostacyclin (PGI₂). Arachidonic acid (A. acid). Cyclic adenosine monophosphate (cAMP). **B) Endothelin Pathway.** ET-1 is produced in ECs. It has two receptor subtypes, ET_A and ET_B. Both receptors are found on SMCs. This pathway causes vessel narrowing through the constriction of PSMCs and proliferation of PSMCs and PAECs. Endothelin-1 (ET-1). Endothelin receptor A (ET_A). Endothelin receptor B (ET_B). **C) Nitric Oxide Pathway.** NO is produced from L-arginine via nitric oxide synthase enzymes in the ECs. NO activates sGC which catalyzes the formation of cGMP. PDE5 is responsible for the degradation of cGMP. Nitric oxide (NO). Soluble guanylate cyclase (sGC). Phosphodiesterase type 5 (PDE5). L-arginine (L-a). Guanosine triphosphate (GTP). Cyclic guanosine monophosphate (cGMP). Guanosine monophosphate (GMP).

1.4 Genetics & Genomics of PAH

Gene regulation is controlled by a number of factors including promoters, enhancers, super enhancers and epigenomic regulation. These work together to tightly control the spatial and temporal expression of target genes. A promoter is a region of the genome typically

located next to the transcription start site of a gene where regulatory complexes bind to start transcription of that gene. They are responsible for turning the transcription of a gene on or off. They are different to enhancers which fine tune the expression of a gene through bound transcription factors and complexes. Enhancers are typically found further away from gene transcription start sites than promoters. However, they are often located inside a topologically associated domains (TADs) which also contains their target gene (if acting in a cis- fashion). Regions inside TADs are more likely to interact with other locations inside the same TAD than with regions outside the TAD (for full review see: [Szabo et al., 2019]). Super enhancers are regions of the genome which have a high number of enhancer regions and high transcription factor binding (for full review: [Pott and Lieb, 2014]). Epigenetic regulation refers to regulation conferred by the epigenome, reversible changes to the DNA sequence which affects gene transcription. This could be hypermethylation which can cause a decrease in expression if it occurs over promoter or enhancer regions. Epigenomic regulation also encompasses non-coding RNA mechanisms such as long non-coding RNAs and microRNAs. These can affect translation of genes by contributing to cellular RNA interference and the breakdown of target mRNA before it can be translated into protein.

To assess how genetic changes can affect the regulation of genes and their protein products, GWAS can be used. These are large scale studies which are used to investigate genetic variation within a disease population and assess its association with disease risk by comparing the genomes of a disease population with a control population. They often discover single nucleotide polymorphisms (SNPs) which are associated with a higher susceptibility for disease. These SNPs can be located within genes and cause changes to the output of transcription, such as a truncated protein. However, they can also be found within the regulatory regions discussed previously. Many SNPs associated with human disease and found through GWAS are in non-coding regions [Albert and Kruglyak, 2015]. SNPs have been found to be linked to an increase in susceptibility to different cardiovascular diseases which act through changes to enhancer regions and subsequent gene expression changes (for full review see: [AneneNzulu et al., 2021]).

Mutations in several different genes, most in the TGF-beta/BMP signalling pathway, have been discovered to be associated with PAH (Table 1.2) [Southgate et al., 2020]. Many studies focus on rare genetic variants located in and affecting the function of a single gene. However, approximately 75% of iPAH cases cannot be explained by these variants and have another cause. The importance of the contribution of common variation to disease compared to rare mutations has been widely debated [Gibson, 2012] and is beginning to be assessed in PAH. Common variation has a lower effect size and a higher frequency in the population, whereas rare mutations tend to have a large effect size and by definition are found at a much lower frequency. It is likely that the combination of all variation

in the patient's genome, common and rare, alongside environmental and lifestyle factors, will determine the speed of progression of PAH and their disease-state, be that healthy, mild or more severe.

Gene	Entrez ID	References
Rare Pathogenic Variants		
<i>BMPR2</i>	659	[Deng et al., 2000, Lane et al., 2000]
<i>ACVRL1</i> (encodes <i>ALK1</i>)	94	[Trembath et al., 2001]
<i>BMPR1B</i>	658	[Chida et al., 2012]
<i>SMAD9</i> (encodes <i>SMAD8</i>)	4093	[Nasim et al., 2011, Shintani et al., 2009]
<i>SMAD4</i>	4089	
<i>SMAD1</i>	4086	
<i>ENG</i>	2022	[Harrison et al., 2005]
<i>GDF2</i> (encodes <i>BMP9</i>)	2658	[Wang et al., 2016, Gräf et al., 2018]
<i>KCNK3</i>	3777	[Ma et al., 2013]
<i>SOX17</i>	64321	[Zhu et al., 2018, Gräf et al., 2018]
<i>ATP13A3</i>	79572	[Gräf et al., 2018]
<i>AQP1</i>	358	[Gräf et al., 2018]
<i>KLF2</i>	10365	[Eichstaedt et al., 2017]
<i>TBX4</i>	9496	[Kerstjens-Frederikse et al., 2013]
<i>CAV1</i>	857	[Austin et al., 2012]
<i>FIGN</i>	55137	[Puigdevall et al., 2019]
<i>EIF2AK4</i>	440275	[Best et al., 2017]
<i>KCNA5</i>	3741	[Remillard et al., 2007]
Common Pathogenic Variants		
<i>CBLN2</i>	147381	[Germain et al., 2013]
<i>PDE1A/DNAJC10</i>	5136/54431	[Kimura et al. [2017] [Rhodes et al., 2019]
<i>HLA-DPA1/B1</i>	3113/3115	
<i>SOX17</i>	64321	

Table 1.2: Genes which are affected by rare and common deleterious genetic variants and have been associated with PAH.

1.4.1 TGF β Superfamily & BMPR2 Signaling

First discovered in the early 2000s [Deng et al., 2000, Lane et al., 2000], mutations in bone morphogenetic protein receptor 2 (*BMPR2*) are the most common genetic cause of PAH (Table 1.2). They have been found to affect around 75% of hPAH [Soubrier et al., 2013] and between 11% and 40% of iPAH cases [Koehler et al., 2004, Morisaki et al., 2004]. Currently, there have also been mutations found in the following members of the TGF β signalling pathway, and associated proteins: *ACVRL1*, *BMPR1B*, *BMP9*, *mothers against decapentaplegic homologue (SMAD)1/4/9* (Table 1.2). This highlights the relevance of this signalling pathway in PAH and the importance of further study into its mechanisms.

The TGF β superfamily has two parallel arms, bone morphogenetic protein (BMP) signalling and TGF β signalling. These require different ligands and receptors and have downstream signalling pathways which target genes that affect cell function differently. However, both pathways function through phosphorylation of intracellular SMAD transcription factors and the cooperation with cofactor SMAD4. In BMP signalling, *BMPR2* encodes a serine-threonine receptor which, when bound to by a ligand (BMP 2, 4, 6, 7, 9 or 10) forms a complex with a type 1 receptor (*BMPR1A*, *BMPR1B*, *ACVRL1* or *ACVRL2*) whose intracellular domain becomes phosphorylated. This results in a SMAD phosphorylation signal cascade involving SMAD4, which leads to SMAD (1/5/8) complex translocation to the nucleus and target gene transcription (Figure 1.3A). Common target genes are inhibitors of DNA binding genes [Shi and Massagué, 2003] which are important for cell migration, proliferation and apoptosis. BMP signalling is also tissue-specific and therefore, mutations in different receptor and ligand types are found in numerous different human diseases. In TGF β signalling, the TGF β ligand binds to the transforming growth factor β receptor 2 (TGFBR2) which forms a complex with the type 1 receptor *ALK5*. The intracellular domain of this complex becomes phosphorylated. This results in a SMAD phosphorylation signal cascade involving the SMAD (2/3) complex and SMAD4, which leads to translocation to the nucleus and targets specific gene transcription (Figure 1.3B).

It is hypothesized that a loss of function and/or decrease of *BMPR2* and members of this pathway is important in PAH aetiology. However, work has suggested that the effect on the signalling pathway is highly dependent on the location and type of mutation. For example, PAH patients with mutations affecting the cytoplasmic region of *BMPR2* have been found with intact downstream SMAD signalling [Nishihara et al., 2002], whereas other mutations lead to a decrease of the signalling pathway [Yang et al., 2008]. Mutations in the 5' untranslated region of *BMPR2* have also been shown to affect its expression [Pousada et al., 2017]. This suggests the possibility of genetic and/or environmental modulators of this pathway and the need for more research into other genetic and environmental causes of PAH and how they might interact.

Dysfunction of the BMP signalling pathway has been implicated in mechanisms known to contribute to PAH. One example is aberrant EndoMT. It is shown both in-vitro and in-vivo, that a reduction in BMPR2 signalling leads to an increase in EndoMT [Reynolds et al., 2012, Hopper et al., 2016]. It has also been shown that this increase in EndoMT can be ameliorated by stimulating BMPR2 signalling through rapamycin exposure [Ranchoux et al., 2015] or targeting with ligands [Reynolds et al., 2012]. Dysfunction of this pathway has also been linked to apoptosis and proliferation changes seen in pulmonary artery cells in PAH. Endothelial cells with knockdown of *BMPR2* via siRNA have been shown to have increased apoptosis [Teichert-Kuliszewska et al., 2006] and mutations affecting this gene could, therefore, be part of the mechanism that initiates the creation of the anti-apoptotic and pro-proliferative phenotype.

Only 20% of patients with *BMPR2* mutations develop PAH [Evans et al., 2016]. This reduced penetrance suggests a role for unknown disease modifiers or genetic mutations. Mutations have been found in genes that are believed to modify the severity of BMP signalling dysfunction in PAH. One example is in a region reported to affect the expression of *FIGN* which encodes fidgetin [Puigdevall et al., 2019] (Table 1.2). Fidgetin is involved in microtubule-severing ATPase activity and mutations in which are linked to lower folate levels in the cardiovascular system [Wang et al., 2017]. The region discussed and affecting *FIGN* is in linkage disequilibrium with *BMPR2* and therefore inherited with it. It has been hypothesized that *FIGN* can modulate the impact of a *BMPR2* mutation as mutations that increase the levels of fidgetin are known to have a protective role against congenital heart disease [Wang et al., 2017] and could therefore modulate PAH penetrance for patients with mutations affecting *BMPR2*.

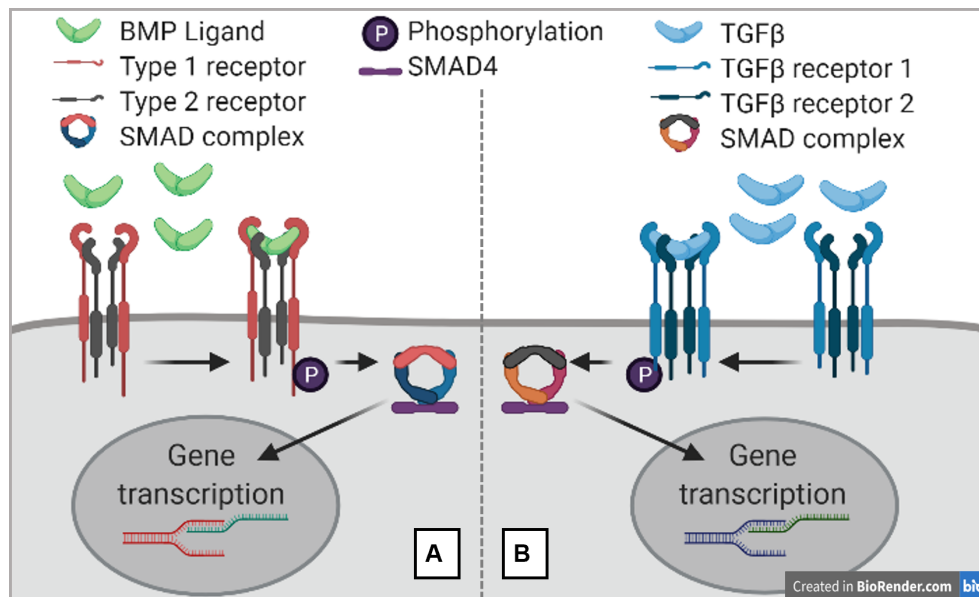


Figure 1.3

The TGF β superfamily. A) The BMP pathway. BMPRII is a serine-threonine type 2 receptor which, when bound to by a ligand (BMPs 2, 4, 6, 7, 9 and 10) forms a complex with a type 1 receptor (BMPRI, BMPRII, ACVRL1 or ACVRL2) whose intracellular domain becomes phosphorylated. This results in a SMAD phosphorylation signal cascade involving the SMAD complex (SMAD 1/5/8) and SMAD 4, which translocates to the nucleus and causes target gene transcription. Bone morphogenetic protein (BMP). Bone morphogenetic protein receptor (BMPR). Mothers against decapentaplegic homologue (SMAD). **B) The TGF β pathway.** The TGF β ligand binds to TGFBR2 which forms a complex with type 1 receptor (ALK5). The intracellular domain of this complex becomes phosphorylated. This results in a SMAD phosphorylation signal cascade involving the SMAD complex (SMAD 2/3) and SMAD 4, which translocates to the nucleus and targets gene transcription. Transforming growth factor β (TGF β). Transforming growth factor β receptor 2 (TGFBR2).

1.4.2 Other Genes Associated with hPAH

Rare mutations in other genes have also been discovered that are thought to be causal in PAH. These genes include *KCNK3*, *CAV1*, *KLF2*, *TBX4*, *KCNA5*, *ATP13A3*, *APQ1* and *SOX17* (Table 1.2). *KCNK3* encodes for a member of the potassium channel sub-family, and it has been found that heterozygous missense variants are disease-causing in several independent families. These mutations cause a loss of function of the potassium channel which is important in the regulation of resting membrane potential and maintaining vascular tone [Ma et al., 2013]. *KCNA5* is another potassium channel protein in which mutations associated with PAH have been found [Remillard et al., 2007]. These genes show the importance of regulation of resting potential and maintaining vascular tone and how the disruption of this can result in PAH. Several transcription factors have also been implicated in PAH aetiology. *KLF2* is a member of the Krüppel-like family of zinc-finger transcription factors and *TBX4* is a member of the T-box gene transcription factor family. Mutations in both these genes have been associated with PAH with *TBX4* mutations

being more common in child-onset disease [Eichstaedt et al., 2017, Kerstjens-Frederikse et al., 2013]. With increased collaboration enabling larger genetic studies, the knowledge of these genes and novel factors will improve and novel pathways will be discovered which could be targeted by therapeutics.

Graf, et al 2018 performed whole genome sequencing of 1038 PAH cases (86.7% iPAH, 5.5% hPAH, 5.7% drug-associated PAH, 2.1% Pulmonary Veno-occlusive Disease (PVOD)) and 6385 controls with other rare diseases and showed a significant association between rare mutations in *ATP13A3*, *AQP1* and *SOX17* (Figure 1.4) and PAH cases when known PAH genes (*BMPR2*, *EIF2AK4*, *ENG*, *ACVRL1*, *TBX4*, *SMAD9*, *KCNK3*) were excluded from the analysis. They showed 19.9% of the PAH cohort had mutations in these known PAH-associated genes and provided an independent validation of the role of *GDF2* (encoding BMP9) in PAH. Familial segregation of mutations in *AQP1* and *SOX17* were also found, which increases the likelihood of these genes being causal in PAH. It was also found that those with mutations in *AQP1* and *SOX17* were significantly younger at diagnosis (*AQP1*, 36.9 ± 14.3 years. *SOX17*, 32.8 ± 16.2 years) when compared with other cases with no mutations in known genes (51.7 ± 16.6 years). *ATP13A3* was found to have several protein-truncating variants which were predicted to affect the protein's function in polyamine transport. Additional work from this study suggested that loss of *ATP13A3* inhibits proliferation and increases apoptosis [Gräf et al., 2018], a possible trigger for the initiation of PAH.

Another recent study also found mutations in *SOX17* are associated with PAH (Figure 1.4). Zhu, et al 2018 performed whole-exome sequencing of PAH–congenital heart diseases (CHD) patients and i/hPAH and identified *SOX17* as a risk-associated gene for both groups. It was estimated that 3.2% of PAH–CHD and 0.7% of PAH cases included in the study possessed rare deleterious mutations affecting *SOX17*. Many of the mutations are located in the DNA binding domain of the Sox17 protein, suggesting a loss of function and a disruption of the transcription factor role of *SOX17* in the aetiology of PAH. Interestingly, 163 rare deleterious mutations in 149 putative *SOX17* target genes, although not reaching genome-wide significance, were also discovered [Zhu et al., 2018], adding further evidence as to the importance of Sox17 in PAH aetiology. This analysis was performed on a very small population size of only 256 patients (15 hPAH cases, 160 iPAH, 61 trios and 20 duos) which was compared to gnomAD whole genome sequencing subjects as population controls and then validated against 1319 European control subjects (unaffected people). To improve on the statistical analysis presented in this research, larger disease populations should be used.

1.4.3 Common Variants associated with PAH

Although rare mutations have historically been focused on for genetic association discoveries, in recent years the role of common variants in disease has been more widely explored and several common variations have been discovered which are associated with PAH. At the time of writing, these include SNPs affecting *CBLN2*, *PDE1A/DNAJC10*, *HLA-DPA1/HLA-DPB1* and *SOX17* (Table 1.2) [Germain et al., 2013, Kimura et al., 2017, Rhodes et al., 2019].

A SNP believed to affect the *CBLN2* gene was discovered through a multinational GWAS study that did not include patients with known *BMPR2* mutation. The locus 18q22.3 (rs2217560) was found to be associated with both hPAH and iPAH (odds ratio – 1.97) and is located 52kb downstream from *CBLN2* which encodes for the cerebellin-2 precursor. It was found that although it is expressed in healthy lung ECs, *CBLN2* had increased mRNA expression in PAECs cultured from PAH patients when compared to controls from lung cancer surgeries. This work did not find evidence of *CBLN2* expression in PSMCs although there are adhesion molecules present on PSMCs which can bind to *CBLN2*. Although the small numbers of patients (625 cases, 1525 healthy controls) recruited for this study may have hampered their ability to discover common variants with a smaller effect size, this study provided the first common variant to be potentially associated with PAH [Germain et al., 2013].

In 2017 a GWAS of Japanese patients discovered a locus on chromosome 2 which is associated with PAH (odds ratio – 5.18). Unlike the study by Germain et al, 2013, this study included those patients who suffered from *BMPR2* mutations to assess how common variations may influence patients with and/or without *BMPR2* mutations. In total, there were 8 patients with known *BMPR2* mutations in the discovery cohort and 3 in validation cohorts. Similarly to the 2013 GWAS, this study was also relatively small with only 44 patients and 2,993 control, again hampering its ability to find common variants with a smaller effect size. A locus affecting *PDE1A/DNAJC10* was discovered. It was also found that the level of PDE1A protein was higher in distal pulmonary artery walls in iPAH patients than controls and previous work has shown an increase in this protein in the PSMCs of PAH patients [Murray et al., 2007]. It is important to note that this finding has not been replicated in more recent GWAS studies. However, this may be due to the different ethnicity investigated in this study, whereas more recent studies are made up of European populations.

In 2019, the largest GWAS investigating PAH to date using data from 11,744 European individuals (2085 PAH patients) was published. To achieve these large numbers, data from 4 individual studies (UK National Institute for Health Research BioResource study (UK NIHRBR), the US PAH Biobank study (PAHB), the French PH Allele-Associated

Risk study (PHAAR), and the King’s University-led British Heart Foundation PAH study (BHFPAH)) were combined. To perform discovery analysis, these four studies were split into two arms based on how the data was collected, whole-genome sequencing (UK NIHRBR) or genome-wide genotyping array (PAHB, PHAAR, BHFPAH), which gave approximately equal power to the two arms. These discovery analyses were then cross-validated using the other arm and any discoveries confirmed with a final meta-analysis containing the four original studies. *HLA-DPB1* encodes the major histocompatibility complex class II DP β –chain and a common variant associated with specific *HLA-DPB1* functional alleles was found to increase the risk for PAH (rs2856830, odds ratio, 1.56). Interestingly, once the disease has developed, individuals possessing a C/C homozygous genotype have increased median survival of 13.50 years [12.07 – >13.50] when compared to those with a T/T homozygous genotype (6.97 years [6.02–8.05]) [Rhodes et al., 2019]. It is so far unclear why this common variation affects survival so remarkably and whether it is a possible target that can be utilised in future therapies. This study also discovered PAH-associated common variants located upstream of the *SOX17* gene; the same gene known to harbour rare mutations which are associated with PAH (see section 1.4.2).

1.4.4 Epigenetic Regulation and PAH

Epigenetic regulation has also been found to affect PAH risk through DNA and histone methylation and acetylation. For example, hypermethylation of an enhancer which is thought to affect *SOD2* has been found to be responsible for tissue-specific *SOD2* deficiency in some hPAH cases [Archer et al., 2010]. The methylation pattern of hPAECs derived from patients with iPAH has been assessed via unsupervised hierarchical clustering [Hautefort et al., 2017]. This analysis showed two clusters which discriminated between controls and patients with iPAH. They included hypermethylation of gene promoters related to lipid metabolism and transport. However, this investigation only focused on gene promoters and not non-coding regions such as enhancers or regions which can control histone modification. To fully understand epigenetic regulation due to methylation in hPAECs from iPAH patients, these regions will also need to be assessed. There have also been advances in non-coding RNA mechanisms and their contribution(s) to PAH physiology (for full review see [Boucherat et al., 2015]). One example is the microRNA family 130/301 [Bertero et al., 2018]. An increase in the expression of members of this microRNA family in hPAECs and hPSMCs has been shown to affect the regulation of the ECM and drove downstream PAH progression.

1.5 *SOX17*

1.5.1 Members of the SOXF Family

The *SOX17* gene is located on chromosome 8 (8q11.23) and encodes the transcription factor Sox17 which is localised to the nucleoplasm of endothelial cells [Thul et al., 2017]. The Sox17 transcription factor is a member of the Sox (SRY box transcription factor) protein family. Sox proteins are important regulators of cell fate and differentiation [Kamachi and Kondoh, 2013]. *SOX17* is in the SoxF subgroup, which also includes *SOX7* and *SOX18* (Figure 1.6). Mouse models suggest that SoxF proteins have some functional redundancy although the complete loss of function of one cannot be compensated for by the other proteins [Kamachi and Kondoh, 2013, Corada et al., 2013, Zhou et al., 2015]. This subgroup has a key role in cardiovascular development [Francois et al., 2010], with Sox17 being important for systemic arterial fate specification and maintenance [Corada et al., 2013, Zhou et al., 2015]. Low Sox17 and higher Sox7 and Sox18 levels are essential for systemic vein EC development and maintenance Duong et al. [2014]. Sox18 alone is thought to be essential for lymphatic EC development and maintenance [François et al., 2008]. Although there has been work published that shows Sox17's role in the systemic vasculature, the specific role of the different SoxF proteins in PAECs and their roles in downstream pathways in this cell type have yet to be investigated.

SoxF proteins contain a highly conserved high mobility group (HMG) domain (Figure 1.4A) which binds to the minor groove of the promoter of target DNA and bends it by 80°. This allows other proteins to bind to the DNA and have their target effects [Palasingam et al., 2009]. Although the HMG domain is conserved and the SOXF factors have very similar DNA binding domains [Lange et al., 2009] (Figure 1.4B), they target different genes and can have effects on mutually exclusive pathways. This occurs through the recruitment of different co-factors. Rare pathogenic mutations have been discovered in PAH patients which are located in this binding domain and are thought to disrupt the transcription function of *SOX17* [Gräf et al., 2018]. The Sox17 protein also contains a transactivation domain which binds to the co-factors such as β -catenin (Figure 1.5). In *Xenopus* cap tissue, it has been shown that a decrease in β -catenin results in the repression of some Sox17 target genes *Hnf1 β* and *Foxa2* highlighting the importance of β -catenin as a co-factor. Interestingly, not all Sox17 target genes were affected in this study, *Foxa1* levels remained unchanged. This is believed to be due to the recruitment of different co-activators by Sox17 [Sinner et al., 2004].

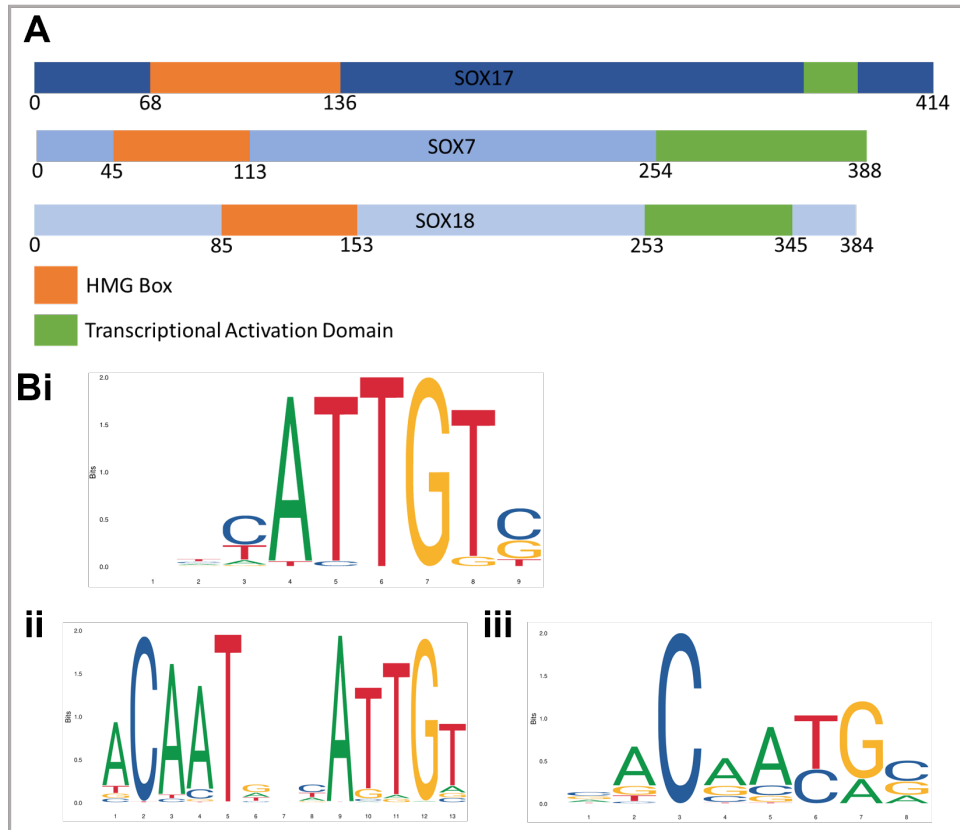


Figure 1.4

The SOXF transcription factor family. A) Basic structure of SOXF proteins. Each SOXF factor has a conserved HMG DNA binding domain and a transcriptional activation domain. Numbers represent amino acid number. B) SOXF consensus binding sequences. The consensus binding sequences for i. SOX17 ii. SOX7 and iii. SOX18. All Obtained from Jaspargenereg.net, high mobility group (HMG)

1.5.2 *SOX17* Location, Expression & Function

Sox17 has a well-documented role in the development of the cardiovascular system. For example, it has been shown that EC deletion of *SOX17* in mouse models leads to embryonic lethality due to underdeveloped arteries and a complete lack of large arteries [Corada et al., 2013]. It has also been shown that conditional deletion of *SOX17* in splanchnic mesenchyme-derivatives, a cell type essential for early cardiovascular development, leads to severe vascular abnormalities and lethality after 3 weeks due to dilated cardiomyopathy [Lange et al., 2014]. These mice developed enlarged pulmonary arteries and biventricular hypertrophy. These studies highlight the importance of *SOX17* in the early development of the cardiovascular system and its vital role in the tissues primarily affect in PH.

In zebrafish, Sox7 and Sox18 regulate the expression of nuclear receptor subfamily 2 group F member 2 (*NR2F2*), an essential transcription factor for venous fate determination [Swift et al., 2014] which prevents arterial cell fate determination [Chen et al., 2012]. When *SOX17* expression is deleted in mouse ECs via tamoxifen-inducible knockdown,

there is an increase in venous EC markers such as *NR2F2* and endomucin, a protein that prevents the migration of ECs. This, coupled with the decrease in arterial markers CXCR4 and Notch4 suggest that arterial fate determination has been affected by the loss of Sox17 [Corada et al., 2013]. It was also shown that Sox17 has an important role in the maintenance of EC arterial identity through increased Notch signalling [Corada et al., 2013]. This shows a continued role of Sox17 after development and throughout adult life and is discussed further in 1.5.2.

In ECs, Sox17 is important in the maintenance of barrier function and proliferation which causes EC regeneration following injury [Liu et al., 2019]. In this study, mouse EC were exposed to endotoxins to induce a hypoxic environment. There are three hypoxia response elements (HRE) in the promoter of *SOX17* which HIF1 α can bind to. This study suggests that binding of HIF1 α to the third HRE induces the expression of *SOX17*, which targets cyclin-E1, a gene whose translated protein increases proliferation and maintains the barrier function of ECs. Sox17 negative ECs had increased endothelial permeability and lower proliferation than controls, both features which are known to be affected in PAH [Vaillancourt et al., 2015]. Other studies have also found evidence that Sox17 induces proliferation through transcriptional targeting of cell cycle genes [Lange et al., 2009]. Sox17 is a proangiogenic regulator through its interaction with vascular endothelial growth factor (VEGF) signalling (see section 1.5.2). Sox17 also promotes EC migration and angiogenesis in mice and human ECs [Yang et al., 2013, Lee et al., 2014].

Multiple genetic studies have predicted both rare and common mutations involving *SOX17* to be associated with PAH [Zhu et al., 2018, Gräf et al., 2018, Rhodes et al., 2019] (Table 1.2, Table 1.3). These studies predict either a decrease in Sox17 levels or a loss of function results in disease. Interestingly, it has previously been shown that altered *SOX17* expression increases vascular fragility in arteries of the brain, increasing the risk of aneurysm [Bilguvar et al., 2008, Foroud et al., 2012], adding to the growing body of evidence for Sox17's role in artery EC function. In PAECs, this could be a cause of endothelial cell injury, one of the first hallmarks of PAH.

The loss of *SOX17* expression in PAECs could be a factor that contributes to this loss of barrier function. Although the roles of Sox17 in cardiovascular development and EC function have been studied in several organisms and cell types, there is little work on its specific role in PAECs especially with regards to PAH. This work needs addressing to discover more about the role of Sox17 in PAH.

1.5.3 Sox17 Related Signalling Pathways

Sox17 is involved in several signalling pathways. One example is the VEGF signalling pathway, an important pathway for stimulating angiogenesis. Sox17 is a positive feedback

regulator of VEGF signalling (Figure 1.5A). VEGF signalling is a complicated process with multiple vascular endothelial growth factor receptor 2 (VEGFR2) ligands which can form different homo/heterodimers or complexes with other surface proteins depending on the specific EC subtype [Simons et al., 2016]. Regardless of the receptor type, activation via VEGF binding results in trans-phosphorylation of tyrosine residues on the intracellular domain of VEGFR2. Which domains become phosphorylated will result in one of several signal cascades in the cells. The two best known to occur in ECs are the phospholipase C γ (PLC γ)-ERK1/2 pathway, which is important for vascular development, and the PI3K-AKT-mTOR pathway which is important in cell survival and regulation of the barrier function. The PLC γ -ERK1/2 pathway involves phosphorylation of VEGFR2 Y1173 which leads to internalization of VEGFR2 into the early endosome antigen 1 (EEA-1) and downstream calcium-dependent signalling pathway activation which affects EC migration and proliferation through activation of ERK1 and ERK2 signalling [Wei et al., 2009, Wythe et al., 2013, Simons et al., 2016] (Figure 1.5A). The PI3K-AKT-mTOR pathway involves the activation of SRC through the phosphorylation of VEGFR2 Y949. One of the downstream effects of this pathway is PASMC vasodilation through activation of endothelial nitric oxide synthase [Jin et al., 2003] (Figure 1.5A). VEGF signalling is regulated in ECs by the delta-like protein (DLL)4-notch pathway. VEGF promotes DLL4 expression which in turn activates notch signalling and suppresses VEGFR2 expression; this is required to stabilize newly formed vessels [Suchting et al., 2007, Thurston et al., 2008].

It has been shown in both mice sprouting ECs or the retina and human umbilical vein endothelial cells (HUVECs) that *SOX17* expression is promoted by VEGF [Kim et al., 2016] and is thought to act as a positive regulator of VEGF signalling in tumor-derived ECs through upregulating VEGFR2 [Yang et al., 2013]. Work by Corada, et al 2013 however disputes this claim as they found no evidence in mouse ECs that VEGF activates *SOX17*. However, this could likely be due to a difference in origin of the cells the studies were completed in, for example tumor and non-tumor and/or mouse and human. VEGF-induced angiogenesis is negatively regulated by notch signalling, which also negatively regulates Sox17. This suggests a role for the loss of Sox17 in disrupted angiogenesis. VEGF plasma levels are increased in severe PAH patients and animal models and VEGF and VEGFR2 expression are high in vascular lesions of PAH patients, highlighting the importance of this pathway in the disease [Tuder et al., 1994, Papaioannou et al., 2009, Selimovic et al., 2009].

Another signalling pathway affected by Sox17 is the canonical Wnt pathway which, in human ECs, controls the expression of target genes such as interleukin 8 (*IL8*), an important angiogenic regulator [Néstor et al., 2005]. Other genes such as *c-myc*, *cyclin-D1*, *VEGF*, *survivin* and *axin2* are also Wnt targets and are involved in cell maintenance, pro-

liferation and survival; important functions in ECs [Bain et al., 2003, Rawadi et al., 2003, Chen et al., 2005]. However, the downstream effects of canonical Wnt signalling are dependent on the organism and cell type and therefore need further study in PAECs. In the canonical Wnt signalling pathway (Figure 1.5B), the frizzled receptors (FZD) modulate phosphorylation of low-density lipoprotein receptor-related protein co-receptor (LPR)5 and LPR6 to form a complex which is located on the plasma membrane of the cell [He et al., 2004]. In the absence of a ligand, this complex does not repress glycogen synthase kinase 3 β (GSK3 β), whose main function in this pathway is to target β -catenin for proteasome-mediated degradation through phosphorylation of serine residues. When the FZD receptor is bound by a canonical Wnt ligand, it represses the activity of GSK3 β which then allows β -catenin to stabilize. β -catenin can then move to the nucleus where it forms a complex with T-cell factor/lymphoid enhancer factor (TEF–LEF) through binding with the TEF–LEF HMG box. Legless family docking proteins (BCL9, BCL9L) and coactivators (PYGO-1, PYGO-2) are also important in this complex and allow it to have its downstream effect on target genes such as *cyclinD1* [Katoh and Katoh, 2007]. The Sox17 protein has a similar HMG binding domain structure to TEF–LEF and is well known to interact with β -catenin. In colon carcinoma cells, Sox17 has been shown to inhibit the canonical Wnt pathway by forming a complex with TEF–LEF and β -catenin and targeting them for GSK3 β -independent degradation. Sox17 can also bind to β -catenin and use it as a co-factor [Sinner et al., 2007]. These different outcomes for β -catenin can change the output of the canonical Wnt signalling pathway and regulation of this pathway is essential for normal development of the cardiovascular system and in adults for the regulation of pulmonary vascular homeostasis through PAEC proliferation, survival and motility [Perez et al., 2009]. This study also found that the activation of both the canonical Wnt pathway and the Wnt/planar cell polarity (Wnt–PCP) pathway was required for the preservation of pulmonary vascular homeostasis. The Wnt–PCP pathway is important for cell migration around injury sites through the activation of small GTPases Rac, Rho and cdc42 [Shulman et al., 1998]. If there is a loss of signalling through the canonical Wnt signalling pathway, it has been hypothesised that ECs will have increased signalling through the Wnt–PCP pathway due to a loss of the balance between these signalling pathways. This is thought to lead to abnormal migration and repair by ECs, causing obstruction of small vessels and possibly leading to PAH [Perez et al., 2014]. It may be the case that if there is a loss of balance of these pathways through inhibition of the canonical Wnt signalling pathway via a loss of Sox17, PAEC proliferation may increase and obliteration of the vessels occur.

Diseases that affect the cardiovascular system have been found to have an association with mutations affecting the Wnt signalling pathway, for example, atherosclerosis and stroke [Dejana, 2010, Marinou et al., 2012]. The canonical Wnt pathway has also been

implicated in PAH. Studies in PAECs have shown that an increase in β -catenin transcriptional activity following BMP2 exposure leads to increased proliferation, survival, and angiogenesis. Evidence suggests that BMP2 inhibits GSK3 β and therefore allows for the increase of stable β -catenin. This study also found that BMP2 can also upregulate the non-canonical Wnt signalling pathway. This pathway does not rely on β -catenin for downstream signalling but uses small GTPases to have its downstream effect [Perez et al., 2009]. It is hypothesised that changes in either the canonical or non-canonical Wnt signalling pathway could contribute to PAH via the disruption of vessel regeneration following injury. The Wnt–PCP signalling pathway alone has also been implicated in PAH. Plexiform lesions, a hallmark of iPAH, have been found to contain elevated levels of Wnt signalling target genes such as Wnt11 and ROCK [Laumanns et al., 2009].

Sox17 is also involved in the maintenance of notch1 signalling (Figure 1.5C), a pathway that is important in arterial fate determination by repressing venous differentiation in developing arteries [Lawson et al., 2001]. It is well characterised in the systemic vasculature but not in the pulmonary vasculature. In mouse intra-aortic hematopoietic cluster cells, Sox17 causes an increase in expression of notch1 through binding to its promoter region [Saito et al., 2018]. Whether this interaction is present in PAECs has not yet been studied. The receptor can be bound by notch ligands jagged (JAG)1 and JAG2 and DLL1, DLL3, and DLL4 [Fleming, 1998]. These ligands participate in short-range communication through contact. This cell-to-cell interaction promotes the double cleavage of the notch receptor extracellularly by A disintegrin and metalloproteinase domain-containing protein 10 (*ADAM10*) and intracellularly by γ -secretase which leads to the release of the notch intracellular domain (NICD). The NICD then moves into the nucleus where it binds to the DNA-binding transcription factor RBPJ. This activates genes that are important in maintaining a normal cell cycle through proliferation, apoptosis and differentiation, for example, *Hes1* and *Runx1* [Saito et al., 2018] (Figure 1.5C). Earlier findings in human HUVECs also suggest notch signalling downregulates Sox17 at the protein level but not at the transcript level [Lee et al., 2014]. These studies highlight the complex relationship between Sox17 and notch1 signalling in different organisms and cell types and that this relationship may ensure correct cell cycle progression and therefore be essential for normal cell function. Therefore, if there is a loss of *SOX17* expression, this may hamper this signalling pathway and have effects on normal cell cycle progression. However, this needs further targeted research in specific cell types of interest.

Notch1 signalling has been shown to have a potential link to PAH progression. Notch1 expression and protein levels have found to be higher in iPAH lungs than in controls [Dabral et al., 2016]. Notch1 knockout in mice is lethal at the embryonic stage with severe cardiovascular abnormalities [Krebs et al., 2000]. This highlights the importance of Notch1 signalling in the development of a functional cardiovascular system. In other

studies, hypoxia-exposed PAECs had higher notch1 expression, suggested to be due to a hypoxia-inducible factor (HIF) binding to HRE in the notch1 promoter. This increase of notch1 caused a decrease in p21 (a cell cycle inhibitor) and an increase in proliferation. A decrease in apoptosis due to increased survivin and Bcl-2 expression was also found to be a result of an increase in notch 1 signalling. These findings were lost with notch1 knockdown in these cells. The study suggests that this could be important in the creation or maintenance of the pro-proliferative, antiapoptotic PAECs seen in PAH [Dabral et al., 2016].

VEGF, WNT and Notch signalling pathways can all have varying downstream effects depending on the cell type they are activated in. Therefore, the effect which the loss of Sox17 has on these pathways in PAECs requires further investigation. This will allow a deeper investigation into the link between Sox17 and PAH.

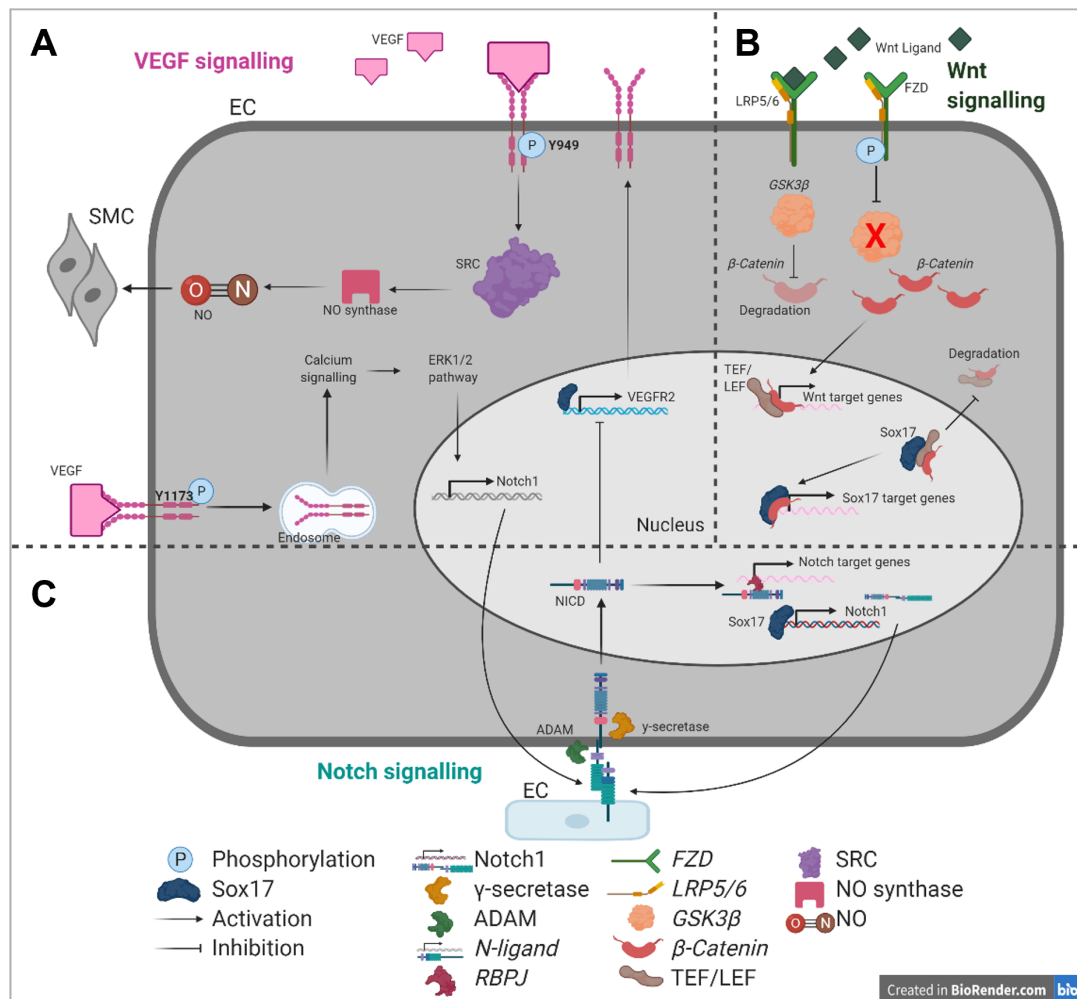


Figure 1.5

***SOX17* related signalling pathways.** endothelial cell (EC). smooth muscle cell (SMC).

A) The VEGF signalling pathway. Activation via VEGF binding results in trans-phosphorylation of tyrosine residues on the intracellular domain of VEGFR2. The PLC γ -ERK1/2 pathway involves phosphorylation of VEGFR2 Y1173 which leads to internalization of VEGFR2 into an endosome and downstream Ca²⁺-dependent activation of ERK1 and ERK2 signalling. The PI3K-AKT-mTOR pathway involves the activation of SRC through the phosphorylation of VEGFR2 Y949. VEGF signalling is regulated by the Dll4-notch pathway. VEGF promotes Dll4 expression activates notch signalling and suppresses VEGFR2 expression by which *SOX17* is potentially also negatively regulated. *SOX17* expression is promoted by VEGF and acts as a positive regulator of VEGF signalling through upregulating VEGFR2. VEGF, vascular endothelial growth factor. VEGFR2, vascular endothelial growth factor receptor 2. PLC γ , Phosphoinositide phospholipase-gamma. ERK1/2, extracellular signal-regulated protein kinase 1/2. PI3K-AKT-mTOR, phosphatidylinositol -3-kinase-serine/threonine protein kinase B-mechanistic target of rapamycin. Delta-like protein (DLL)4. Nitric oxide (NO). **Continued...**

Figure 1.5

Continued... B) The Wnt signalling pathway. FZD modulate phosphorylation of LRP5 and LRP6 to form a plasma membrane complex. In the absence of a ligand, this complex does not repress GSK3 β , which will target β -catenin for proteasome-mediated degradation. When the FZD receptor is bound by a canonical Wnt ligand, it represses the activity of GSK3 β which then allows β -catenin to be stabilised. β -catenin moves to the nucleus where it binds to TEF/LEF. The Sox17 protein interacts with β -catenin and can inhibit the canonical Wnt pathway by forming a complex with TEF/LEF and β -Catenin and targeting them for GSK3 β -independent degradation. Sox17 can also bind to β -catenin and use it as a co-factor. FZD, frizzled receptor. LRP5/6, Low-density lipoprotein receptor-related protein co-receptor (LPR)5/6. Glycogen synthase kinase 3 β (GSK3 β). T-cell factor/lymphoid enhancer factor (TEF–LEF). **C) The Notch1 signalling pathway.** Sox17 causes an increase in expression of notch1 through binding to its promoter region. The receptor is bound by notch ligands Jag1 and Jag2 and DLL-1, DLL3 and DLL44. This promotes the double cleavage of the notch receptor extracellularly by ADAM10 and intracellularly by γ -secretase which leads to the release of the NICD into the nucleus where it binds to the DNA-binding transcription factor RBPJ. Notch signalling can also downregulate Sox17 at the protein level and ADAM, a disintegrin and metalloproteinase domain containing protein 10. Notch intracellular domain (NICD). Jagged (JAG)1/2. Delta-like protein (DLL)1/3/4.

1.5.4 Sox17 & the TGF β Superfamily

As discussed in section 1.4.1, the TGF β superfamily is important in the onset and progression of PAH and multiple genetic aberrations in different members of this superfamily have been found to be associated with an increase in PAH risk (Table 1.2). There is growing evidence for a link between the two arms of this signalling pathway (TGF β signalling and BMP signalling) and Sox17, with a particular focus on pulmonary artery development and common signalling pathways. As discussed in section 1.5.2, Sox17 has an important role in the Notch, VEGF and Wnt signalling pathways. However, the BMP signalling pathway can have opposite effects on these pathways to Sox17. For example, whilst Sox17 has been shown to both promote and be promoted by VEGF signalling (Figure 1.7A), BMP9 signalling has been shown to block VEGF-induced angiogenesis in bovine aortic ECs [Scharpfenecker et al., 2007]. Work by [Lange et al., 2009] in mouse lung endothelial cells shows the role of Sox17 as an effector of activated TGF β signalling, inhibiting downstream cell-cycle inhibitors and activating the cell cycle.

1.5.5 Common Variants Affecting *SOX17*

The GWAS study previously described [Rhodes et al., 2019] highlighted a novel PAH-associated genomic region located 106–200kb upstream of the *SOX17* gene promoter. This region contained two statistically and locationally independent signals, which will be referred to as *SOX17*-signal 1 and *SOX17*-signal 2 (Figure 1.6). *SOX17*-signal 1 contains 4 SNPs that are potentially causal and is located 100–103kb upstream of *SOX17*. The

lead SNP of *SOX17*-signal 1 is rs13266183 (odds ratio, 1.36). *SOX17*-signal 2 is located 106–142kb upstream of *SOX17* and has 31 potentially causal SNPs, with the lead SNP being rs10103692 (odds ratio, 1.80). These SNPs are common in the population tested with 59% of patients homozygous for the risk allele (Table 1.3) of both SNPs compared to 46% of controls. The likelihood for causality of these SNPs, for both *SOX17*-signal 1 and 2, were determined via a Bayesian credible set analysis, which statistically prioritises the most likely variants to be causal from a set of associated variants.

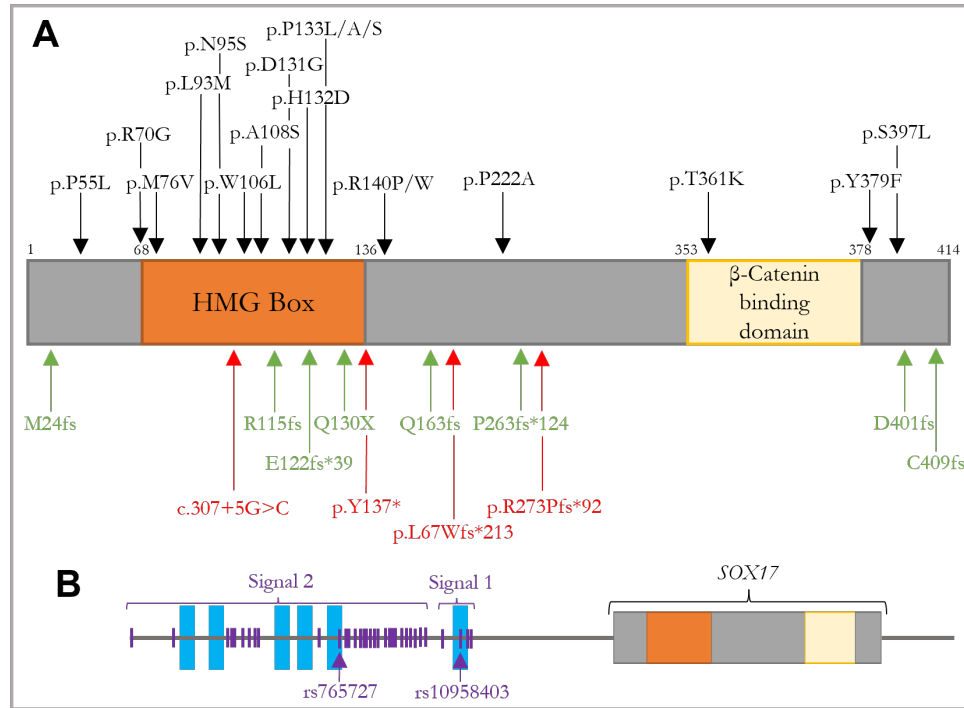


Figure 1.6

***SOX17* genetic variations associated with PAH.** Not to scale. **A) Schematic of the *SOX17* gene showing the location of known PAH associated variants.** Numbers indicate amino acid number. Black Arrows; missense variants. Red arrows; protein truncating variants. Green arrows; likely gene disrupting variants. Adapted from [Wu et al., 2021]. **B) Schematic of the region upstream of the *SOX17* gene.** Showing single nucleotide variants (purple bars & arrows) located in an putative enhancer regions (blue boxes) which may affect *SOX17* expression. [Rhodes et al., 2019].

Mining of publicly available epigenomic datasets of vascular endothelial cells (hPAECs, endothelial progenitor cells and HUVECs) allowed the assessment of the regulatory abilities of these regions (Figure 1.7A). Regions sensitive to DNase digestion are known to be areas of open, accessible chromatin, where transcription factors and other protein complexes can bind and are known as active regulatory regions of the genome. Histone modifications such as histone H3 lysine 4 monomethylation, histone H3 lysine 4 trimethylation and histone H3 lysine 27 acetylation suggest enhancer regions, promoter regions

and areas of active regulatory regions respectively and can be used to identify regions of interest. Using these epigenetic markers, it was found that the SNPs rs10958403 (within *SOX17*-signal 1) and rs765727 (within *SOX17*-signal 2) were in separate putative enhancer regions. Using publicly available chromatin conformation Hi-C data it was shown that both rs10958403 and rs765727 are located in a TAD in which the only gene present is *SOX17*, making it the most likely target for the enhancer region (Figure 1.7A).

Through a luciferase reporter assay (Figure 1.7B), the enhancer activity of the regions surrounding both SNPs was shown. An allelic difference between the risk and non-risk alleles in both SNPs was also highlighted; the non-risk allele having a stronger enhancer effect in both (Table 1.3, Figure 1.7B). To assess the effect of genetic manipulation on the region defined by *SOX17*-signal 1 and to elucidate the connection between this region and *SOX17* expression, CRISPR-inhibition of the signal region was performed. This resulted in downregulation of *SOX17* expression, providing in-vitro evidence that the *SOX17*-signal 1 region does have some effect on *SOX17* expression [Rhodes et al., 2019]. The specific effect of *SOX17*-signal 2 region's enhancer ability on *SOX17* has so far not been demonstrated. If an effect is found, this would show that both signals' enhancer function is targeted towards *SOX17* expression and would highlight variants located in a non-coding region that affect the expression of a PAH-linked gene.

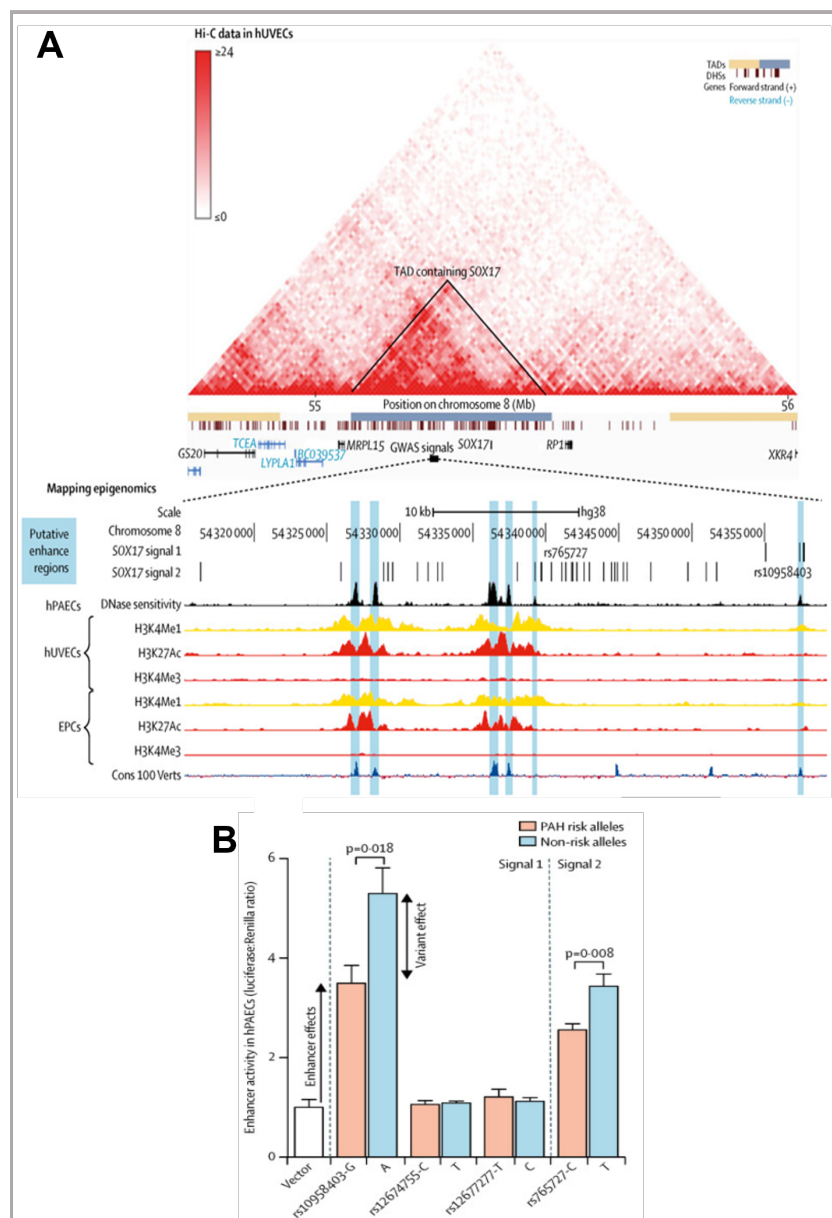


Figure 1.7

Common *SOX17* genetic variation associated with PAH. **A) Epigenetic mapping of the region surrounding GWAS hits.** GWAS hits are located upstream of *SOX17*. The black box highlights the region containing the GWAS hits and below is this region zoomed in. The hits corresponding to *SOX17*-signal 1 and 2 and their location are shown by black bars. Publicly available data showing DNase 1 sensitivity (showing areas of open and accessible chromatin) from human pulmonary arterial endothelial cells (hPAECs), histone 3 lysine 4 mono- (H3K4Me1, found in enhancer regions) and tri-methylation (H3K4Me3, found in promoter regions) and histone 3 lysine 27 acetylation (H3K27Ac, found in active regulatory regions) from both human pulmonary venous endothelial (hUVECs) and epithelial progenitor cells (EPCs) is shown. Putative enhancer regions suggested by this data are highlighted in light blue. All the GWAS signals are in a topologically associated domain which also contains *SOX17*. **B) Luciferase enhancer assay.** A luciferase enhancer assay in hPAECs showing the enhancer activity of a control vector (white), the risk allele (orange) and the non-risk alleles (blue) of each locus of *SOX17*-signal 1 (rs10958403, rs12674755, rs12677277) and one locus of *SOX17*-signal 2 (rs765727). Luciferase:renilla ratio of both alleles at rs10958403 (*SOX17*-signal 1) and rs765727 (*SOX17*-signal 2) show increased enhancer activity when compared to the control vector. A variant specific enhancer effect is also shown between the non-risk and risk allele of these loci. Figures taken from [Rhodes et al., 2019].

Name	Lead SNP	Location (chr:pos)	PAH Risk Allele	Frequency	Non-Risk Allele	Frequency
Signal 1	rs10958403	8:55269940	G	0.74(0.75)	A	0.26(0.25)
Signal 2	rs765727	8:55251803	C	0.91(0.87)	T	0.09(0.13)

Table 1.3: Common variants associated with PAH. Frequency shows the frequency of the SNP in GWAS population (general population) [Rhodes et al., 2019]. chromosome (Chr). position (pos).

There is a growing body of evidence of a causal link between *SOX17* and PAH with both rare mutations and common variants having been discovered which affect *SOX17* (Figure 1.4 and 1.5) and are over-represented in PAH cohorts. Additionally, the rare mutations analysis showed the mutations to have a familial inheritance pattern [Gräf et al., 2018] which suggests a causal role for *SOX17*. The mechanism of *SOX17* dysfunction in PAH is not understood and shows the need for more investigation of *SOX17* in PAH.

1.6 Summary

In summary, hPAEC injury is thought to be a key factor in PAH. The cause of this injury is unknown and could be attributed to both environmental and genetic factors. A recent GWAS of PAH has found SNPs located in a putative enhancer upstream of *SOX17*, an endothelial TF. *SOX17* has many known roles in PAH-relevant functions and its dysfunction could lead to PAH development and progression. However, this requires further investigation. Furthermore, current PAH therapies have disadvantages for the patient and often do not reverse symptoms. In-silico analysis of *SOX17* dysfunction in hPAECs could lead to novel compounds being discovered which can improve symptoms and outcomes for patients with PAH.

1.7 Hypotheses

Genetic variation in PAH causes reduced *SOX17* signalling and pathological endothelial dysfunction. Specifically:

1. There is allele-specific transcription factor binding at a putative enhancer region upstream of *SOX17* and variation in this region affects *SOX17* expression.
2. Changes in *SOX17* expression contribute to PAH risk through downstream pathway and subsequent endothelial cell changes.
3. Compounds that correct downstream effects of *SOX17* dysfunction will reverse the morphological, functional and genetic effects of genetic manipulation of *SOX17*.

1.8 Aims

1. To investigate the binding of transcription factors to rs765727 (*SOX17*-signal 2).
2. To investigate the effect of a putative enhancer surrounding *SOX17*-signal 2 on *SOX17* expression.
3. To investigate the role of Sox17 in hPAECs function and signalling pathways.
4. To assess drug compounds for correcting Sox17 dysfunction.

2 Methods

2.1 Declaration of Responsibilities

To analyse the effect of *SOX17* knockdown on hPAEC proliferation, apoptosis and cell viability, members of my lab (Eleni Vasilaki, EV. Yookyee Wu, YW) performed MTT, Caspase-Glo 3/7 and CellTitre-Glo assays respectively. I analysed the raw data from these experiments.

To analyse the effect of knockdown of *SOX17* by siRNA on genes and proteins involved in the extracellular matrix, Liliana Som (LS) performed quantitative-PCR (qPCR) on samples from *SOX17*-siRNA experiments. I analysed the raw data from these experiments.

To analyse the differential expression patterns which occur due to CRISPR inhibition (CRISPR–INH) of *SOX17*-signal 1, members of my lab (Chris Rhodes, CR. Jurjan Aman, JA) performed RNA-sequencing of CRISPR-genetically modified hPAECs, the LOG fold change genetic data from which was used in chapter 6.

The remaining data was produced and analysed by myself.

2.2 Materials

A list of the all the materials, manufacturers and catalogue codes is provided in supplemental table 7.1.

2.3 Cell Culture

Human embryonic kidney 293FT (HEK293FT) (Invitrogen, US) and hPAECs (PromoCell GmbH, Germany and Lonza, USA) were cultured at 37°C in 5% CO₂. HEK293FTs were cultured in Dulbecco’s Modified Eagle Medium (DMEM, Sigma Aldrich) supplemented with 10% fetal bovine serum (FBS) (v/v), 1% L–Glutamate, 1% Sodium pyruvate and 1% essential amino acids (all Sigma Aldrich, all (v/v)). Fibronectin (Sigma Aldrich) was used as a seeding agent for hPAECs which were used at passage 4-8 and cultured in 10% FBS (v/v) endothelial growth medium 2 (EGM2), (PromoCell or Lonza), with additional growth factor and antibiotic/antimycotic supplements provided by the manufacturer.

2.4 Primer & Probe Design

All oligonucleotide primers and probes were ordered from Life Technologies Limited (UK) and resuspended according to the manufacturer’s instructions.

Biotin-labelled and ‘cold’ (unlabeled) probes for the EMSA (Table 2.1) were designed based on the 21 base pairs centered on rs765727 with either the non-risk (T) or risk (C)

allele as the 11th base. Competitor probes to investigate transcription factor binding to the target probes (Table 2.2) were designed based on replacing the 6/8 base pairs at the center with the consensus sequence of the transcription factor in question.

Allele	FWD Probe (5'-3')	REV Probe (5'-3')
Biotin-labelled Probes		
C	-ACTGGCCCAGCTCAAGTTACT-bio	-AGTAACTTGAGCTGGGCCAGT-bio
T	-ACTGGCCCAGTTCAAGTTACT-bio	-AGTAACTTGAAGTGGGCCAGT-bio
Unlabeled (cold) Probes		
C	-ACTGGCCCAGCTCAAGTTACT-	-AGTAACTTGAGCTGGGCCAGT-
T	-ACTGGCCCAGTTCAAGTTACT-	-AGTAACTTGAAGTGGGCCAGT-

Table 2.1: EMSA biotin (-bio) labelled and unlabeled probe sequences.

Transcription Factor	FWD Probe (5'-3')	REV Probe (5'-3')
ZFX	-ACTGGCTCAAGTTAAGTTACT-	-AGTAACTTAACTTGAGCCAGT-
LIN54	-ACTGGCGCCAGTTAGTTACT-	-AGTAACTAACTGGGCGCCAGT-
ROR-alpha2	-ACTGGCCAGGTCAAAGTTACT-	-AGTAACTTTGACCTGGCCAGT-
RAR	-ACTGGCAGTTCAAGAGTTACT-	-AGTAACTCTTGAAGTGCCAGT-

Table 2.2: Transcription factor competition EMSA probes. FWD, forward probe. REV, reverse probe.

CRISPR–INH guides for inhibiting *SOX17*-signal 2 (Table 2.3) were designed using Tefor-CRISPOR software (version 4.7).

Target	FWD Guide (5'-3')	REV Guide (5'-3')
SOX17 promoter	-CACCGTCTGGTCTACAGCGTACCC-	-AAACGGGTACGCTGTAGACCAGAC-
rs765727 guide 1	- CACCGCTGGCCCAGCTCAAGTTACT -	- AAACAGTAACTTGAGCTGGGCCAGC -
rs765727 guide 2	- CACCGGAGGCCAAGTAACTTGAGC -	- AAACGCTCAAGTTACTTGGCCTCC -
rs765727 guide 3	- CACCGAGGCCAAGTAACTTGAGCT -	- AAACAGCTCAAGTTACTTGGCCTC -
rs765727 guide 4	- CACCGCCTGTGGGTGGATGACTAAC -	- AAACGTTAGTCATCCACCCACAGGC -

Table 2.3: CRISPR guide RNA sequences. FWD, forward guide. REV, reverse guide.

Primers for use in q-PCR (Table 2.5) were designed using PrimerBlast (NCBI, USA) using default parameters. The primer efficiency was tested for all qPCR primers by performing a qPCR on a serial dilution of cDNA and inputting the -slope of the log(DNA copy number) and Ct value into the ThermoFisher qPCR calculator efficiency calculator to obtain the efficiency percentage. Only primers with an efficiency percentage of between

90 - 110% were used for qPCR. This ensures that all qPCR primers for genes of interest have a similar efficiency to the housekeeping gene (*Actβ*) and can therefore be compared.

Gene	FWD Primer (5'-3')	REV Primer (5'-3')
<i>SOX17</i>	-GGACCGCACGGAATTTGAAC-	-GGACACCACCGAGGAAATGG-
<i>MRPL15</i>	-GCGGATCCTGCCAAATTTCC-	-AACTCTGCTTCCTTGGACGG-
<i>TMEM68</i>	-GCCGTTTGGCATGGTTATGA-	-GGGTGCGCTAAATAGGTCCG-
<i>ADAMTS12</i>	-TTGCTTTGAAGGCGGCAAC-	-CCCAATGTCAACATAACCAGATCC-
<i>MMP17</i>	-CCTCAGCCTGGGAGTGGAGTG-	-GAGCAGCGTGGGGTTTTTCAT-
<i>LAMB3</i>	-CAAACATGCGGTGGTGATGT-	-GTGTAGTCAGACGAGCACTCA-
<i>COL4A6</i>	-GCCACTTCGTCCACCTACTC-	-AGGCACAACGTAACCAGGAG-
<i>BMP9</i>	-GCGGGTAAACTTCGAGGACA-	-CTGCACGATAGCGTGTTTCG-
<i>BMPR2</i>	-GGCTGAAC TTATGATGATTTGGGAA-	-CACGCCTATTATGTGACAGGTTGC-
<i>CYCIN E1</i>	-AGCCAGCCTTGGGACAATAA-	-TGGGTAAACCCGGTCATCAT-
<i>NOTCH 1</i>	-GTCAACGCCGTAGATGACC-	-TTGTTAGCCCCGTTCTTCAG-
<i>NOTCH 4</i>	-GAGGACAGCATTGGTCTCAAGG-	-CAACTCCATCCTCATCAACTTCTG-
<i>CDH5</i>	-CTTCACCCAGACCAAGTACACA-	-AATGGTGAAAGCGTCCTGGT-
<i>SOX7</i>	-TGTCGGATGGACAATCGCC-	-TTCCACGACTTTCCCAGCA-
<i>SOX18</i>	-TTCATGGTGTGGGCAAAGGA-	-TTCAGCTCCTTCCACGCTTT-
<i>ACTB1</i>	-GCACCACACCTTCTACAATGA -	-GTCATCTTCTCGCGGTTGGC-

Table 2.4: qPCR gene primers. FWD, forward primer. REV, reverse primer.

2.5 Electromobility Shift Assay

2.5.1 In-Silico Analysis of TF binding at *SOX17*-signal 2

To investigate TF whose binding may be affected by rs765727 (chr8:55251803, hg19), in-silico analysis was used. The CIS-BP database (Build 1.02. Dec 2018), PROMO (TRANSFAC, version 8.3. Dec 2018) and ConSite (Dec 2018) were used to assess differential binding of TFs to a sequence containing the risk (C) or non-risk (T) allele. Those with differential binding were assessed for expression in cultured hPAECs using RNAseq (Unpublished data, Christopher Rhodes, Jurjan Aman). TFs with detectable expression in hPAECs were selected for further investigations.

2.5.2 Electromobility Shift Assay Protocol

To investigate the differential binding of TFs to the risk (C) and non-risk (T) allele present at rs765727, an EMSA was performed. Nucleic proteins were extracted from hPAECs using the NE-PER Nuclear and Cytoplasmic Extraction Reagents kit (ThermoScientific) following the manufacturer's instructions. Proteins were concentrated using an Amicon centrifugal filter (300 Kd NMWL, MERCK, Germany) and protein concentration was assessed using a Pierce Coomassie (Bradford) protein assay kit (ThermoScientific) according to the manufacturer's instructions.

Biotin-labelled probes (bio-probes), cold-probes and transcription factor cold-probes (table 2.1, table 2.2) were annealed with 1x annealing buffer (10x: TRIS [100 mM], EDTA [10 mM], NaCl [500 mM], all Sigma Aldrich) by heating for 5 minutes at 95°C followed by gentle cooling at room temperature for two hours. Competition EMSA were performed using the LightShift Chemiluminescent EMSA protocol (ThermoScientific) and optimisation discussed in section 3.2.1 and with both the risk bio-probe and the non-risk bio-probe and their corresponding specific and non-specific competition probes. A specific competition probe is unlabeled and has the same sequence as the bio-probe in question whereas the non-specific competitor probe has a different sequence as the bio-probe and is also unlabeled. The specific competition probe for the (T) allele was used as the non-specific probe for the (C) allele and vice versa.

To further assess the binding of specific transcription factors to the target sequences, a TF competition EMSA was used. Transcription factor competition cold probes (Table 2.2) were designed based on replacing the 6/8 base pairs at the center of the unlabeled probe with the consensus binding sequence of the transcription factor in question. The consensus binding sequences of these TFs were found through mining the publicly available database JASPAR (jaspar.genereg.net).

2.5.3 Supershift EMSA assay

To investigate which transcription factor binds to rs765727, a supershift assay was used. The methods, materials and reagents discussed in section 2.5.2 were utilised with the addition of a RAR-related orphan receptor alpha (ROR-*alpha*) antibody (sc-518081 X, Santa Cruz, USA) or rabbit IgG (sc-2025, Santa Cruz) as a control.

2.5.4 EMSA and Supershift Imaging

All images for EMSA and supershift EMSA were taken on an ChemiDoc MP imaging system (Universal Hood 3, BioRad, US) using the ImageLab software (BioRad).

2.6 CRISPR-Inhibition

2.6.1 Inhibition of putative upstream enhancer of *SOX17*

To inhibit the putative enhancer upstream of *SOX17*, CRISPR-INH was used. Guides (Table 2.3) were cloned into a lentiCRISPRi vector (Addgene #118154), transformed into Stbl3 E. coli and grown on ampicillin positive LB-agar (Sigma Aldrich) plates for 24 hours. Two colonies were selected, and plasmid DNA extracted using the GenElute Plasmid MiniPrep Kit (Sigma Aldrich). After sequence verification using Genewiz (Leipzig, Germany) universal U6 primer, colonies which contained the lentiCRISPRi vector were used for Pureyield Plasmid Midiprep (Promega, USA).

Co-transfection of the purified plasmids into 70% confluent HEK293FT cells was performed with PEIpro transfection reagent (Polyplus Transfection, France) according to the manufacturer's instructions. The purified plasmids were transfected with lentiviral packaging plasmids pMDLg/pRRe, pRSV-Rev and pMD2.G (Addgene #12251, #12253 and #12259). Sequences targeting blue fluorescent protein (BFP) (Addgene, #118157) or green fluorescent protein (GFP) (Addgene #118158) cloned into a lentiCRISPRi vector (Addgene #11815) were used as negative controls. A guide targeting the promoter of *SOX17* (Table 2.3) cloned into a lentiCRISPRi vector (Addgene #11815) was used as a positive control. 96 hours after transfection, lentiviral particles were harvested using Lenti-X concentrator (Clontech, USA) following the manufacturer's protocol.

hPAECs plated in 6-well plates and at over 80% confluency were transduced with 10 μ l of concentrated lentiviral particles and, after 48 hours antibiotic selection was performed with blasticidin [10 μ g/ml] (ThermoFisher) for 72 hours. Following this, the blasticidin was removed and complete EGM2 was added for a minimum of 2 hours before continuing onto mRNA extraction using the RNeasy Plus kit (Qiagen, UK) according to the manufacturer's instructions.

2.6.2 Relative Expression of *SOX17*

To investigate the change in expression of *SOX17* and nearby genes *mitochondrial ribosomal protein L15 (MRPL15)* and *transmembrane protein 68 (TMEM68)*, reverse transcription-PCR (RT-PCR) and qPCR were performed with *actin beta (ACTB)* used as a reference gene. Primers are shown in table 2.5. RT-PCR was performed using Multiscribe Reverse Transcriptase (Applied Biosystems, ThermoFisher) and the resulting complimentary DNA (cDNA) used for qPCR using PowerSYBRgreen mastermix (Applied Biosystems, ThermoFisher). The relative expression of *SOX17*, *MRPL15* and *TMEM68* was calculated as

$$2^{-(\delta CT[gene] - \delta CT[ACTB])} \quad (1)$$

and compared to cells transfected with either a BFP or GFP control vector. δ CT refers to the change in cycle threshold of either *ACTB* or the target gene.

2.7 siRNA

2.7.1 siRNA of SOXF Factors

siRNA experiments were performed using silencer select siRNA (ThermoFisher) targeting *SOX17*, *SOX7* and *SOX18* (#s34626, #s38029 and #s28927 respectively). Comparable control siRNA was also used and included a negative control of scrambled-siRNA (#ASO2FOQH, ThermoFisher) and a positive control targeting *GAPDH* (#ASO2FLIC,

ThermoFisher). hPAECs were seeded in fibronectin ([10 $\mu\text{g}/\text{ml}$]Sigma Aldrich) coated 6 well plates in EGM2 media (10% (v/v) FBS) and left to grow for 24 hours until confluency reached 95%.

On the day of the experiment, for each well of a 6 well plate, 9 μl of Lipofectamine RNAiMAX (ThermoFisher) was diluted in 150 μl OPTIMEM medium (PromoCell). 3 μl (30pmol) of the appropriate siRNA (10 μM) was diluted in 150 μl of OPTIMEM medium. 150 μl of the Lipofectamine RNAiMAX/OPTIMEM mix and 150 μl of the siRNA/OPTIMEM mix were combined and incubated at room temperature for 5 minutes. During this time, the EGM2 (10% (v/v) FBS) was removed from the wells and 1ml phosphate buffered saline (PBS) (Sigma Aldrich) per well was used to wash the cells. 1ml of OPTIMEM medium and 250 μl of the appropriate siRNA was added to each well. Wells which contained only cells and OPTIMEM medium were also created as used as negative controls.

6 hours after the addition of siRNA, wells were washed with 1ml PBS and 1ml of EGM2 (10% (v/v) FBS) was added. 24 hours later, this was removed, the wells washed with 1ml PBS and 1ml EGM2 (0.2% (v/v) FBS) was added to the cells which were incubated overnight to quiesce and synchronise the cells for maximal response to stimuli in following protocols. At 48 hours after siRNA exposure, transfected cells were either used for RNA extraction and RT-qPCR (section 2.6.2, whole cell protein extraction and western blotting (Section 2.7.3) or for functional experiments (Section 2.8).

2.7.2 Relative Expression of SOXF Factors

To investigate the change in mRNA levels of *SOX17*, *SOX7* and *SOX18*, RTPCR and qPCR were performed with *ACT β* used as a reference gene. Primers are shown in table 2.5 and RNA extraction, RT-PCR and qPCR was performed according to section 2.6.2. However, the relative expression of all genes was calculated as previously described (Section 2.6) and compared to the positive siRNA control (#ASO2FLIC).

2.7.3 Protein Quantification by Immunoblotting

To assess the levels of SoxF protein following siRNA transduction, total protein was extracted from cells using RIPA buffer (10X, Sigma) supplemented with protease and phosphatase inhibitor cocktail (ThermoFisher). Proteins were diluted 1:10 and quantified using a Pierce Coomassie (Bradford) assay (ThermoFisher) according to the manufacturer's instructions. Proteins were denatured for 5 minutes at 95°C in Laemmli SDS sample buffer (ThermoFisher) before separating on a NuPAGE 4 to 12% (w/v), Nis-TRIS mini protein gel (ThermoFisher) in NuPAGE MES-SDS running buffer (1X, ThermoFisher). Dry transfer was completed using a trans-blot turbo mini PVDF transfer pack and a

trans-blot turbo transfer system (both Biorad, USA). Membranes were then blocked in 5% (w/v) milk (Sigma) in 1x TBST (ThermoFisher) for 1 hour before incubating in primary antibodies (Table 2.5) overnight at 4°C. The membranes were washed three times for 15 minutes in 1x TBST before incubating for 1 hour at room temperature in the relevant secondary antibody (Table 2.5). Membranes were then washed three times for 10 minutes in 1x TBST before adding 750 μ l Immobilon Crescendo Western HRP substrate (MERCK) and imaging on an ChemiDoc MP imaging system (Universal Hood 3, BioRad, US) using the ImageLab software (BioRad). Protein quantification analysis was performed in ImageJ (Version 1.52a, ImageJ, USA).

Protein	Experiment	Antibody Number	Company	Concentration
Sox17	Western blotting	ab224637	Abcam	1:500
Sox7	Western blotting	AF2766	RnD Systems	1:500
Sox18	Western blotting	SC-166025	Santa Cruz	1:500
ROR- α	Western blotting	sc-518081	Santa Cruz	1:100
β -Actin	Western blotting	SC-47778	Santa Cruz	1:250
Vinculin	Western blotting	V9131	Sigma	1:200
Anti-Mouse	Western blotting	ab6728	Abcam	1:2000
Anti-Goat	Western blotting	ab6741	Abcam	1:2000
Anti-Rabbit	Western blotting	A6154	Sigma	1:5000
ROR-alpha	EMSA, ChIP	sc-518081X	Santa Cruz	1 μ g, 4 μ g

Table 2.5: Antibodies

2.8 Functional Assays

To assess the effect of siRNA-*SOX17* on hPAEC function, assays investigating proliferation, apoptosis, cell viability and barrier function were used and are discussed below. These assays were chosen as they investigate functions which are known to be impacted in PAH (discussed in section 1.2.2) and have been shown to be affected by loss of *SOX17* expression in other cell types in previous work (Section 1.5.2). The MTT, caspase 3/7 and CellTitre-Glo assays were carried out by EV and YW and the results were analysed by myself.

2.8.1 MTT Assay

To assess proliferation of cells, an MTT assay was used. This colorimetric assay measures cellular metabolism to give a readout on proliferation. In the reaction, yellow tetrazolium salt (MTT) is converted to purple formazan crystals by metabolically active cells. These crystals are then dissolved and the change in colour measured on a plate reader. Here, cells were plated in EGM2 (10% (v/v) FBS) in 96-well plates at 10,000/well and allowed

to grow overnight before commencing with siRNA as described in section 2.7.1. Following siRNA exposure, VEGF [150 ng/ml] was added for 24 hours. On the day of the MTT, the EGM2 was replaced and blank wells were created. 10 μ l of MTT [5 mg/ml] reagent was added and the cells incubated for 4 hours. Following this, 100 μ l of MTT detergent (94.65% isopropanol, 5% NP40, 0.35% HCL (v/v) [1 M]) was added to each well and the plate shaken for 10 minutes. A plate reader was then used to test absorbance at 570nm.

2.8.2 Caspase 3/7 Assay

To assess apoptosis of cells, the Promega Caspase-Glo 3/7 assay was used according to the manufacturer's instructions. This assay measures both caspase-3 and caspase-7 levels in cells as a readout of caspase or apoptosis activity. Cells were plated in EGM2 (10% (v/v) FBS) in 96-well plates at 15,000/well and allowed to grow overnight before commencing with siRNA as described in section 2.7.1. Following siRNA exposure, lipopolysaccharide (LPS) [10 μ g/ml] or TNF- α [10 ng/ml] was added for 24 hours. Following this, the caspase assay was performed.

2.8.3 CellTitre-Glo Assay

To assess viability of cells, the CellTitre-Glo cell viability kit (Promega) was used according to the manufacturer's instructions. This assay measures ATP as a indicator of cell viability and produces a luminescent readout. Here, cells were plated in EGM2 (10% (v/v) FBS) in 96-well plates at 15,000/well and allowed to grow overnight before commencing with siRNA as described in section 2.7.1. Following siRNA exposure, LPS [10 μ g/ml] or TNF- α [10 ng/ml] was added for 24hours. Following this, the CellTitre-Glo assay was performed.

2.8.4 Boyden Chamber Permeability Assay

To assess permeability, a Boyden chamber assay was used. Cells which had undergone siRNA-*SOX17* were plated into corning 24-well inserts (1 μ m) (Corning, US) and left to acclimatise for 24 hours. The Boyden chamber was then started with the addition of 40 kDa FITC-dextran (ThermoFisher) to the insert. After 1, 2, 3, 4 and 24 hours, 100 μ l of EGM2 in duplicate was taken from the bottom chamber and placed into a 96 well plate and the fluorescence measured with a plate reader. Blank wells, of only EGM2, were also measured and subtracted from the other measurements.

2.9 RNA Sequencing

To assess whole transcriptomic effects of knockdown of SOXFs by siRNA and CRISPR-INH of *SOX17*-signal 2, RNA sequencing (RNAseq) was carried out by the Imperial College

BRC Genomics Facility. To obtain transcript abundance values, Salmon software (version 1.2.1, <https://combine-lab.github.io/salmon>) and GENCODE reference transcriptome (release 28, <https://www.gencodegenes.org/human/>) was used to analyse fastq files [Patro et al., 2017]. These were converted to gene expression data using the Bioconductor package tximport [Soneson et al., 2016] (version 1.2.5042, R studio, US) [Team, 2021]. For *SOX17*-signal 1, this step was carried out by CR and JA. For siRNA samples, gene expression was compared between the negative control condition and the relevant genetic manipulation condition with (experimental batches, donor) as covariates using the R package EdgeR [Robinson et al., 2010]. Principal component analysis (PCA) to identify potential sources of global variation in the dataset (PCs 1-3 included as covariates to adjust for sample/experimental variability). For CRISPR–INH samples, gene expression was compared to the BFP control. and correction using FDR method was completed in (R, R studio) [Team, 2021]. P-values and LogFC were plotted using the EnhancedVolcano package [Blighe et al., 2018] in R studio (version 1.2.5042, RStudio Inc, USA).

2.9.1 Gene Ontology Analysis

To assess the large-scale gene ontology and pathway changes resulting from either siRNA of *SOX17* or CRISPR–INH of *SOX17*-signal 2, over-representation analysis was performed using WEB-based Gene SeTAnaLysis Toolkit (Webgestalt). For siRNA samples, genes were only included if they met an FDR cut off ($q < 0.05$) in each condition when compared to the negative control experiments. Due to the low sample size of the CRISPR–INH analysis, a cut off of $p < 0.05$ was used instead of $q < 0.05$ as an exploratory analysis. The method of interest used was over-representation analysis using the functional database category gene ontology and sub-categories biological processes and molecular function.

2.10 The Connectivity Map: prediction of rescue compounds from RNA signatures

2.10.1 RNA-Sequencing of Genetically Modified hPAECs

To analyse the differential expression patterns which occur when *SOX17* expression is manipulated, members of my lab (CR, JA) performed RNA-sequencing of genetically modified hPAECs and all work described in this section was performed by CR and JA. Guides (Table 2.6) targeting the *SOX17* promoter (chr8:55,370,524-55,370,543) and a SNP (chr8: 55,235,000- 55,270,000) in a putative upstream enhancer (rs10958403, *SOX17*-signal 1) were inserted into either a lentiCRISPRi vector coupled with a KRAB inhibitor (Lenti-KRAB, Addgene #118154) or a SAM enhancing element (Lenti-SAM, Addgene #89567) to either inhibit or activate the target area respectively. Three test conditions

Target	FWD Guide ('5-3')	REV Guide ('5-3')
SOX17 promoter	-CACCGTCTGGTCTACAGCGTACCC-	-AAACGGGTACGCTGTAGACCAGAC-
<i>SOX17</i> -signal 1	-CACCGCTGGCCCAGCTCAAGTTACT-	-AAACAGTAACTTGAGCTGGGCCAGC-

Table 2.6: CRISPR guides for *SOX17*-signal 1 and *SOX17* promoter. FWD, forward guide. REV, reverse guide.

were produced; *SOX17*-promoter repression, *SOX17*-promoter activation and *SOX17*-signal 1 repression. A vector containing a guide targeting BFP (Addgene, #118157) was used as a negative control. Transfection, transduction, RNA extraction and RTPCR were completed as outlined in section 2.5.1 and 2.5.2.

RNAseq was carried out by the Imperial BRC Genomics Facility as discussed in section 2.8. Gene expression was compared between BFP condition and the relevant genetic manipulation condition using the R package edgeR. For example, *SOX17*-promoter activation expression was compared to BFP-Lenti-SAM expression reads to assess changes in expression levels.

2.10.2 In-Silico Analysis of Genetically Modified *SOX17* RNA-Sequencing Data

All further analysis and experiments detailed in the following sections of 2.10 were carried out by me (RW) using Microsoft Excel (Version 1908, Microsoft Office, Microsoft, USA). Lists of the genes which were most differentially expressed between the genetically modified condition (*SOX17*-promoter repression and activation and *SOX17*-signal 1 repression) and the BFP control were created. Only genes which had (log)fold change of over 0.25 (or below -0.25) and $p < 0.05$ were used. False discovery rate (FDR)-corrected P-values could not be used due to the high number of genes tested (12704) and low number of repeats of each condition (3-6) decreasing the power of these values. Gene names were changed to HUGO-approved names using Genenames.org [Yates et al., 2017] before compiling lists of the top up-regulated and down-regulated genes for each of the three conditions, resulting in 6 lists in total (Tables 7.2, 7.3 and 7.4).

2.10.3 Using The Connectivity Map to Predict Compounds to Reverse *SOX17* Dysfunction

To discover compounds which could be repurposed for treatment of *SOX17* dysfunction, the connectivity map (CMap) was used [Subramanian et al., 2017]. The lists (described in section 2.10.2) were used to create queries for the three conditions, *SOX17*-promoter activation, *SOX17*-promoter repression, and *SOX17*-signal 1 repression. Candidate compounds were selected based on their tau score, with over +90 or under -90 being the cut-offs for selection. For *SOX17*-promoter activation, compounds were chosen with a score of over +90. For *SOX17*-promoter repression and signal-1 repression, compounds

were chosen with a score of under -90. These scores were chosen as they represent a match with increase in *SOX17* expression or the opposite of a match with a decrease. The compounds selected were Sirolimus, Aminopurvalanol-a and YK-4279.

2.10.4 Cell Culture

For detailed cell culture methods see section 2.2. 300,000 hPAEC per/well of a 6-well plate were plated in EGM2 (10% (v/v) FBS) and left to grow for 1 day. On day two, media was changed to 0.2% (v/v) FBS to quiesce the cells. On the day of exposure, media was replaced with EGM2 (10% (v/v) FBS) before beginning drug exposures.

2.10.5 Drug Exposure

All drugs were resuspended to [1 M] in dimethyl sulfoxide (DMSO) and kept at -20°C. For cell exposure experiments, Aminopurvalanol-A (B6935, APEXBIO, Houston, Texas) and YK-4279 (A3946, APEXBIO) were used at [10 nM] and [50 nM]. Sirolimus (A8314, APEXBIO) was used at [0.1 nM] and [10 nM]. Cells were exposed to the drugs for both 6 and 24 hours. These time points and concentrations were chosen as they were used to create the compound reference signatures in the CMap or found through a literature search. Vehicle (DMSO) controls were also created at the relevant concentration and time points.

2.10.6 Analysis of Candidate Gene Expression

Total messenger ribonucleic acid (mRNA) was extracted from hPAEC using the RNeasy mini kit (Qiagen). RTPCR was performed using Multiscribe Reverse Transcriptase (Applied Biosystems, ThermoFisher) and the resulting cDNA used for qPCR using Power-SYBRgreen mastermix (Applied Biosystems, ThermoFisher). The signature of each compound was compared with the condition signature which had been used to select them, for example, the sirolimus signature was compared with *SOX17*-promoter repression. The genes which were selected for qPCR were consistently perturbed (had a high z-score) in the compound signature across five cell lines or more and were present in the condition gene list. This enabled the selection of highly correlated genes whose gene expression could be assessed in the future to investigate the accuracy of the CMap and allow us to assess compound effect in PAECs. Several genes were also selected (when present in both lists) as they are biologically relevant in PAH. These genes are *BMPRI1A* and *IL1R1A*. *BMPRI1A* is an important receptor within the BMP signalling pathway whose relevance to PAH is discussed in section 1.4.1. *IL1R1A* expresses the interleukin-1 receptor 1-alpha and has been shown to have higher expression levels in iPAH patients [Parpaleix et al., 2016]. The relative expression of candidate genes (Table 7.5) compared to *ACTβ*, was

calculated as previously described (Section 2.5.2) and compared to the corresponding time and media control.

2.11 Graphing & Statistical Analysis

All specific statistical tests and graphing used are discussed in the methods sections of each chapter (Chapters 3, 4 and 5).

GraphPad Prism (version 8.0.2, GraphPad Software, San Diego, California, USA) was used for all graphing and statistical analysis except those discussed below. All graphs show mean ($\pm$ SEM).

R studio (version 1.2.5042, RStudio Inc) was used for all RNAseq analysis and graphing of PCA and volcano plots.

3 Defining Upstream Regulators of *SOX17* & Effects of Common Variation at the *SOX17* Locus in PAH

3.1 Introduction

A international PAH GWAS (discussed in full in section 1.5.5) defined associated common genetic variation upstream of the *SOX17* gene located in a putative enhancer region. This work also showed a variant effect on the enhancer ability of *SOX17*-signal 2 (rs765727), with the common, non-risk allele (T) having a higher enhancer ability in luciferase assays. This suggests there is a difference in the enhancer activity on *SOX17* due to a SNP at the rs765727 locus but does not provide evidence for the mechanism which causes it. Rhodes et al also provided evidence from CRISPR–INH work that the enhancer region containing *SOX17*-signal 1 affects *SOX17* expression but provided no information regarding the enhancer containing *SOX17*-signal 2 and its effect on *SOX17* expression. As discussed (Section 1.5.5), Rhodes, et al (2019) used chromatin conformation Hi-C data to show *SOX17*-signal 2 is contained within a TAD where the only gene present is *SOX17* (Figure 1.5). TADs are 3D chromosomal structures which are linked to functional processes such as replication, transcription and TF binding. Loci or regulatory regions within TADs are more likely to interact with each other than with loci or regulatory regions outside the TAD. Furthermore, enhancer and promoter regions often only interact with targets within the same TAD (for full review see: [Szabo et al., 2019]). This suggests that the most likely target for the enhancer region containing *SOX17*-signal 2 is also *SOX17*.

Enhancers are sections of the genome which affect the expression of target genes. They can function in a cis- or trans-fashion to enhance the expression of nearby or distal (e.g. >1Mb) target genes in a tissue- and cell-specific manner. Enhancers require active factors to act through chromatin looping and must be in direct contact with target gene promoters to have an effect. Chromatin looping must occur to bring the enhancer region into contact with the target region. For example, enhancers commonly act through the binding of transcription factors to target sequences within the enhancer which then effect nearby gene expression. TFs can only bind to accessible chromatin (euchromatin) to have their target effects and each transcription factor recognises a consensus binding sequence which dictates where it can bind. However, it has been shown that the secondary and tertiary conformation of DNA all affect TF binding [Gordân et al., 2013]. Many mutations and SNPs associated with human disease and found through GWAS are in non-coding regions [Albert and Kruglyak, 2015] and genetic variation in enhancers has been linked to asthma, Crohn’s disease, cancer and many others [Hrdlickova et al., 2014]. Because enhancers affect the level of gene expression, it is likely the cause of disease in these cases is linked to changes in spatial and temporal gene expression.

To assess which transcription factors bind to the region containing rs765727 and whether there is differential binding between the reference and variant sequence, EMSA and in-silico analysis were used. This was to predict transcription factors which are likely to bind based on multiple variables such as the sequence itself and the target tissue or cell type. To ascertain if there is differential binding to the sequence surrounding rs765727 depending on the base present at the SNP locus an electromobility shift assay (EMSA) was used. EMSA require extracted nuclear protein from the cell type of interest, biotin-labelled probes which correspond to the sequences of interest and unlabeled probes of the same sequences. As shown in figure 3.1 after co-incubation an electrophoresis assay separates the free probes from probe:protein complexes, resulting in the shift pattern which gives the assay its name. If the labelled probes have sequences which differ by a single base pair and there are different shift patterns this suggests different proteins (for example TFs) are binding to the two sequences. If the same band is present with both sequences but has a difference in intensity, this can also suggest preferential binding. This can be further investigated through competition EMSA which requires unlabeled or "cold" probes which have the same sequences as the labelled probes. By combining labelled probes with either matching or unmatching cold probes it is possible to run a competitive or non-competitive EMSA respectively (Figure 3.1). Once a differential shift pattern is discovered, 'TF-binding probes' can be used to assess which TF may be responsible. These are cold probes which have the consensus binding sequence of the TF at the center and can be used in a competitive EMSA to ascertain which TF is responsible for the differential shift seen. In-silico analysis using TF databases can be used to investigate which TF may bind to the different sequences. Furthermore, antibodies corresponding to the TF predicted to bind can be added to the assay to investigate a supershift pattern (Figure 3.1). If this TF is responsible for the differential binding, then it will bind to the probe:protein complex, making it larger and causing it to travel less far through the gel, creating a supershift pattern and removing the shift pattern seen before. This provides good evidence that this TF binds to one of the two sequences and may be responsible for any difference in binding.

EMSA are used to investigate differential protein or TF binding to different sequences of DNA. EMSA were used to investigate if there is differential transcription factor binding to the sequence surrounding and including rs765727 when compared to the reference sequence. 21base pair (bp) DNA probes were designed with the risk or non-risk allele as the central base. This is to ensure any transcription factor binding site which includes the allele is contained within the probe. Bio-probes are labelled with a 3' biotin label which can be visualized whilst imaging. Other probes are also created with the same sequence, without the 3'-biotin label. These are known as cold/competition probes and are used in competition EMSA reactions. If a cold probe with a different sequence is

added to a reaction and removes the shift from the image, it suggests that the sequence of the cold 'non-competitive' probe has a higher affinity for that transcription factor than the sequence of the biotin-labelled probe. PAEC nuclear extract was combined with probes as this best represents both the disease-affected cells in PAH and the location where differential binding to rs765727 will have the greatest effect. To narrow down the possible TFs which could be responsible for the additional binding seen, *in-silico* analysis of TF prediction sites can be performed. These sites predict which TF will bind to a sequence. It is possible in some databases to assess two different sequences to see differential TF binding. Supershift EMSA can then be used to further investigate the binding of specific TFs to the sequence of interest. An antibody specific to the TF of interest is added to the binding reaction. If the TF responsible for creating the shift seen in previous EMSAs is bound by this antibody, the probe-TF-antibody complex will be larger and create a supershift pattern in the resulting image. Because of the specific binding between antibodies and their target proteins, this supershift band suggests that the TF responsible for the binding to the probe sequence is the specific antibody target.

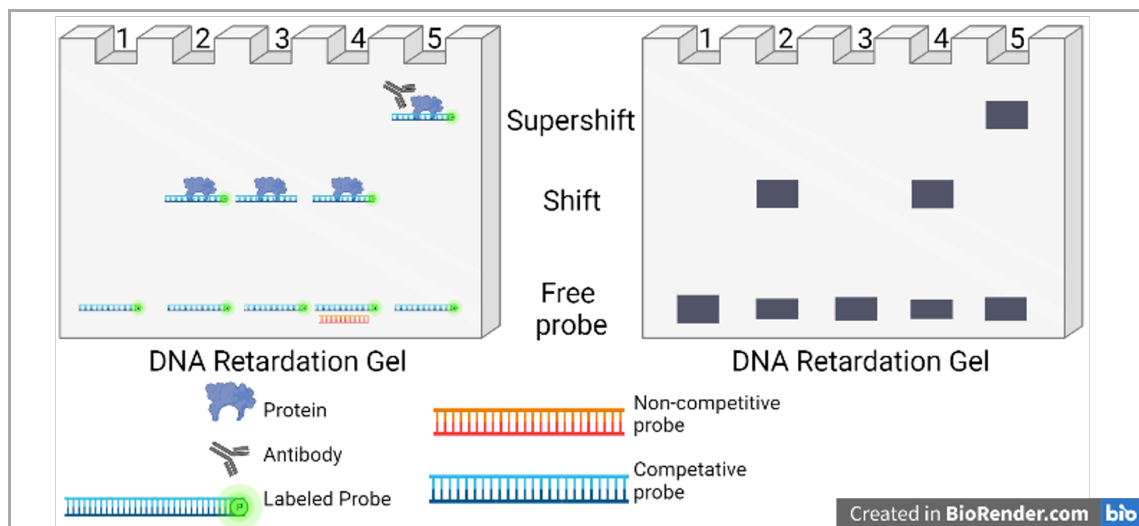


Figure 3.1

Electromobilty shift and supershift assay. A representation of an EMSA, competition EMSA and a supershift assay with reagents shown on the left and the corresponding expected image of the membrane. Well 1 contains only nuclear protein. Well 2 contains nuclear protein and a biotin-labeled probe. Well 3 contains nuclear protein, a biotin-labeled probe and a non-labelled, competitive probe. Well 4 contains nuclear protein, a biotin-labeled probe and a non-labelled, non-competitive probe. Well 5 contains nuclear protein, a biotin-labeled probe and an antibody targeting the TF of interest. Adapted from ThermoFisher.

To assess the effect of the enhancer region containing *SOX17*-signal 2 on *SOX17* expression, CRISPR–INH was used. CRISPR–INH is the targeted silencing of transcription through blocking of the binding of TFs or elongation complexes to the target area. This physical prevention of binding factors from their target sequence inhibits their action. The CRISPR complex contains a dead Cas9 (dCas9) which is deficient in nuclease activity and

therefore cannot create strand breaks but attaches to target DNA and blocks the binding of other proteins or complexes. To increase the inhibition capability of the dCas9, it is often combined with a repressor domain such as Kruppel associated box (KRAB) which recruits co-repressors KP1, SETDB1 and histone deacetylases to increase inhibition [Gilbert et al., 2013]. When targeting a section of an enhancer region, the CRISPR–INH will block binding of TF and prevent the enhancer’s target gene expression effects. To investigate the putative enhancer’s role on *SOX17* expression, CRISPR inhibition (CRISPR–INH) was used to inhibit the *SOX17*-signal 2 region surrounding rs765727. CRISPR–INH utilises a dCAS9-CRISPR construct coupled with a transcriptional repressor KRAB construct which allows the complex to attach to the target DNA and block the binding of other complexes such as transcription factors. Specific regions are targeted through guide sequences. In this thesis, CRISPR–INH has been performed using lentiviral packaging plasmids pMDLg/pRRe, pRSV-Rev and pMD2.G (Addgene #12251, #12253 and #12259) which are third generation packaging plasmids and an envelope expressing plasmid (pMD2.G) respectively. These are used to insert CRISPR-inhibition vectors into cells to cause the inhibition of transcription at a targeted DNA/genomic region (for a full review on CRISPR-inhibition see: [Larson et al., 2013]). HEK293FT cells were used to produce lentiviral particles. These cells were selected as a fast-growing variant of human embryonic kidney 293T which are designed for high production of lentiviruses. These cells stably express the SV40 large T antigen at a high level under the control of the human cytomegalovirus (CMV) promoter. The T antigen allows high level of protein to be expressed from the lentiviral vectors used, and a higher titre of lentivirus is produced.

SOX17 has been shown to act as a tumour suppressor gene in different cancer cell lines [Zhang et al., 2016, Zhou et al., 2019], and has a wide variety of known gene targets including *CCND1* [Liao et al., 2019] but the important targets in hPAEC are not known. As *SOX17* dysfunction through rare and common variants are associated with PAH development, it is important to establish these. Assessing the wider transcriptomic changes which occur through inhibiting the *SOX17* enhancer region in hPAEC will give more insight into the role of this region in hPAEC function and biology and potentially assess its link to PAH. RNAseq analyses typically provide the fold change of detected genes in a sample when compared to a control. The differentially expressed genes (DEGs) can then be assessed for wide-scale changes in cell function, pathways and biological processes through gene ontology (GO) analysis. This will allow further investigation into specific functions or pathways which can be targeted for investigation and novel therapies.

3.1.1 Hypotheses

There is allele-specific transcription factor binding at a putative enhancer region containing rs765727 upstream of *SOX17*.

Variation at the rs765727 locus within an enhancer region affects *SOX17* expression and has downstream effects on function gene pathways in hPAECs.

3.1.2 Aims

1. To investigate the binding of transcription factors to the genomic region containing rs765727.
2. To investigate how different alleles present at this locus affect transcription factor binding.
3. To investigate the effect of inhibition of the PAH genomic locus containing rs765727 on *SOX17* expression.
4. To investigate the downstream transcriptomic and functional pathway changes which occur due to inhibition of *SOX17*-signal 2.

3.2 Methods

3.2.1 Optimisation of EMSA protocol

To optimise the LightShift Chemiluminescent EMSA protocol (ThermoScientific) optimisation EMSA were run to assess which additional reagents were required to obtain clear shift bands in the resulting images (Figure 3.2). Glycerol is added to increase the strength of the bands, but too much can cause vertical streaks in the gel (as seen in wells 7 and 8: Figure 3.2). NP-40 is a protein detergent which is used to maintain the solubility of proteins. The pH of the binding buffer used and the effect of the additional reagents on this can determine the strength and specificity of the probe:protein binding. Additional cations (MgCl_2 , ZnSO_4) will affect the overall pH of the binding reaction. If the optimal pH of the binding reaction is unknown, different concentrations of these cations can be used to assess the effect of different pH binding buffers on the protein:DNA probe binding. A combination of different reagents was added to binding reactions (Table 3.1) and an EMSA was run according to the manufacturer's instructions (ThermoFisher). The binding reaction from wells three and four were selected for use in future EMSA as they showed the clearest bands.

Reagents	Final Conc.	Wells (μl)										
		1	2	3	4	5	6	7	8	10	11	12
ddH ₂ O	-	9	9	10.6	10.6	9.6	9.6	4.6	4.6	12	11	9
BB (1)	-	5	5									
BB (kit)	1x			2	2	2	2	2	2	2	2	2
Poly(didC)	50 ng/ μl	1	1	1	1	1	1	1	1	1	1	1
Glycerol	15%/0.5% (v/v)							6	6	1	1	1
NP-40	0.05% (v/v)					1	1			1	1	1
MgCl_2	5 mM			0.4	0.4	0.4	0.4	0.4	0.4	1	1	1
ZnSO_4	5 mM			1	1	1	1	1	1			
Unlabeled Probe	1000 nM											2
Protein	10 ng	4	4	4	4	4	4	4	4		1	1
Labelled Probe	1 nM	1	1	1	1	1	1	1	1	2	2	2
Allele		C	T	C	T	C	T	C	T	CNTL	CNTL	CNTL

Table 3.1: Binding reactions for EMSA optimisation. Conc, concentration. ddH₂O, double distilled H₂O. BB, binding buffer. BB (1), Binding buffer 1 (HEPES [20 mM], KCL [90 mM], MgCl_2 [5 mM] and NP-40 [40%]). CNTL, Control EBNA Kit (all ThermoFisher).

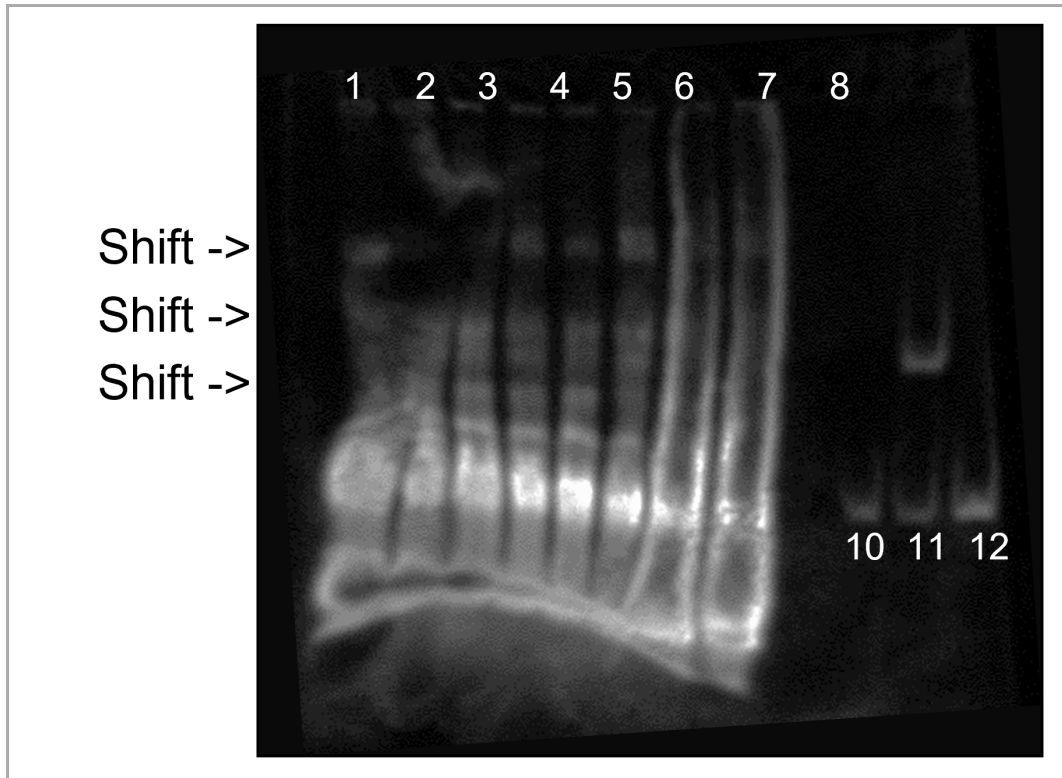


Figure 3.2

Optimisation of EMSA protocol. Wells contain nuclear extract (hPAEC, [10 ug]) and the biotin-labelled probe for the non-risk allele (T, wells 2, 4, 6 and 8) or the risk allele (C, wells 1, 3, 5 and 7). Wells 1 and 2 contains only a binding buffer (HEPES [20 mM], KCL [90 mM], MgCl₂[5 mM] and NP-40 [40%]). All other wells use the binding buffer from the LightShift Chemiluminescent EMSA kit (ThermoFisher). Wells 3 and 4 contain both MgCl₂ [100 mM] and ZnSO₄ [2 mM]. Wells 5 and 6 also contain NP-40 [1%] (v/v). Wells 7 and 8 contain 15% (v/v) Glycerol and no NP40. Wells 10, 11, 12 contain the control kit from the LightShift Chemiluminescent EMSA kit (ThermoFisher). All wells contain distilled water to make up to 20 μ l. Representative image of 2 runs (n=2)

3.2.2 Imaging of EMSA, competition EMSA and supershift assays

All imaging of EMSA, competition EMSA and supershift assays was performed on a ChemiDoc MP imaging system (Universal Hood 3, BioRad, US) using the ImageLab software (BioRad).

3.2.3 Creation of sgRNA for CRISP-INH of rs765727

To investigate the role of *SOX17*-signal 2 in an enhancer's role in *SOX17* expression, CRISPR-INH was used. Four sgRNA guides were created using Tefor-CRISPOR software (version 4.7) (Table 2.3) to target the region surrounding rs765727. When designing CRISPR-INH experiments, it is common practice to design multiple guide RNA for the region of interest, each containing the region of interest but spanning different sections. In this case, four guides have been designed which all contain rs765727 at different points

along to guide. This is to increase the likelihood of the guide having the effect desired. Guides can be ineffective or less effective for several reasons. First, the GC content of a guide will affect its stability, with a higher GC content conferring a higher stability. Second, the production of off target effects. These occur when the guide binds to an unintended region of the genome. The length of the guide can determine the chance of off-target inhibition effects. Most guide design software will provide a specificity score which gives information on the likelihood of off target effects so the guides with lower scores can be selected. The endogenous state of the chromatin can also affect the effectiveness of CRISPR–INH guides. If the target region is in a section of closed chromatin, the guide will not be able to attach to the DNA and have its effect. To minimise all these effects in this work, guides were created with a high GC content (56.5%) and were chosen because their specificity scores were high. The location of *SOX17*-signal 2 is in an open section of chromatin (Figure 1.5) and so there was not predicted to be any effects which can occur due to closed chromatin.

3.2.4 Statistical analysis

To assess the change in relative expression of the target genes *SOX17*, *TMEM58* or *MRPL15* following CRISPR–INH of rs765727, the expression of target genes relative to *ACTB* was graphed with all conditions relative to the GFP-guide control. To establish statistical significance, an ordinary, one-way analysis of variance (ANOVA) with Dunnett’s multiple comparisons test was performed on the average of three experiments between a control group (BFP) and the groups *SOX17* promoter, guide 3 and guide 4 using GraphPad Prism (version 8.0.2, GraphPad Software, San Diego, California, USA).

3.3 Results

3.3.1 Differential binding of hPAEC nuclear protein to the non-risk allele

To assess differential binding of TFs to rs765727, an EMSA was used. Additional binding of nuclear proteins from hPAECs was observed as additional shift bands with the non-risk (T) allele-containing 21bp DNA probe which were not present with the risk allele (C)-containing 21bp DNA (Figure 3.3). A competition EMSA showed removal of one of these shift bands with the competitive (T) cold probe but not with the non-competitive (C) cold probe (Figure 3.3).

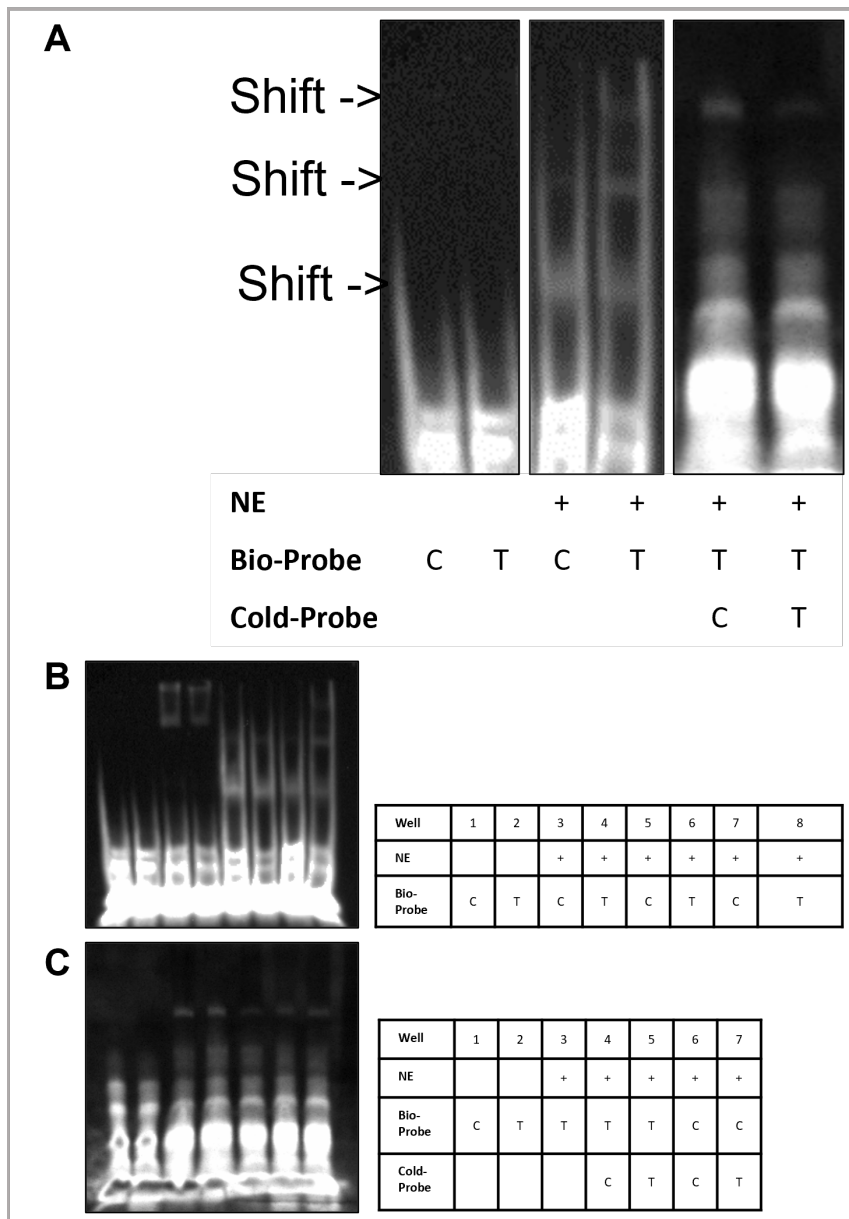


Figure 3.3

Differential binding of proteins to rs765727. **A)** EMSA assay showing binding of hPAEC nuclear proteins to 21bp DNA probes containing the sequence at the rs765727 region. There are additional bands present with the T bio-probe which are either not present with the C bio-probe, or are present at a lower intensity. These bands are lost with the addition of the T cold probe (competitive probe) but not the C probe (non-competitive). The shift is highlight by an arrow. **B)** Full image of left and middle EMSA image. **C)** Full image of right EMSA image. NE, nuclear extract from hPAECs. Bio-probe, biotin-labeled probe. Cold-Probe, unlabeled probe. Cold-probe: R, ROR-alpha; Rh, ROR-alpha at higher concentration. AB, antibody: M, mouse IgG; R, ROR-alpha; Rh, ROR-alpha at higher concentration. Representative images (n=3).

3.3.2 In-silico analysis suggests multiple TFs which could preferentially bind to the non-risk allele

In-silico analysis using TF databases CIS-BP, PROMO and ConSite was used to assess the potential TFs which are responsible for the preferential binding of nuclear proteins seen with the non-risk (T) allele. This identified 15 transcription factors with different likelihoods of binding being affected by the variant (binding score). RNAseq data of gene expression from hPAECs (Figure 3.4) shows robust expression of 10 TFs *RXR α* , *E47*, *ROR α -2*, *ZFX*, *RARB*, *NKX2-5*, *RARG*, *LIN54* and *RARA* and no expression of *NKX-2*, *-1*, *-4*, *-2*, *-3* or *HNF1A*. TFs with no detectable expression in hPAECs were not investigated further. E47 was not selected despite having the highest expression in hPAECs (Figure 3.4) as a splice variant (E12) was shown to bind to the risk allele (C) through *in-silico* analysis of this sequence. The members of the NKX2 family were not selected as they either had very low expression or no detectable expression (Figure 3.4). RAR- α was selected as three sub-types with the same binding sequence (alpha, beta and gamma) all showed expression in hPAECs and were found to bind to only the non-risk (T) allele in in-silico analysis (Figure 3.4). RXR-alpha was not selected for the first round of competitive EMSA as it forms a complex with RAR during binding and this was chosen first. ZFX was also selected due to its expression in hPAECs and in-silico analysis. It was also the TF found through Cis-BP that had the highest binding score, suggesting strong affinity between it and the region surrounding rs765727 (Figure 3.4). Lin-54 was selected as it had the second lowest binding score, suggesting a low affinity. ROR α was selected as the only TF to come out of the analysis from PROMO and because it has a well-documented function in the activation of HIF-1 and hypoxic signalling, an important pathway in PAH [Kim et al., 2008].

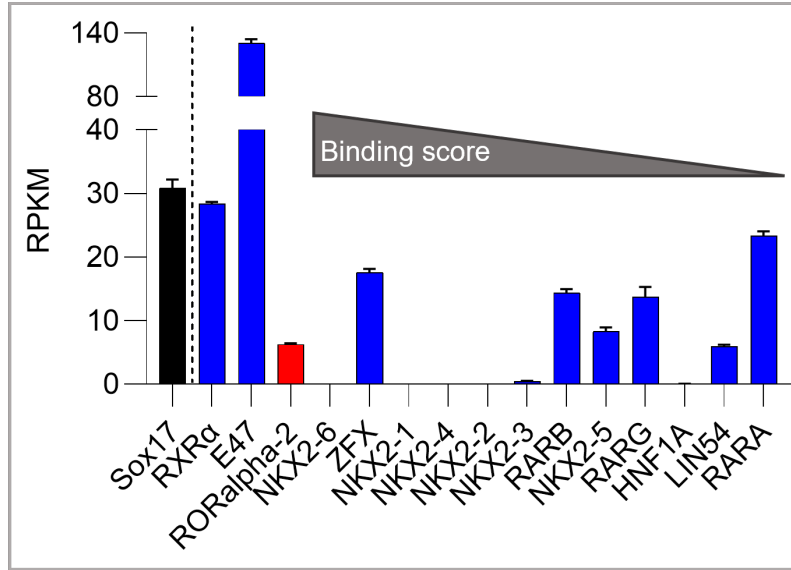


Figure 3.4

Relative expression of transcription factors of interest in hPAECs. The expression (RPKM) of TFs of interest in hPAECs. *SOX17* is shown in black for reference. *ROR-alpha* is shown in red. All other transcription factors are shown in blue. Binding score refers to the likelihood of binding occurring based on the consensus binding sequence of the TF and the sequence in question and is taken from Cis-BP. It decreases from left to right beginning with *NKX2-6*. Bars show mean ($\pm$ SEM). RPKM, reads per kilobase of transcript per million mapped reads.

3.3.3 Competition EMSA show loss of shift with ROR α TF probes

Competition TF cold probes were used in a competition EMSA to assess the binding of Lin-54, ZFX, RAR- α and ROR α to the non-risk (T) biotin labelled probe. Whilst there was no difference in band intensity with Lin-54, ZFX or RAR- α , the shift band was removed with the ROR α cold probe (Figure 3.5A). Analysis of the consensus binding sequence of ROR α (Figure 3.5B) shows the potential binding of this TF around rs765727. This suggests that this TF is responsible for the additional binding found with the non-risk biotin-labelled probe. Additionally, by investigating the consensus binding sequence of ROR- α , and the sequence surrounding rs765727, it is possible to see where this TF may bind to the genome (Figure 3.5B).

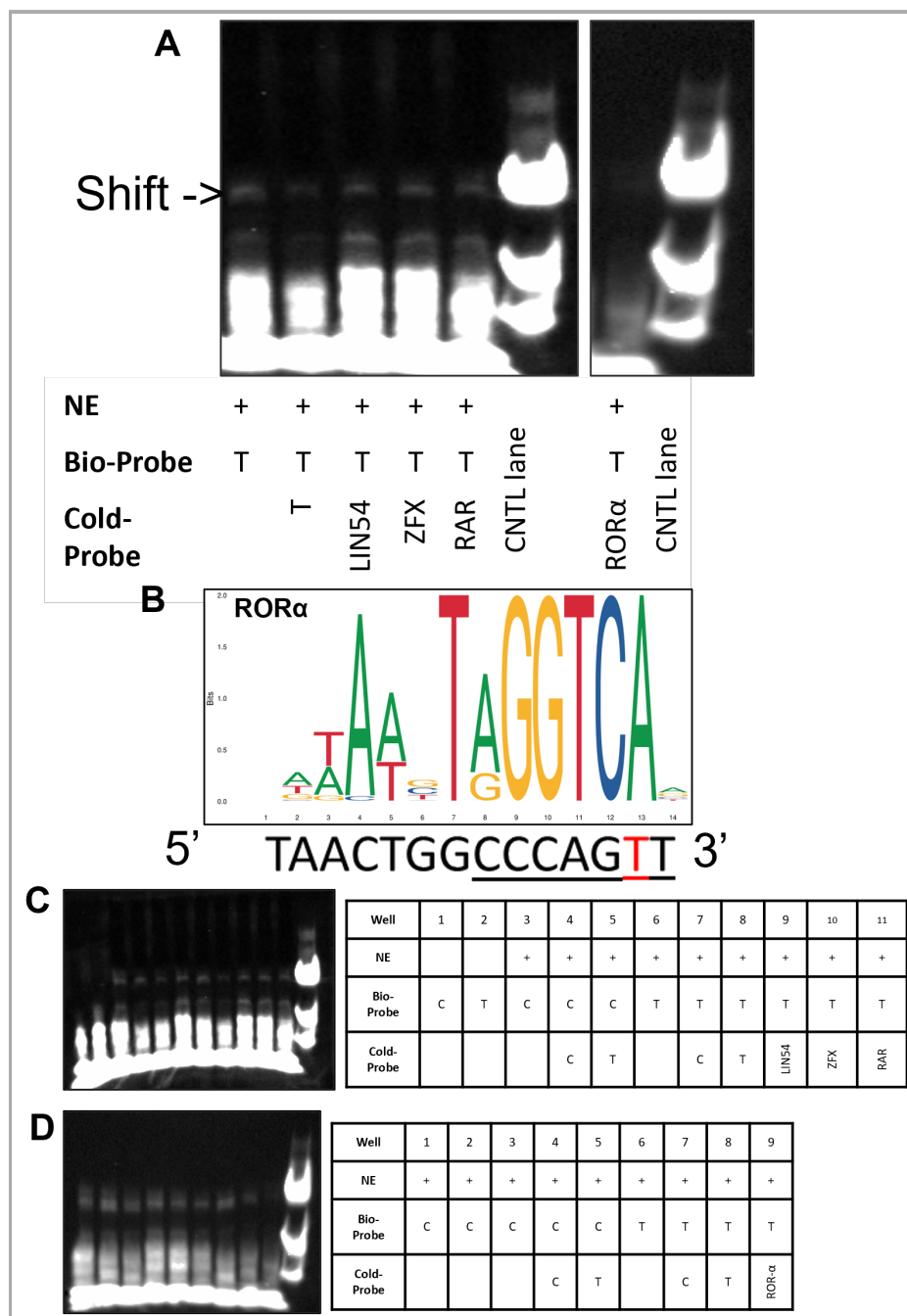


Figure 3.5

Differential binding of competition TF consensus binding sequence probes to rs765727. **A)** Shift bands seen with the non-risk (T) probe are removed with the T competitive probe but not removed with TF binding probes for Lin-54, ZFX, RAR- α . A shift band is removed with the addition of the TF binding probe for ROR α binding. Control lanes from both EMSA are shown. These are controls from the LightShift Chemiluminescent EMSA protocol (ThermoScientific). Shift is highlighted by an arrow. Representative images (n=3). **B)** The consensus binding sequence of ROR α taken from JASPAR (jaspar.genereg.net) shown with sequence of the region surrounding rs765727 taken from <https://genome.ucsc.edu/>. The underlined sequence refers to the potential binding location of ROR α and the T in red font is the site of rs765727. CNTL lane, Control lane from LightShift Chemiluminescent EMSA protocol (ThermoScientific). **C)** Full image of left EMSA image. **D)** Full image of right EMSA image. Bio-probe, biotin-labeled probe. Cold-Probe, unlabeled probe. Cold-probe: R, ROR-alpha; Rh, ROR-alpha at higher concentration. AB, antibody: M, mouse IgG; R, ROR-alpha; Rh, ROR-alpha at higher concentration.

3.3.4 Supershift EMSA suggest ROR- α is responsible for the differential binding at rs765727

To further assess the binding of ROR α to rs765727, a supershift EMSA was used. As shown in figure 3.5, the TF binding probe for ROR α removes the shift seen with the non-risk probe. This shift band is also removed with the addition of an antibody specific for ROR α but not with mouse IgG (Figure 3.6). This suggests that the ROR α antibody is binding to the non-risk biotin-probe:ROR α complex, making it larger and causing it to travel less far through the gel.

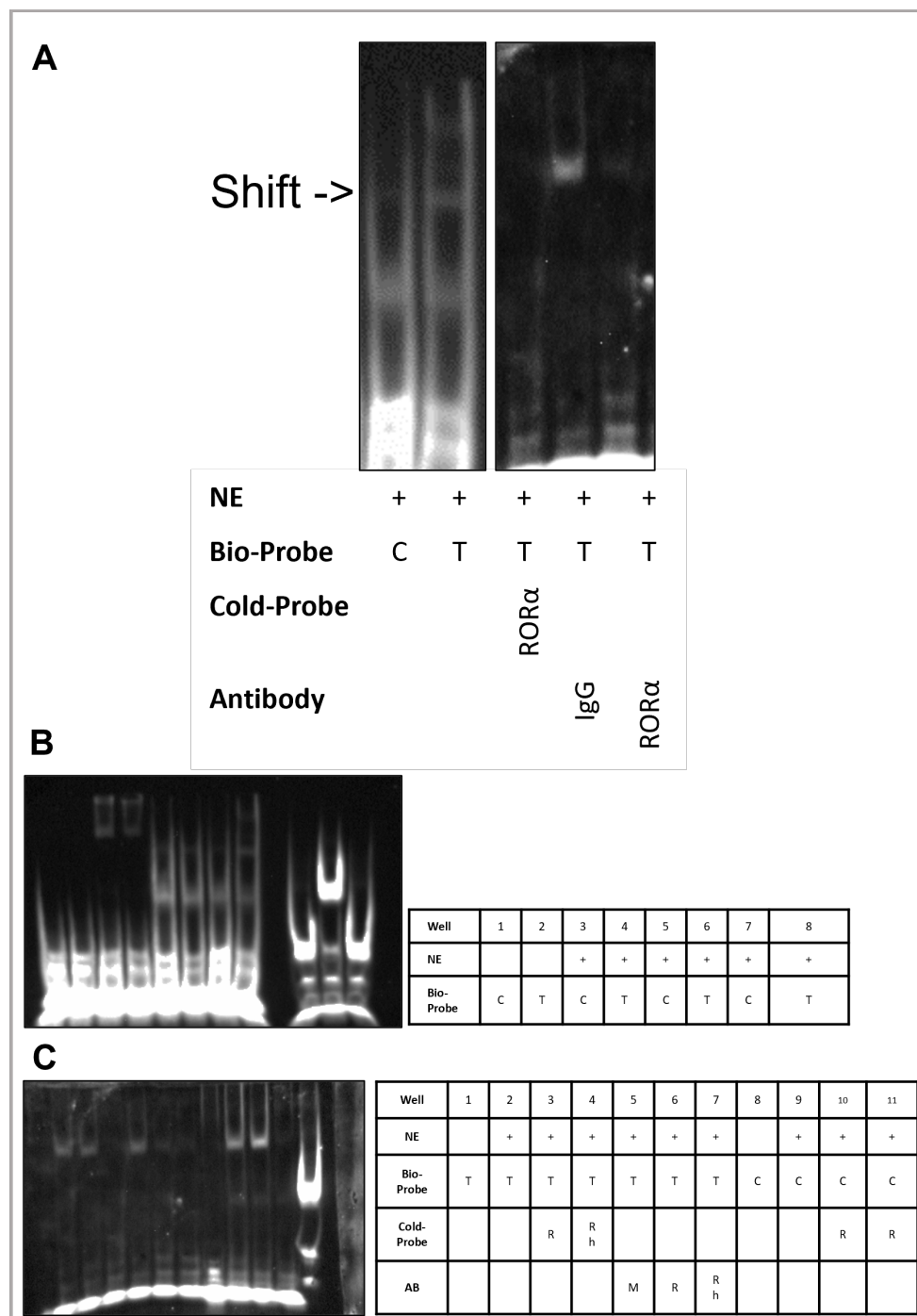


Figure 3.6

Supershift EMSA. **A)** Supershift EMSA showing the removal of non-risk allele-specific nuclear protein binding with both the ROR α TF binding probe and a specific anti-body for ROR α . NE, nuclear extract from hPAECs. Shift is highlighted by an arrow. **B)** Full image of left EMSA image. **C)** Full image of right EMSA image. Bio-probe, biotin-labeled probe. Cold-Probe, unlabeled probe. Cold-probe: R, ROR-alpha; Rh, ROR-alpha at higher concentration. AB, antibody: M, mouse IgG; R, ROR-alpha; Rh, ROR-alpha at higher concentration. Representative images (n=2/3).

3.3.5 CRISPR-INH of the region surrounding rs765727 affects *SOX17* expression

To assess the effect of *SOX17*-signal 2 activity on *SOX17* expression, CRISPR-INH was used to target this region. CRISPR-INH of the region surrounding rs765727 with guides 3 and 4 in hPAECs led to a significant decrease in *SOX17* expression, as did inhibition of the *SOX17* promoter. (Figure 3.7A). No guide significantly affected the expression of *TMEM68* (Figure 3.7Bi), however guide 4 lead to a significant decrease in the nearby gene *MRPL15* (Figure 3.7Bii).

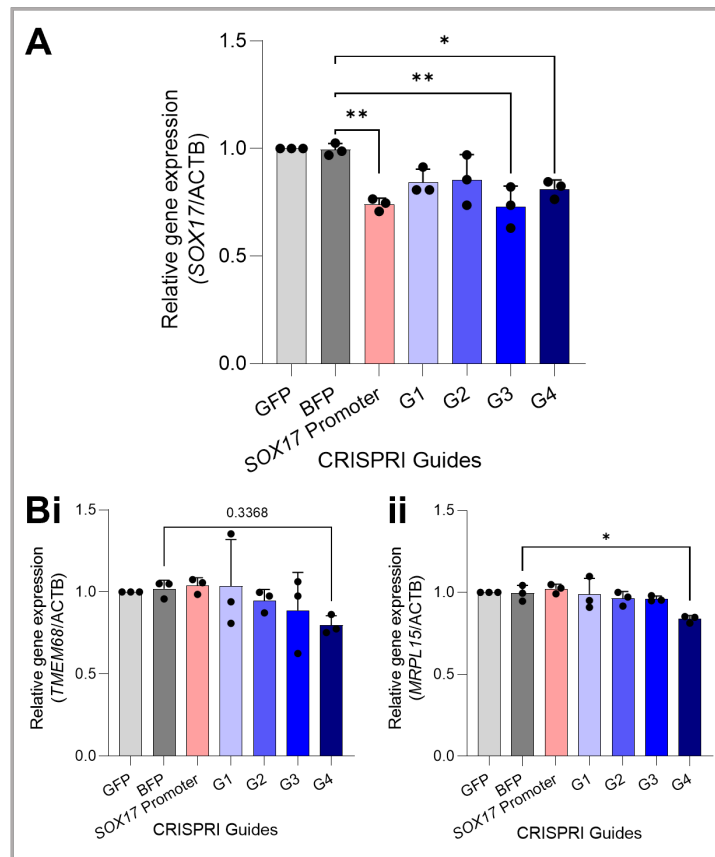


Figure 3.7

Knockdown of *SOX17* through CRISPR-Inhibition of *SOX17*-signal 2. **A)** Relative gene expression of *SOX17* compared to *ACTβ* in hPAECs. **B)** Relative expression of **i.** *TMEM68* and **ii.** *MRPL15* compared to *ACTβ* in hPAECs. Ordinary 1-way ANOVA of all conditions (except GFP) compared to BFP condition with Dunnett's multiple comparisons test. * - p < 0.05, ** - p < 0.01. Average of three separate experiments (n=3). Bars show mean ($\pm$ SEM). G1-4, *SOX17*-signal 2 guides 1-4. GFP, green fluorescent protein. BFP, blue fluorescent protein.

3.3.6 RNAseq of hPAEC targeted by CRISPR-INH

To assess the transcriptomic changes which occur following CRISPR-INH of *SOX17*-signal 2, RNAseq analysis was undertaken. Prior to differential expression analysis, principal component analysis was performed and as principle components 1, 2 and 3

showed dispersion of samples, they were used as co-variates in the subsequent analysis (Figure 3.8A/B). RNAseq analysis showed 312 significant DEGs (12271 variables, 191 up-regulated, 121 down-regulated) following CRISPR–INH of *SOX17*-signal 2 with guides 3 and 4 (Figure 3.8). *SOX17* expression was nominally decreased by -0.361 but didn't reach significance ($p=0.197$, $f=0.864$). The top 50 most up and down significantly regulated genes are shown in table (table 3.2). Gene ontology analysis of these DEGs shows enrichment for processes and functions with links to TNF activity (TNF activity, NF-kappa beta signalling pathway) (Figure 3.9), though few reach $q<0.05$ following corrections for multiple testing.

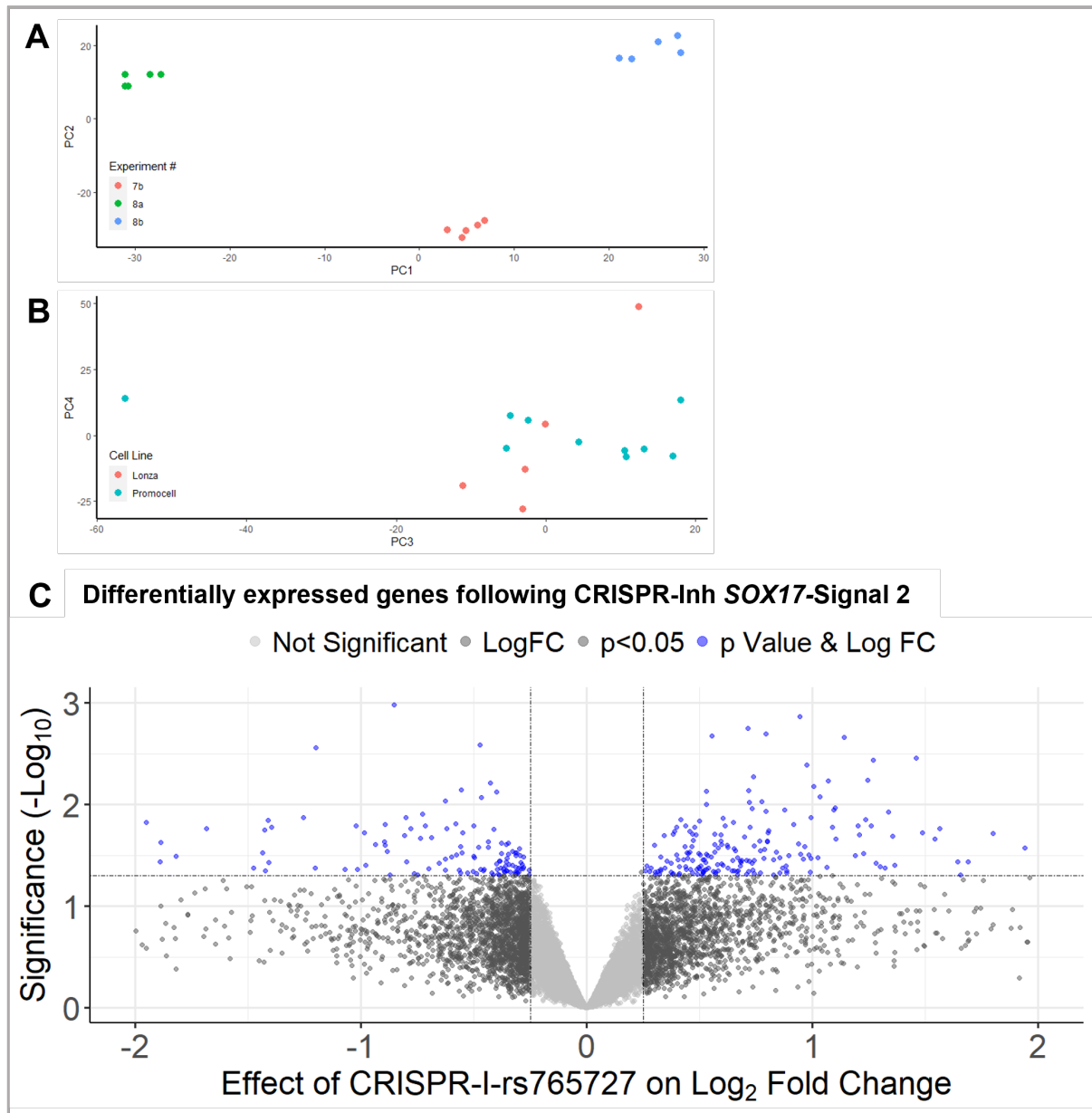


Figure 3.8

Differentially expressed genes following CRISPR-inhibition of *SOX17*-signal 2 in hPAECs. **A)** Principle components 1 and 2 coloured by experiment number. Experiments cluster closely together. **B)** Principle components 3 and 4 coloured by cell company origin (Lonza or Promocell). There is no clustering due to cell origin. **C)** Volcano plot of the $\log_2$ fold change in gene expression between hPAEC targeted with CRISPR-inhibition of *SOX17*-Signal 2 and a BFP control and the negative $\log_{10}$ p-value. Differentially expressed genes shown in blue meet the cut off points: $p < 0.05$ and $\log_2 FC < -0.25$ or > 0.25 . $n=4$, 12271 variables.

Gene Name	NCBI entrez ID	Log Fold Change	Statistical Significance (p)
AC010084.1	ENSG000 00229308.1	2.99	0.002462
ADAMTS12	81792	2.98	0.003538
TXLNG2P	246126	2.87	0.022713
PRKY	5616	2.79	0.042818
FBXL13	222235	1.94	0.026866
AK4P1	206	1.80	0.0193
MUC1	4582	1.69	0.036699
MYCN	4613	1.66	0.049297
ZNF814	730051	1.64	0.036625
CCDC144A	9720	1.56	0.01723
RP5-991G20.1	ENST000 00620814.1	1.54	0.021675
SEC31B	25956	1.49	0.018951
DNAH1	25981	1.46	0.003511
ATF7IP2	80063	1.36	0.039752
SPINT1	6692	1.35	0.020427
CTC-457L16.1	ENST0000 0625108.1	1.34	0.011858
VAV3	10451	1.32	0.042282
PILRB	29990	1.30	0.04078
LRRC17	10234	1.28	0.037756
DISP2	85455	1.27	0.003675
AC007952.5	ENSG000 00228157	1.26	0.01631
RP11-734K23.9	ENSG000 00277701	1.24	0.005808
FAP	2191	1.23	0.014152
JMJD7-PLA2G4B	8681	1.23	0.030324
SNCAIP	9627	1.21	0.019847
SHC3	53358	-0.98	0.019113

Gene Name	NCBI entrez ID	Log Fold Change	Statistical Significance (p)
OSGIN2	734	-1.02	0.043395
ZDHHC2	51201	-1.02	0.01621
ADRBK2	157	-1.07	0.043647
SUGCT	79783	-1.20	0.002768
ZEB2	9839	-1.20	0.042205
FAM46A	55603	-1.25	0.013394
RP11-435I10.4	ENSG000 00261832	-1.40	0.016749
SLC7A2	6542	-1.41	0.037231
EIF3CL	728689	-1.41	0.014307
LOXL1	4016	-1.42	0.044589
IL6	3569	-1.43	0.017772
VEGFC	7424	-1.44	0.029647
POPDC3	64208	-1.47	0.041957
SERPINB2	5055	-1.68	0.017183
CXCL2	2920	-1.82	0.032239
CXCL8	3576	-1.89	0.023774
SEMA3D	223117	-1.89	0.036597
CXCL1	2919	-1.95	0.014977
UNC79	57578	-2.22	0.002101
TMSL6	7120	-2.46	0.038126
IL21R	50615	-2.63	0.035318
CXCL6	6372	-4.02	0.001826
RP11-96D1.11	ENSG00 000260441	-4.15	0.047302
RP11-609D21.3	ENSG00 000279296	-4.28	0.004919

Table 3.2: Top 50 significantly differentially expressed genes following *SOX17*-signal 2 inhibition. All genes have NCBI Entrez IDs shown except *RP5-991G20.1*, *CTC-457L16.1*, *AC007952.5*, *RP11-734K23.9*, *RP11-435I10.4*, *RP11-96D1.11* and *RP11-609D21.3* which have Ensemble ID shown.

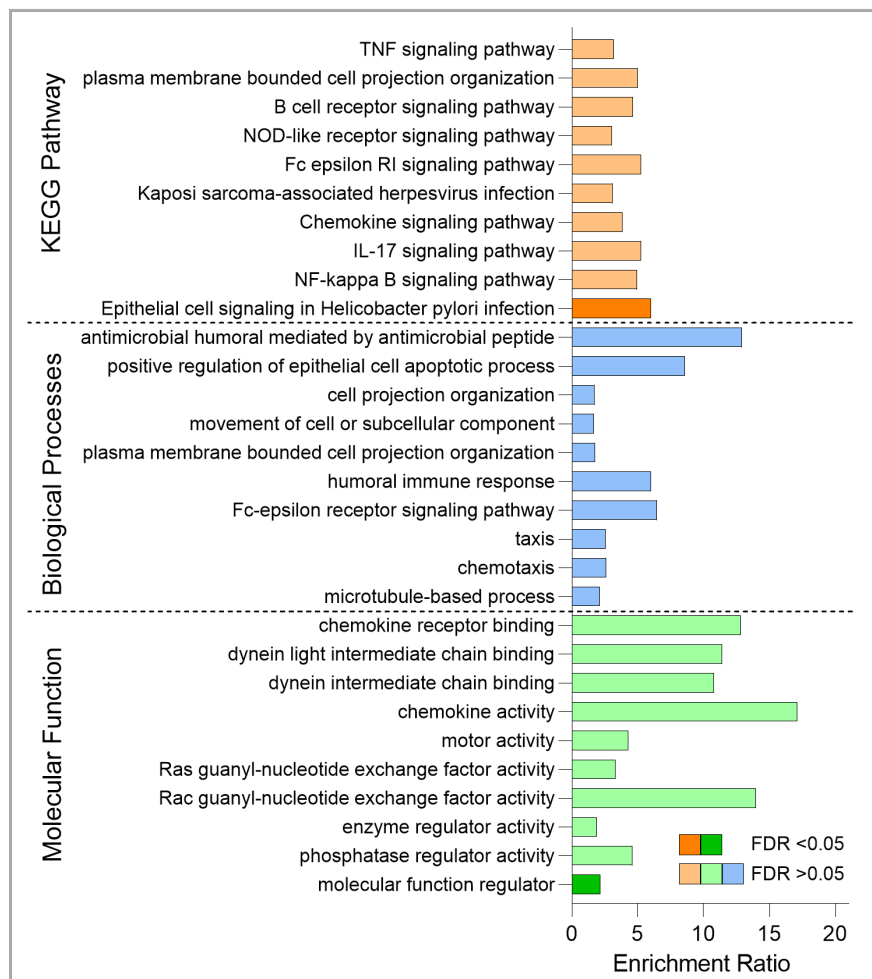


Figure 3.9

Over-representation analysis for enriched pathways, functions and molecular processes following CRISPR-inhibition of *SOX17*-Signal 2. Gene ontology analysis of KEGG pathway (orange), biological process (blue) and molecular function (green) enrichment following CRISPR *SOX17*-Signal 2 in hPAECs. Darker colours indicate FDR $q < 0.05$. Lighter colours indicate FDR $q > 0.05$. Enrichment ratios obtained from WebGestalt.

3.4 Discussion

The first aim of the work highlighted in this chapter is to investigate the binding of TFs to *SOX17*-signal 2 and whether the rs765727 non-risk or risk-alleles cause differential TF binding. The second aim is to investigate if a putative enhancer upstream of the *SOX17* gene affects *SOX17* expression and the transcriptomic and downstream pathway changes which occur following this inhibition. EMSA have been used to demonstrate the differential binding of hPAEC nuclear proteins to genomic sequences containing the rs765727 non-risk (T) and risk (C) alleles. In silico analysis of the sequence surrounding rs765727 has been used to identify multiple candidate TFs whose binding to this region could be affected by the PAH variant. Transcriptomic analysis of hPAEC was used to confirm which of these candidates was robustly expressed in the cells principally

involved in PAH pathology and expressing *SOX17*. Finally, competition and supershift EMSA demonstrated that ROR-alpha was the TF binding differentially to the rs765727 sequence. To address the second main aim, CRISPR-INH has been used to inhibit the region surrounding *SOX17*-signal 2 (rs765727) and the expression of *SOX17* was shown to be down-regulated specifically using RT-qPCR. Downstream gene expression alterations and pathway changes were assessed using RNAseq and GO, though the statistical power was limited.

Previous work using luciferase assay reporters in hPAEC has shown a difference in enhancer ability between the risk and non-risk allele present at rs765727 [Rhodes et al., 2019]. However, that work did not investigate whether the SNP present at rs765727 caused differential binding of proteins at this region or which TFs might be affected. The EMSA shown in this chapter suggest that not only does the non-risk allele (T) have additional binding of proteins, but that the TF ROR-alpha is a key factor affected. The supershift pattern is not present in this EMSA image and further supershift EMSA with a longer running time may show this supershift band and would add to the evidence highlighted here that there is additional binding at this region when the non-risk allele (T) is present and that ROR-alpha is responsible for this binding. Although EMSA do provide evidence of differential binding to the non-risk (T) allele of ROR-alpha, to further build on this work, a chromatin immunoprecipitation (ChIP)-PCR should be performed to assess the binding of ROR-alpha to rs765727 within hPAECs. ChIP-PCR investigates the interactions between proteins and DNA within the cell through DNA:protein crosslinking, sonication, immunoprecipitation and PCR. By using an antibody specific for ROR-alpha, it would be possible to assess the regions of the genome where ROR-alpha binds in hPAECs and investigate the genomic region surrounding rs765727 endogenously. An allelic imbalance assay could also be used to quantify allele-specific gene expression and transcription factor binding of ROR-alpha to rs765727. An allelic imbalance assay requires genotyped cells to investigate the regulatory activity of a region on a gene and as it requires a heterozygous individual two alleles can be compared within in the same sample (or individual). It is based on the knowledge that SNPs which affect the expression of a gene in a heterozygous individual will cause a difference in expression of that gene depending on which chromosome is being evaluated by qPCR of in-phase SNPs. For example, for rs765727, I hypothesize that the non-risk (T) allele will cause a increase in expression of *SOX17* when compared to the risk (C) allele in a heterozygous individual. Therefore, in an allelic imbalance I would expect the chromosome containing the non-risk allele to have higher *SOX17* expression.

ROR-alpha is a transcription factor which has been found to play roles in various mechanisms, the most investigated being inflammation, circadian rhythm and metabolism [Delerive et al., 2001, Duez and Staels, 2008, Oh et al., 2019]. It has robust expression in

both human vascular endothelial and smooth muscle cells [Besnard et al., 2002]. There is little research into its role in the vasculature or in cardiovascular diseases. However, some work suggests it has a protective role against high-fat diet induced cardiac abnormalities in mice and its deficiency impairs mitochondrial biogenesis [chao Zhao et al., 2017]. On the other hand, other work shows a potential detrimental affect in pathological retinal angiogenesis in mice. ROR-alpha deficiency in a mouse model of proliferative retinopathy showed an increase in ROR-alpha lead to an increase in neovascularisation [Sun et al., 2015]. These two examples show the conflicting roles ROR-alpha can have in the vasculature. However, there is (at the time of writing) no literature on the role of ROR-alpha in PAH or in hPAECs.

EMSA are commonly used to investigate the impact of variation in regulatory regions on human disease, for example type 2 diabetes and ovarian cancer [Fogarty et al., 2014, Yodsurang et al., 2018]. In the study by Fogarty, et al (2014), the risk allele (rs11257655) linked to *CDC123/CAMK1D* expression had higher transcriptional activity and a higher binding affinity for nuclear proteins extracted from multiple relevant cell types. Using supershift assays, they showed specific binding of FOXA1 and FOXA2 to the risk allele, suggesting the binding of these TFs to the risk allele (rs11257655) increases the enhancer activity of this region. This study highlights how additional binding of TFs to a risk allele can increase target gene expression which can lead to an increased risk of disease. Techniques have been proposed which could be used to rescue the normal function of these enhancers or, in the case of the work described in this thesis, increase the expression of the target gene of the enhancer containing the protective/non-risk allele. The CRISPR-Cas9 system provides the potential for both activation and inhibition of enhancer regions and could be adapted for each variant and disease.

Previous work has shown that *SOX17*-signal 2 is located in a putative enhancer region inside a TAD, within which the only gene is *SOX17* [Rhodes et al., 2019]. Because of this, it is likely that *SOX17* is affected by this enhancer region. The work presented here shows a significant decrease in *SOX17* expression following CRISPR-INH of the region surrounding *SOX17*-signal 2 for the first time, establishing that this genomic signal associated with PAH development also functions through regulation of *SOX17*. Furthermore, there is no effect on the expression of a nearby gene *TMEM68* with any guide (Figure 3.6), demonstrating the specificity of this finding. The decrease in the expression of the nearby gene *MRLP15* with guide 4 is likely to be a false positive as the four guides used target very similar sections surrounding rs765727 and none of the other three guides show an effect on *TMEM68* expression. These findings, coupled with previous work from Rhodes et al (2019), suggest that the region surrounding *SOX17*-signal 2 is part of an enhancer region which affects *SOX17* expression. Furthermore, when combined with the EMSA investigations discussed in chapter 3, it suggests that the binding of ROR-alpha to the

non-risk (T) allele present at rs765727 (*SOX17*-signal 2) underpins the enhancer ability of the region, increasing *SOX17* expression and thereby decreasing PAH risk.

To further assess the role of *SOX17*-signal 2, ROR-alpha binding to the non-risk (T) allele and the effect on *SOX17* expression, investigation into the effect of *ROR-α* overexpression on *SOX17* expression could be performed. This could be achieved with adenoviral overexpression of *ROR-α* but would need to be performed in cells genotyped to ensure the non-risk allele is present. This work would provide further evidence that ROR-alpha selectively binds to rs765727 and causes an increase in the expression of *SOX17* through the effect of an enhancer region.

The decrease in *SOX17* expression seen with CRISPR–INH guides 3 and 4 are like that seen with inhibition of the *SOX17* promoter, suggesting a robust inhibition of the *SOX17* gene with these three constructs. Due to the small effects enhancers generally have on target genes, it is likely that inhibiting an enhancer will have a smaller effect on the target gene than inhibiting the promoter of the gene. A possible cause for the small changes seen with the *SOX17* promoter guide is the potential for cells which have not been successfully transduced with lentivirus containing the CRISPR–INH construct (low efficiency) being collected for downstream experiments. Although a selection step was included in the experiment which should kill all the cells which do not have a blasticidin-resistance gene (conferred by the successful transduction with lentivirus), it is possible this was not 100% effective. To remove this possibility, colony expansion can be performed from single cells. This produces expanded populations which should all contain the CRISPR–INH constructs as they are grown from cells which contain it. However, this could not be performed in hPAECs for two main reasons. Firstly, hPAEC are primary cells obtained from donors and can only be used for a certain number of passages before they stop proliferating, senesce and can no longer be used. Clonal expansion of cells requires cell lines which can be expanded through many passages. Secondly, transduction with lentiviruses can be very toxic to hPAECs and it was found that cells with low confluency would not survive transduction with lentiviruses. For these reasons, clonal expansion was not attempted in this work. There are immortal endothelial cell lines such as EA.hy926 which could be used in the future to produce more stable CRISPR–INH lines. However, recent work has suggested large differences in gene expression and signalling pathways between these and primary lines used in most vascular biology investigations [Wang et al., 2021]. Furthermore, these specific cells are derived from human umbilical vein endothelial cells and are therefore less relevant to PAH research than hPAECs. As discussed throughout chapter 1, different endothelial cells have widely different physiology which can impact the research which they should be used for. hPAECs provide the best disease cell model for PAH and so were used in this work.

Analysis of the significantly DEGs following CRISPR–INH shows that although *SOX17* expression did not change significantly in this analysis, it did trend towards a decrease. This non-significant and small change can be attributed to the relatively small effect enhancer regions have on gene expression which was discussed earlier. The relatively small number of samples which were analysed through RNAseq (n=4 per condition) would also have played a role in the low power of the statistical analysis performed. GO analysis of significantly DEG following CRISPR–INH of *SOX17*-signal 2 showed pathways linked to the TNF-signalling pathway (TNF-signalling and NF-Kappa beta signalling) were enriched. The TNF-signalling pathway is especially important in immune cell response to injury and has previously been implicated in PAH as a potential regulator of *BMP2* and notch signalling [Hurst et al., 2017]. Further work into the genes associated with these enriched pathways needs performing including better powered transcriptomics, robust validation by qPCR and assessment of the effect of *SOX17*-signal 2 CRISPR–INH on their protein level and the affected (e.g. TNF signalling) pathway(s).

3.5 Conclusion

The work discussed in this chapter shows binding of the TF ROR-alpha to the rs765727 non-risk (T) allele only and suggests it is responsible for the allele-specific enhancer ability highlighted in previous work [Rhodes et al., 2019]. Evidence has also been presented that suggests inhibiting this region via CRISPR–INH decreases the expression of *SOX17* in hPAECs. This suggests the polymorphism leads to preferential binding of ROR-alpha to the non-risk (T) allele present at rs765727 specifically increases the expression of *SOX17*. *SOX17* enhancer activity affects gene expression including a possible enrichment of genes which are involved in the TNF signalling pathway which requires validation. In chapter 4, further RNAseq analyses of different manipulations of *SOX17* are compared. Chapter 4 also focuses on the functional role of *SOX17* in hPAECs.

4 Functional Analysis of Sox17 Manipulation in Pulmonary Vascular Cells

4.1 Introduction

Rare, loss of function genetic variants in *SOX17* are pathogenic in PAH, suggesting loss of *SOX17* activity by any mechanism, including reduced enhancer activity due to common variants, is associated with the development of PAH (Section 1.4.2). *SOX17* is a TF with documented roles in arterial EC fate determination and regulation of the cell cycle. A key early pathogenic feature in PAH is endothelial cell injury which leads to a hyper-proliferative and anti-apoptotic state. In the field of cancer biology, *SOX17* is a well-established tumour suppressor gene [Zhang et al., 2016, Zhou et al., 2019] and controls the expression of cell cycle genes such as *CCDN1* (encodes cyclin D1) [Liao et al., 2019]. In these studies, a loss of *SOX17* contributes to hyper-proliferative cells which are typical of cancer phenotypes. This suggests that if *SOX17* is lost in cell types relevant to PAH such as hPAECs it may lead to the hyper-proliferative and anti-apoptotic state seen in this disease. Furthermore, as discussed in section 1.5, *SOX17* has been shown (in systemic endothelial and non-endothelial cells) to have many important roles in functions that are relevant to PAH including proliferation, permeability and apoptosis of hPAECs, and signalling pathways that are known to affect disease such as VEGF, WNT and Notch (Section 1.5.2). In ECs, Sox17 is important in the maintenance of barrier function and proliferation which causes EC regeneration following injury [Liu et al., 2019]. Interestingly, it has previously been shown that altered *SOX17* expression increases vascular fragility in arteries of the brain, increasing the risk of aneurysm through unregulated EC barrier function [Bilguvar et al., 2008, Foroud et al., 2012]. The loss of *SOX17* expression could therefore be a factor that contributes to the loss of barrier function seen in hPAECs in PAH. Downstream signalling and function affected by the loss of *SOX17* in pulmonary vascular cells is not well established and further studies are required to ascertain how these changes could contribute to PAH. In this chapter, the effect of knocking down *SOX17* through siRNA and CRISPR–INH on downstream pathways and functions in hPAECs will be assessed to investigate the role of *SOX17* in hPAECs and the link between its dysfunction and PAH.

siRNA is typically used to knock down mRNA levels of target genes. They are double stranded RNA, typically under 30 bps with each strand has a 2 bp overhang at the 3' end and work through the RNA interference (RNAi) pathway. Inside the cell, the siRNA construct is cleaved into 2 single RNA strands by Dicer, a helicase enzyme with an RNase motif. These strands are known as the passenger and leader strands. With the siRNA constructs used in this thesis (silencer select, ThermoFisher), the passenger strand is targeted for degradation by the cell. This minimises off-target effects without compromising

the efficiency of the leading strand which combines with co-factors to form the RNA induced silencing complex (RISC). The siRNA of the siRNA:RISC complex guides the complex to the target mRNA which is then cleaved by RISC. This leads to a decrease in the target gene at the mRNA (and usually protein) level which can be assessed by qRT-PCR and Western blotting respectively.

Only a small number of downstream targets of *SOX17* have been identified. *SOX17* has been shown to affect the expression of *CDH5*, *CYCLINe1*, *NOTCH4* and *NOTCH1* either through direct binding to their promoter or an unknown intermediary [Corada et al., 2013, Clarke et al., 2013, Dejana et al., 1999]. *SOX17* and the BMP signalling pathway are closely linked as discussed in section 1.5.4. However, these potential *SOX17* downstream targets have not been assessed in hPAEC.

Recently, large-scale transcriptomic analysis of human PAH tissue and gene knockdown in PAH-relevant cells have been completed such as the largest PAH-lung transcriptomic analysis to date by Stearman et al (2019). Here, the differences between control and PAH lung tissue transcriptomes were assessed to provide disease-specific gene expression classifiers. Specific pathways such as TNF and Wnt signalling pathways were shown to be especially important in upstream immune response and PAH-sex differences respectively. One limitation of this study was the use of whole lung tissue samples which doesn't allow the specific role of each cell type captured in the study to be individually assessed. To avoid this, transcriptomic datasets can be obtained from cell culture of one type of cell (e.g. hPAECs) or single-cell RNAseq of patient cells can be performed. A recent study investigated the heterogeneity of hPAECs derived from PAH patients and controls [Asosingh et al., 2021]. 629 genes were found to be differentially regulated between control and PAH samples as well as distinct clusters of hPAECs within patients which are responsible for important pathological pathways and functions. These studies highlight the importance of transcriptomic analysis to uncover the basis of overall tissue dysfunction and hPAEC dysfunction in PAH. Defining the transcriptomic changes driven by *SOX17* dysfunction in hPAECs will provide information on the functional pathway changes and function disruption which may occur in PAH patients with pathogenic variations that cause a decrease in *SOX17*. This will provide insight into the mechanism(s) which are contributing to PAH in these patients and potential targets for therapeutic rescue circumventing the loss of *SOX17* itself.

4.1.1 Hypothesis

Changes in *SOX17* expression contribute to PAH risk through downstream effects on signalling pathways relevant to hPAEC function changes.

4.1.2 Aims

1. To investigate the downstream targets of Sox17 in hPAEC.
2. To investigate the role of Sox17 in hPAEC functions (proliferation, viability, apoptosis and barrier function).

4.2 Methods

4.2.1 RNAseq and principal component analysis

Principal component analysis of RNAseq data was undertaken to identify patterns in the datasets. This allows patterns of global variation to be used as co-variates in the subsequent analysis. In hPAECs samples which had undergone *SOX17*, *SOX7* and *SOX18*-siRNA, principal components 1 and 2 showed dispersion of samples and were used as covariates in the subsequent analysis. This was required to correct for non-specific factors affecting broad gene expression such as sample quality and donor differences (Figure 4.2).

4.2.2 Selection of *SOX17*-related genes

The genes *BMP9*, *BMPR2*, *NOTCH1*, *NOTCH4*, *CYCLINe1* and *CDH5* were selected for further validation following RNAseq analysis of *SOX17*-siRNA knockdown hPAECs as they have been shown to have a previous connection with *SOX17* which is discussed in section 4.1.

4.2.3 Statistical Analysis

To assess the change in relative expression of genes following siRNA, the expression of the gene relative to *ACTB* was graphed with all conditions relative to the positive siRNA control. To assess statistical significance, an unpaired, one-way student's t test was performed on the average of four experiments between a negative siRNA control group and the *SOX17*-siRNA group. To assess the change in Sox17 protein following siRNA, the protein level relative to Vinculin was analysed with all conditions relative to the negative siRNA control. To test statistical significance, an unpaired, one-way student's t test was performed on the average of four experiments between a negative siRNA control group and the *SOX17*-siRNA group.

To assess the change in relative expression of SOXF factors following siRNA for all SOXF TFs, the expression of SOXF factors relative to *ACTB* was analysed with all conditions relative to the positive siRNA control. To establish statistical significance for both, an ordinary, one-way ANOVA with Dunnett's multiple comparisons test was performed on the average of four experiments between a negative siRNA control group and the SOXF factors groups. The same graphing and statistical tests were used to assess the effect of siRNA on SOXF factors protein levels.

To assess the effect of *SOX17*-siRNA and treatments on hPAEC proliferation, viability and apoptosis, an ordinary, two-way ANOVA with Dunnett's multiple comparisons test was performed on the average for three experiments between the *SOX17*-siRNA condition and the negative siRNA control using GraphPad Prism (version 8.0.2, GraphPad

Software, San Diego, California, USA). To assess the effect of *SOX17*-siRNA on hPAEC barrier function, an ordinary, one-way ANOVA with Dunnett's multiple comparisons test was performed on the average for three experiments between the *SOX17*-siRNA condition and the negative siRNA control.

All graphing and statistical analysis was completed using GraphPad Prism (version 8.0.2, GraphPad Software, San Diego, California, USA).

4.3 Results

4.3.1 *SOX17*-siRNA successfully knocks down *SOX17* at the mRNA and protein level

To assess the effect of a decrease of *SOX17* in hPAECs, siRNA was used to knock down *SOX17* mRNA. This caused a significant decrease in relative mRNA and protein levels of *SOX17* when compared with a negative control siRNA (Figure 4.1). mRNA levels fell to 0.222 (± 0.06) ($p < 0.0001$) relative to the negative siRNA control and protein levels fell to 0.035 (± 0.05) ($p < 0.0001$) relative to the negative siRNA control.

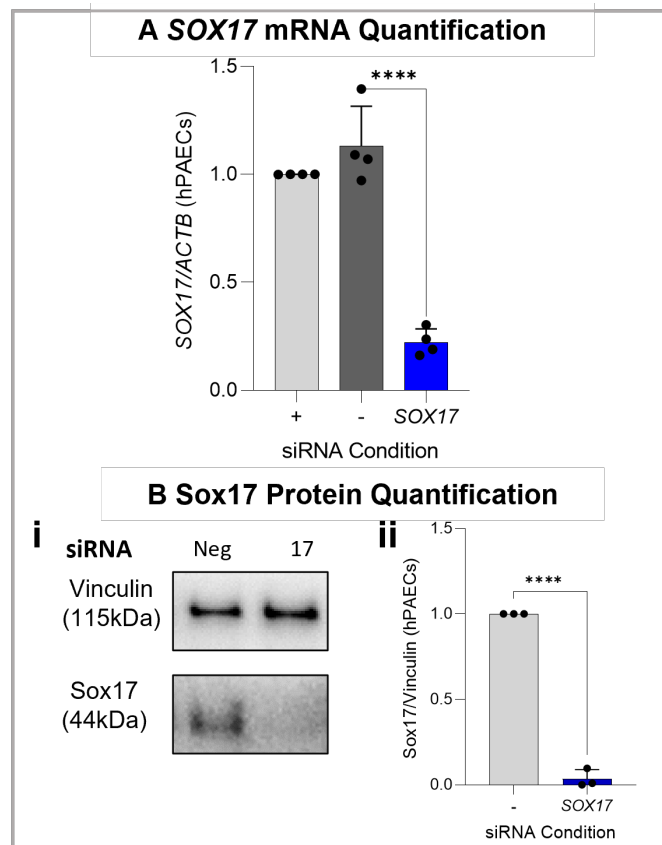


Figure 4.1

Knockdown of *SOX17* following siRNA. *SOX17* mRNA and protein levels fell significantly following siRNA. **A)** Relative mRNA quantification of *SOX17* by qPCR. The change in *SOX17* expression is normalised to *ACTB* and all siRNA conditions are relative to the positive control. **B)i.** Representative image of a western blot of Sox17 and vinculin protein in siRNA-*SOX17* and siRNA-negative samples. **ii.** Quantification of western blotting of Sox17. Sox17 protein level is normalised to vinculin and relative to the negative control. +, positive control. -, negative control. All statistical tests shown are paired, one-way, student's t test. **** - $p < 0.001$. Points are averages of either three (protein) or four (mRNA) separate experiments ($n=3/4$). Bars show mean ($\pm$ SEM). All in hPAECs following *SOX17*-siRNA treatment for 48 hours. Full western blot images are shown in supplemental figure 7.3.

4.3.2 Knockdown of *SOX17* leads to downstream genetic and molecular pathway changes

To assess the transcriptomic changes which occur due to *SOX17* inhibition, RNAseq analysis of siRNA-*SOX17* and siRNA-control treated hPAEC samples was undertaken. Prior to differential expression analysis, principle component analysis was performed and as principal components 1 and 2 showed dispersion of samples, they were used as covariates in the subsequent analysis (Figure 4.2A/B). 738 genes were identified which were significantly down- or up-regulated ($p < 0.05$) by more than 0.25 Log_2 fold change following *SOX17*-siRNA (Figure 4.2). The top 50 most significantly (FDR $q < 0.05$) differentially expressed genes are shown in table 4.1. GO using these 738 genes shows significant enrichment in biological processes, molecular functions and pathways involving adhesion and extracellular matrix (ECM) (adherens junction, focal adhesion, cell adhesion, biological adhesion, ECM organisation, extracellular structure constituent, organisation and interaction and ECM-receptor interaction) (Figure 4.3).

Gene Name	NCBI Entrez ID	Log Fold Change	Statistical Significance (p)	FDR (q)
<i>RPS4Y1</i>	6192	9.63	3.27E-08	2.42E-04
<i>TXLNG2P</i>	246126	9.47	1.02E-07	3.12E-04
<i>DDX3Y</i>	8653	8.44	6.10E-07	6.20E-04
<i>ZFY</i>	7544	8.40	2.38E-06	1.11E-03
<i>RASAL3</i>	64926	7.92	3.85E-07	5.78E-04
<i>EIF1AY</i>	9086	7.68	2.01E-06	1.10E-03
<i>USP9Y</i>	8287	6.33	1.49E-07	3.63E-04
<i>TTY15</i>	64595	5.39	8.17E-08	3.12E-04
<i>CCNA1</i>	8900	5.08	5.91E-06	2.00E-03
<i>KCNAB1</i>	7881	4.71	4.97E-06	1.78E-03
<i>MMP17</i>	4326	4.53	5.74E-07	6.20E-04
<i>TYW1B</i>	441250	4.01	2.99E-06	1.35E-03
<i>ZNF415</i>	55786	3.86	1.73E-06	1.05E-03
<i>TPM2</i>	7169	3.48	2.20E-06	1.11E-03
<i>VAMP8</i>	8673	3.37	1.56E-06	1.00E-03
<i>IGF2BP1</i>	10642	3.32	4.71E-07	5.78E-04
<i>FAP</i>	2191	3.13	4.16E-06	1.58E-03
<i>PTK7</i>	5754	2.86	3.79E-06	1.54E-03
<i>SLC38A5</i>	92745	2.82	1.22E-05	2.97E-03
<i>ERV3-1</i>	2086	2.55	2.07E-06	1.10E-03
<i>GSTM3</i>	2947	2.48	1.39E-06	1.00E-03
<i>RPL22L1</i>	200916	2.35	4.74E-07	5.78E-04
<i>GSTT2B</i>	653689	2.28	1.03E-05	2.60E-03
<i>ITGA1</i>	3672	2.27	7.44E-06	2.27E-03
<i>ADORA2B</i>	136	2.08	3.28E-06	1.38E-03
<i>PTPRU</i>	10076	1.85	6.84E-06	2.17E-03
<i>APOBEC3C</i>	27350	1.76	4.03E-06	1.58E-03
<i>DDR2</i>	4921	1.49	2.30E-06	1.11E-03
<i>ZFX</i>	7543	-1.18	6.43E-06	2.12E-03
<i>CDH13</i>	1012	-1.41	1.01E-05	2.60E-03
<i>NDN</i>	4692	-1.49	1.05E-05	2.60E-03
<i>PTGS1</i>	5742	-1.55	9.27E-06	2.57E-03
<i>TFF3</i>	7033	-1.66	8.40E-06	2.44E-03
<i>GBP3</i>	2635	-2.29	1.01E-05	2.60E-03
<i>ITM2A</i>	9452	-2.43	4.31E-07	5.78E-04
<i>RCN3</i>	57333	-2.74	4.45E-06	1.64E-03

Gene Name	NCBI Entrez ID	Log Fold Change	Statistical Significance (p)	FDR (q)
<i>COL8A2</i>	1296	-3.00	1.02E-05	2.60E-03
<i>PEG10</i>	23089	-3.04	9.03E-06	2.56E-03
<i>CLIC2</i>	1193	-3.10	6.96E-06	2.17E-03
<i>HSD17B1</i>	3292	-3.70	4.68E-07	5.78E-04
<i>SFMBT2</i>	57713	-4.06	1.95E-06	1.10E-03
<i>HOPX</i>	84525	-4.66	5.83E-06	2.00E-03
<i>FAM86DP</i>	692099	-5.05	7.76E-06	2.31E-03
<i>HLA-J</i>	3137	-5.37	1.49E-06	1.00E-03
<i>GSTM1</i>	2944	-6.34	1.35E-06	1.00E-03
<i>RPL13P12</i>	388344	-6.41	7.17E-07	6.73E-04
<i>XXyac-R12DG2.3</i>	ENSG00000170919	-6.52	3.97E-08	2.42E-04
<i>EGFLAM</i>	133584	-6.58	3.24E-06	1.38E-03
<i>MTRNR2L8</i>	100463486	-6.78	9.00E-07	7.84E-04
<i>RP11-411H5.1</i>	ENSG00000226715	-7.44	1.40E-06	1.00E-03

Table 4.1: Top 50 significantly differentially expressed genes following *SOX17*-siRNA. All genes have NCBI Entrez IDs shown except *XXyac-R12DG2.3* and *RP11-411H5.1* which have Ensemble ID shown.

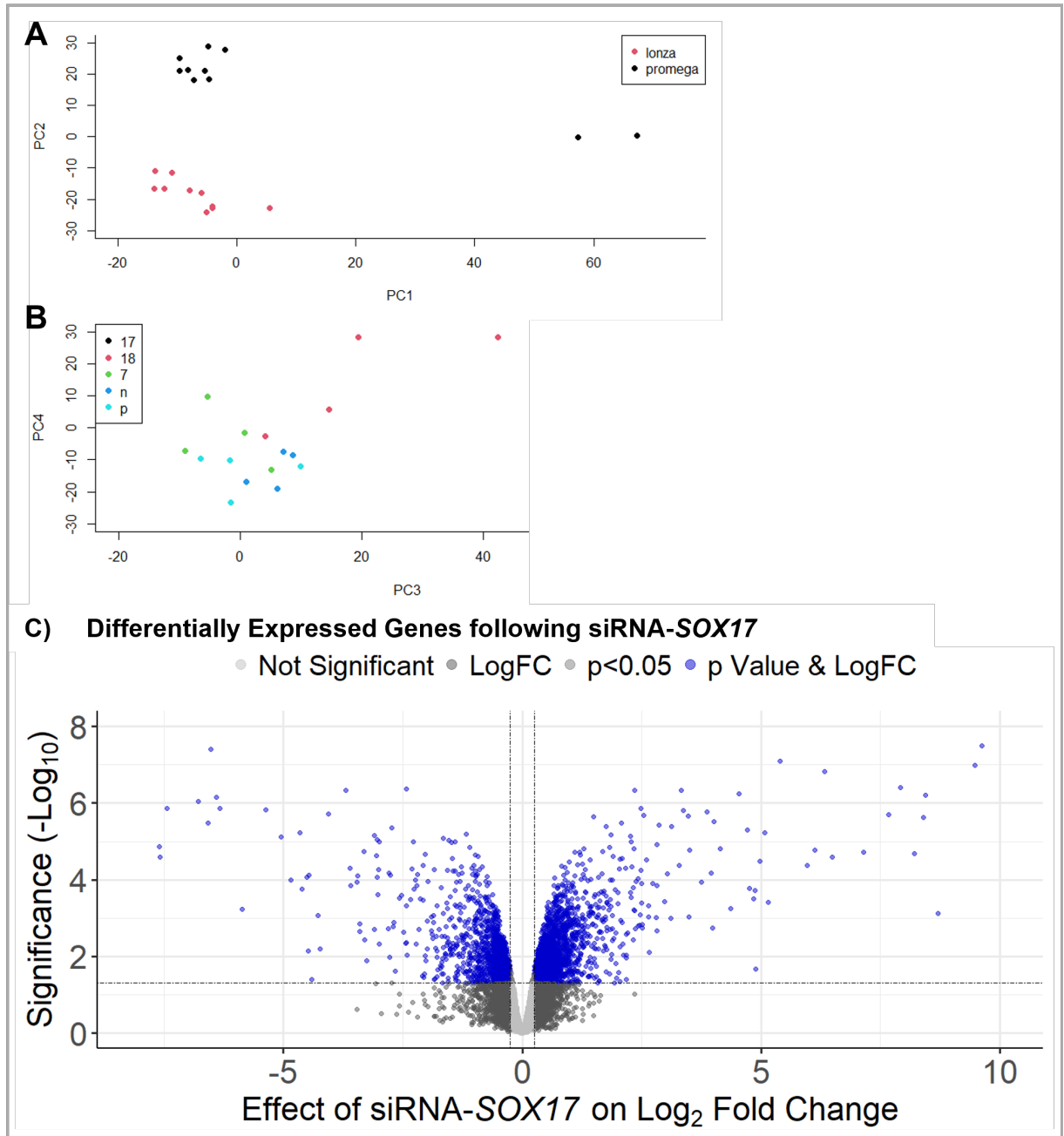


Figure 4.2

Differentially expressed genes following siRNA-SOX17 in hPAECs. **A)** Principle components 1 and 2 coloured by cell company origin (Lonza or Promocell). Samples cluster by cell company origin. **B)** Principle components 3 and 4 coloured by siRNA condition. 17 - *SOX17*, 7 - *SOX7*, 18 - *sox18*, n - negative control, p - positive control. There is no clustering due to siRNA condition. **C)** Volcano plot of the $\log_2$ fold change (FC) between siRNA-*SOX17* and siRNA-Negative Control and the negative $\log_{10}$ p-value. Differentially expressed genes shown in blue met the cutoff points: $p < 0.05$ and Log fold change < -0.25 or > 0.25 . n=4, 12196 variables.

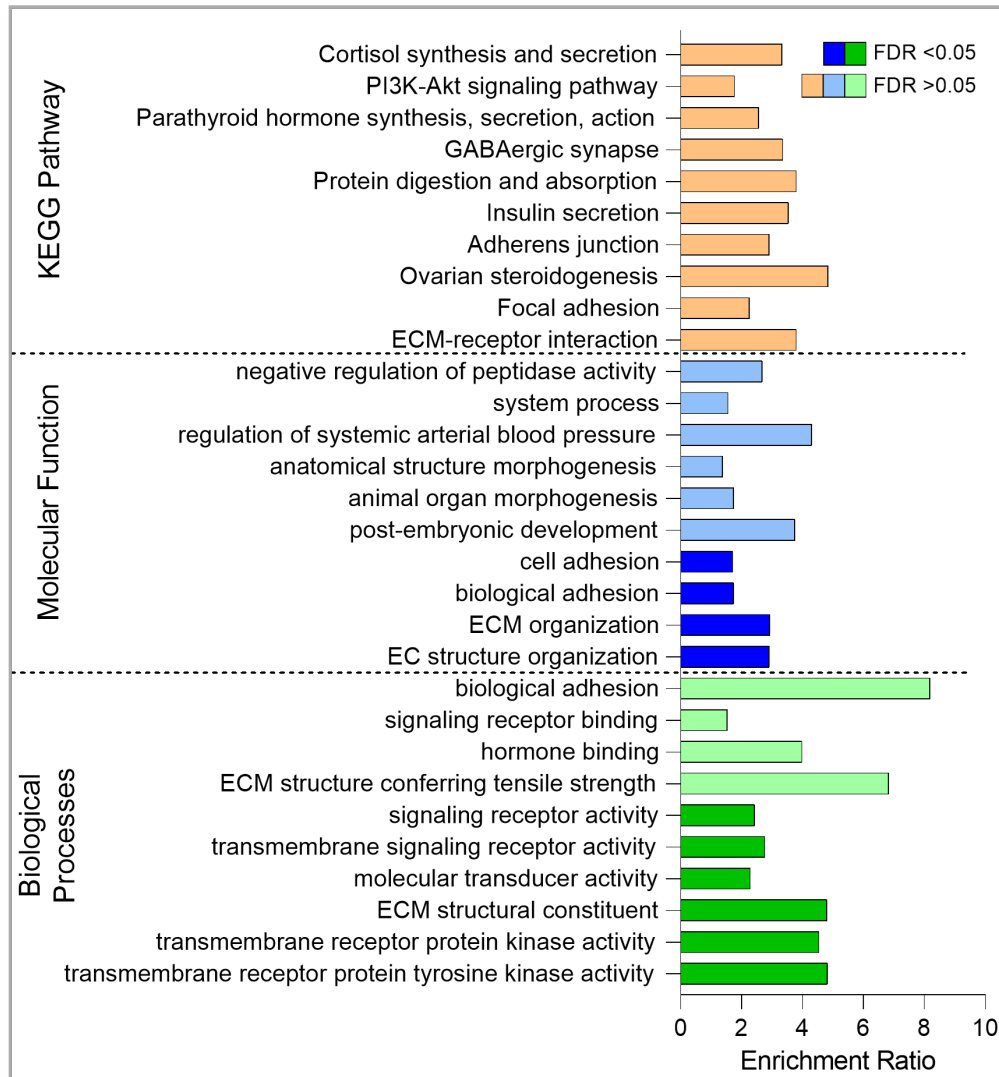


Figure 4.3

Over-representation analysis for enriched pathways and functions following siRNA-*SOX17*. Gene ontology analysis of KEGG pathway, biological process and molecular function enrichment following siRNA-*SOX17* in hPAECs. Darker colours indicate FDR $q < 0.05$. Lighter colours indicate FDR $q > 0.05$. Enrichment ratios were obtained from WebGestalt.

4.3.3 Knockdown of *SOX17* influences downstream targets in hPAECs

To assess the effect of *SOX17*-siRNA on downstream targets and validate targets found through RNAseq analysis, qPCR was used to assess gene expression following siRNA treatment of independent hPAEC cultures. Enriched gene targets were selected from the ECM and adhesion processes (taken from the GO analysis discussed in section 4.3.2). *ADAMTS12*, *MMP17*, *LAMB3* and *COL4A6* expression significantly increased when compared to a negative control siRNA (Figure 4.4), 8 others did not significantly change (Figure 4.4). When analysed through RNAseq, all of the expression changes of these

genes met the FDR cut off of less than 0.05 and changed in the same direction as shown via qPCR (Table 4.2).

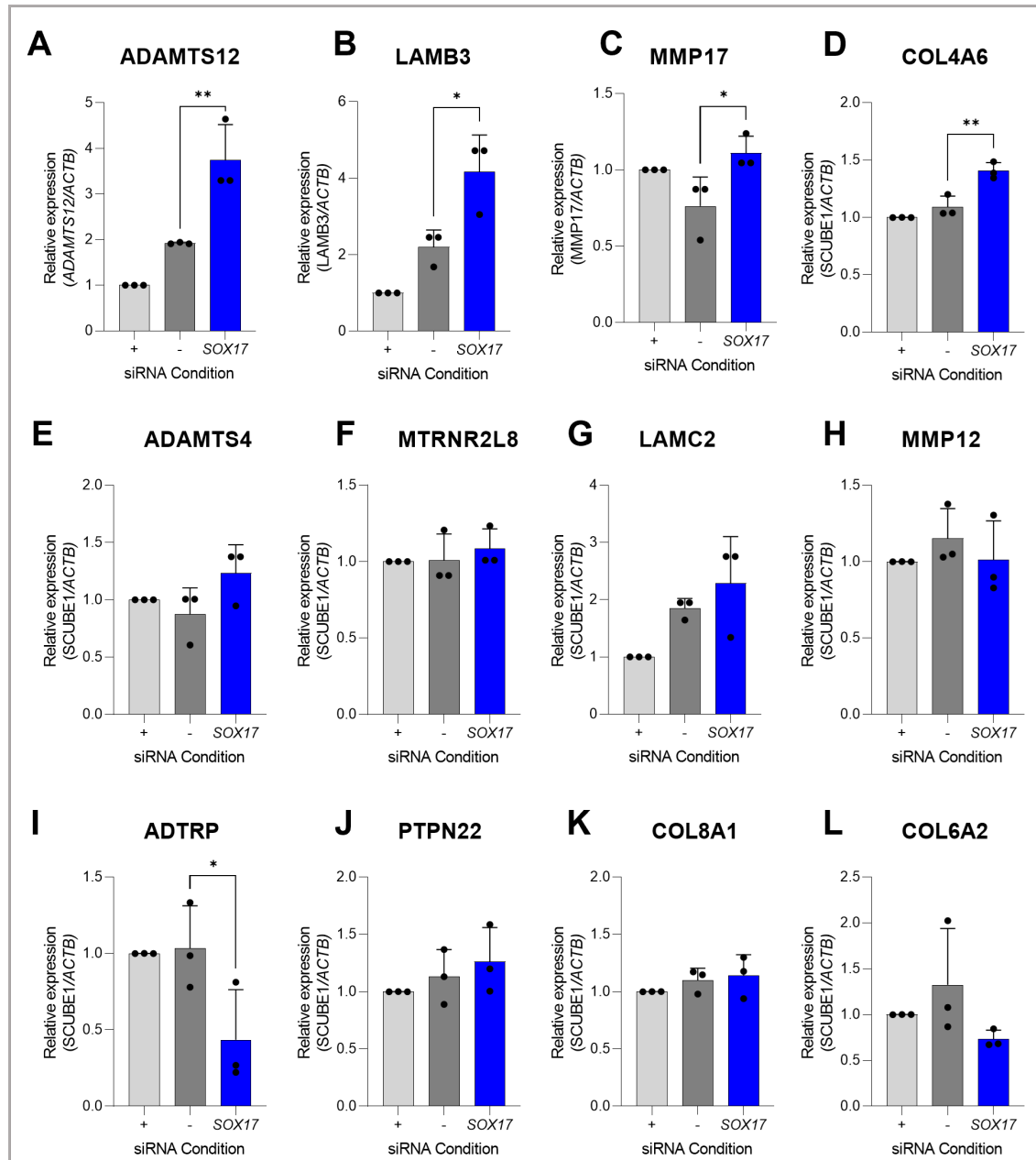


Figure 4.4

Validation of gene ontology target genes. Relative mRNA quantification of target genes by qPCR. The change in target gene expression is normalised to *ACTB* and all siRNA conditions are relative to the positive control. Target genes are **A)** *ADAMTS12*, **B)** *LAMB3*, **C)** *MMP17*, **D)** *COL4A6*, **E)** *ADAMTS4*, **F)** *MTRNR2L8*, **G)** *LAMC2*, **H)** *MMP12*, **I)** *ADTRP*, **J)** *PTPN22*, **K)** *COL8A1* and **L)** *COL6A2*. A, B and C are from ECM processes and functions. D, E, F, G and H are from adhesion processes and functions. The statistical test paired, one-way, student's t test was run on all and all are not significant. Points are averages of three separate experiments (n=3).

To assess the effect of *SOX17*-siRNA on genes known to be relevant to *SOX17* (from literature) expression, qPCR was used to assess hPAEC gene expression following siRNA treatment. *NOTCH4* and *CDH5* expression significantly decreased when compared to a negative control siRNA. The expression of *CYCLINE1* trended towards an increase but was non-significant. The expression of *BMP9*, *BMPR2* and *NOTCH1* did not significantly change (Figure 4.5). When analysed through RNAseq, the expression changes of *CYCLINE1* and *NOTCH1* are significant (Table 4.2).

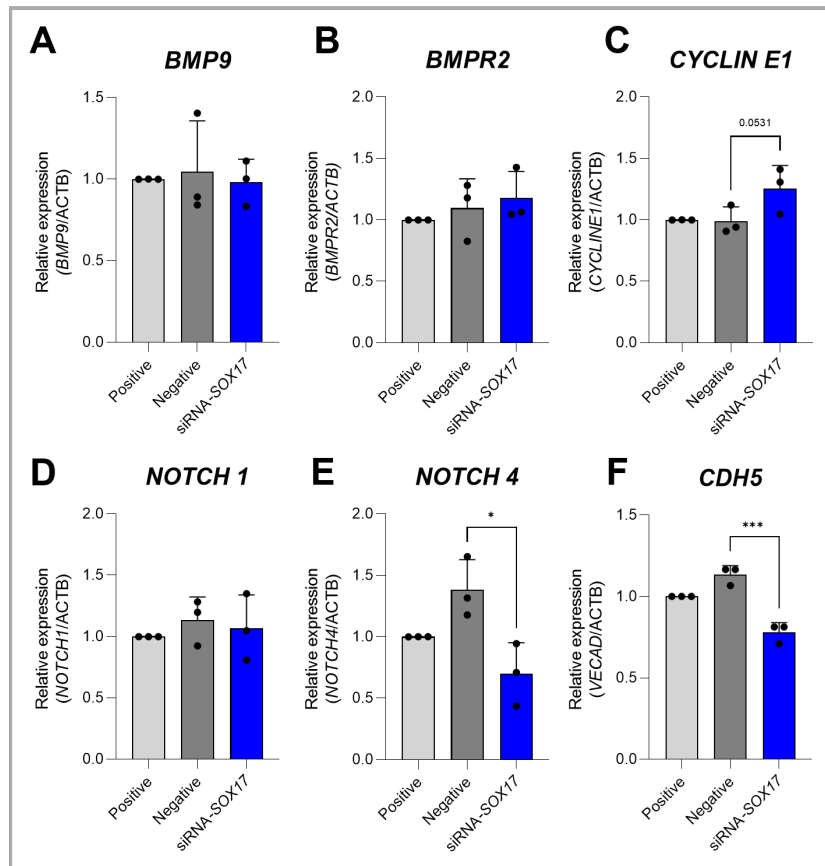


Figure 4.5

Validation of potential *SOX17*-associated genes. Relative mRNA quantification of target genes by qPCR. The change in target gene expression is normalised to *ACTB* and all siRNA conditions are relative to the positive control. Target genes are **A)** *BMP9*, **B)** *BMPR2*, **C)** *CYCLINE1*, **D)** *NOTCH1*, **E)** *NOTCH4* and **F)** *CDH5*. +, positive control. -, negative control. All statistical tests shown are paired, one-way, student's t test. * - $p < 0.05$. ** - $p < 0.01$. Points are averages of three separate experiments ($n=3$). Bars show mean ($\pm$ SEM). All in hPAECs following *SOX17*-siRNA treatment for 48 hours.

Gene Name	Log Fold Change	p value (p)	FDR (q)	Direction changes in qPCR with SOX17-siRNA
<i>ADAMTS12</i>	10.954	0.00003	0.004	Up
<i>MMP17</i>	4.533	0.0000006	0.001	Up
<i>LAMB3</i>	2.506	0.002	0.038	Up
<i>COL4A6</i>	2.297	0.00005	0.006	Up
<i>BMP9</i> (<i>GDF2</i>)	NF	NF	NF	NC
<i>BMPR2</i>	-1.054	0.094	0.279	NC
<i>CYCLINE1</i> (<i>CCNE1</i>)	0.813	0.012	0.095	Up
<i>NOTCH1</i>	-0.26	0.034	0.159	NC
<i>NOTCH4</i>	-0.033	0.776	0.885	NC
<i>CDH5</i> (<i>CDH5</i>)	-0.032	0.806	0.903	Down

Table 4.2: GO analysis and *SOX17*-associated gene changes in RNAseq and qPCR analysis. NF, not found. NC, no change. Bold, significance value <0.05.

4.3.4 Knockdown of *SOX17* influences hPAEC functions

To determine the functional impact of *SOX17* dysfunction in hPAEC, the functional responses to relevant stimuli on hPAEC in culture were investigated following siRNA treatment. hPAEC proliferation in response to VEGF was increased as determined by MTT assays by *SOX17* knockdown ($p < 0.001$) (Figure 4.6A). However, this result is difficult to interpret as the positive control (VEGF) did not induce an increase in proliferation in control (siRNA-GAPDH) cells. Knockdown of *SOX17* led to an increase in hPAEC monolayer permeability measured by fluorescence in a Boyden Chamber assay when compared to a control siRNA condition ($p < 0.05$) (Figure 4.6B). hPAEC apoptosis in response to either TNF- α or LPS increased when compared to a siRNA control ($p < 0.001$) (Figure 4.6C). This was consistent with hPAEC viability which decreased when compared to a siRNA control with *SOX17*-siRNA with the addition of either TNF- α or LPS treatment (Figure 4.6D).

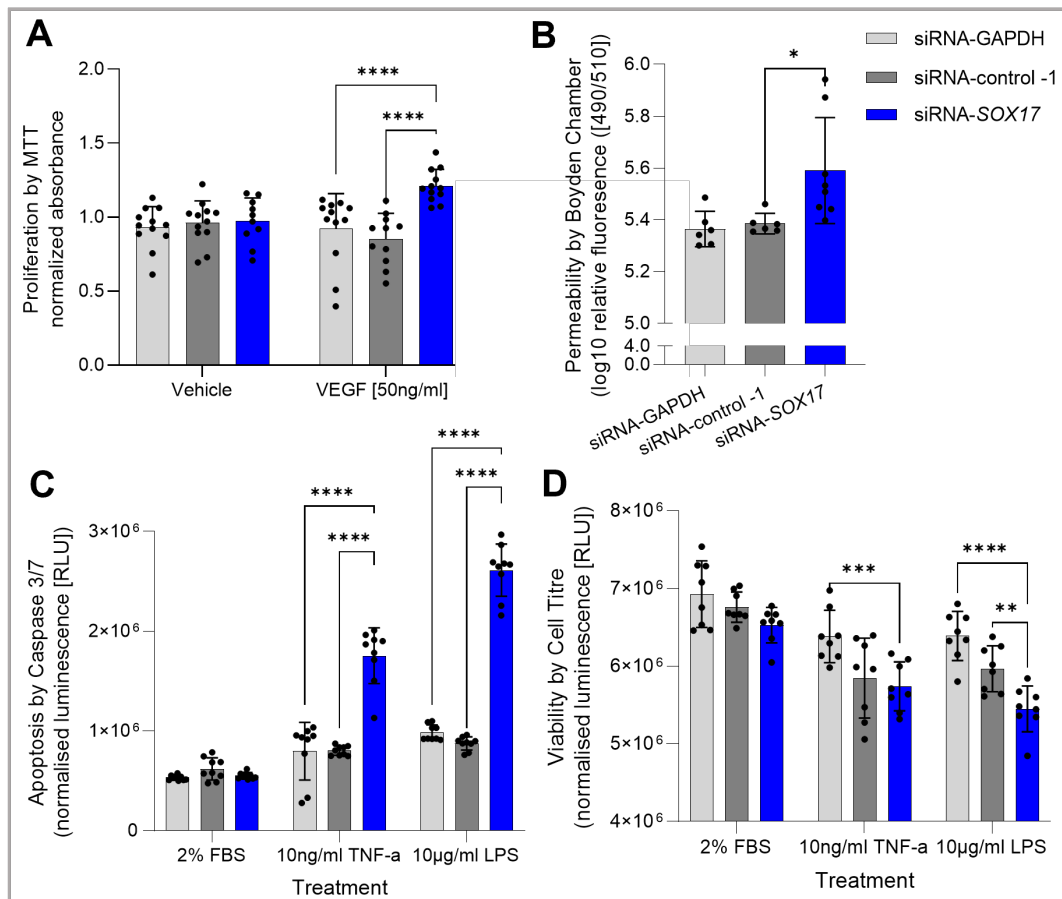


Figure 4.6

Effect of *SOX17*-siRNA on hPAEC function. A) MTT proliferation assay. B) Permeability barrier function assay. C) Caspase 3/7 apoptosis assay. D) Cell Titre viability assay. All are performed in hPAEC 48 hours after *SOX17*-siRNA. Statistical tests used in A, C and D are ordinary two-way ANOVA with Dunnett's multiple comparisons test. The statistical test used in B is an ordinary one-way ANOVA with Dunnett's multiple comparison test. * - $p < 0.05$. ** - $p < 0.01$. *** - $p < 0.005$. **** - $p < 0.001$. All show 3 separate experiments ($n=3$). Bars show mean ($\pm$ SEM).

4.3.5 Comparison of differentially expressed genes following CRISPR-INH of *SOX17*-signal 1 and 2 and *SOX17*-siRNA

To assess the similarities of DEG following transcriptional repression of *SOX17*-signal 1 and 2 and *SOX17*-siRNA the DEG lists were compared from hPAEC targeted by siRNA or CRISPR-INH of the *SOX17* enhancer regions. This was performed to narrow down genes which are important targets of *SOX17* and might be functionally relevant to the changes in hPAEC gene expression and function following *SOX17* loss. Up-regulated genes following *SOX17*-siRNA and CRISPR-INH of *SOX17*-signal 2 were found to have the most genes in common (43 from 191 potentially overlapping genes, 22.5%) (Figure 4.7A)(Table 4.3). The up-regulated genes from *SOX17*-siRNA and CRISPR-INH of *SOX17*-signal 1 had the second most genes in common (14). GO analysis of the 24

genes both up-regulated following *SOX17*-siRNA and CRISPR-INH of *SOX17*-signal 2 shows significant enrichment in different cancer pathways (microRNAs in cancer, bladder cancer, central carbon metabolism in cancer, endometrial cancer, non-small cell lung cancer) (Figure 4.7B).

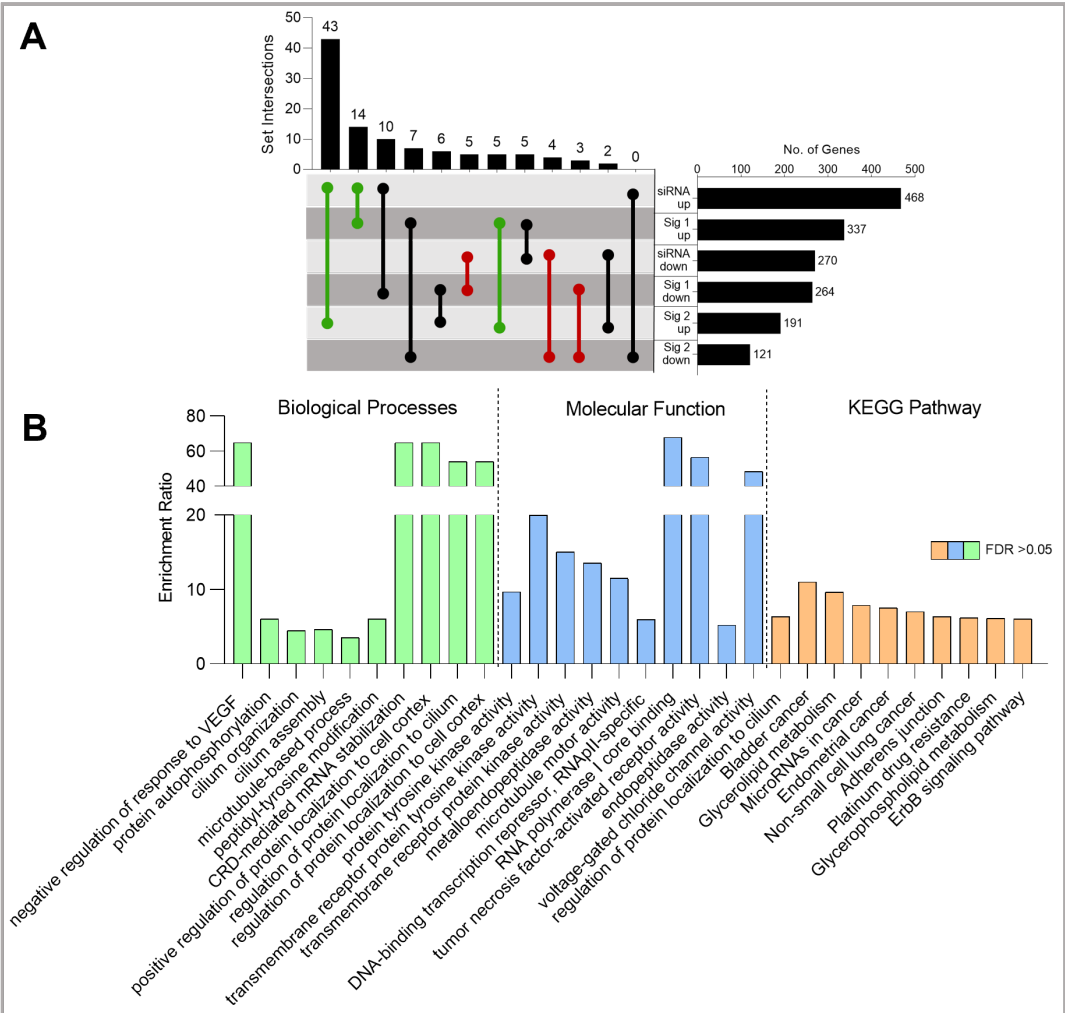


Figure 4.7
Comparison of DEG following CRISPR-INH of *SOX17*-signal 1 and 2 and *SOX17*-siRNA. **A)** Upset plot showing the interactions between differentially expressed gene lists from CRISPR-INH of *SOX17*-signal 1 and 2 and *SOX17*-siRNA. The gene lists with the most genes in common are *SOX17*-siRNA and CRISPR-INH of *SOX17*-signal 2 up-regulated genes (43 genes). **B)** Gene ontology analysis of KEGG pathway, biological process, and molecular function enrichment of 24 genes which are common between the *SOX17*-siRNA and CRISPR-INH of *SOX17*-signal 2 up-regulated lists. Lighter colours indicate FDR $q > 0.05$. Enrichment ratios were obtained from WebGestalt.

Gene Name	Entrez ID	LogFC in siRNA-SOX17	Statistical significance (p)	LogFC in SOX17-signal 2	Statistical significance (p)
<i>SNAPC4</i>	6621	0.61	0.01	0.35	0.07
<i>AP3M2</i>	10947	0.64	0.01	0.48	0.02
<i>CLCN6</i>	1185	0.66	0.01	0.50	0.03
<i>CROCC</i>	9696	0.67	0.01	0.47	0.06
<i>PNISR</i>	25957	0.69	0.01	0.53	0.01
<i>CCNL2</i>	81669	0.70	0.01	0.36	0.05
<i>IFI16</i>	3428	0.73	0.01	0.39	0.06
<i>SRSF5</i>	6430	0.74	0.01	0.27	0.07
<i>DEAF1</i>	10522	0.76	0.01	0.41	0.04
<i>NINL</i>	22981	0.77	0.01	0.74	0.04
<i>TPST2</i>	8459	0.81	0.01	0.54	0.05
<i>AHRR</i>	57491	0.83	0.01	0.52	0.04
<i>CLK4</i>	57396	0.88	0.01	0.44	0.04
<i>EPB41L2</i>	2037	0.9	0.01	0.42	0.05
<i>ZNF354B</i>	117608	0.93	0.01	0.44	0.08
<i>ARMC9</i>	80210	0.94	0.01	0.60	0.02
<i>LENG8</i>	114823	0.96	0.01	0.60	0.04
<i>TMEM246</i>	84302	1.02	0.01	0.74	0.06
<i>PKD1P1</i>	339044	1.03	0.01	1.31	0.05
<i>TNFRSF25</i>	8718	1.13	0.01	0.59	0.06
<i>KIFC2</i>	90990	1.2	0.01	0.73	0.04
<i>ZNF224</i>	7767	1.23	0.01	0.50	0.04
<i>ZNF767</i>	79970	1.25	0.01	0.67	0.09
<i>SIAE</i>	54414	1.27	0.01	0.49	0.04
<i>DNAH1</i>	25981	1.29	0.01	1.46	0.01
<i>NBPF8P</i>	728841	1.42	0.01	0.48	0.06
<i>DDR2</i>	4921	1.49	0.01	0.72	0.06
<i>RP11-1223D19.3</i>	-	1.54	0.01	0.88	0.01
<i>CCDC188</i>	388849	1.77	0.01	0.84	0.03
<i>ERBB2</i>	2064	2.05	0.01	0.91	0.07
<i>GPAT3</i>	84803	2.51	0.01	1.13	0.07
<i>ERV3-1</i>	2086	2.55	0.01	0.82	0.06
<i>FAP</i>	2191	3.13	0.01	1.23	0.01
<i>IGF2BP1</i>	10642	3.32	0.01	0.78	0.08
<i>MMP17</i>	4326	4.53	0.01	0.96	0.03
<i>PRKY</i>	5616	6.12	0.01	2.79	0.04
<i>USP9Y</i>	8287	6.33	0.01	2.40	0.06
<i>EIF1AY</i>	9086	7.68	0.01	2.14	0.10
<i>ZFY</i>	7544	8.40	0.01	2.62	0.13
<i>TXLNG2P</i>	246126	9.47	0.01	2.87	0.02
<i>RPS4Y1</i>	6192	9.63	0.01	3.12	0.08
<i>AC010084.1</i>	ENSG00000 229308.1	10.11	0.01	2.99	0.01
<i>ADAMTS12</i>	81792	10.95	0.01	2.98	0.01

Table 4.3: Differentially expressed genes in common between siRNA-SOX17 and SOX17-signal 2 hPAECs. All show entrez ID except *AC010084.1* which shows ensemble ID.

4.3.6 Knockdown of individual SOXF TFs by siRNA has an effect on other SOXF TFs in hPAECs

To determine if SOXF TFs have redundant gene targets in hPAECs the similarities between the DEGs following knockdown of each SOXF was assessed. To assess the effect of knockdown of each SOXF TF on the expression of the other SOXF TF, siRNA of each SOXF factor was used and the mRNA and protein levels were measured using qPCR and western blots (Figure 4.8). *SOX17* expression was significantly decreased with the addition of *SOX17*-siRNA and significantly increased with the addition of *SOX7*-siRNA (Figure 4.8Ai). The protein level of Sox17 was significantly decreased with *SOX17*-siRNA but did not significantly change with the addition of either *SOX7* or *SOX18*-siRNA (Figure 4.8Aii, iii). *SOX7* expression was significantly decreased with the addition of *SOX7* and *SOX17*-siRNA (Figure 4.8Bi). The protein level of Sox7 was significantly decreased with *SOX7* and *SOX17*-siRNA but did not significantly change with the addition of *SOX18*-siRNA (Figure 4.8Bii, iii). *SOX18* expression was significantly decreased with the addition of *SOX18* and *SOX7*-siRNA (Figure 4.8Ci). The protein level of Sox18 was significantly decreased with *SOX18* and *SOX7*-siRNA but did not significantly change with the addition of *SOX17*-siRNA (Figure 4.8Cii, iii).

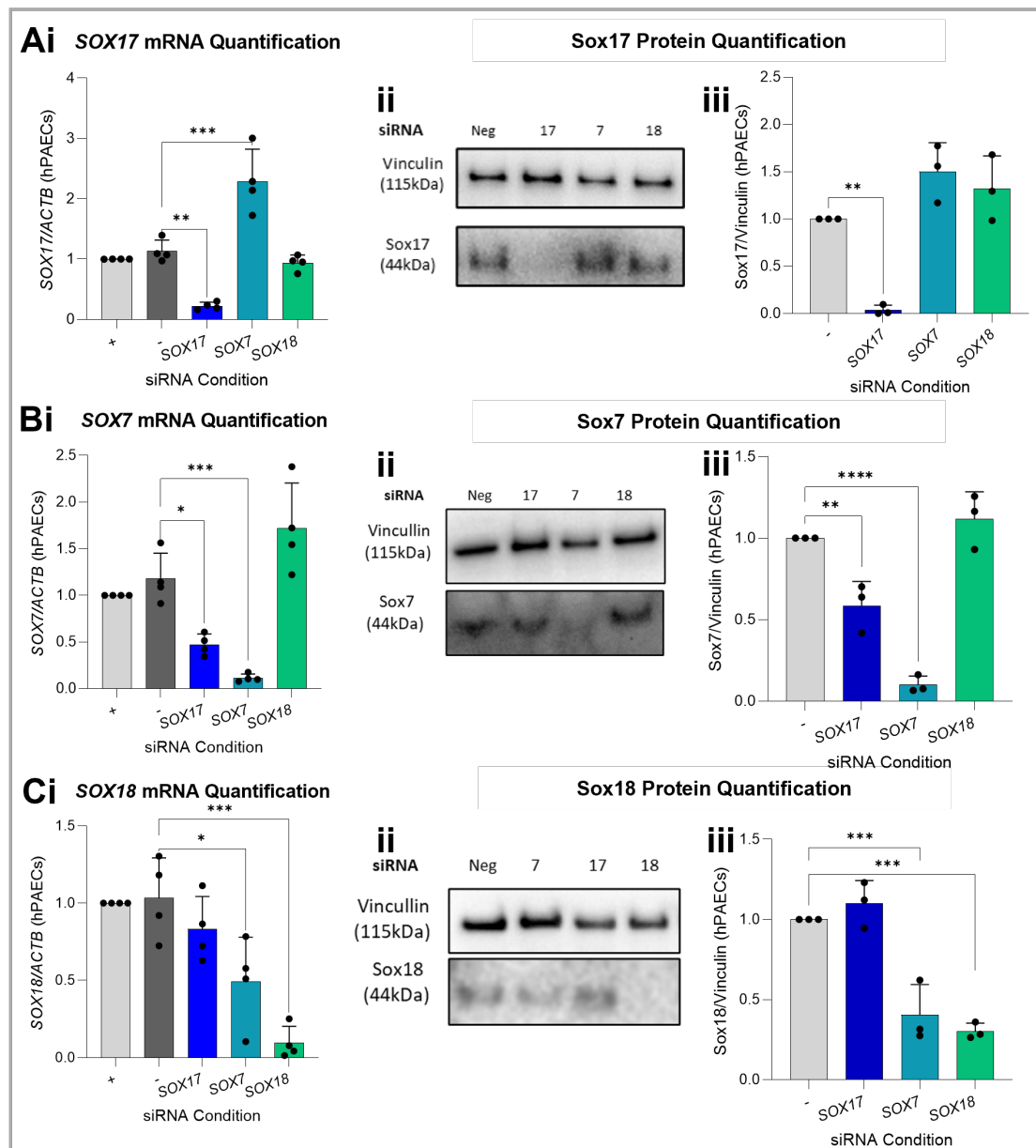


Figure 4.8

Knockdown of SOXF factors. A)i. Relative expression of *SOX17*. The change in *SOX17* expression is normalised to *ACTB* and all siRNA conditions are relative to the positive control. ii. Representative image of a western blot of Sox17 and vinculin protein in siRNA-*SOX17* and siRNA-negative samples. iii. Quantification of western blotting of Sox17. Sox17 protein level is normalised to vinculin and relative to the negative control. All in hPAECs following *SOX17*-siRNA treatment for 48 hours. B)i. Relative expression of *SOX7*. The change in *SOX7* expression is normalised to *ACTB* and all siRNA conditions are relative to the positive control. ii. Representative image of a western blot of Sox7 and vinculin protein in siRNA-*SOX7* and siRNA-negative samples. iii. Quantification of western blotting of Sox7. Sox7 protein level is normalised to vinculin and relative to the negative control. +, positive control. -, negative control. All in hPAECs following *SOX7*-siRNA treatment for 48 hours. **Continued...**

Figure 4.8

C)i. Relative expression of *SOX18*. The change in *SOX18* expression is normalised to *ACTB* and all siRNA conditions are relative to the positive control. **ii.** Representative image of a western blot of Sox18 and vinculin protein in siRNA-*SOX18* and siRNA-negative samples. **iii.** Quantification of western blotting of Sox18. Sox18 protein level is normalised to vinculin and relative to the negative control. All in hPAECs following *SOX18*-siRNA treatment for 48 hours. All tests show ordinary 1-way ANOVA with Dunnett's correction for multiple comparisons. * - $p < 0.05$. ** - $p < 0.01$. *** - $p < 0.005$, **** - $p < 0.001$. Averages of either three (protein) or four (mRNA) separate experiments ($n=3/4$). Bars show mean ($\pm$ SEM). Full western blot images are shown in supplemental figure 7.3.

To assess if different SOXF factors have overlapping or distinct downstream targets in hPAEC, DEG from hPAEC treated with siRNA for *SOX17*, *SOX18* and *SOX7* (Supplemental figure 7.1 and 7.2) were compared. Up-regulated genes following *SOX17* and *SOX18*-siRNA were found to have the most genes in common (265 from 1386 total genes, 19.12%) (Figure 4.9A) (Table 4.4). The down-regulated genes from these two lists had the second most genes in common (146). GO analysis of the 265 genes from the up-regulated following *SOX17* and *SOX18*-siRNA shows significant enrichment in biological processes, molecular functions and pathways involving adhesion and ECM (adherens junction, focal adhesion, biological adhesion, ECM organisation, extracellular structure constituent, organisation and interaction and ECM-receptor interaction) (Figure 4.9B).

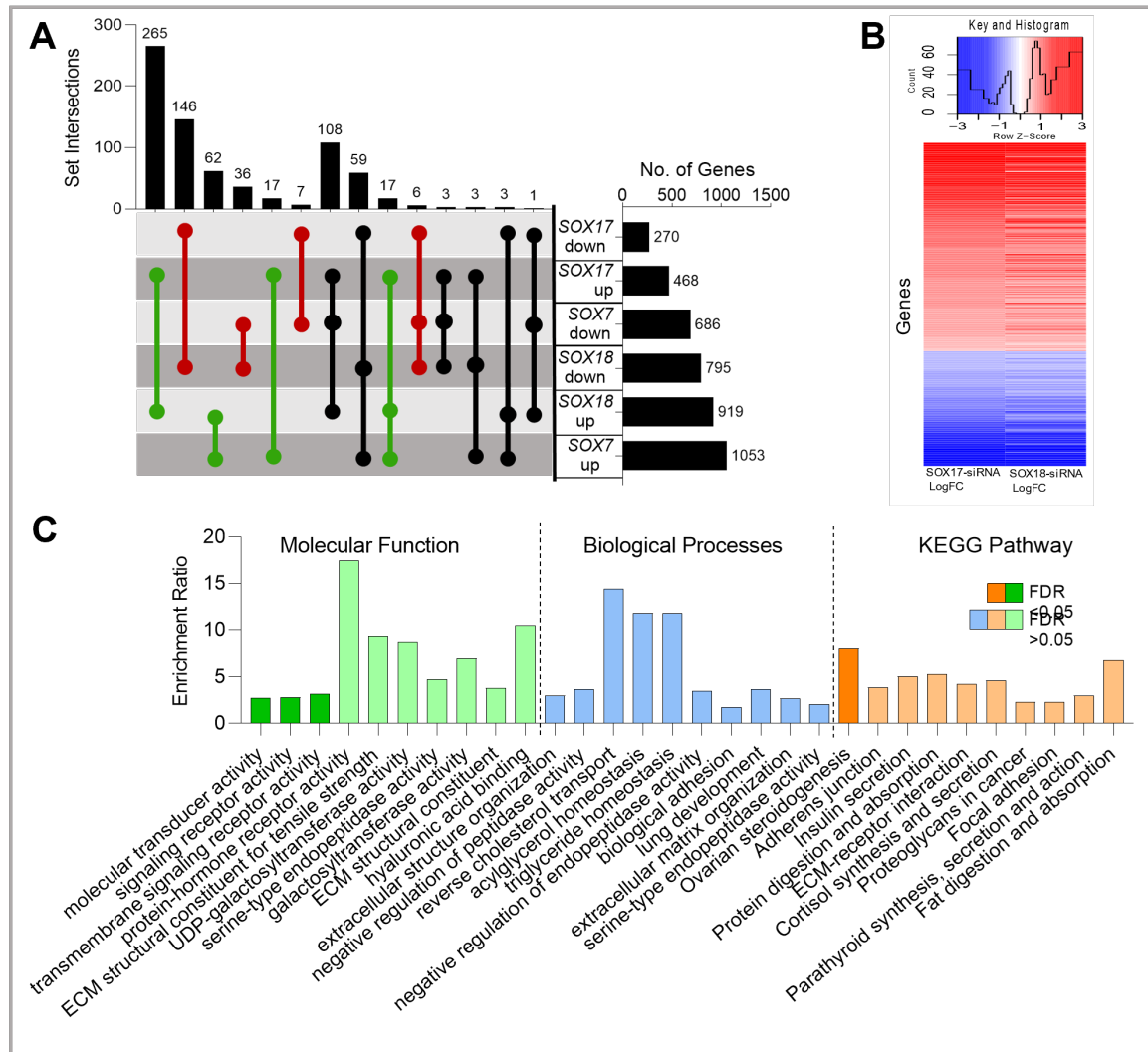


Figure 4.9

Comparison of knockdown of SOXF factors. **A)** Upset plot showing the interactions between differentially expressed gene lists from *SOX17*, *SOX18*, and *SOX7*-siRNA. The gene lists with the most genes in common are *SOX17*-siRNA and *SOX18*-siRNA up-regulated genes (265 genes). **B)** A heatmap of the 411 DEGs in common between *SOX17* and *SOX18*-siRNA in hPAECs. Red shows an increase in Log fold change (Log FC) and blue shows a decrease. **C)** Gene ontology analysis of KEGG pathway, biological process and molecular function enrichment of 265 genes which are common between the *SOX17*-siRNA and *SOX18*-siRNA up-regulated lists. Darker colours indicate FDR $q < 0.05$. Lighter colours indicate FDR $q > 0.05$. Enrichment ratios were obtained from WebGestalt.

Comparison Lists	Up regulated Genes	Down regulated Genes
SOX17 and SOX18 siRNA	<i>HOXB7,HMG20B,GTF2F1,MAP4K5,IVD,PLCB3,GLMP,UBXN1,HCLS1,ACAA1,ANAPC5,RPL10A,GATAD1,NOB1,PLD2,CARD8,RNF7,PLA2G15,CPSF4,PBX2,SLC25A43,SCARB1,NOA1,SNAPC4,KIAA0556,B4GALT6,NAE1,EEF1A1P5,PLEKHO1,MT-RNR1,C7orf43,CLCN6,RHOB1,ALAD,PDPR,MRI1,VAMP2,CMTM3,ETV5,ANKRD13D,P4HA3,CFAP36,CCS,SARS2,PCSK1,PNISR,RP11-164A7.1,CCNL2,GMFG,VAV2,TCTN3,ST3GAL4,BCL2L2,RP11-117P22.1,SRCEEF1A1,CENPT,IFI16,RRP7A,GATSL3,CERCAM,LIMD2,ICA1,NPEPL1,PEX16,SRSF5,RNASEH2B,MOAP1,NLGN2,DEAF1,IFFO1,NAT9,NINL,SPRED3,ZBTB47,MYL9,CDK11A,TPGS1,POLR1E,TPST2,SLC17A9,ADIRF,CECR1,CBWD1,ATF6B,ORAOV1,AMDHD2,RPAIN,LYSMD4,HDAC9,GPR89B,B3GALNT1,SS18L2,PCSK7,DYNC2LI1,FSD1,RPL9P9,TSEN54,SLC25A28,CTXN1,LIG1,CLK4,HLAC,SNRNP48,AP000695.6,FBXO4,EPB41L2,CAPG,DECR2,SPTBN2,CIRBP,GJC1,ANKRD9,IDUA,ZNF354B,EP400NL,GOLGA8B,ST3GAL6,P3H3,ZNF287,FAM175A,ARMC9,SUSD5,SEPW1,CD276,NAMPTL,DZIP1,FBXW7,EPSS8L1,LENG8,ABC7-42389800N19.1,HNRNPDL,CFAP54,C11orf96,RP11-717F1.2,GCAT,LTBR,CLK1,CENPV,MMP14,CLEC11A,LMBR1L,RHOD,AP000695.4,ZNF266,TRPM4,CARD11,HDAC10,ADCY3,LOXL4,MEIS2,NR5A2,ZNF211,TTLL12,RGS10,COL13A1,SNHG4,ATHL1,NRIP3,TNFRSF25,NUDT12,TMEM17,QPCTL,ZBED3-AS1,</i>	<i>RC3H2,NCOR1,SPEN,ASCC3,SLC39A9,HCFC1,CGGBP1,SMAD4,HUWE1,FAM43A,PPP3CA,PPP1CB,CCNI,SERTAD2,CNOT6,FAM127B,CTDSP2,PANK3,NHLRC2,HEG1,STT3A,STK40,SLC39A10,MED13,KIAA0430,PROS1,SMC1A,CREB1,RP11-395L14.11,SYT11,PTBP3,FOXO1,KMT2D,PPME1,NECTIN3,ZBED6,TAB2,PTP4A1,RNF152,TBL2,KLF12,IGF1R,ZXDB,PKIA,XBP1,PELI1,NBDY,NELFCD,CENPN,UGDH,CYB5R3,YBX3,CNOT6L,CA5BP,SP4,DIAPH3,HELZ,PLA2R1,CYFIP2,DDAH1,FAM13A,ABLIM1,KCNK6,NFIB,IGF2BP2,SGCE,RP11-470M1.2,C7orf13,OPN3,KIAA1715,RPL9,NOMO1,TFPI2,ATP5G3,NTN4,ATP11B,WDFY3,ZFH3,CDK6,ARID5B,PSPH,CRIM1,EGFR,PCDHGB5,C22orf34,AHNAK,SYNGR2,C9orf64,INSR,XRRA1,EDA2R,TMEM99,C1orf115,CDH13,PNPLA4,AC011325.1,GCNT1,NDN,TMEM88,AP005901.1,HIST1H2BD,RP11-46B11.5,MAMLD1,USP32P1,ASNS,POMZP3,NEGR1,ADM2,HSD17B2,CASP1,LAMP3,CTD-2303H24.2,COL8A1,AIF1L,AC009948.3,GBP3,CCDC144B,SDHAF3,ITM2A,ROBO1,LMO7,GALNT15,RCN3,HLA-L,TOX,AC084219.4,TIMP3,PEG10,ECHDC3,CLIC2,GALNT8,HSD17B1,FBN2,RP11-178A10.1,SMOC1,HOPX,FAM86DP,HLA-J,RPL13P12,RP11-411H5.1</i>

Comparison Lists	Up regulated Genes	Down regulated Genes
SOX17 and SOX18 siRNA	SMAGP,RP11-252A24.2, GOLGA6L9,KLHDC8B,AGAP6, MMP11,ZNF224,LY6G5C,NAAA, KLF2,SRPX,RPL32P29,CCT6P3, B3GNT8,ANPEP,KLHL6, AC084018.1,DPP4,LAYN, RP11-536O18.2,C15orf52, FAM110D,SDC1,STRIP2,SRPX2, RP1-222E13.10,HLA-H,CD44, RENBP,PSG4,IL27RA,RGS14, NFE4,CORO6,PLCXD1,GLB1L, SHANK1,ADAMTS10,ABCA1, AC005682.5,MTTP,BAHCC1, GPR162,ZNF135,IFI44,LY6G5B, AC004837.3,COL4A5,PTPRU, PDE1C,C17orf97, CETP,PLPP2,LGR4,PLEKHG4, TPD52L1,CGREF1,ERBB2 ,ADORA2B,GRIK4,GPR3, ADAMTS4,NPIPA5,DRP2,ITGA1, COL4A6,RP11-371A22.1,PRSS3, NSG1,RPL22L1,MR1,TMEM163, RP11-152F13.3,GSTM3,ZNF154, GFPT2,LL22NC03-2H8.5,LAMB3, ERV3-1,MYO5B,AC011306.2,FAP, PARP15,IGF2BP1,VAMP8,TPM2, MFAP2,SCUBE1,ZNF415,GPAT2, TYW1B,CDH2,TNFSF4,CYP1B1, ARHGEF10L,RGS7BP,PTPN22, CYTL1,RASAL3,G0S2,TBC1D3E	
SOX17 and SOX7 siRNA	NOSIP,SUCLG2,BCAR3,SMOX, BNC1,NIPAL1,NPR2,CPNE7, RP11-577H5.5,GSTT2B,WNK4, TRPV2,SLIT2,SLC38A5,FGF5, CCNA1,PAX8-AS1	NXN,PDE4D,COL6A2, TSHZ3,PELI2,LITAF,SPINT2

Comparison Lists	Up regulated Genes	Down regulated Genes
<i>SOX18</i> and <i>SOX7</i> siRNA	<i>IQSEC1, EXT2, MMAB, PLSCR3, GALK1, DHCR7, SRSF2, GLIS2, KIRREL, RNF180, HACD3, STARD4, NFIX, NSDHL, INSIG1, CYP51A1, ACAT2, MVD, SGPP1, CXXC5, DNAJC15, BASP1, FDFT1, PRPH2, EMP3, SHC3, CAPN5, MCM6, ADAMTSL1, TMEM97, LOXL1, ZNF703, STEAP3, NBPFL4, PODXL2, KRT7, NUDT11, KRT19, PTPRJ, LAMC2, ABI3BP, SEMA3D, NNAT, SFRP1, EDIL3</i>	<i>RASGRF2, CELSR1, CBS, TMEM246, PPP1R9A, LXN, CLEC2B, DLL4, RNF144A, ADD3, CYBA, AIG1, ADGRF5, PTPRN2, HOXB5, SORT1, APOD, TSPAN12, HOXD1, PPP1R16B, GAB1, FRAS1, VWA5A, GRK5, LRCH2, CADPS2, DIAPH2, CTD-3113P16.11, MYH10, RNF40</i>

Table 4.4: Differentially expressed genes in common between siRNA conditions.

4.4 Discussion

The chapter aimed to investigate the effect of loss of *SOX17* in hPAEC on gene expression and cell function to assess the possible role(s) of *SOX17* dysfunction in PAH. *SOX17* was knocked down through siRNA and CRISPR–INH and the effect on downstream pathways and functions in hPAECs was assessed. RNAseq and GO analysis demonstrated that *SOX17* regulates functional molecular pathways and processes involved in regulating ECM, adhesion and cancer development in hPAEC. Potential downstream targets of *SOX17* knockdown which might affect these pathways (*ADAMTS12*, *MMP17*, *LAMB3* and *COL4A6*) and other pathways important to PAH (*NOTCH4*, *CYCLINE1* and *CDH5*) have also been validated in independent siRNA experiments by qPCR. Knock-down of *SOX17* via siRNA was shown to affect proliferation, cell viability, apoptosis, and to a small degree, barrier function in hPAECs; functions that are known to be important in PAH. The work highlighted here also identifies crossover between genes regulated by different SOXF TFs in hPAECs.

SOX17 has established roles in systemic artery endothelial cells, however, there is little known about its role in the pulmonary vasculature. In this chapter, it was shown that a loss of *SOX17* affects the expression of genes involved in cell adhesion and different functions of the ECM (structural constituent, organisation, receptor interaction) in hPAEC. ECM has recently been investigated as an important factor in PAH. It plays an important role in the organisation and regulation of the basement membrane in vessels and is in direct contact with ECs in some smaller vessels. The basement membrane is thicker in the lungs of iPAH patients than controls [Jandl et al., 2020]. In this work, genes that were differentially expressed following *SOX17* knockdown and highly enriched in GO analysis

for ECM functions were shown to be affected at the mRNA level in a second validation study. More work is required to validate these at a protein level, but this work suggests an important role for *SOX17* in the maintenance of the ECM and potentially regulation of the basement membrane in PAH. The genes (*LAMB3*, *COL4A6*, *MMP17* and *ADAMTS12*) have previously been shown to have an interaction with *SOX17* [Niakan et al., 2010, Schachterle et al., 2017, Ha et al., 2019, Takahashi et al., 2020]. *LAMB3* produces the laminin subunit beta 3, a key protein of the ECM. *COL4A6* encodes for one of the six subunits of collagen type 4. Both type 4 collagen and laminin have been shown to strengthen EC barrier integrity, with an increase in collagen type 4 also increasing cell adhesion in PAH-relevant cell types [Jandl et al., 2020]. This study does not investigate the possible cause of increased laminin or collagen type 4 in iPAH samples but in this chapter, it was shown in two independent sets of experiments that a decrease in *SOX17* leads to a significant increase in both *LAMB3* and *COL4A6*, potentially elucidating part of a mechanism which explains the potential change in adhesion and ECM functions.

Work in this chapter showed a significant increase in both *ADAMTS12* and *MMP17* expression following a loss of *SOX17*, suggesting they may be important contributors to the functional effects of *SOX17* loss in hPAEC. *ADAMTS12* is a disintegrin and matrix metalloproteinase gene with an important role in ECM composition (for full review: [Mohamedi et al., 2021]). Coupled with *MMP17* it is part of a large family of matrix metalloproteinase genes which may have an important role in PH (for full review: [Chelladurai et al., 2012]). MMPs have important roles in pathophysiological functions in hPAECs in PAH. Although *MMP17* has not been directly linked to PAH, loss of function variants, found through in familial studies, confer a greater risk of aortic aneurysm in mice through dysfunctional ECM filament deposition and an enlarged aortic lumen [Alonso et al., 2015]. In this thesis it was shown that an increase in *MMP17* occurs after loss of *SOX17*, suggesting up regulation of *MMP17* in patients with PAH with *SOX17* dysfunction may affect artery walls and regulation of this gene is important in vessel wall homeostasis. However, further work is required to establish the link between *SOX17* and these genes. For example, the promoter regions of *ADAMTS12*, *MMP17*, *LAMB3* and *COL4A6* should be assessed for Sox17 binding motifs to investigate if they are directly influenced by Sox17 levels or if they are part of a wider pathway. The effect of Sox17 levels on hPAEC adhesion should also be assessed using functional assays. If adhesion is significantly affected following *SOX17* knockdown, it will provide a further function which could be targeted with therapeutics to correct for *SOX17* dysfunction.

SOX17 has been linked to several PAH-relevant genes and pathways (Section 1.5.2 and 1.5.3). Of the *SOX17*-related genes tested, only *NOTCH4* and *CDH5* showed a significant difference after *SOX17*-siRNA knockdown. *CYCLINE1* showed a nominal but not significant increase following *SOX17* knockdown. The lack of effect on the expression of

BMP9, *BMPR2*, *SOX2* and *SOX4* is not unexpected as, at the time of writing, there has been no previous work suggesting a direct link between *SOX17* and their expression and no work searching for *SOX17* motifs in the regulatory regions affecting these genes has been published. However, it is well established that *SOX17* directly binds to regulatory loci which affect *NOTCH4* expression and this relationship leads to regulation of arterial fate identity [Corada et al., 2013, Clarke et al., 2013]. *SOX17* knockdown significantly affected *NOTCH4* expression here, suggesting that this relationship is relevant in hPAECs.

Previous work in mouse intra-aortic hematopoietic cluster cells has shown a direct link between *SOX17* and *CDH5* through binding of *SOX17* to the promoter of *CDH5*, which encodes a protein (VE-cadherin) that is essential for forming EC adherens junctions [Dejana et al., 1999]. Loss of VE-cadherin in endothelial cells causes apoptosis and an increase in EC barrier permeability [Corada et al., 2001]. Barrier permeability functional work carried out in this thesis show that a decrease in *SOX17* in hPAECs negatively affects endothelial barrier formation, creating a leaky monolayer. However, although there was a negative effect on barrier function with *SOX17* knockdown, it was very small. Therefore, in the future, it would be beneficial to assess barrier function with a control such as a weak detergent to investigate further the role of *SOX17* in barrier function. hPAEC apoptosis and survival shows an increase in apoptosis and decrease in cell viability with the addition of either TNF- α or LPS stimuli in hPAEC with reduced *SOX17* levels. Interestingly, it has also been shown that mice ECs which lack VE-cadherin are unable to respond to the anti-apoptotic, survival effects of VEGF [Corada et al., 2001, Hebda et al., 2013]. Although the effects of VEGF on apoptosis and survival following *SOX17* knockdown were not assessed in this thesis, it would be an interesting avenue of enquiry to investigate whether the loss of response to VEGF due to a decrease in VE-cadherin occurs in hPAECs. *CYCLINE1* encodes a protein that regulates members of the cyclin-kinase family, well-documented regulators of the cell cycle. It has been previously shown in a mouse vascular injury model that *SOX17* increases *CYCLINE1* expression which leads to increased EC regeneration and viability [Liu et al., 2019]. Interestingly, the data shown in this thesis, a knockdown of *SOX17* leads to a non-significant increase in *CYCLINE1* in hPAECs. The discrepancy between these findings could be contributed to the use of mouse ECs vrs human hPAEC. Different endothelial cell types within an organism have different signalling pathways and react to stimuli in different ways depending on their immediate environment [Conway and Carmeliet, 2004]. For example, HUVECS and brain microvascular ECs have been shown to have different tight junction formation which affects their barrier function [Uwamori et al., 2019]. Furthermore, there are species-specific differences in ECs between mice and humans [Song et al., 2020]. Although there is consistent up-regulation of *CYCLINE1* in both qPCR and siRNA work shown in this thesis,

additional repeats would need to be completed to investigate the link between *SOX17* knockdown and *CYCLINE1* in hPAEC. Additionally, to further investigate the effect of *SOX17* on *CYCLINE1* in hPAEC, the promoter of *CYCLINE1* could be explored for an open Sox17 binding motif to assess its active binding.

SOX17 has been well investigated as a tumour suppressor gene (for full review: [Tan et al., 2020]) and it is shown here that loss of *SOX17* through CRISPR–INH of upstream enhancer regions or knockdown of mRNA levels affects pathways relating to cancer (MicroRNAs in cancer, bladder, endometrial and non-small cell lung cancer). It has been shown that endometrial cancer patients with longer recurrence-free survival have higher *SOX17* levels and an overexpression of *SOX17* causes tumour cell proliferation suppression [Zhang et al., 2016]. In the work shown in this thesis, *SOX17* knockdown in hPAECs causes an increase in proliferation with the addition of VEGF. More recent work has also shown the *SOX17* promoter to be under hypermethylation in lung cancer [Liao et al., 2019]. This causes decreased expression of the gene, proliferation of cancer cells and a decrease in *SOX17* target genes. However, this hypermethylation of the *SOX17* promoter was not found in a recent study of promoter hypermethylation in iPAH pulmonary ECs [Hautefort et al., 2017]. Although the genes which influence the enriched pathways shown in this thesis need validation in separate experiments, it is reassuring that pathways that are already heavily linked to *SOX17* are enriched in hPAECs as it adds additional validation to this work.

Previous research has shown redundancy between the three members of the SOXF subgroup of TFs. It has been shown that although there may be some functional redundancy, especially between Sox17 and Sox7, complete loss of one will not be completely compensated for by another [Kamachi and Kondoh, 2013, Corada et al., 2013, Zhou et al., 2015]. Interestingly, the work shown in this thesis suggests a loss of *SOX7* through siRNA knockdown results in an increase in *SOX17* expression, although this is not coupled with a significant increase of Sox17 protein at the time points investigated in this thesis. Furthermore, a decrease in *SOX17* by siRNA results in lower levels of *SOX7* at the mRNA and protein level. *SOX18* at both the mRNA and protein level is decreased by siRNA targeting *SOX7*. These effects are unlikely to be caused by off-target effects of the siRNA constructs as they target specific regions of each SOXF which are not shared by the others. Therefore this suggests a potential feedback mechanism in hPAECs, especially between *SOX7* and the other two members of the family. More work is needed to fully investigate the relationship between these TFs and their possible role in the effectiveness of compounds used to rescue *SOX17* dysfunction. It has been shown that although the HMG domain is conserved and the SOXF factors have very similar DNA binding domains [Lange et al., 2009], they target different genes and can have effects on mutually exclusive pathways. When comparing the differentially regulated genes following knockdown of SOXF

TFs, similar genes are differentially regulated following *SOX17* and *SOX18* knockdown (411 DEG), but the number of genes that are similar between *SOX7* and either *SOX17* or *SOX18* are much lower (97 and 24 DEG respectively). The genes which are similar between *SOX17* and *SOX18* knockdown hPAECs show enrichment for pathways, functions and processes involved in ECM and adhesion, the same pathways which are enriched when only *SOX17*-siRNA DEG are considered. This implies that *SOX17* and *SOX18* regulate similar pathways in hPAECs but further work is needed to assess the direct link between the SOXF factors, these 411 genes and the functional role this plays. However, it is possible that *SOX18* could compensate for loss of *SOX17*. This has important implications for therapies which aim to correct for *SOX17* dysfunction and provides a potential second target (*SOX18*) for these studies. If *SOX17* and *SOX18*, through further study, are shown to affect similar pathways and functions, it could become a druggable target in the future.

4.5 Conclusion

The work discussed in this chapter begins to define the transcriptomic, functional and pathway consequences of *SOX17* dysfunction in hPAECs. Gene pathways related to ECM and adhesion were confirmed in an independent validation, establishing important roles of *SOX17* in hPAECs. Proliferation, viability, and apoptosis of hPAEC are affected by the loss of *SOX17*. The chapter also suggests an overlap or redundancy in the gene regulatory relationship between the different members of the SOXF family of TFs. Overall, this work suggests *SOX17* has an important effect on hPAEC genes and functions which are relevant to PAH.

Chapter 5 focuses on using transcriptional data from this thesis to predict existing compounds for repurposing in PAH, specifically looking for compounds to correct *SOX17* dysfunction.

5 Repurposing of Drug Compounds to Rescue Loss of SOX17 Function in PAH

5.1 Introduction

Discovery and investigation of a novel compound or therapeutic for any disease can take many years, significant investment and less than 10% of drugs entering Phase 1 are awarded a licence. For diseases that are less well studied or require more complex investigation to discover 'druggable' targets, this percentage can be even higher [Hingorani et al., 2019]. Clinical trials can fail at any stage but the most common reasons for failure are safety concerns, minimal or no efficacy on changing disease outcome(s) and financial impact/costs. To minimise the first and last of these issues, compounds that have already undergone trials for other diseases can be repurposed for new diseases. This can reduce possible failures due to safety concerns when the compound has already been demonstrated as safe in humans and reduces costs. There are current therapies used in PAH which have been repurposed from other diseases. One example is Sildenafil which was originally tested (unsuccessfully) for use in angina patients, became a well-used erectile dysfunction drug, and was eventually tested in PAH patients and approved for use in 2006 [Galiè et al., 2009, Wang et al., 2014]. However, even with faster and cheaper approval of compounds for use in PAH which repurposing can bring about, current therapies available for the treatment of PAH often do not reverse pulmonary artery remodelling. Furthermore, there is well recognised inter-individual variation in response to licensed compounds.

Personalised medicine involves the tailoring of therapies to a specific patient. This can be achieved by the combination of -omics analysis to provide the most complete picture of a patient's disease. By having more information on the genetic and molecular cause(s) of a patient's disease, therapies can be used which target these causes and therapies which do not (and would likely not affect the patient's outcome) will not be prescribed. In the future, this could allow for a more personalised approach to PAH diagnosis and treatment. Knowledge of specific genetic and molecular cause(s) of disease can also provide a better understanding of the druggable targets that should be investigated. The CMap offers a novel pathway of drug repurposing coupled with transcriptomic analysis and will be discussed in this chapter [Lamb et al., 2006, Subramanian et al., 2017].

The CMap is an online database of genetic signatures which can be used to find connections between diseases, compounds, genetic manipulations and cellular processes [Lamb et al., 2006]. It has previously been used to improve biological understanding of pathways, diseases, or compounds, discover drug pathways and mechanisms of action, investigate novel compounds and foster discovery of drugs and compounds which can be re-purposed

(for a full and comprehensive review see: [Qu and Rajpal, 2012]). It has also been used to investigate existing transcriptomic datasets to identify a compound to reverse PAH [Gu et al., 2021].

Using multiple cell types, the creators have assessed transcriptomic changes resulting from exposure to different 'perturbagens'. A perturbagen is a reagent used for cell treatment and determination of resulting biological response and covers both gene over-expression and knock-down as well as small compounds such as drugs. Genetic manipulation is achieved using CRISPR/Cas9 constructs, short hairpin RNA and open reading frames for editing, knock down and overexpression respectively. The resulting genetic changes are grouped together from multiple replicates to create a genetic signature. Therefore, one perturbagen will have multiple signatures, one for each cell line and set of conditions used. To assess gene expression changes resulting from a perturbagen, it is compared to a control population which is a representation of all the other perturbagens which were tested on the same physical plate. This shows how a single perturbagen affects the cell line when compared to other perturbagens tested alongside it. Relevant controls for genetic manipulation and compound perturbagens were also included.

The CMap enables the assessment of similarity of any genetic signature with the signatures of all the perturbagens so far tested (reference signatures). Lists of up- or down-regulated genes can be queried against all reference signatures, producing a 'connectivity' or tau score from -100 to +100 which reflects the correlation between the supplied signature and the registered perturbagen effects. A query results in connectivity scores for all comparisons between your signature and the entirety of the reference signatures. This score refers to how strongly your signature is connected to the signature of the specific perturbagen in the specific cell type and an overall connectivity score is provided (Figure 5.1). For example, a score of +95 means that only 5% of reference signatures are more strongly positively correlated with this perturbagen than the user's input signature. A positive score indicates that the signature of this perturbagen is similar, whereas a negative score suggests they are opposing. No connection indicates there is no similarity in genetic changes between the two signatures (Figure 5.1).

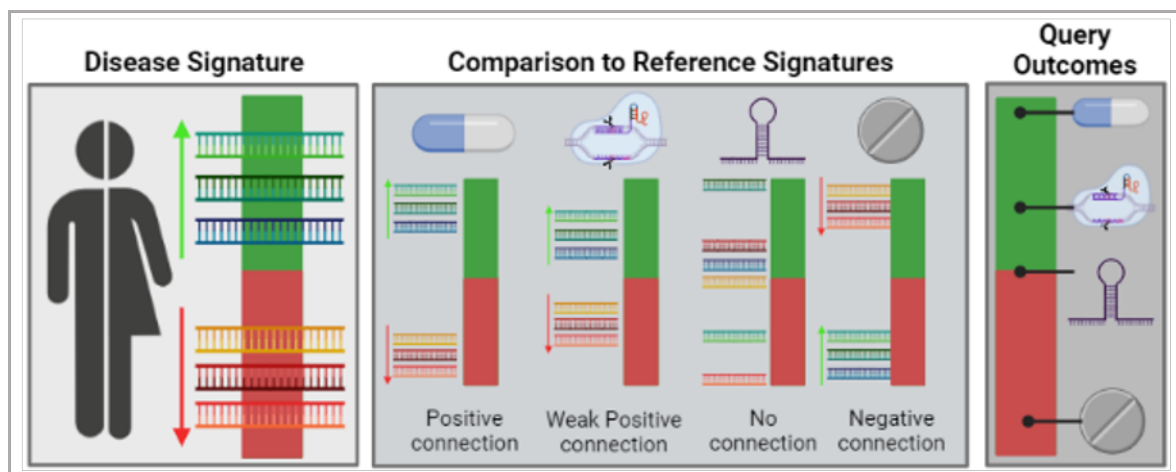


Figure 5.1

Connectivity map workflow. A representation of the workflow of the Connectivity Map. A disease signature contains lists of differentially regulated genes (represented by RNA ladders coloured by the direction of change) and is compared to all reference signatures within the CMap. The query outcome shows how connected all reference signatures are to the disease signature. Adapted from [Lamb et al., 2006].

Previous work has utilised the CMap for drug discovery in PAH [Gu et al., 2021]. AG1296 was identified as a potential compound for use in PAH and was demonstrated to reverse several PAH molecular and functional phenotypes in both human cells and the PAH Sugen/hypoxia rat model. This study demonstrated the potential of identifying drug candidates at the -omics level, in comparison to more traditional ‘single target’ approaches and highlights how in silico studies can be used to re-purpose compounds for PAH and how this has been applied to PAH therapeutic research.

Previous chapters in this thesis have discussed evidence for both the role of *SOX17* in hPAECs and how the dysfunction of this gene could lead to PAH. *SOX17* is a well-documented TF with known roles in EC functions including proliferation, apoptosis and barrier function. This thesis has provided evidence of downstream targets which might be affected by *SOX17* dysfunction in hPAECs such as *CDH5*, *NOTCH4*, *ADAMTS12*, *MMP17*, *LAMB3* and *COL4a6*. Changes in these functions or genes following *SOX17* dysfunction is a potential novel therapeutic avenue for investigation in PAH. By utilising the CMap and investigating compounds that could correct for *SOX17* dysfunction, it may be possible to provide therapies for repurposing in patients with *SOX17* dysfunction in PAH.

5.1.1 Hypothesis

Compounds with opposing effects on gene expression profiles to loss of *SOX17* will reverse the effects of *SOX17* dysfunction in hPAEC.

5.1.2 Aims

1. To investigate the CMap for novel compounds for use in correcting *SOX17*-dysfunction in PAH.
2. To expose PAECs to candidate compounds and assess changes in *SOX17*-associated gene expression.

5.2 Methods

5.2.1 Connectivity Map Query and Analysis

To interrogate the CMap, DEG lists were created and used to query the connectivity map for compounds which could cause an opposite effect to the loss of *SOX17*. This included compounds which had a positive connection to an increase in *SOX17* (CRISPR activation (CRISPR–ACT) of the *SOX17* promoter) and compounds which had a negative connection to a decrease in *SOX17* (CRISPR–INH of the *SOX17* promoter and/or *SOX17*-signal 1). The tau score ranges from -100 to +100 and reflects the correlation between the supplied signature and the registered perturbagen effects. A query results in connectivity scores for all comparisons between the input signature and all the reference signatures within the CMap. This score refers to how strongly the input signature is connected to the signature of the specific perturbagen in the specific cell type and an overall connectivity score is provided. A positive score indicates that the signature of this perturbagen is similar, whereas a negative score suggests they are opposing. From the queries, sirolimus, aminopurvalanol-a and YK-4279 were selected. To assess the effect of these compounds on the genetic changes which occur following *SOX17* dysfunction genes were selected from the CMap compound signature. The genes selected were consistently perturbed (had a high z-score) in the compound signature across five cell lines or more and were present in the condition gene list. This enabled the selection of highly correlated genes whose gene expression could be assessed in the future to investigate the accuracy of the CMap and allow us to assess compound effect in PAECs.

5.2.2 Statistical Analysis

To assess the change in the relative expression of genes following exposure to CMap perturbagens, the expression of the gene relative to *ACTB* was graphed with all conditions relative to the DMSO control. To establish statistical significance, an unpaired, one-way student's t test was performed on the average of three experiments between the DMSO control group and the exposure groups using GraphPad Prism (version 8.0.2, GraphPad Software, San Diego, California, USA).

5.3 Results

5.3.1 Analysis of differentially expressed genes following *SOX17* promoter and signal 1 manipulation

To create DEG lists for use in the CMap, RNAseq analysis of hPAEC following CRISPR–INH of the *SOX17* promoter and *SOX17*-signal 1 (rs10958403) and CRISPR–ACT of the *SOX17* promoter was undertaken by previous members of my lab. 621, 519 and 579 genes were identified which were significantly ($p < 0.05$) down- or up-regulated by more than 0.25 Log_2 fold change following CRISPR–ACT of the *SOX17* promoter (Figure 5.2), CRISPR–INH of the *SOX17* promoter (Figure 5.3) or following CRISPR–INH of *SOX17*-signal 1 (rs10958403) (Figure 5.4), respectively. The most differentially expressed genes (maximum 150 up and 150 down as defined by CMap input requirements) from these lists were selected to query the CMap and are shown in supplemental tables 7.3, 7.4 and 7.5.

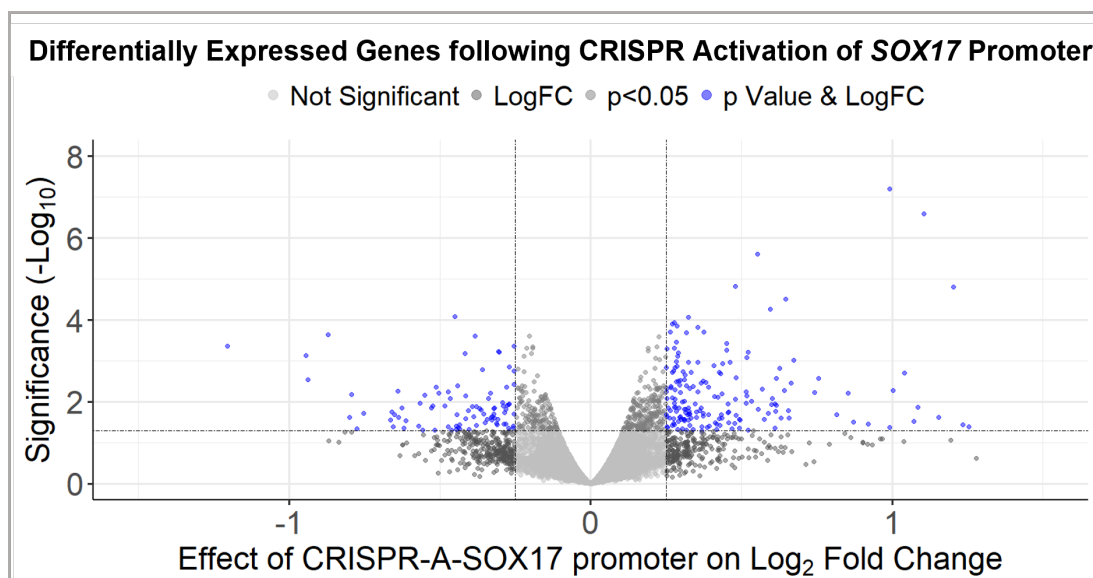


Figure 5.2

Differentially expressed genes following CRISPR-activation of *SOX17* promoter in hPAECs. Volcano plot of the log_2 fold change (FC) between CRISPR-ACT of *SOX17* Promoter and BFP Control and the negative log_{10} p-value. Differentially expressed genes shown in blue met the cutoff points: $p < 0.05$ and Log_2 fold change < -0.25 or > 0.25 . $n=4$, 12707 variables.

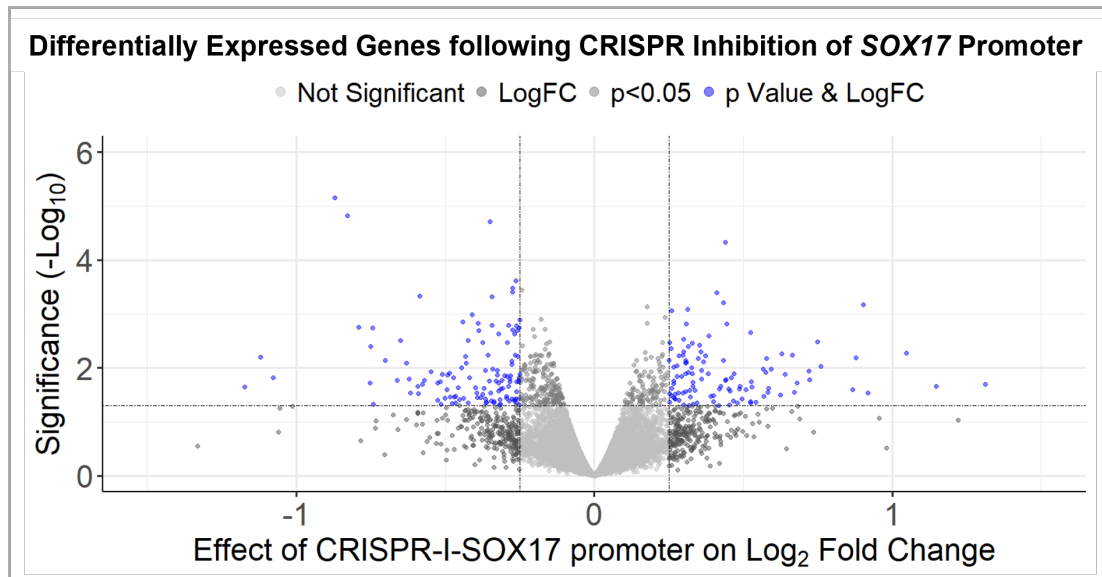


Figure 5.3

Differentially expressed genes following CRISPR-inhibition of *SOX17* promoter in hPAECs. Volcano plot of the $\log_2$ fold change (FC) between CRISPR-INH of *SOX17* Promoter and BFP Control and the negative $\log_{10}$ p-value. Differentially expressed genes shown in blue met the cutoff points: $p < 0.05$ and $\log_2$ fold change < -0.25 or > 0.25 . $n=4$, 12629 variables.

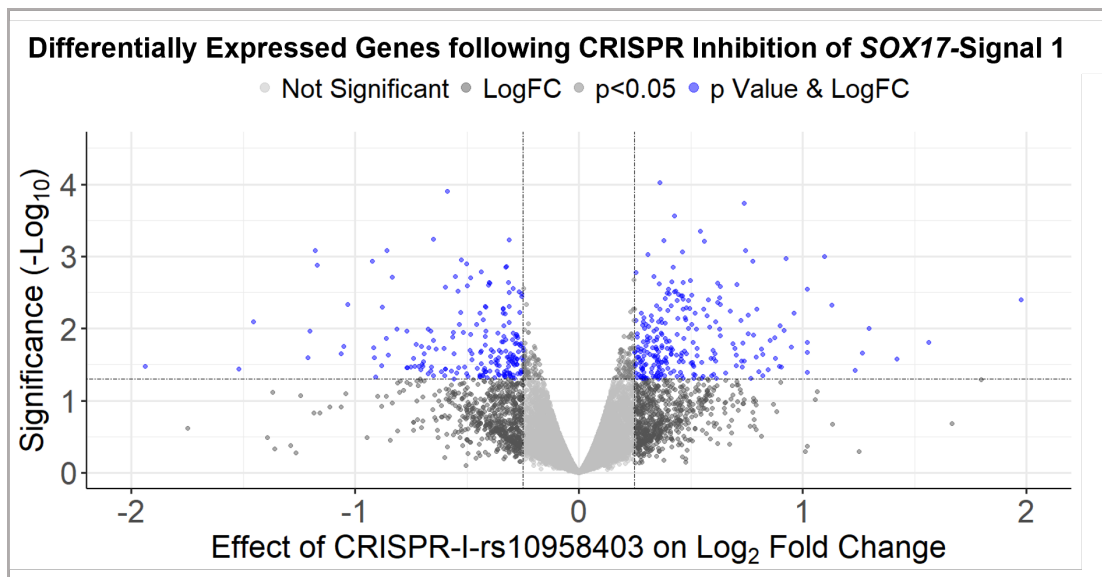


Figure 5.4

Differentially expressed genes following CRISPR-inhibition of *SOX17*-Signal 1 in hPAECs. Volcano plot of the $\log_2$ fold change (FC) between CRISPR-INH of *SOX17*-Signal 1 and BFP Control and the negative $\log_{10}$ p-value. Differentially expressed genes shown in blue met the cutoff points: $p < 0.05$ and $\log_2$ fold change < -0.25 or > 0.25 . $n=4$, 12629 variables.

5.3.2 Comparison of connectivity map perturbagen signatures with RNAseq analysis of *SOX17* inhibition

To predict if commercially available compounds which could be re-purposed would be able to reverse the gene changes associated with *SOX17* inhibition, the CMap database was interrogated. The top 5 compounds found through each query are shown in table 5.1. From these queries, three compounds were selected which showed either a high positive connection with an increase in *SOX17* (meaning they mimic the effect of CRISPR–ACT of the *SOX17* promoter) or a high negative correlation with a decrease in *SOX17* (opposing the effect of CRISPR–INH of the *SOX17* promoter or *SOX17*-signal 1). These compounds were sirolimus (*SOX17* promoter CRISPR–INH: tau -96.94), aminopurvalanol-a (*SOX17* promoter CRISPR–INH: tau -94.17, *SOX17*-signal 1 CRISPR–INH: tau -99.65) and YK-4279 (*SOX17* promoter CRISPR–ACT: tau +93.88). Opposing and mimicking effects can be observed in the perturbagen signatures and DEG following CRISPR–INH of *SOX17*-signal 1 and the *SOX17* promoter and CRISPR–ACT of the *SOX17* promoter, by comparing z-scores and log₂ fold change (respectively) (Figure 5.5). This is especially visible with the comparison between aminopurvalanol-a and *SOX17*-signal 1 although none of the comparisons are complete matches. This however is to be expected due to the limitations of the CMap. For each compound, a panel of genes was selected which showed consistent directional changes across multiple cell lines in the CMap for further testing by qPCR following compound exposure in hPAEC.

Query	Compound	Tau Score	Compound Family
<i>SOX17</i> promoter repression	Sirolimus	-96.94	mTOR inhibitor
	BX-795	-96.49	IKK inhibitor
	Aminopurvalanol-a	-94.17	CDK inhibitor
	SB-218078	-93.62	CHK inhibitor
	Lestaurtinib	-93.11	FLT3 inhibitor
<i>SOX17</i> -signal 1 repression	Aminopurvalanol-a	-99.65	CDK inhibitor
	Palbociclib	-99.58	CDK inhibitor
	Purvalanol-a	-99.37	CDK Inhibitor
	HG-5-113-01	-99.30	Protein kinase inhibitor
	Idarubicin	-99.12	Topoisomerase inhibitor
<i>SOX17</i> promoter activation	YK-4279	93.88	Apoptosis stimulant
	Quinine	91.97	Cytochrome inhibitor
	Taurocholic acid	91.46	Bile Acid

Table 5.1: Top compounds which reached a tau score cutoff of +/-90 in CMap queries. Bold compounds were used in further investigations.

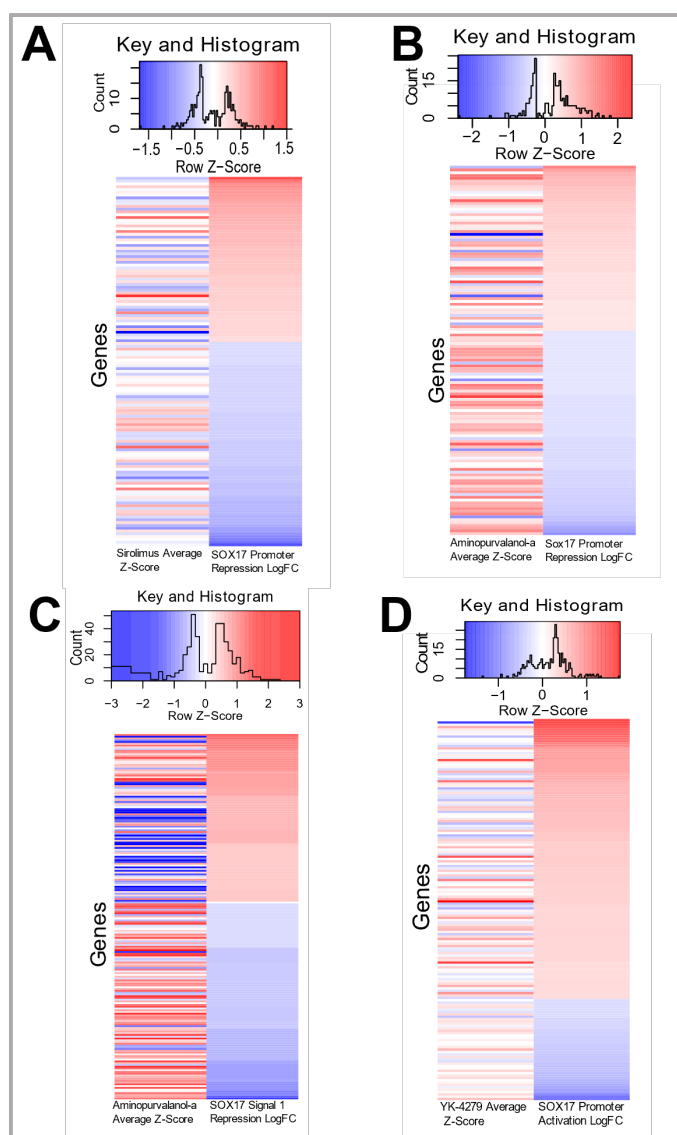


Figure 5.5

Heatmaps of CRISPR-siRNA fold change and CMap perturbagen Z-scores. Comparison of DEG following **A)** CRISPR-INH of the *SOX17* promoter in hPAECs and sirolimus exposure, **B)** CRISPR-INH of the *SOX17* promoter in hPAECs and aminopurvalanol-a exposure, **C)** CRISPR-INH of *SOX17*-signal 1 in hPAECs and aminopurvalanol-a exposure and **D)** CRISPR-ACT of the *SOX17* promoter in hPAECs and YK-4279 exposure. Blue represents a decrease in expression, red represents an increase in expression. Rows are each gene from CMap query. LogFC, Log₁₀ fold change. Average Z-score, average of z-scores taken from the CMap perturbagen query for each compound.

5.3.3 Exposure of hPAEC to CMap Compounds

The expression changes in hPAECs in the chosen genes following exposure to the compounds predicted to correct *SOX17*-associated gene expression by CMap (sirolimus, YK-4278, and aminopurvalanol-a) was assessed via qPCR to investigate the relevance of these compounds to the primary cell type affected by *SOX17* loss in PAH. For sirolimus, all

(7/7) of the genes tested showed the same directionality of expression change between the CMap signature and the qPCR data at the 6 hour time point, with *PITPNM1*, *MXD3*, *IL1RL1*, *FAH* and *RAB8A* having a significant change in expression with at least one concentration. For the 24 hour time point, 4/7 show the opposite directionality, with *PKD1P1*, *PITPNM1* and *MXD3* having a significant decrease in expression with at least one concentration (Figure 5.6). For Aminopurvalanol-a, only half (4/7 and 2/5) of the genes tested changed in the direction predicted by the CMap signature for this compound at the 6 hour time point. Of the gene tested for changes following aminopurvalanol-a exposure, *PLXNA2*, *POLE2*, *KIF20a*, *CCNA2*, *SHQ1* and *NQO1* had a significant change in expression with at least one concentration (Figure 5.7 and 5.8). For the 24 hour time point, only *BMPR1A*, *SORT1* and *LAMB2* did not follow the prediction of the CMap with the 50 μ m concentration being significant for *BMPR1A* and both concentrations being significant for *SORT1*. For YK4279, the predictions made in the CMap correctly predict all (5/5) of the gene expression changes tested following YK4279 exposure in hPAECs at the 6 hour time point although only *AURKA* and *DEPDC1* have significant changes. For the 24 hour time point 4/5 did not follow the CMap prediction with *KIF20a* and *CENPE* being significant (Figure 5.9). The small number of significant results following YK-4279 treatment may be due to high levels of cell death which occurred with this compound at all concentrations. These data confirmed that CMap-predicted compounds can cause opposing gene expression changes to those associated with *SOX17* dysfunction in hPAEC, albeit to a varying extent and in a potentially time-dependent manner.

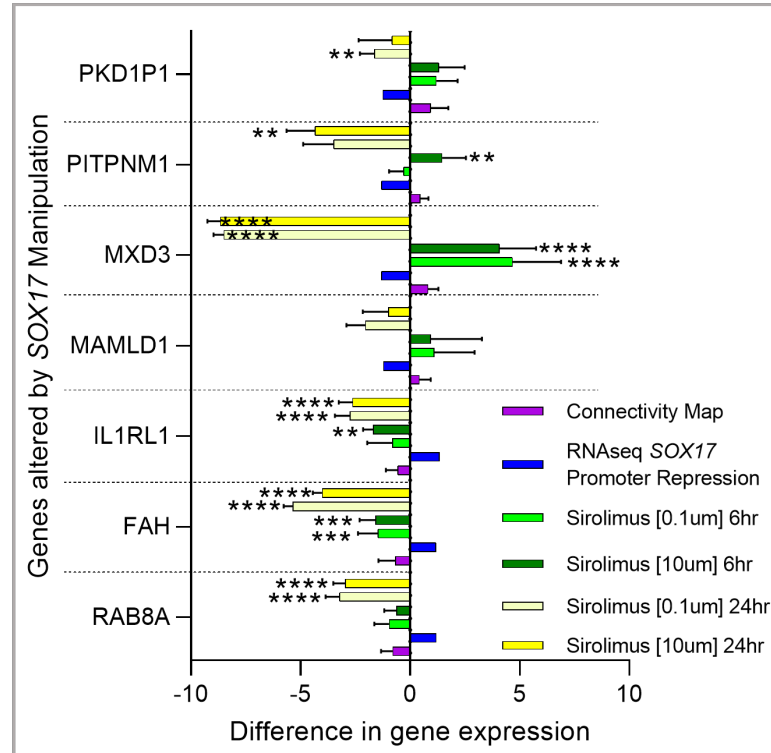


Figure 5.6

Differentially expressed genes following Sirolimus exposure. Gene expression of *PKD1P1*, *PITPNM1*, *MXD3*, *MAMLD1*, *IL1RL1*, *FAH* and *RAB8A*. Connectivity map, perturbagen z-score taken from the CMap. RNAseq *SOX17* promoter repression, fold change from RNAseq analysis of DEG following CRISPR–INH of the *SOX17* promoter. Sirolimus[0.1 μm /10 μm], relative expression in hPAECs to controls following sirolimus exposure at the stated concentrations. Statistics show sirolimus compared to vehicle (DMSO) control (not shown for clarity). Ordinary, one-way ANOVA. ** - $p<0.01$. *** - $p<0.005$. **** - $p<0.001$. Points are averages of three separate experiments ($n=3$). Bars show mean ($\pm$ SEM).

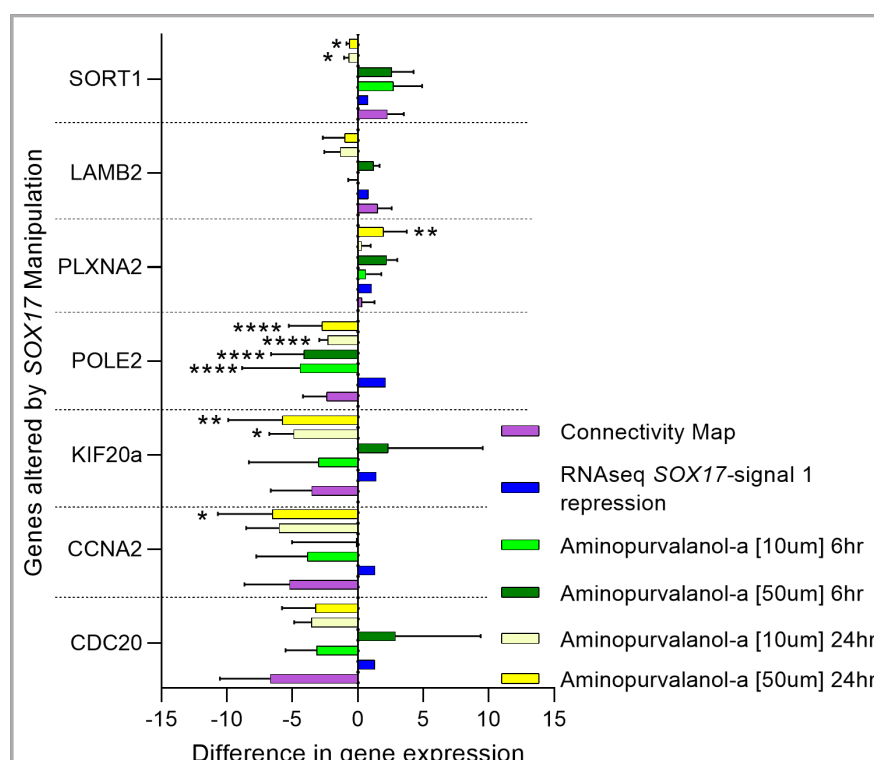


Figure 5.7

Differentially expressed genes following Aminopurvalanol-a exposure. Gene expression of *SORT1*, *LAMB2*, *PLXNA2*, *POLE2*, *KIF20A*, *CCNA2* and *CDC20*. Connectivity map, perturbation z-score taken from the CMap. RNAseq *SOX17*-signal 1 repression, fold change from RNAseq analysis of DEG following CRISPR-INH of *SOX17*-signal 1. Aminopurvalanol-a [10 μ m/50 μ m], relative expression in hPAECs to controls following aminopurvalanol-a exposure at the stated concentrations. Statistics show aminopurvalanol-a compared to vehicle (DMSO) control (not shown for clarity). Ordinary, one-way ANOVA. * - $p < 0.05$. ** - $p < 0.01$. *** - $p < 0.005$. **** - $p < 0.001$. Points are averages of three separate experiments ($n=3$). Bars show mean ($\pm$ SEM).

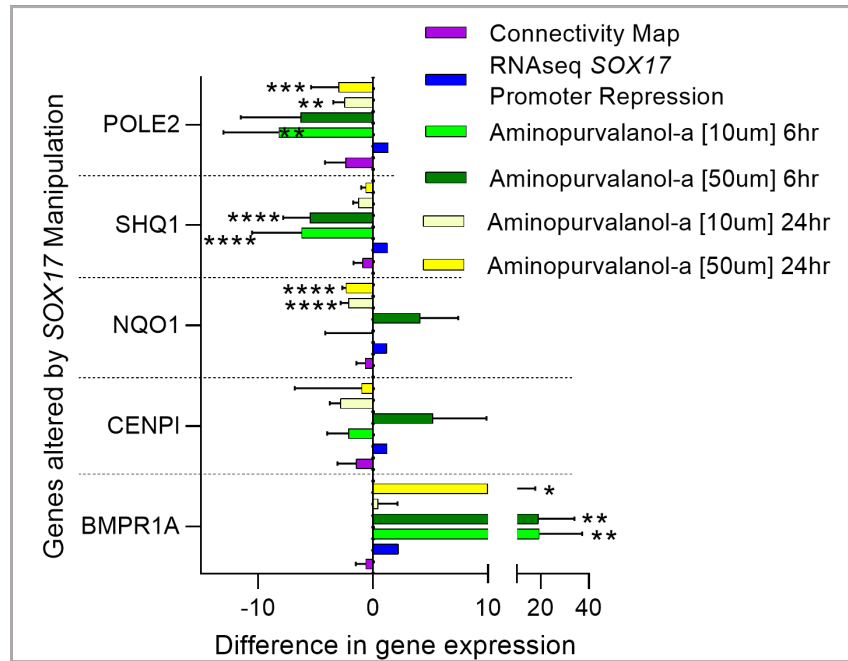


Figure 5.8

Differentially expressed genes following Aminopurvalanol-a exposure. Gene expression of *POLE2*, *SHQ1*, *NQO1*, *CENPI* and *BMPR1A*. Connectivity map, perturbation z-score taken from the CMap. RNAseq *SOX17* promoter repression, fold change from RNAseq analysis of DEG following CRISPR-INH of the *SOX17* promoter. Aminopurvalanol-a [10 μ m/50 μ m], relative expression in hPAECs to controls following aminopurvalanol-a exposure at the stated concentrations. Statistics show aminopurvalanol-a compared to vehicle (DMSO) control (not shown for clarity). Ordinary, one-way ANOVA. * - $p < 0.05$. ** - $p < 0.01$. *** - $p < 0.005$. **** - $p < 0.001$. Points are averages of three separate experiments ($n=3$). Bars show mean ($\pm$ SEM).

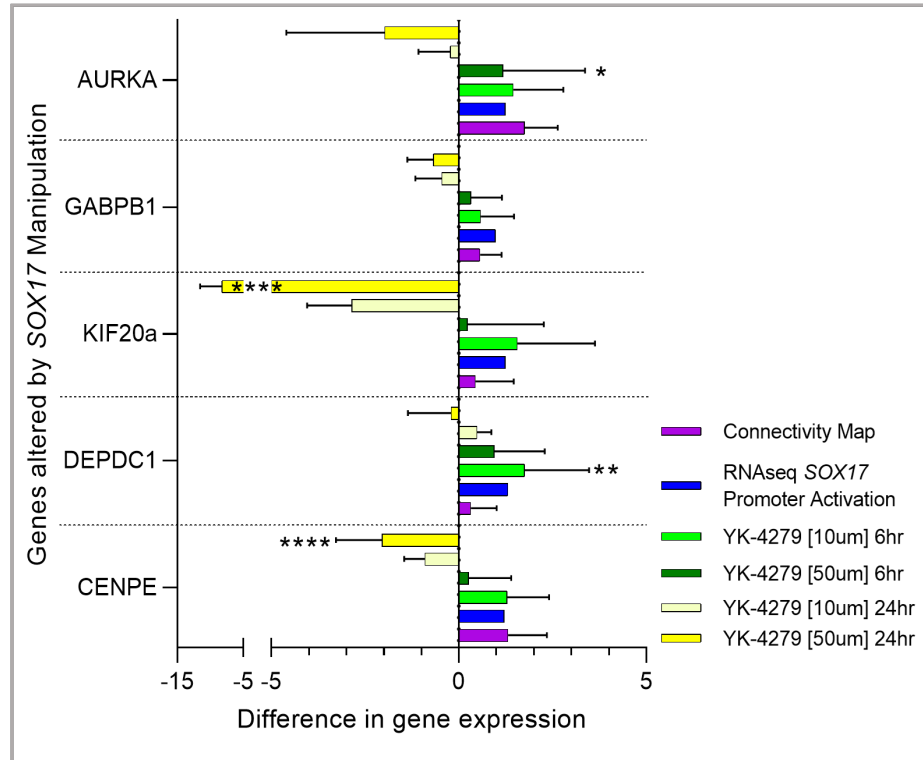


Figure 5.9

Differentially expressed genes following YK-4279 exposure. Gene expression of *AURKA*, *GABPB1*, *KIF20A*, *DEPDC1* and *CENPE*. Connectivity map, perturbagen z-score taken from the CMap. RNAseq *SOX17* promoter activation, fold change from RNAseq analysis of DEG following CRISPR–ACT of the *SOX17* promoter. YK-4279 [10 μ m/50 μ m], relative expression in hPAECs to controls following YK-4279 exposure at the stated concentrations. Statistics show YK-4279 compared to vehicle (DMSO) control (not shown for clarity). Ordinary, one-way ANOVA. * - $p < 0.05$. ** - $p < 0.01$. **** - $p < 0.001$. Points are averages of three separate experiments ($n=3$). Bars show mean ($\pm$ SEM).

5.4 Discussion

This chapter aimed to investigate the CMap for novel compounds for use in correcting *SOX17* dysfunction in PAH, specifically in hPAECs. The compounds sirolimus, aminopurvalanol-a and YK-4279 were tested and showed they can correct for some genetic changes associated with *SOX17* dysfunction in hPAECs.

To date, there has been a single study utilising the CMap for drug discovery in PAH [Gu et al., 2021]. Here, 6 individual PAH transcriptomic databases were combined to give a “PAH lung transcriptomic signature” which was used to identify AG1296 as a potential compound for use in PAH. AG1296 is an ATP-competitive selective inhibitor of platelet-derived growth factor (*PDGF*) receptors. It was demonstrated to reverse several PAH molecular and functional phenotypes in both human cells and the PAH Sugen/hypoxia rat model. Plans are now underway to bring this compound to clinical trial in PAH patients. Although this study does provide a good PAH therapy candidate, because induced pluripotent stem cell-derived ECs from multiple PAH patient lungs were used, it has a number of limitations. Firstly, by using lung induced pluripotent stem cell-derived ECs this work is based on cells which are not the primary cell type in PAH such as hPAECs or SMCs. Using more a more relevant cell type would improve the relevance of the signature used to query the CMap which would increase the number of potential positive "hits" found. Secondly, using a signature created through a mix of multiple patient ECs leads to a non-personalised signature. Because of this, any differences between the cause of their disease (such as different genetic causes) will be diluted and the resulting compound(s) found through the CMap query may not be relevant to personalised therapies, even though they may be useful for treating the disease as a whole. This limitation is acknowledged by the investigators and they suggest using their platform for patient specific cell signature testing or testing of specific groups of PAH sufferers.

Creating a list of genes for use in the CMap highlights one limitation of the CMap, that some genes are not recognised and therefore cannot be included in DEG lists. These include, but are not limited to pseudogenes, read-through transcripts, putative protein-coding genes or long non-coding RNA. Depending on the investigation, this may limit the size and quality of the lists created. However, for most studies it will mean the removal of a small number of hits from the DEG list and a small effect on the resulting queries and tau scores. Another limitation of the CMap is the limited number of cell lines which are used to create the L1000, and the full CMap. During a query, it is possible to select only specific cell lines which are the most relevant to the research. However, if the study being undertaken does not directly relate to any of the cell lines testing in the CMap, it will remain a limitation of this dataset. This can result in incorrect perturbagen scores, and therefore the selection of truly “un-connected” perturbagens. By using the largest disease signatures possible in queries and testing the compounds in human hPAECs, a relevant

cell type to PAH, I aimed to minimize this limitation as far as possible. Recent work has questioned the reliability of the CMap and the method of perturbagen prioritization. Some have suggested that variability between sample replicate gene expression data used to create the reference database affects the reliability of prioritization of reference signatures [Lim and Pavlidis, 2019]. It was found that compound concentration and cell target affinity is highly predictive of the strength of a compounds profile within the dataset. Therefore, if the compound is not at the optimal concentration, the genetic response may be inaccurate when measured, leading to incorrect connectivity scores and false positives and negatives when selecting candidate compounds. However, other comparisons have highlighted the reproducibility and broad use of this platform [Abrew et al., 2016]. When assessing genetic changes, post perturbagen exposure, the creators of the original CMap strived to maximise the coverage of the genome whilst minimizing cost and increasing throughput. Previous studies [Eisen et al., 1998] had suggested that a selection of genes can have similar expression patterns across different cell types and that these clusters of genes are highly correlated with transcriptome-wide expression data. It was found that approximately 1000 landmark genes were enough to recover 82% of the expected connections from the original CMap [Subramanian et al., 2017]. This next-generation CMap uses these landmark genes (L1000) to predict the whole transcriptome, and this is now used to assess genetic changes after perturbagen exposure. This has allowed the creation of a much larger dataset, with 400,000 profiles at the time of writing which enables a greater number of perturbagens to be assessed and improves the power of the CMap.

Sirolimus, also known as rapamycin, is a specific inhibitor of mammalian target of rapamycin (mTOR) that has been successfully used as an immunosuppressive agent in kidney transplant patients (for full review see: [Li et al., 2014]). It acts as an allosteric inhibitor of mTOR complex 1 (mTORC1), an important signal regulator of many pathways with roles in diseases such as cancer, age-related diseases and diseases related to metabolism [Liu and Sabatini, 2020]. Sirolimus can inhibit mTORC2 in some cell types, but only through long-term exposure [Laplane and Sabatini, 2012]. A similar immunosuppressor, tacrolimus, has been assessed in hPAEC from iPAH patients, and it was found to reverse dysfunctional BMPR2 signaling and prevented exaggerated chronic hypoxic PAH in mice with conditional *BMPR2* deletions [Spiekerkoetter et al., 2013]. mTOR signalling and PAH are linked in pulmonary artery smooth muscle cells, where it plays a role in hypoxia-induced pulmonary artery remodelling through smooth muscle cell proliferation [He et al., 2020], and when coupled with hypoxia, mTOR signalling increases the proliferation of rat ECs [Li et al., 2007]. Work has also shown a connection between aberrant mTOR signalling and EndoMT, a process through which hPAECs potentially differentiate into cells that form part of the neointima in PAH (Section 1.2.2) (For a full review: [Babicheva et al., 2021]). This shows a relationship between mTOR signalling and

functions and cell types which are relevant in PAH. Due to the increasing evidence of an mTOR signalling link to PAH, clinical trials have begun which aim to inhibit the mTOR complexes. An albumin-bound nanoparticle form of sirolimus is currently undergoing a phase I/II clinical trial and has so far shown no safety concerns and a positive effect on clinical outputs such as 6 minute walk distance and pulmonary vascular resistance [Simon et al., 2019]. The work shown in this chapter identifies and investigates sirolimus as a compound with a highly negatively connected transcriptome to the *SOX17* dysfunction transcriptome, a known PAH risk gene. Although not tested for here, it would be interesting to see the effect of sirolimus on functions which are well known to be affected by mTOR signalling such as proliferation, in hPAECs. In chapter 4, it was shown that a loss of *SOX17* promotes increased proliferation, a hallmark phenotype of hPAECs in PAH. I would hypothesize, the addition of sirolimus to these cells would cause a reversal of this phenotype, representing a possible rescue of *SOX17* dysfunction in hPAEC. This future work could potentially lend more evidence for the use of sirolimus as a therapy in PAH. Previous work in HUVEC has suggested mTOR signalling is involved in the regulation of *SOX17* through VEGF and that overexpression of mTOR leads to an increase in *SOX17* [Kim et al., 2016]. However, Zhou, et al (2009) showed the opposite, that an increase in *SOX17* expression occurs in response to sirolimus exposure (mTOR inhibition) in human embryonic stem cells. This suggests a complicated relationship between mTOR and *SOX17* which can have different outcomes depending on the cell type being investigated. The investigations carried out by Kim, et al (2016) may be more relevant to the work discussed in this thesis as HUVEC are a more relevant cell type to hPAEC. Although, the relationship between mTOR and *SOX17* requires more investigation in hPAEC to tease out the effect of mTOR on *SOX17* in this cell type. To build on the work by both these studies, the effect of both overexpression and downregulation of mTOR signalling on *SOX17* levels in hPAECs should be assessed. Furthermore, *SOX17* expression following sirolimus exposure was not investigated in this thesis but it would be interesting to see if an effect is found in hPAECs.

Aminopurvalanol-a (also known as purvalanol-a) is a cyclin dependent kinase (CDK)1/cyclin B inhibitor which arrests proliferating cells in the G2/M stage of the cell cycle and prevents proliferation. Its known targets are CDK1, CDK2, CDK5 and CDK6 [Knockaert et al., 2002, Breton et al., 2003]. It has been investigated in human microvascular endothelial cells as an anti-angiogenic drug and was shown to inhibit proliferation through G2/M stage arrest, increased apoptosis and preventing tube formation (an assay typically used to measure angiogenesis in ECs) [Zahler et al., 2010]. This drug was chosen from the investigation of the CMap as it showed a highly negative connection with *SOX17* loss through CRISPR-INH of both the *SOX17* promoter and *SOX17*-signal 1. Interestingly, the compound class "CDK inhibitor" was also negatively linked to a loss in *SOX17* and

a overexpression of CDK6 was highly positively connected with a loss of *SOX17* (data not shown). This links with both work from chapter 4 which shows a link between members of the cyclin kinase family and *SOX17* and wider research which shows some CDKs are upregulated following *SOX17* silencing in artery ECs [Lee et al., 2015]. These findings suggest that an increase in some CDK relates to a decrease in *SOX17* and that CDK inhibitors such as aminopurvalanol-a could be a useful compound in correcting for it. Concerning PAH, CDKs have very recently been shown to be relevant to disease aetiology and could be a druggable target in the future [Weiss et al., 2019]. In SMCs, it was shown that CDK inhibitors (dinaciclib, inhibits CDK1, 2, 5 and 9 and palbociclib, inhibits CDK4 and 6) reduced proliferation in SMCs and both the monocrotaline and Sugden/Hypoxia rat models reversed disease phenotypes including a reduction in pulmonary vascular remodelling [Weiss et al., 2019]. The potential of this drug class to affect not just hPAECs but hPSMC too is beneficial in its use in PAH. Overall, CDK inhibitors need further investigation for use in PAH.

YK-4279 is an ETS family inhibitor that has been undergone pre-clinical efficiency trials as an anti-lymphoma drug [Lamhamedi et al., 2015]. However, it has not been tested in PAH-relevant tissue or cells such as hPAECs or hPSMC. ETS family members are TFs and are known oncogenes whose aberrant expression is found in many solid tumours [Size-more et al., 2017]. They also have roles in vascular development and maintenance (for full review: [Craig and Sumanas, 2016]). YK-4279 specifically inhibits ERG transcription and ERG-mediated cell migration and proliferation in prostate cancer [Wei et al., 2009]. The ERG TF is essential for EC homeostasis and has recently been shown to bind to a super enhancer upstream of the *SOX17* gene in HUVEC, suggesting a possible role for ERG in the regulation of *SOX17* in this cell type [Kalna et al., 2019]. A super-enhancer is a location within the genome which contains a cluster of enhancers and often preferentially targets cell lineage regulators, such as *SOX17*. iPAH patient lung ECs have lower levels of nuclear ERG protein, with some samples having complete loss of ERG [Looney et al., 2017]. However, it is not clear if this is a cause of disease or a response to vascular injury. The work discussed in this chapter suggests a positive connection between YK-4279, an ERG inhibitor, and the transcriptional changes which occur following an increase in *SOX17*.

To build upon the work discussed in this chapter, the effect of sirolimus, aminopurvalanol-a and YK-4279 on hPAEC dysfunction following *SOX17* knockdown must be assessed. In chapter 4, the effect of knockdown of *SOX17* on hPAEC functions (proliferation, apoptosis, cell viability and barrier function) was discussed, with *SOX17* knockdown increasing them all except barrier function, which significantly decreased. As these compounds have been selected as potential rescuers of the effects of *SOX17* dysfunction, it would be interesting to assess their effect on these functions in hPAEC following *SOX17* knockdown.

These studies could be used to prioritise compounds to be taken forward into in-vivo work in mouse or rat models of PAH. In these investigations, the prioritised compounds could be assessed for reversal or prevention of PAH haemodynamics and vascular remodelling in these animal models. Due to the high levels of apoptosis seen with YK-4279 in this work, this compound would not be prioritised for this work until the ideal concentration could be found. If used in disease, a concentration which is too high could have unintended side effects such as hyper-apoptosis of ECs. In the future, this experiment could be repeated with lower compound concentrations, to assess if this decreases the amount of apoptosis which occurs whilst still maintaining the desired gene expression effects.

5.5 Conclusion

The work discussed in this chapter investigates repurposing compounds to correct for *SOX17* dysfunction in hPAECs. Sirolimus, aminopurvalanol-a and YK-4279 were selected from an interrogation of the CMap and it was shown they can successfully produce opposing gene changes to those associated with *SOX17* dysfunction in hPAEC. Based on the investigations shown here and previous work discussed in section 5.4 aminopurvalanol-a and YK-4279 could be good candidates for further study, first in PH animal models to assess their effect on disease phenotypes. If the results of these in-vivo studies are promising aminopurvalanol-a and YK-4279 could be used in human clinical trials, with the added benefit of being repurposed therapies and the time and cost saving which this entails. As sirolimus is already being investigated in human clinical trials, the work discussed here suggests it could be a good candidate for patients with *SOX17*-dysfunction related PAH.

6 Conclusions & Future Work

GWAS often result in the discovery of causal SNPs in non-coding regions of the genome. These are typically very difficult to interpret. The investigations detailed in this thesis provide an example of a workflow which can be used to interpret these non-coding SNPs and discover functions and downstream pathways affected by the genes these SNPs affect. This workflow could be applied to other diseases. It also highlights a mechanism for the discovery of novel compounds which can be repurposed for different diseases based on the downstream transcriptional changes resulting from causal SNPs.

In chapter 3 of this thesis, the role of a SNP upstream of the *SOX17* gene within a putative enhancer region was assessed in the context of TF binding. It was shown that the non-risk (T) allele of *SOX17*-signal 2 (rs765727) conferred binding of the TF ROR-alpha to this enhancer region. Furthermore, it was shown that this enhancer region, when inhibited by CRISPR-INH, caused a decrease in *SOX17* expression. Together, these findings suggest the PAH-associated SNP rs765727 is located in a putative enhancer region that targets *SOX17* and when the non-risk allele is present, there is preferential binding of ROR-alpha and potentially a higher expression of *SOX17*.

The 4th chapter of this thesis the role of *SOX17* in hPAEC, an important cell type in the aetiology of PAH was investigated. It was shown that knocking down *SOX17* via siRNA affected hPAEC proliferation, apoptosis, viability, and barrier function. Previously known downstream targets of *SOX17* were also shown to be affected by its knockdown in hPAEC. This chapter also investigated pathways and functions which were affected by *SOX17* knock down and found those involving the ECM and adhesion were affected. Additionally, genes important for ECM and adhesion were shown to be affected in *SOX17*-siRNA hPAEC.

To discover compounds for repurposing in PAH to correct the knockdown of *SOX17*, chapter 5 of this thesis investigated the CMap and highlighted aminopurvalanol-a, sirolimus and YK-4279 as compounds that could be utilised. These drugs were shown to affect hPAEC gene expression in a similar way to that predicted by the CMap and could be taken further to assess their effect on hPAEC function and potentially as PAH therapies.

6.1 Defining *SOX17*-signal 2 & regulators of *SOX17*

Both loss of function mutations within the *SOX17* gene and two PAH GWAS signals located upstream of it are associated with PAH (Figure 1.4) [Zhu et al., 2018, Gräf et al., 2018, Rhodes et al., 2019]. In the largest PAH GWAS to date it was shown that *SOX17*-signal 1 inhibition in hPAECs caused a decrease in *SOX17* expression [Rhodes et al., 2019] and work discussed in this thesis (Chapter 3) shows the same occurs with the inhibition

of *SOX17*-signal 2. Furthermore, this thesis showed preferential binding of the TF ROR-alpha to the non-risk allele (T) present at *SOX17*-signal 2. This suggests binding to the non-risk allele by ROR-alpha leads to higher expression of *SOX17* and a protective effect of *SOX17* in PAH. This complements previous work in PAH, which show rare mutations which cause a loss of function of *SOX17* are associated with a higher risk of PAH [Zhu et al., 2018, Gräf et al., 2018]. Additionally, previous in-vivo work has shown non-specific *SOX17* knockdown in mice results in embryonic lethality with heart defects and enlarged veins [Azuma et al., 2002] and endothelial-specific knockdown in either embryonic or adult mice causes defects in artery specification and in-utero lethality [Clarke et al., 2013, Corada et al., 2013]. These studies support a role for *SOX17* in artery endothelial cells, an important cell type in PAH but do not characterise the effect of *SOX17*-signal 1 or signal-2 in-vivo (discussed below).

To further investigate the role of this enhancer region in *SOX17* expression and the link between *SOX17*-signal 2 and ROR-alpha binding, a ChIP-PCR of ROR-alpha binding regions in hPAECs could be performed. This would provide in-vitro evidence of the binding of ROR-alpha to *SOX17*-signal 2. If this is found, knockdown and/or overexpression of *ROR-α* could be performed in hPAECs to assess the effect of changing ROR-alpha levels on *SOX17* expression. Additional work investigating the differential expression of *SOX17* in genotyped patient blood outgrowth endothelial cells (BOECs) could be beneficial with the addition of an allelic imbalance assay. This would provide information on the differential gene expression of the two potential alleles present at *SOX17*-signal 2. One issue with this approach is it requires patient blood donations from many donors because of the low frequency of the risk-allele (C, allele frequency: 0.09). To have enough samples to perform biological replicates, many patients will have to donate and be genotyped for this region. Having additional evidence that ROR-alpha selectively binds to rs765727 and causes increased expression of *SOX17* would help us understand the cause of the association between *SOX17*-signal 2 and PAH.

To further characterise *SOX17*-signal 1 and 2 in PAH models in-vivo. Rats (Sprague Dawley) with the deletion of both *SOX17*-signal 1 and 2 are being used to assess the total loss of this region in PAH. Mice (C57BL/6 background) with deletion of either *SOX17*-signal 1 or 2 have also been created to assess the regions individually. These animals are being assessed for changes to pulmonary morphology and PAH hemodynamics such as right ventricular pressure and functions which are known to be affected by *SOX17* dysfunction, such as barrier function through an Evans Blue extravasation assay. Work is currently ongoing with these animals in the lab of Professor Lan Zhao (NHLI, Imperial College London). Although deleting the *SOX17* gene in previous in-vivo work has produced severe changes to the vasculature, I would hypothesise the effects from either deleting an enhancer or changing a single base pair within an enhancer region would

lead to a milder phenotype due the modulatory nature of enhancer regions compared to deleting whole genes. However, this in-vivo work will give additional information about the effect of these enhancers in this disease in a whole organism.

6.2 The role of *SOX17* in hPAECs and PAH

At the time of writing, there have been no studies investigating the role of Sox17 in hPAECs and why its dysfunction could lead to PAH. As discussed previously, hPAECs are an important cell type in PAH and their injury is thought to be a catalyst for disease onset and progression. For this reason, primary hPAECs obtained from healthy donors were used throughout this research. Furthermore, in the pulmonary system, *SOX17* is expressed most highly in these cells [Corada et al., 2013], therefore it is the likely cell type for an effect of loss of Sox17 to take effect.

SOX17 has an important role in cell cycle regulation, proliferation, apoptosis, and barrier function. Previous work has shown its importance for the maintenance of barrier function and proliferation following injury of the mouse endothelial layer [Liu et al., 2019]. Other studies have found evidence that Sox17 induces proliferation through transcriptional targeting of cell cycle genes [Lange et al., 2009]. There is little evidence of a direct link between *SOX17* and apoptosis, although some recent work does suggest *SOX17* regulates apoptosis in colorectal cancer cells [Hu et al., 2021]. These functions have been shown in this thesis to be affected in hPAECs by the knockdown of *SOX17*. Proliferation and apoptosis increased, whereas cell viability and barrier function decreased with a knockdown of *SOX17*. In the context of PAH onset and progression these finding highlight a role for *SOX17* as an important susceptibility factor in EC injury and PAH progression through it's roles in these functions.

This thesis has also shown a link between *SOX17* dysfunction in hPAECs and dysfunction of the ECM and cell adhesion, both of which are affected in PAH. For example, the basement membrane (which ECM is an essential part of) is thicker in the lungs of iPAH patients than controls [Jandl et al., 2020]. Knockdown of *SOX17* has previously been shown to affect the ECM and drive PAH through epigenetic regulation via the microRNA family 130/301 (miR0130/301) [Bertero et al., 2018]. It was found that knockdown of *SOX17* increased the expression of members of this microRNA family in hPAECs and hPSMCs providing a potential link between the knockdown of *SOX17* and regulation of the ECM which has been shown in this thesis.

To further investigate the role of *SOX17* in hPAEC, the role of *SOX17* in ECM and adhesion should be studied further. *SOX17* knockdown hPAECs should be assessed for changes in adhesion function using assays looking at the difference in adhesion between control and *SOX17*-knock down hPAECs. As different members of the collagen family

are affected differently by *SOX17* knockdown, their protein expression and the overall collagen production of *SOX17* knockdown hPAECs should also be assessed. Additionally, genes which are involved in adhesion and most affected by *SOX17* downregulation such as members of the collagen family, should be assessed in rescue experiments by first knocking down *SOX17* and then supplementing with collagen to assess its affect on adhesion. This will evaluate the functional mechanisms through which *SOX17* affects hPAEC adhesion function.

6.3 The role of *SOX17* in intracranial aneurysms and cancer

To assess how *SOX17* could play a role in PAH onset and progression, and how the pathways and functions which it affects could be linked to PAH, it is important to assess the role of *SOX17* in other cell types and diseases. By comparing its functions and related pathways in PAH-relevant cell types to those in others, it is possible to gain further understanding of its possible role in PAH. Previous investigations into *SOX17* dysfunction in cancer have shown it plays a role in tumour growth through promoting proliferation and barrier function and is associated with an increased risk for intracranial aneurysms (IA) [Tan et al., 2020, Foroud et al., 2012].

IA occur due to weakening of the walls of arteries and veins in the brain which causes a ballooning of the vessel and can lead to subarachnoid haemorrhage. Similarly to PAH, it is thought to occur due to EC injury and a breakdown in vessel barrier function. There have been links found between mutations affecting the *SOX17* protein and IA [Foroud et al., 2012]. These mutations and ones found in previous studies [Bilguvar et al., 2008, Yasuno et al., 2010] affect the 3' and 5' ends of the Sox17 protein and are not in linkage disequilibrium, meaning they are likely to be inherited separately. This study [Foroud et al., 2012] did not find any signals associated with IA through 2 discovery cohorts and only found a link between mutations affecting *SOX17* in a meta-analysis of these two discovery cohorts. Their findings also did not find support for other previously found signals associated with IA, including the 5' mutations affecting *SOX17*. The low number of patients used in the discovery cohorts and subsequent meta-analysis may be responsible for this, leading to a low power GWAS. SNPs which are in linkage disequilibrium with a large enhancer region upstream of the *SOX17* gene, have also been shown through chromatin conformation capture to associate with the *SOX17* promoter [Laarman et al., 2019] have also been linked to this disease. This study took previously identified SNPs from 6 GWAS-identified regions most strongly associated with IA and assessed their enhancer ability and the expression of the target genes and found *SOX17* to be expressed in all sections of the vessel wall in the circle of Willis (the circulatory system that supplies blood to the brain). Human IA samples have lower Sox17 levels, shown via immunostaining, than control samples (5 patients and 7 controls) [Lee et al., 2015]. Furthermore, mice

with endothelial cell-specific *SOX17*-deficiency exhibit phenotypes that are very similar to human IAs such as vessel dilation and wall thinning. These studies suggest a link between IA and decreased *SOX17* in vascular endothelial cells and suggest *SOX17* may be (at least partially) responsible for the EC injury which causes IA. It is well known that EC injury is a hallmark of PAH, and it is thought to be an initial factor in disease aetiology. This thesis provides evidence that a decrease in *SOX17* affects hPAEC ability to form tight barriers and causes the EC monolayer to become leaky, a phenomenon which is also found in IAs [Lee et al., 2015]. The creation of an endothelial-specific *SOX17* knockout mouse provides a platform to test novel compounds to target a loss of *SOX17* in IA and could provide useful knowledge that could be relevant to *SOX17* loss in PAH. Due to the similarities between PAH and IA, it is likely that compounds targeting *SOX17* loss in one disease could be relevant to the other. The continuing investigations into the role of *SOX17* and IA needs monitoring in the future for potential cross-overs with PAH.

SOX17 has been shown to play a tumour suppressor role in many cancers (for full review: [Tan et al., 2020]) and this thesis shows complimentary findings. For example, a loss of *SOX17* through CRISPR–INH of upstream enhancer regions or knockdown of mRNA levels affects pathways relating to cancer (MicroRNAs in cancer, bladder, endometrial and non-small cell lung cancer). *SOX17* has previously been shown to be especially important in endometrial cancer [Zhang et al., 2016, Liao et al., 2019, Zhou et al., 2019] where the loss of *SOX17* has been shown to inhibit migration and lead to increased proliferation in relevant cell types. In the work shown in this thesis, *SOX17* knockdown in hPAECs causes an increase in proliferation with the addition of VEGF, suggesting a similar role for *SOX17* in PAH relevant cells and cancer cells. Cancer has a similarity to PAH as the hyper-proliferative, anti-apoptosis phenotype which is stereotypical of hPAEC/PASMC in PAH is also found in cancer and leads to tumour growth and metastasis. In the work shown in this thesis, *SOX17* knockdown in hPAECs causes a potential increase in proliferation and a decrease in apoptosis, a phenomenon previously seen in *SOX17*-associated cancer studies [Zhang et al., 2016, Liao et al., 2019, Zhou et al., 2019]. The loss of *SOX17* in PAH, either through mutations within the gene or SNPs affecting its regulatory regions could be an initiating factor for the hyper-proliferative state seen in PAH, as it is in some cancers. At the time of writing, there have been no published investigations into therapies targeting *SOX17* dysfunction in cancer but, like in IA, this requires monitoring as it could have implications for therapies for *SOX17* dysfunction in PAH.

6.4 *SOX17* and therapies in PAH

As discussed in chapter 5, current therapies for PAH are not effective in every patient, they do not reverse pulmonary artery remodelling and even if they show some effect, this

can disappear after a few years. For this reason, it is extremely important to investigate new compounds that can be used in PAH. There are novel therapeutic avenues for PAH being explored which *SOX17* dysfunction may have an effect on including hypoxic signalling, BMP9 analogues and personalised medicine [Ghataorhe et al., 2017, Wilkins et al., 2018, Prins et al., 2019, Sommer et al., 2020].

Hypoxic signalling controls the cell's response to low oxygen environments which can occur as a result of environmental stimuli, or as a part of normal physiology. HIF1 and HIF2 regulate transcriptional responses to hypoxia through hypoxia response elements on target gene promoters. During hypoxia, the process of HIF1 α (a HIF1 subunit) degradation by prolyl hydroxylase (PHD2) is inhibited, allowing the HIF1 complex to form and activate target genes [Huang et al., 1998]. Dysfunctional hypoxic signalling is believed to have a role in PAH development and inhibition of HIF1 α has been shown to affect PAEC proliferation, VEGF expression and energy metabolism [Iyer et al., 1998, Carmeliet et al., 1998]. The level of PHD2 is reduced in PAECs in plexiform lesions suggesting a limit to HIF1 α degradation in these cells and structures [Dai et al., 2016]. These findings show inhibiting HIF could be a potential therapy for PAH and several studies have explored targeting mTOR to target hypoxic signalling. It has been shown that inhibition of mTOR reverses hypoxia-induced pulmonary vascular remodelling in hypoxia-PH mouse models [Padden-berg et al., 2007] and attenuates neointimal formation and possibly reverses EndoMT in monocrotaline rats [Nishimura et al., 2001, Ranchoux et al., 2015]. Rapamycin is an mTOR inhibitor that is currently undergoing a phase 1 clinical trial (ClinicalTrials.gov ID: NCT02587325) to test the safety of a rapamycin derivative albumin-bound mTOR inhibitor (ABI-009) in PAH patients. Results (at the time of writing) suggest an improvement in 6-minute walk distance and hemodynamic features with no safety issues reported and it has been suggested that this therapeutic merits further investigation [Simon et al., 2019]. However, recent work has suggested HIF1 α directly influences *SOX17* expression through HIF responsive elements in the *SOX17* promoter [Liu et al., 2019]. The binding of HIF1 α to the *SOX17* promoter increases *SOX17* expression which causes an increase in cyclin E1 expression and endothelial regeneration in mice models of endothelial cell injury. This contradicts the findings discussed in this thesis which suggest a loss of *SOX17* leads to increased hPAEC proliferation in the presence of VEGF. Additional work needs to be done to fully assess this difference in findings and to allow further investigation of the role of *SOX17* in hypoxic signalling and therapies targeting it.

Mutations in the gene encoding *BMPR2* are the most common genetic cause of PAH and mutations which affect proteins of the BMP signalling pathway are discussed further in section 1.4.1. The loss of this signalling pathway is one of the most studied causes of PAH. Recent work using recombinant BMP9 (a *BMPR2* ligand) has shown an increase in barrier integrity of PAECs and reversal of PAH in monocrotaline rats and *BMPR2* knockout

mice models [Long et al., 2015]. However, further work has shown that inhibiting BMP9 with neutralizing antibodies or knocking it down in mice models has beneficial effects against PAH progression [Tu et al., 2019]. This highlights the complicated nature of this signalling pathway and it has been suggested that knockdown of endogenous BMP9 and giving BMP9 as a therapeutic may not lead to the same results and further investigation is warranted [Ormiston et al., 2019]. There is evidence for a link between BMP signalling and Sox17. In extra-embryonic endoderm cells (reprogrammed pluripotent stem cells), BMP signalling knockdown with small molecule inhibitors decreases *SOX17* transcription. This is reversed with the addition of recombinant BMP2 and BMP4 [Yang et al., 2020]. However, this work has not been performed in cell types relevant to PAH and therefore needs investigation. In this thesis, it was shown that knocking down *SOX17* does not affect *BMP9* or *BMPR2* expression, important members of the BMP signalling pathways, however the effect of the BMP signalling pathway on *SOX17* was not investigated. It would be of interest to assess the effect of BMP signalling on *SOX17* to ascertain if targeting this pathway with therapies such as BMP9 peptides (currently under development as a therapeutic, [Long et al., 2015]) would have an effect on *SOX17*.

Whilst there have been many breakthroughs in understanding the molecular and physiological background of PAH, there still remains an unmet need for therapeutics which have long-term positive effects for patients and/or reverse pulmonary artery remodelling. To increase the speed at which this occurs, some have been investigating repurposing existing therapies for use in PAH [Prins et al., 2019]. One example is sildenafil (section 1.3.4) which was originally used for erectile dysfunction [Goldstein et al., 1998] and has been successfully repurposed for PAH [Galiè et al., 2005]. Another example is rapamycin, which was originally used as a tumor suppressor and in cancer treatment and previously discussed. Repurposing drugs may not only increase the rate at which therapies are made available for PAH patients, but it will also decrease the cost of bringing these therapies to the market, as many have already undergone safety and drug interaction trials in humans. Additionally, current therapeutics for PAH produce highly variable responses in terms of improvement in symptoms or outcome. This is perhaps the most important manifestation of the heterogeneous nature of PAH. To better understand this, it is firstly important to understand the genetic mutations which contribute to disease and how they differ between patients. This will give us a better chance of prescribing the most suitable therapy, or combination of therapies, for an individual. It has long been known that there are essential differences between individuals which directly impact their response to therapies. The completion of the human genome project [Lander et al., 2001, Venter et al., 2001] came with a promise to make personalised medicine a reality for many patients. Over the last two decades, genomics research has rapidly expanded, not only allowing for the discovery of disease-associated risk variants but also expanding the field of therapy discovery and

prescription.

There are several methods to discover novel drugs and/or repurpose existing ones. Large-scale CRISPR screens are commonly used for drug discovery and to give insight into the genetic background of various diseases and biological pathways [Fellmann et al., 2017, Ahmad and Amiji, 2018]. However, this method is both time and reagent consuming and therefore expensive to run. In silico approaches such as the CMap, provide a rapid, low-cost alternative method to perform similar research.

The CMap is a tool that can be used to integrate -omics data with repurposing compounds to select medicines depending on the cause(s) of a disease. At the time of writing, the CMap has been used to investigate novel uses for drugs for a wide range of diseases such as ovarian cancer, Gaucher disease and anti-aging [Rho et al., 2011, Yuen et al., 2012, Gillbro et al., 2017, Musa et al., 2018]. It has been used in previous work to discover a compound for use in PAH [Gu et al., 2021]. Here, 6 individual PAH transcriptomic databases were combined to give a “PAH lung transcriptomic signature” which was used to identify AG1296, an ATP-competitive selective inhibitor of *PDGF* receptors which reversed several PAH molecular and functional phenotypes in both human cells and the Sugen/hypoxia rat model of PAH. This study demonstrated the potential of identifying drug candidates at the -omics level, in comparison to more traditional ‘single target’ approaches and highlights how in silico studies can be used to re-purpose compounds for new diseases and how this has been applied to PAH therapeutic research.

In this thesis, transcriptomic data from *SOX17* knockdown hPAECs was used to query the CMap to discover compounds that could be repurposed for treating patients with PAH which is caused (at least in part) by *SOX17* dysfunction (rare pathogenic variants or risk SNPs). Three compounds were investigated further as it was suggested they could cause an opposite transcriptomic signature to *SOX17* dysfunction. Based on the work shown in this thesis and previous work on mTOR signalling and the CDK family in PAH, the most promising of the three are probably sirolimus and aminopurvalanol-a. To build on the work shown here, sirolimus and aminopurvalanol-a should be assessed for their effect on hPAEC function to see if they can rescue the changes caused by *SOX17* knock-down. For example, in chapter 4 it was shown that loss of *SOX17* in hPAECs leads to an increase in proliferation and apoptosis and a decrease in barrier function, leading to a leaky monolayer. Sirolimus and aminopurvalanol-a should be used in rescue experiments to assess whether they can rescue these functions in hPAEC with *SOX17* loss. To continue the development of these compounds as potential therapies it is important to assess them in animal models. Because they have been selected as *SOX17* dysfunction therapies it is important they are assessed in animal models which have *SOX17* dysfunction such as those discussed previously from Professor Zhao and/or the endothelial-specific *SOX17*

knockouts used in IA research [Lee et al., 2015].

Although it was used in this thesis to select compounds which successfully caused the opposite transcriptomic effects to *SOX17* knockdown and suggested a compound which is currently being used in clinical trials for PAH there are limitations to the CMap, especially in relation to this work. For example, the CMap contains the transcriptomic profiles of a small selection of cells which have been exposed to perturbagens. None of these cells are relevant to PAH as they are not pulmonary vessel cell types such as hPAECs or SMCs and it does not contain the profiles of any endothelial cells. To remove this limitation, future updates to the CMap should contain a more diverse collection of cell transcriptomes or novel databases should be created. This would allow a comparison between a PAH disease signature within a relevant cell type or cell types which are similar. In this work, the transcriptomic profiles of hPAECs following *SOX17*-signal 1 or promoter inhibition or *SOX17* promoter activation were used to create the gene lists. To improve on the work of this thesis, the gene lists used to query the CMap should be refined. For example, the DEG in common between *SOX17* dysfunction following siRNA and CRISPR–INH could be used as it represents the most important genes following *SOX17* dysfunction.

This thesis has focused on *SOX17* dysfunction and provided evidence that a loss of *SOX17* is relevant to PAH through downstream targets and pathways. The work discussed based on the investigation of the CMap was performed on the hypothesis that loss of *SOX17* is detrimental in PAH and compounds which oppose the downstream effects of *SOX17* dysfunction will be beneficial. However, an increase in *SOX17* above normal physiological levels has been shown to be detrimental in a study of airway inflammation [Ha et al., 2019]. Here, in an mouse asthma model, *SOX17* was higher in challenged mice than controls and deletion of *SOX17* in adult mice alleviated allergic airway responses. In this study, the authors predict that increased Sox17 expression leads to increased recruitment of inflammatory cells which could be responsible for the increased allergic response. This could be a concern for therapies which aim to boost *SOX17* in PAH as increased inflammation is a key hallmark in the disease. Investigations of the effect of increased *SOX17* in both hPAECs and in-vivo models of PAH will need to be undertaken to assess whether the effect seen in lung endothelial cells occurs in hPAECs.

6.5 Summary

In summary, the body of work presented here supports a role for *SOX17* in the development of PAH by providing a molecular mechanism through which *SOX17* dysfunction affects hPAEC homeostasis. *SOX17* dysfunction in PAH is driven by both common genetic variation in enhancer regions upstream of the *SOX17* gene which affect binding of ROR- α and rare variation which causes a loss of function of Sox17. This *SOX17*

dysfunction leads to signalling and functional changes in hPAECs which could be a cause of endothelial dysfunction in PAH. *SOX17* dysfunction can, in part, be opposed by compounds predicted to be useful for repurposing using transcriptomic data and the CMap. These compounds could be priorities for therapeutic development in PAH, identifying *SOX17* as a potentially druggable pathway which could be harnessed as part of a personalized medicine approach for this crippling disease.

References

- K. N. D. Abrew, R. M. Kainkaryam, Y. K. Shan, G. J. Overmann, R. S. Settivari, X. Wang, J. Xu, R. L. Adams, J. P. Tiesman, E. W. Carney, J. M. Naciff, and G. P. Daston. Grouping 34 chemicals based on mode of action using connectivity mapping. *Toxicological sciences : an official journal of the Society of Toxicology*, 151:447–61, 2016. ISSN 1096-0929. doi: 10.1093/toxsci/kfw058. URL <http://www.ncbi.nlm.nih.gov/pubmed/27026708>.
- G. Ahmad and M. Amiji. Use of crispr/cas9 gene-editing tools for developing models in drug discovery, 3 2018. ISSN 18785832.
- F. W. Albert and L. Kruglyak. The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics* 2015 16:4, 16:197–212, 2 2015. ISSN 1471-0064. doi: 10.1038/nrg3891. URL <https://www.nature.com/articles/nrg3891>.
- M. Alonso, A. B. García-Redondo, D. Guo, E. Camafeita, F. Martínez, A. Alfranca, N. Méndez-Barbero, Ángela Pollán, C. Sánchez-Camacho, D. T. Denhardt, M. Seiki, J. Vázquez, M. Salaices, J. M. Redondo, D. Milewicz, and A. G. Arroyo. Deficiency of mmp17/mt4-mmp proteolytic activity predisposes to aortic aneurysm in mice. *Circulation Research*, 117:e13–e26, 7 2015. ISSN 15244571. doi: 10.1161/CIRCRESAHA.117.305108. URL <https://www.ahajournals.org/doi/abs/10.1161/CIRCRESAHA.117.305108>.
- C. G. AneneNzelu, M. C. Lee, W. L. Tan, A. Dashi, and R. S. Foo. Genomic enhancers in cardiac development and disease. *Nature Reviews Cardiology* 2021 19:1, 19:7–25, 8 2021. ISSN 1759-5010. doi: 10.1038/s41569-021-00597-2. URL <https://www.nature.com/articles/s41569-021-00597-2>.
- S. L. Archer, G. Marsboom, G. H. Kim, H. J. Zhang, P. T. Toth, E. C. Svensson, J. R. Dyck, M. Gomberg-Maitland, B. Thébaud, A. N. Husain, N. Cipriani, and J. Rehman. Epigenetic attenuation of mitochondrial superoxide dismutase 2 in pulmonary arterial hypertension: A basis for excessive cell proliferation and a new therapeutic target. *Circulation*, 121:2661–2671, 6 2010. ISSN 00097322. doi: 10.1161/CIRCULATIONAHA.109.916098. URL <https://www.ahajournals.org/doi/abs/10.1161/circulationaha.109.916098>.
- E. Arciniegas, C. Y. Neves, L. M. Carrillo, E. A. Zambrano, and R. Ramírez. Endothelial-mesenchymal transition occurs during embryonic pulmonary artery development. *Endothelium: Journal of Endothelial Cell Research*, 12:193–200, 7 2005. ISSN 10623329. doi: 10.1080/10623320500227283.

- K. Asosingh, S. Comhair, L. Mavrakis, W. Xu, D. Horton, I. Taylor, S. Tkachenko, B. Hu, and S. Erzurum. Single-cell transcriptomic profile of human pulmonary artery endothelial cells in health and pulmonary arterial hypertension. *Scientific Reports* 2021 11:1, 11:1–13, 7 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-94163-y. URL <https://www.nature.com/articles/s41598-021-94163-y>.
- E. D. Austin, L. Ma, C. LeDuc, E. B. Rosenzweig, A. Borczuk, J. A. Phillips, T. Palomero, P. Sumazin, H. R. Kim, M. H. Talati, J. West, J. E. Loyd, and W. K. Chung. Whole exome sequencing to identify a novel gene (caveolin-1) associated with human pulmonary arterial hypertension. *Circulation: Cardiovascular Genetics*, 5:336–343, 6 2012. ISSN 1942325X. doi: 10.1161/CIRCGENETICS.111.961888.
- M. Azuma, Y. Kanai, J. M. Gad, Y. Tajima, C. Taya, M. Kurohmaru, Y. Sanai, H. Yonekawa, K. Yazaki, P. P. L. Tam, and Y. Hayashi. Depletion of definitive gut endoderm in sox17-null mutant mice. *Development (Cambridge, England)*, 129:2367–79, 5 2002. ISSN 0950-1991.
- A. Babicheva, A. Makino, and J. X. Yuan. mtor signaling in pulmonary vascular disease: Pathogenic role and therapeutic target. *International journal of molecular sciences*, 22: 1–29, 2 2021. ISSN 1422-0067. doi: 10.3390/IJMS22042144. URL <https://pubmed.ncbi.nlm.nih.gov/33670032/>.
- D. B. Badesch, G. E. Raskob, C. G. Elliott, A. M. Krichman, H. W. Farber, A. E. Frost, R. J. Barst, R. L. Benza, T. G. Liou, M. Turner, S. Giles, K. Feldkircher, D. P. Miller, and M. D. McGoon. Pulmonary arterial hypertension: Baseline characteristics from the reveal registry. *Chest*, 137:376–387, 2 2010. ISSN 19313543. doi: 10.1378/chest.09-1140.
- G. Bain, T. Müller, X. Wang, and J. Papkoff. Activated -catenin induces osteoblast differentiation of c3h10t1/2 cells and participates in bmp2 mediated signal transduction. *Biochemical and Biophysical Research Communications*, 301:84–91, 1 2003. ISSN 0006291X. doi: 10.1016/S0006-291X(02)02951-0.
- R. J. Barst, L. J. Rubin, W. A. Long, M. D. McGoon, S. Rich, D. B. Badesch, B. M. Groves, V. F. Tapson, R. C. Bourge, B. H. Brundage, S. K. Koerner, D. Langleben, C. A. Keller, S. Murali, B. F. Uretsky, L. M. Clayton, M. M. Jöbsis, S. D. Blackburn, D. Shortino, and J. W. Crow. A comparison of continuous intravenous epoprostenol (prostacyclin) with conventional therapy for primary pulmonary hypertension, 2 1996. ISSN 00284793.
- O. Ben-Yehuda, D. Pizzuti, A. Brown, M. Littman, H. Gillies, N. Henig, and T. Peschel. Long-term hepatic safety of ambrisentan in patients with pulmonary arterial hypertension, 7 2012. ISSN 07351097.

- R. L. Benza, D. P. Miller, M. Gomberg-Maitland, R. P. Frantz, A. J. Foreman, C. S. Coffey, A. Frost, R. J. Barst, D. B. Badesch, C. G. Elliott, T. G. Liou, and M. D. McGoon. Predicting survival in pulmonary arterial hypertension: Insights from the registry to evaluate early and long-term pulmonary arterial hypertension disease management (reveal). *Circulation*, 122:164–172, 7 2010. ISSN 00097322. doi: 10.1161/CIRCULATIONAHA.109.898122.
- T. Bertero, A. L. Handen, and S. Y. Chan. Factors associated with heritable pulmonary arterial hypertension exert convergent actions on the mir-130/301-vascular matrix feedback loop. *International Journal of Molecular Sciences*, 19, 8 2018. ISSN 14220067. doi: 10.3390/IJMS19082289. URL [/pmc/articles/PMC6121519/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6121519/)[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6121519/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6121519/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6121519/).
- S. Besnard, C. Heymes, R. Merval, M. Rodriguez, J. P. Galizzi, J. A. Boutin, J. Mariani, and A. Tedgui. Expression and regulation of the nuclear receptor ror in human vascular cells. *FEBS Letters*, 511:36–40, 1 2002. ISSN 1873-3468. doi: 10.1016/S0014-5793(01)03275-6. URL [https://onlinelibrary.wiley.com/doi/full/10.1016/S0014-5793%2801%2903275-6https://onlinelibrary.wiley.com/doi/abs/10.1016/S0014-5793%2801%2903275-6https://febs.onlinelibrary.wiley.com/doi/10.1016/S0014-5793\(01\)03275-6](https://onlinelibrary.wiley.com/doi/full/10.1016/S0014-5793%2801%2903275-6https://onlinelibrary.wiley.com/doi/abs/10.1016/S0014-5793%2801%2903275-6https://febs.onlinelibrary.wiley.com/doi/10.1016/S0014-5793(01)03275-6).
- D. H. Best, K. L. Sumner, B. P. Smith, K. Damjanovich-Colmenares, I. Nakayama, L. M. Brown, Y. Ha, E. Paul, A. Morris, M. A. Jama, M. W. Dodson, P. Bayrak-Toydemir, and C. G. Elliott. Eif2ak4 mutations in patients diagnosed with pulmonary arterial hypertension. *Chest*, 151:821–828, 4 2017. ISSN 19313543. doi: 10.1016/j.chest.2016.11.014. URL [http://journal.chestnet.org/article/S0012369216623966/fulltexthttp://journal.chestnet.org/article/S0012369216623966/abstracthttps://journal.chestnet.org/article/S0012-3692\(16\)62396-6/abstract](http://journal.chestnet.org/article/S0012369216623966/fulltexthttp://journal.chestnet.org/article/S0012369216623966/abstracthttps://journal.chestnet.org/article/S0012-3692(16)62396-6/abstract).
- K. Bilguvar, K. Yasuno, M. Niemelä, Y. M. Ruigrok, M. V. U. Z. Fraunberg, C. M. V. Duijn, L. H. V. D. Berg, S. Mane, C. E. Mason, M. Choi, E. Gaál, Y. Bayri, L. Kolb, Z. Arlier, S. Ravuri, A. Ronkainen, A. Tajima, A. Laakso, A. Hata, H. Kasuya, T. Koivisto, J. Rinne, J. Öhman, M. M. Breteler, C. Wijmenga, M. W. State, G. J. Rinkel, J. Hernesniemi, J. E. Jääskeläinen, A. Palotie, I. Inoue, R. P. Lifton, and M. Günel. Susceptibility loci for intracranial aneurysm in european and japanese populations. *Nature Genetics*, 40, 12 2008. ISSN 10614036. URL [/pmc/articles/PMC2682433/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2682433/)[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2682433/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2682433/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2682433/).

- BJC. Pah module 3: the treatment of pulmonary hypertension - the british journal of cardiology, 2020. URL <https://bjcardio.co.uk/2016/12/revision-pah-module-3-the-treatment-of-pulmonary-hypertension/3/>.
- K. Blighe, R. Sharmila, and L. Myles. Enhancedvolcano: Publication-ready volcano plots with enhanced colouring and labeling. *GitHub*, 1.7.14, 2018.
- O. Boucherat, F. Potus, and S. Bonnet. microrna and pulmonary hypertension. *Advances in Experimental Medicine and Biology*, 888:237–252, 2015. ISSN 22148019. doi: 10.1007/978-3-319-22671-2_12. URL https://link.springer.com/chapter/10.1007/978-3-319-22671-2_12.
- M. Breton, R. Bellé, P. Cormier, O. Mulner-Lorillon, and J. Morales. M-phase regulation of the recruitment of mrnas onto polysomes using the cdk1/cyclin b inhibitor aminopurvalanol. *Biochemical and Biophysical Research Communications*, 306:880–886, 7 2003. ISSN 0006-291X. doi: 10.1016/S0006-291X(03)01083-0.
- A. I. Campbell, Y. Zhao, R. Sandhu, and D. J. Stewart. Cell-based gene transfer of vascular endothelial growth factor attenuates monocrotaline-induced pulmonary hypertension. *Circulation*, 104:2242–2248, 10 2001. ISSN 00097322. doi: 10.1161/hc4201.097838.
- P. Carmeliet, Y. Dor, J. M. Herber, D. Fukumura, K. Brusselmans, M. Dewerchin, M. Neeman, F. Bono, R. Abramovitch, P. Maxwell, C. J. Koch, P. Ratcliffe, L. Moons, R. K. Jain, D. Collen, and E. Keshet. Role of hif-1 in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. *Nature*, 394:485–490, 7 1998. ISSN 00280836. doi: 10.1038/28867.
- Y. chao Zhao, L. wei Xu, S. Ding, Q. qi Ji, N. Lin, Q. He, L. chen Gao, Y. yuan Su, J. Pu, and B. He. Nuclear receptor retinoid-related orphan receptor deficiency exacerbates high-fat diet-induced cardiac dysfunction despite improving metabolic abnormality. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1863: 1991–2000, 8 2017. ISSN 0925-4439. doi: 10.1016/J.BBADIS.2016.10.029.
- P. Chelladurai, W. Seeger, and S. S. Pullamsetti. Matrix metalloproteinases and their inhibitors in pulmonary hypertension. *European Respiratory Journal*, 40: 766–782, 9 2012. ISSN 0903-1936. doi: 10.1183/09031936.00209911. URL <https://erj.ersjournals.com/content/40/3/766><https://erj.ersjournals.com/content/40/3/766.abstract>.
- C. Chen, H. Chen, J. Sun, P. Bringas, Y. Chen, D. Warburton, and W. Shi. Smad1 expression and function during mouse embryonic lung branching morphogenesis. *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 288, 6 2005. ISSN 10400605. doi: 10.1152/ajplung.00277.2004.

- X. Chen, J. Qin, C.-M. Cheng, M.-J. Tsai, and S. Y. Tsai. Coup-tfii is a major regulator of cell cycle and notch signaling pathways. *Molecular endocrinology (Baltimore, Md.)*, 26:1268–77, 8 2012. ISSN 1944-9917. doi: 10.1210/me.2011-1305. URL <http://www.ncbi.nlm.nih.gov/pubmed/22734039><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3404301>.
- J. C. Cheng, H. M. Chang, and P. C. Leung. Egr-1 mediates epidermal growth factor-induced downregulation of e-cadherin expression via slug in human ovarian cancer cells. *Oncogene*, 32:1041–1049, 2 2013. ISSN 09509232. doi: 10.1038/onc.2012.127.
- A. Chida, M. Shintani, T. Nakayama, Y. Furutani, E. Hayama, K. Inai, T. Saji, S. Nonoyama, and T. Nakanishi. Missense mutations of the *bmpr1b* (*alk6*) gene in childhood idiopathic pulmonary arterial hypertension. *Circulation Journal*, 76:1501–1508, 6 2012. ISSN 1346-9843. doi: 10.1253/circj.CJ-11-1281. URL <http://japanlinkcenter.org/JST.JSTAGE/circj/CJ-11-1281?lang=en&from=CrossRef&type=abstract>.
- K. M. Chin and L. J. Rubin. Pulmonary arterial hypertension, 4 2008. ISSN 07351097.
- B. W. Christman, C. D. McPherson, J. H. Newman, G. A. King, G. R. Bernard, B. M. Groves, and J. E. Loyd. An imbalance between the excretion of thromboxane and prostacyclin metabolites in pulmonary hypertension. *New England Journal of Medicine*, 327:70–75, 7 1992. ISSN 0028-4793. doi: 10.1056/NEJM199207093270202. URL <http://www.nejm.org/doi/abs/10.1056/NEJM199207093270202>.
- L. H. Clapp, P. Finney, S. Turcato, S. Tran, L. J. Rubin, and A. Tinker. Differential effects of stable prostacyclin analogs on smooth muscle proliferation and cyclic amp generation in human pulmonary artery. *American Journal of Respiratory Cell and Molecular Biology*, 26:194–201, 2 2002. ISSN 10441549. doi: 10.1165/ajrcmb.26.2.4695. URL <http://www.atsjournals.org/doi/abs/10.1165/ajrcmb.26.2.4695>.
- R. L. Clarke, A. D. Yzaguirre, Y. Yashiro-Ohtani, A. Bondue, C. Blanpain, W. S. Pear, N. A. Speck, and G. Keller. The expression of *sox17* identifies and regulates haemogenic endothelium. *Nature Cell Biology*, 15:502–510, 5 2013. ISSN 1465-7392. doi: 10.1038/ncb2724. URL <http://www.nature.com/articles/ncb2724><https://www.nature.com/articles/ncb2724>.
- E. M. Conway and P. Carmeliet. The diversity of endothelial cells: A challenge for therapeutic angiogenesis. *Genome Biology*, 5:1–5, 1 2004. ISSN 14656906. doi: 10.1186/GB-2004-5-2-207/METRICS. URL <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2004-5-2-207>.
- C. D. Cool, D. Kennedy, N. F. Voelkel, and R. M. Tuder. Pathogenesis and evolution of plexiform lesions in pulmonary hypertension associated with scleroderma and hu-

- man immunodeficiency virus infection. *Human Pathology*, 28:434–442, 1997. ISSN 00468177. doi: 10.1016/S0046-8177(97)90032-0. URL <https://pubmed.ncbi.nlm.nih.gov/9104943/https://pubmed.ncbi.nlm.nih.gov/9104943/?dopt=Abstract>.
- C. D. Cool, J. S. Stewart, P. Werahera, G. J. Miller, R. L. Williams, N. F. Voelkel, and R. M. Tudor. Three-dimensional reconstruction of pulmonary arteries in plexiform pulmonary hypertension using cell-specific markers: Evidence for a dynamic and heterogeneous process of pulmonary endothelial cell growth. *American Journal of Pathology*, 155:411–419, 1999. ISSN 00029440. doi: 10.1016/S0002-9440(10)65137-1.
- M. Corada, F. Liao, M. Lindgren, M. G. Lampugnani, F. Breviario, R. Frank, W. A. Muller, D. J. Hicklin, P. Bohlen, and E. Dejana. Monoclonal antibodies directed to different regions of vascular endothelial cadherin extracellular domain affect adhesion and clustering of the protein and modulate endothelial permeability. *Blood*, 97:1679–1684, 3 2001. ISSN 0006-4971. doi: 10.1182/BLOOD.V97.6.1679. URL <http://ashpublications.org/blood/article-pdf/97/6/1679/1672733/h8060101679.pdf>.
- M. Corada, F. Orsenigo, M. F. Morini, M. E. Pitulescu, G. Bhat, D. Nyqvist, F. Breviario, V. Conti, A. Briot, M. L. Iruela-Arispe, R. H. Adams, and E. Dejana. Sox17 is indispensable for acquisition and maintenance of arterial identity. *Nature Communications*, 4:2609, 12 2013. ISSN 2041-1723. doi: 10.1038/ncomms3609. URL <http://www.ncbi.nlm.nih.gov/pubmed/24153254http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3826640http://www.nature.com/articles/ncomms3609>.
- M. P. Craig and S. Sumanas. Ets transcription factors in embryonic vascular development. *Angiogenesis*, 19:275, 7 2016. ISSN 15737209. doi: 10.1007/S10456-016-9511-Z. URL </pmc/articles/PMC4937796//pmc/articles/PMC4937796/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4937796/>.
- S. Dabral, X. Tian, B. Kojonazarov, R. Savai, H. A. Ghofrani, N. Weissmann, M. Florio, J. Sun, D. Jonigk, L. Maegel, F. Grimminger, W. Seeger, S. S. Pullamsetti, and R. T. Schermuly. Notch1 signalling regulates endothelial proliferation and apoptosis in pulmonary arterial hypertension. *European Respiratory Journal*, 48:1137–1149, 10 2016. ISSN 13993003. doi: 10.1183/13993003.00773-2015.
- Z. Dai, M. Li, J. Wharton, M. M. Zhu, and Y. Y. Zhao. Prolyl-4 hydroxylase 2 (phd2) deficiency in endothelial cells and hematopoietic cells induces obliterative vascular remodeling and severe pulmonary arterial hypertension in mice and humans through hypoxia-inducible factor-2. *Circulation*, 133:2447–2458, 6 2016. ISSN 15244539. doi: 10.1161/CIRCULATIONAHA.116.021494.

- G. D'Alonzo, R. Barst, S. Ayres, E. Bergofsky, B. Brundage, K. Detre, A. Fishman, R. Goldring, B. Groves, J. Kernis, P. Levy, G. Pietra, L. Reid, J. Reid, S. Reeves, C. Vreim, G. Williams, and M. Wu. Survival in patients with primary pulmonary hypertension. *Annals of Internal Medicine*, 115:343, 9 1991. ISSN 0003-4819. doi: 10.7326/0003-4819-115-5-343.
- E. Dejana. The role of wnt signaling in physiological and pathological angiogenesis, 10 2010. ISSN 00097330. URL <https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.110.223750>.
- E. Dejana, G. Bazzoni, and M. G. Lampugnani. Vascular endothelial (ve)-cadherin: Only an intercellular glue? *Experimental Cell Research*, 252:13–19, 10 1999. ISSN 0014-4827. doi: 10.1006/EXCR.1999.4601.
- P. Delerive, D. Monté, G. Dubois, F. Trottein, J. Fruchart-Najib, J. Mariani, J. C. Fruchart, and B. Staels. The orphan nuclear receptor ror is a negative regulator of the inflammatory response. *EMBO Reports*, 2:42, 2001. ISSN 1469221X. doi: 10.1093/EMBO-REPORTS/KVE007. URL <https://pmc/articles/PMC1083804/>
[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1083804/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1083804/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1083804/).
- Z. Deng, J. H. Morse, S. L. Slager, N. Cuervo, K. J. Moore, G. Venetos, S. Kalachikov, E. Cayanis, S. G. Fischer, R. J. Barst, S. E. Hodge, and J. A. Knowles. Familial primary pulmonary hypertension (gene pph1) is caused by mutations in the bone morphogenetic protein receptor-ii gene. *American Journal of Human Genetics*, 67:737–744, 2000. ISSN 00029297. doi: 10.1086/303059.
- M. G. Dickinson, B. Bartelds, G. Molema, M. A. Borgdorff, B. Boersma, J. Takens, M. Weij, P. Wichers, H. Sietsma, and R. M. Berger. Egr-1 expression during neonatal development in flow-associated pulmonary hypertension. *American Journal of Pathology*, 179:2199–2209, 11 2011. ISSN 00029440. doi: 10.1016/j.ajpath.2011.07.030.
- H. Duez and B. Staels. The nuclear receptors rev-erbs and rors integrate circadian rhythms and metabolism. *Diabetes vascular disease research*, 5:82–88, 2008. ISSN 1479-1641. doi: 10.3132/DVDR.2008.0014. URL <https://pubmed.ncbi.nlm.nih.gov/18537094/>.
- T. Duong, K. Koltowska, C. Pichol-Thievend, L. L. Guen, F. Fontaine, K. A. Smith, V. Truong, R. Skoczylas, S. A. Stacker, M. G. Achen, P. Koopman, B. M. Hogan, and M. Francois. Vegfd regulates blood vascular development by modulating sox18 activity. *Blood*, 123:1102–1112, 2 2014. ISSN 0006-4971. doi: 10.1182/BLOOD-2013-04-495432. URL <http://www.bloodjournal.org/content/123/7/1102.long?sso-checked=true>.

- C. A. Eichstaedt, J. Song, R. R. Viales, Z. Pan, N. Benjamin, C. Fischer, M. M. Hoeper, S. Ulrich, K. Hinderhofer, and E. Grünig. First identification of krüppel-like factor 2 mutation in heritable pulmonary arterial hypertension. *Clinical Science*, 131:689–698, 4 2017. ISSN 14708736. doi: 10.1042/CS20160930.
- M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95:14863–14868, 12 1998. ISSN 00278424. doi: 10.1073/pnas.95.25.14863.
- J. D. Evans, B. Girerd, D. Montani, X. J. Wang, N. Galiè, E. D. Austin, G. Elliott, K. Asano, E. Grünig, Y. Yan, Z. C. Jing, A. Manes, M. Palazzini, L. A. Wheeler, I. Nakayama, T. Satoh, C. Eichstaedt, K. Hinderhofer, M. Wolf, E. B. Rosenzweig, W. K. Chung, F. Soubrier, G. Simonneau, O. Sitbon, S. Gräf, S. Kaptoge, E. D. Angelantonio, M. Humbert, and N. W. Morrell. Bmpr2 mutations and survival in pulmonary arterial hypertension: An individual participant data meta-analysis. *The Lancet Respiratory Medicine*, 4:129–137, 2 2016. ISSN 22132619. doi: 10.1016/S2213-2600(15)00544-5. URL <http://www.ncbi.nlm.nih.gov/pubmed/26795434><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4737700>.
- C. Fellmann, B. G. Gowen, P. C. Lin, J. A. Doudna, and J. E. Corn. Cornerstones of crispr-cas in drug discovery and therapy, 2 2017. ISSN 14741784. URL <https://www.nature.com/articles/nrd.2016.238><https://www.nature.com/articles/nrd.2016.238/>.
- R. J. Fleming. Structural conservation of notch receptors and ligands. *Seminars in Cell and Developmental Biology*, 9:599–607, 1998. ISSN 10849521. doi: 10.1006/scdb.1998.0260.
- M. P. Fogarty, M. E. Cannon, S. Vadlamudi, K. J. Gaulton, and K. L. Mohlke. Identification of a regulatory variant that binds foxa1 and foxa2 at the cdc123/camk1d type 2 diabetes gwas locus. *PLoS Genetics*, 10:1004633, 9 2014. ISSN 15537404. doi: 10.1371/JOURNAL.PGEN.1004633. URL </pmc/articles/PMC4161327/></pmc/articles/PMC4161327/?report=abstract><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4161327/>.
- T. Foroud, D. L. Koller, D. Lai, L. Sauerbeck, C. Anderson, N. Ko, R. Deka, T. H. Mosley, M. Fornage, D. Woo, C. J. Moomaw, R. Hornung, J. Huston, I. Meissner, J. E. Bailey-wilson, C. Langefeld, G. Rouleau, E. S. Connolly, B. B. Worrall, D. Kleindorfer, M. L. Flaherty, S. Martini, J. Mackey, F. D. L. R. L. Rosa, R. D. Brown, J. P. Broderick, J. E. Bailey-Wilson, C. Langefeld, G. Rouleau, E. S. Connolly, B. B. Worrall, D. Kleindorfer, M. L. Flaherty, S. Martini, J. Mackey, F. D. L. R. L. Rosa, R. D. Brown, J. P. Broderick, J. P. F. S. Investigators, and the FIA Study Investigators. Genome-wide

- association study of intracranial aneurysms confirms role of anril and sox17 in disease risk. *Stroke*, 43:2846–52, 11 2012. ISSN 1524-4628. doi: 10.1161/STROKEAHA.112.656397. URL <http://www.ncbi.nlm.nih.gov/pubmed/22961961><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3752852>.
- M. Francois, P. Koopman, and M. Beltrame. Sox genes: Key players in the development of the cardio-vascular system. *The International Journal of Biochemistry Cell Biology*, 42: 445–448, 3 2010. ISSN 1357-2725. doi: 10.1016/J.BIOCEL.2009.08.017. URL <https://www.sciencedirect.com/science/article/pii/S1357272509002374?via%3Dihub>.
- M. François, A. Caprini, B. Hosking, F. Orsenigo, D. Wilhelm, C. Browne, K. Paavonen, T. Karnezis, R. Shayan, M. Downes, T. Davidson, D. Tutt, K. S. E. Cheah, S. A. Stacker, G. E. O. Muscat, M. G. Achen, E. Dejana, and P. Koopman. Sox18 induces development of the lymphatic vasculature in mice. *Nature*, 456:643–647, 12 2008. ISSN 0028-0836. doi: 10.1038/nature07391. URL <http://www.nature.com/articles/nature07391>.
- N. Galiè, H. A. Ghofrani, A. Torbicki, R. J. Barst, L. J. Rubin, D. Badesch, T. Fleming, T. Parpia, G. Burgess, A. Branzi, F. Grimminger, M. Kurzyna, and G. Simonneau. Sildenafil citrate therapy for pulmonary arterial hypertension. *The New England Journal of Medicine*, 353:2148–2157, 2005. ISSN 0028-4793. URL <https://www.nejm.org/doi/10.1056/NEJMoa050010><http://www.nejm.org/doi/abs/10.1056/NEJMoa050010>.
- N. Galiè, H. Olschewski, R. J. Oudiz, F. Torres, A. Frost, H. A. Ghofrani, D. B. Badesch, M. D. McGoon, V. V. McLaughlin, E. B. Roecker, M. J. Gerber, C. Dufton, B. L. Wiens, and L. J. Rubin. Ambrisentan for the treatment of pulmonary arterial hypertension: Results of the ambrisentan in pulmonary arterial hypertension, randomized, double-blind, placebo-controlled, multicenter, efficacy (aries) study 1 and 2. *Circulation*, 117: 3010–3019, 6 2008. ISSN 00097322. doi: 10.1161/CIRCULATIONAHA.107.742510.
- N. Galiè, B. H. Brundage, H. A. Ghofrani, R. J. Oudiz, G. Simonneau, Z. Safdar, S. Shapiro, R. J. White, M. Chan, A. Beardsworth, L. Frumkin, and R. J. Barst. Tadalafil therapy for pulmonary arterial hypertension. *Circulation*, 119:2894–2903, 6 2009. ISSN 00097322. doi: 10.1161/CIRCULATIONAHA.108.839274.
- N. Galiè, M. Humbert, J.-L. Vachiery, S. Gibbs, I. Lang, A. Torbicki, G. Simonneau, A. Peacock, A. V. Noordegraaf, M. Beghetti, A. Ghofrani, M. A. G. Sanchez, G. Hansmann, W. Klepetko, P. Lancellotti, M. Matucci, T. McDonagh, L. A. Pierard, P. T. Trindade, M. Zompatori, and M. Hoeper. 2015 esc/ers guidelines for the diagnosis and treatment of pulmonary hypertension. *European Respiratory Journal*, 46:903–975,

- 10 2015a. doi: 10.1183/13993003.01032-2015. URL <http://www.ncbi.nlm.nih.gov/pubmed/26318161>.
- N. Galiè, K. Müller, A. V. Scalise, and E. Grünig. Patent plus: A blinded, randomised and extension study of riociguat plus sildenafil in pulmonary arterial hypertension. *European Respiratory Journal*, 45:1314–1322, 5 2015b. ISSN 13993003. doi: 10.1183/09031936.00105914.
- N. Galiè, V. V. McLaughlin, L. J. Rubin, and G. Simonneau. An overview of the 6th world symposium on pulmonary hypertension, 1 2019. ISSN 13993003. URL <http://erj.ersjournals.com/lookup/doi/10.1183/13993003.02148-2018>.
- M. Germain, M. Eyries, D. Montani, O. Poirier, B. Girerd, P. Dorfmüller, F. Coulet, S. Nadaud, S. Maugenre, C. Guignabert, W. Carpentier, A. Vonk-Noordegraaf, M. Lévy, A. Chaouat, J. C. Lambert, M. Bertrand, A. M. Dupuy, L. Letenneur, M. Lathrop, P. Amouyel, T. J. D. Ravel, M. Delcroix, E. D. Austin, I. M. Robbins, A. R. Hemnes, J. E. Loyd, E. Berman-Rosenzweig, R. J. Barst, W. K. Chung, G. Simonneau, D. A. Tréguët, M. Humbert, and F. Soubrier. Genome-wide association analysis identifies a susceptibility locus for pulmonary arterial hypertension. *Nature Genetics*, 45:518–521, 5 2013. ISSN 10614036. doi: 10.1038/ng.2581.
- P. Ghataorhe, C. J. Rhodes, L. Harbaum, M. Attard, J. Wharton, and M. R. Wilkins. Pulmonary arterial hypertension - progress in understanding the disease and prioritizing strategies for drug development. *Journal of Internal Medicine*, 282:129–141, 8 2017. ISSN 09546820. doi: 10.1111/joim.12623. URL <http://doi.wiley.com/10.1111/joim.12623>.
- H.-A. Ghofrani, A. M. D’Armini, F. Grimminger, M. M. Hoeper, P. Jansa, N. H. Kim, E. Mayer, G. Simonneau, M. R. Wilkins, A. Fritsch, D. Neuser, G. Weimann, and C. Wang. Riociguat for the treatment of chronic thromboembolic pulmonary hypertension. *New England Journal of Medicine*, 369:319–329, 7 2013. ISSN 0028-4793. doi: 10.1056/NEJMoa1209657. URL <http://www.nejm.org/doi/10.1056/NEJMoa1209657>.
- A. Giaid and D. Saleh. Reduced expression of endothelial nitric oxide synthase in the lungs of patients with pulmonary hypertension. *New England Journal of Medicine*, 333:214–221, 7 1995. ISSN 0028-4793. doi: 10.1056/NEJM199507273330403. URL <http://www.nejm.org/doi/abs/10.1056/NEJM199507273330403>.
- A. Giaid, M. Yanagisawa, D. Langleben, R. P. Michel, R. Levy, H. Shennib, S. Kimura, T. Masaki, W. P. Duguid, and D. J. Stewart. Expression of endothelin-1 in the lungs of patients with pulmonary hypertension. *New England Journal of Medicine*, 328:1732–1739, 6 1993. ISSN 15334406. doi: 10.1056/NEJM199306173282402.

- G. Gibson. Rare and common variants: twenty arguments. *Nature reviews. Genetics*, 13:135–45, 1 2012. ISSN 1471-0064. doi: 10.1038/nrg3118. URL <http://www.ncbi.nlm.nih.gov/pubmed/22251874><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4408201>.
- L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, and L. S. Qi. Crispr-mediated modular rna-guided regulation of transcription in eukaryotes. *Cell*, 154:442, 7 2013. ISSN 10974172. doi: 10.1016/J.CELL.2013.06.044. URL </pmc/articles/PMC3770145//pmc/articles/PMC3770145/?report=abstract><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3770145/>.
- J. M. Gillbro, E. Merinville, M. Nilsson, E. Hagforsen, G. Moran, T. Al-Bader, and A. Mavon. Use of "connectivity map" (cmap) to identify novel, safe, effective antiaging ingredients for topical use, 4 2017.
- I. Goldstein, T. F. Lue, H. Padma-Nathan, R. C. Rosen, W. D. Steers, and P. A. Wicker. Oral sildenafil in the treatment of erectile dysfunction. *New England Journal of Medicine*, 338:1397–1404, 5 1998. ISSN 0028-4793. doi: 10.1056/NEJM199805143382001. URL <http://www.nejm.org/doi/abs/10.1056/NEJM199805143382001>.
- R. B. Good, A. J. Gilbane, S. L. Trinder, C. P. Denton, G. Coghlan, D. J. Abraham, and A. M. Holmes. Endothelial to mesenchymal transition contributes to endothelial dysfunction in pulmonary arterial hypertension. *American Journal of Pathology*, 185: 1850–1858, 7 2015. ISSN 15252191. doi: 10.1016/j.ajpath.2015.03.019.
- R. Gordân, N. Shen, I. Dror, T. Zhou, J. Horton, R. Rohs, and M. L. Bullyk. Genomic regions flanking e-box binding sites influence dna binding specificity of bhlh transcription factors through dna shape. *Cell reports*, 3: 1093–104, 4 2013. ISSN 2211-1247. doi: 10.1016/j.celrep.2013.03.014. URL <http://www.ncbi.nlm.nih.gov/pubmed/23562153><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3640701>.
- S. Gräf, M. Haimel, M. Bleda, C. Hadinnapola, L. Southgate, W. Li, J. Hodgson, B. Liu, R. M. Salmon, M. Southwood, R. D. Machado, J. M. Martin, C. M. Treacy, K. Yates, L. C. Daugherty, O. Shamardina, D. Whitehorn, S. Holden, M. Aldred, H. J. Bogaard, C. Church, G. Coghlan, R. Condliffe, P. A. Corris, C. Danesino, M. Eyries, H. Gall, S. Ghio, H.-A. Ghofrani, J. S. R. Gibbs, B. Girerd, A. C. Houweling, L. Howard, M. Humbert, D. G. Kiely, G. Kovacs, R. V. M. Ross, S. Moledina, D. Montani, M. Newnham, A. Olschewski, H. Olschewski, A. J. Peacock, J. Pepke-Zaba, I. Prokopenko, C. J. Rhodes, L. Scelsi, W. Seeger, F. Soubrier, D. F. Stein,

- J. Suntharalingam, E. M. Swietlik, M. R. Toshner, D. A. van Heel, A. V. Noordegraaf, Q. Waisfisz, J. Wharton, S. J. Wort, W. H. Ouwehand, N. Soranzo, A. Lawrie, P. D. Upton, M. R. Wilkins, R. C. Trembath, and N. W. Morrell. Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nature Communications*, 9:1416, 12 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-03672-4. URL <http://www.nature.com/articles/s41467-018-03672-4>.
- M. Gu, M. Donato, M. Guo, N. Wary, Y. Miao, S. Mao, T. Saito, S. Otsuki, L. Wang, R. L. Harper, S. Sa, P. Khatri, and M. Rabinovitch. ipsc-endothelial cell phenotypic drug screening and in silico analyses identify tyrphostin-ag1296 for pulmonary arterial hypertension. *Science Translational Medicine*, 13, 5 2021. ISSN 19466242. doi: 10.1126/scitranslmed.aba6480. URL <https://stm.sciencemag.org/content/13/592/eaba6480><https://stm.sciencemag.org/content/13/592/eaba6480.abstract>.
- E. H. Ha, J. P. Choi, H. S. Kwon, H. J. Park, S. J. Lah, K. A. Moon, S. H. Lee, I. Kim, and Y. S. Cho. Endothelial sox17 promotes allergic airway inflammation. *Journal of Allergy and Clinical Immunology*, 144:561–573.e6, 8 2019. ISSN 0091-6749. doi: 10.1016/J.JACI.2019.02.034. URL <http://www.jacionline.org/article/S0091674919304142/fulltext><http://www.jacionline.org/article/S0091674919304142/abstract>[https://www.jacionline.org/article/S0091-6749\(19\)30414-2/abstract](https://www.jacionline.org/article/S0091-6749(19)30414-2/abstract).
- L. Harbaum, K. M. Baaske, M. Simon, T. Oqueka, C. Sinning, A. Glatzel, N. Lüneburg, K. Sydow, C. Bokemeyer, and H. Klose. Exploratory analysis of the neutrophil to lymphocyte ratio in patients with pulmonary arterial hypertension. *BMC Pulmonary Medicine*, 17:72, 12 2017. ISSN 1471-2466. doi: 10.1186/s12890-017-0407-5. URL <http://bmcpulmed.biomedcentral.com/articles/10.1186/s12890-017-0407-5>.
- R. E. Harrison, R. Berger, S. G. Haworth, R. Tulloh, C. J. Mache, N. W. Morrell, M. A. Aldred, and R. C. Trembath. Transforming growth factor- receptor mutations and pulmonary arterial hypertension in childhood. *Circulation*, 111:435–441, 2 2005. ISSN 0009-7322. doi: 10.1161/01.CIR.0000153798.78540.87. URL <https://www.ahajournals.org/doi/10.1161/01.CIR.0000153798.78540.87>.
- A. Hautefort, J. Chesné, J. Preussner, S. S. Pullamsetti, J. Tost, M. Looso, F. Antigny, B. Girerd, M. Riou, S. Eddahibi, J.-F. Deleuze, W. Seeger, E. Fadel, G. Simonneau, D. Montani, M. Humbert, and F. Perros. Pulmonary endothelial cell dna methylation signature in pulmonary arterial hypertension. *Oncotarget*, 8:52995, 8 2017. ISSN 1949-2553. doi: 10.18632/ONCOTARGET.18031. URL [/pmc/articles/PMC5581088/](https://pmc/articles/PMC5581088/)<https://pmc/articles/PMC5581088/?report=abstract><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5581088/>.

- X. He, M. Semenov, K. Tamai, and X. Zeng. Ldl receptor-related proteins 5 and 6 in wnt/-catenin signaling: Arrows point the way, 4 2004. ISSN 09501991.
- Y. He, C. Zuo, D. Jia, P. Bai, D. Kong, D. Chen, G. Liu, J. Li, Y. Wang, G. Chen, S. Yan, B. Xiao, J. Zhang, L. Piao, Y. Li, Y. Deng, B. Li, P. P. Roux, K. I. Andreasson, R. M. Breyer, Y. Su, J. Wang, A. Lyu, Y. Shen, and Y. Yu. Loss of dp1 aggravates vascular remodeling in pulmonary arterial hypertension via mtorc1 signaling. *American Journal of Respiratory and Critical Care Medicine*, 201:1263–1276, 2020. ISSN 15354970. doi: 10.1164/RCCM.201911-2137OC/SUPPL_FILE/DISCLOSURES.PDF. URL [/pmc/articles/PMC7233340/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233340/)[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233340/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233340/?report=abstract).
- J. K. Hebda, H. M. Leclair, S. Azzi, C. Roussel, M. G. Scott, N. Bidère, and J. Gavard. The c-terminus region of -arrestin1 modulates ve-cadherin expression and endothelial cell permeability. *Cell Communication and Signaling*, 11:1–7, 5 2013. ISSN 1478811X. doi: 10.1186/1478-811X-11-37/FIGURES/3. URL <https://biosignaling.biomedcentral.com/articles/10.1186/1478-811X-11-37>.
- A. D. Hingorani, V. Kuan, C. Finan, F. A. Kruger, A. Gaulton, S. Chopade, R. Sofat, R. J. MacAllister, J. P. Overington, H. Hemingway, S. Denaxas, D. Prieto, and J. P. Casas. Improving the odds of drug development success through human genomics: modelling study. *Scientific Reports 2019 9:1*, 9:1–25, 12 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-54849-w. URL <https://www.nature.com/articles/s41598-019-54849-w>.
- M. M. Hoeper, D. Huscher, H. A. Ghofrani, M. Delcroix, O. Distler, C. Schweiger, E. Grunig, G. Staehler, S. Rosenkranz, M. Halank, M. Held, C. Grohé, T. J. Lange, J. Behr, H. Klose, H. Wilkens, A. Filusch, M. Germann, R. Ewert, H. J. Seyfarth, K. M. Olsson, C. F. Opitz, S. P. Gaine, C. D. Vizza, A. Vonk-Noordegraaf, H. Kaemmerer, J. S. R. Gibbs, and D. Pittrow. Elderly patients diagnosed with idiopathic pulmonary arterial hypertension: Results from the compera registry. *International Journal of Cardiology*, 168:871–880, 9 2013. ISSN 18741754. doi: 10.1016/j.ijcard.2012.10.026.
- M. M. Hoeper, M. Humbert, R. Souza, M. Idrees, S. M. Kawut, K. Sliwa-Hahnle, Z. C. Jing, and J. S. R. Gibbs. A global view of pulmonary hypertension, 4 2016a. ISSN 22132619. URL <https://www.sciencedirect.com/science/article/pii/S2213260015005433?via%3Dihub>.
- M. M. Hoeper, V. V. McLaughlin, A. M. Dalaan, T. Satoh, and N. Galiè. Treatment of pulmonary hypertension, 4 2016b. ISSN 22132619.
- R. K. Hopper, J. R. A. Moonen, I. Diebold, A. Cao, C. J. Rhodes, N. F. Tojais, J. K. Hennigs, M. Gu, L. Wang, and M. Rabinovitch. In pulmonary arterial hy-

- pertension, reduced bmpr2 promotes endothelial-to-mesenchymal transition via hmgal and its target slug. *Circulation*, 133:1783–1794, 5 2016. ISSN 15244539. doi: 10.1161/CIRCULATIONAHA.115.020617.
- B. Hrdlickova, R. C. de Almeida, Z. Borek, and S. Withoff. Genetic variation in the non-coding genome: Involvement of micro-rnas and long non-coding rnas in disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842:1910–1922, 10 2014. ISSN 0925-4439. doi: 10.1016/J.BBADIS.2014.03.011.
- F. Hu, M. Li, L. Mo, Y. Xiao, X. Wang, and B. Xie. Sox-17 is involved in invasion and apoptosis of colorectal cancer cells through regulating mir-302b-3p expression. *Cell biology international*, 45:1296–1305, 6 2021. ISSN 1095-8355. doi: 10.1002/CBIN.11594. URL <https://pubmed.ncbi.nlm.nih.gov/33739578/>.
- L. E. Huang, J. Gu, M. Schau, and H. F. Bunn. Regulation of hypoxia-inducible factor 1 is mediated by an o₂-dependent degradation domain via the ubiquitin-proteasome pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 95:7987–7992, 7 1998. ISSN 00278424. doi: 10.1073/pnas.95.14.7987.
- M. Humbert, O. Sitbon, A. Chaouat, M. Bertocchi, G. Habib, V. Gressin, A. Yaici, E. Weitzenblum, J.-F. Cordier, F. Chabot, C. Dromer, C. Pison, M. Reynaud-Gaubert, A. Haloun, M. Laurent, E. Hachulla, and G. Simonneau. Pulmonary arterial hypertension in france. *American Journal of Respiratory and Critical Care Medicine*, 173:1023–1030, 5 2006. ISSN 1073-449X. doi: 10.1164/rccm.200510-1668OC. URL <http://www.ncbi.nlm.nih.gov/pubmed/16456139><http://www.atsjournals.org/doi/abs/10.1164/rccm.200510-1668OC>.
- M. Humbert, E. S. Segal, D. G. Kiely, J. Carlsen, B. Schwierin, and M. M. Hoeper. Results of european post-marketing surveillance of bosentan in pulmonary hypertension. *European Respiratory Journal*, 30:338–344, 8 2007. ISSN 09031936. doi: 10.1183/09031936.00138706.
- M. Humbert, E. M. Lau, D. Montani, X. Jaïs, O. Sitbon, and G. Simonneau. Advances in therapeutic interventions for patients with pulmonary arterial hypertension. *Circulation*, 130:2189–2208, 12 2014. ISSN 15244539. doi: 10.1161/CIRCULATIONAHA.114.006974. URL <https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.114.006974>.
- M. Humbert, C. Guignabert, S. Bonnet, P. Dorfmüller, J. R. Klinger, M. R. Nicolls, A. J. Olschewski, S. S. Pullamsetti, R. T. Schermuly, K. R. Stenmark, and M. Rabinovitch. Pathology and pathobiology of pulmonary hypertension: state of the art and research perspectives. *The European respiratory journal*, 53, 1 2019. ISSN

13993003. doi: 10.1183/13993003.01887-2018. URL <https://pmc/articles/PMC6351340/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6351340/>.
- L. A. Hurst, B. J. Dunmore, L. Long, A. Crosby, R. Al-Lamki, J. Deighton, M. Southwood, X. Yang, M. Z. Nikolic, B. Herrera, G. J. Inman, J. R. Bradley, A. A. Rana, P. D. Upton, and N. W. Morrell. Tnf drives pulmonary arterial hypertension by suppressing the bmp type-ii receptor and altering notch signalling. *Nature Communications* 2017 8:1, 8:1–14, 1 2017. ISSN 2041-1723. doi: 10.1038/ncomms14079. URL <https://www.nature.com/articles/ncomms14079>.
- M. Iglarz, C. Binkert, K. Morrison, W. Fischli, J. Gatfield, A. Treiber, T. Weller, M. H. Bolli, C. Boss, S. Buchmann, B. Capeleto, P. Hess, C. Qiu, and M. Clozel. Pharmacology of macitentan, an orally active tissue-targeting dual endothelin receptor antagonist. *Journal of Pharmacology and Experimental Therapeutics*, 327:736–745, 12 2008. ISSN 00223565. doi: 10.1124/jpet.108.142976.
- S. Islam, P. Masiakos, J. J. Schnitzer, D. P. Doody, and D. P. Ryan. Diltiazem reduces pulmonary arterial pressures in recurrent pulmonary hypertension associated with pulmonary hypoplasia. volume 34, pages 712–714. W.B. Saunders, 1999. doi: 10.1016/S0022-3468(99)90361-5. URL <https://pubmed.ncbi.nlm.nih.gov/10359169/>.
- N. V. Iyer, L. E. Kotch, F. Agani, S. W. Leung, E. Laughner, R. H. Wenger, M. Gassmann, J. D. Gearhart, A. M. Lawler, A. Y. Yu, and G. L. Semenza. Cellular and developmental control of o₂ homeostasis by hypoxia- inducible factor 1. *Genes and Development*, 12: 149–162, 1 1998. ISSN 08909369. doi: 10.1101/gad.12.2.149.
- K. Jandl, L. M. Marsh, J. Hoffmann, A. C. Mutgan, O. Baum, W. Bloch, H. Thekkekaraputhenparampil, D. Kolb, K. Sinn, W. Klepetko, A. Heinemann, A. Olschewski, H. Olschewski, and G. Kwapiszewska. Basement membrane remodeling controls endothelial function in idiopathic pulmonary arterial hypertension. *American Journal of Respiratory Cell and Molecular Biology*, 63:104–117, 7 2020. ISSN 15354989. doi: 10.1165/RCMB.2019-0303OC/SUPPL_FILE/DISCLOSURES.PDF. URL www.atsjournals.org.
- Z. G. Jin, H. Ueba, T. Tanimoto, A. O. Lungu, M. D. Frame, and B. C. Berk. Ligand-independent activation of vascular endothelial growth factor receptor 2 by fluid shear stress regulates activation of endothelial nitric oxide synthase. *Circulation Research*, 93:354–363, 8 2003. ISSN 00097330. doi: 10.1161/01.RES.0000089257.94002.96.
- Z. C. Jing, X. Jiang, Z. Y. Han, X. Q. Xu, Y. Wang, Y. Wu, H. Lv, C. R. Ma, Y. J. Yang, and J. L. Pu. Iloprost for pulmonary vasodilator testing in idiopathic pulmonary arterial hypertension. *European Respiratory Journal*, 33:1354–1360, 6 2009. ISSN 09031936.

- doi: 10.1183/09031936.00169608. URL <https://erj.ersjournals.com/content/33/6/1354><https://erj.ersjournals.com/content/33/6/1354.abstract>.
- Z. C. Jing, K. Parikh, T. Pulido, C. Jerjes-Sanchez, R. J. White, R. Allen, A. Torbicki, K. F. Xu, D. Yehle, K. Laliberte, C. Arneson, and L. J. Rubin. Efficacy and safety of oral treprostinil monotherapy for the treatment of pulmonary arterial hypertension: A randomized, controlled trial. *Circulation*, 127:624–633, 2 2013. ISSN 00097322. doi: 10.1161/CIRCULATIONAHA.112.124388.
- R. C. Jones, M. Jacobson, and W. Steudel. -smooth-muscle actin and microvascular precursor smooth-muscle cells in pulmonary hypertension. *American Journal of Respiratory Cell and Molecular Biology*, 20:582–594, 4 1999. ISSN 10441549. doi: 10.1165/ajrcmb.20.4.3357. URL <http://www.atsjournals.org/doi/abs/10.1165/ajrcmb.20.4.3357>.
- D. Jonigk, H. Golpon, C. L. Bockmeyer, L. Maegel, M. M. Hoeper, J. Gottlieb, N. Nickel, K. Hussein, U. Maus, U. Lehmann, S. Janciauskiene, T. Welte, A. Haverich, J. Rische, H. Kreipe, and F. Laenger. Plexiform lesions in pulmonary arterial hypertension: Composition, architecture, and microenvironment. *American Journal of Pathology*, 179: 167–179, 7 2011. ISSN 00029440. doi: 10.1016/j.ajpath.2011.03.040.
- V. Kalna, Y. Yang, C. R. Peghaire, K. Frudd, R. Hannah, A. V. Shah, L. O. Almagro, J. J. Boyle, B. Göttgens, J. Ferrer, A. M. Randi, and G. M. Birdsey. The transcription factor *erg* regulates super-enhancers associated with an endothelial-specific gene expression program. *Circulation Research*, 124:1337, 4 2019. ISSN 15244571. doi: 10.1161/CIRCRESAHA.118.313788. URL [/pmc/articles/PMC6493686/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6493686/)[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6493686/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6493686/?report=abstract).
- Y. Kamachi and H. Kondoh. Sox proteins: regulators of cell fate specification and differentiation. *Development (Cambridge, England)*, 140:4129–44, 10 2013. ISSN 1477-9129. doi: 10.1242/dev.091793. URL <http://www.ncbi.nlm.nih.gov/pubmed/24086078>.
- M. Katoh and M. Katoh. Wnt signaling pathway and stem cell signaling network. 2007. doi: 10.1158/1078-0432.CCR-06-2316. URL www.aacrjournals.org.
- W. S. Kerstjens-Frederikse, E. M. Bongers, M. T. Roofthoof, E. M. Leter, M. J. Douwes, A. V. Dijk, A. Vonk-Noordegraaf, K. K. Dijk-Bos, L. H. Hoefsloot, E. S. Hoendermis, J. J. Gille, B. Sikkema-Raddatz, R. M. Hofstra, and R. M. Berger. *Tbx4* mutations (small patella syndrome) are associated with childhood-onset pulmonary arterial hypertension. *Journal of Medical Genetics*, 50:500–506, 2013. ISSN 00222593. doi: 10.1136/jmedgenet-2012-101152.

- E. J. Kim, Y. G. Yoo, W. K. Yang, Y. S. Lim, T. Y. Na, I. K. Lee, and M. O. Lee. Transcriptional activation of hif-1 by ror and its role in hypoxia signaling. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 28:1796–1802, 10 2008. ISSN 10795642. doi: 10.1161/ATVBAHA.108.171546. URL <http://www.ncbi.nlm.nih.gov/pubmed/18658046><https://www.ahajournals.org/doi/10.1161/ATVBAHA.108.171546>.
- K. Kim, I. K. I. Kim, J. M. Yang, E. Lee, B. I. Koh, S. Song, J. Park, S. Lee, C. Choi, J. W. Kim, Y. Kubota, G. Y. Koh, and I. K. I. Kim. Sox transcription factors are positive feedback regulators of vegf signaling. *Circulation Research*, 119:839–852, 9 2016. ISSN 15244571. doi: 10.1161/CIRCRESAHA.116.308483. URL <https://www.ahajournals.org/doi/abs/10.1161/CIRCRESAHA.116.308483>.
- M. Kimura, Y. Tamura, C. Guignabert, M. Takei, K. Kosaki, N. Tanabe, K. Tatsuishi, T. Saji, T. Satoh, M. Kataoka, S. Kamitsuji, N. Kamatani, R. Thuillet, L. Tu, M. Humbert, K. Fukuda, and M. Sano. A genome-wide association analysis identifies pde1a|dnajc10 locus on chromosome 2 associated with idiopathic pulmonary arterial hypertension in a japanese population. *Oncotarget*, 8:74917–74926, 2017. ISSN 19492553. doi: 10.18632/oncotarget.20459.
- M. Knockaert, P. Lenormand, N. Gray, P. Schultz, J. Pouysségur, and L. Meijer. p42/p44 mapks are intracellular targets of the cdk inhibitor purvalanol. *Oncogene 2002 21:42*, 21:6413–6424, 9 2002. ISSN 1476-5594. doi: 10.1038/sj.onc.1205908. URL <https://www.nature.com/articles/1205908>.
- R. Koehler, E. Grünig, M. W. Pauciulo, M. M. Hooper, H. Olschewski, H. Wilkens, M. Halank, J. Winkler, R. Ewert, H. Bremer, S. Kreuscher, and B. Janssen. Low frequency of bmpr2 mutations in a german cohort of patients with sporadic idiopathic pulmonary arterial hypertension. *J Med Genet*, 41:127, 2004. doi: 10.1136/jmg.2004.023101. URL <http://www.jmedgenet.com/cgi/content/full/41/12/e127>.
- G. Kovacs, A. Berghold, S. Scheidl, and H. Olschewski. Pulmonary arterial pressure during rest and exercise in healthy subjects: A systematic review, 10 2009. ISSN 09031936.
- L. T. Krebs, Y. Xue, C. R. Norton, J. R. Shutter, M. Maguire, J. P. Sundberg, D. Gallahan, V. Closson, J. Kitajewski, R. Callahan, G. H. Smith, K. L. Stark, and T. Gridley. Notch signaling is essential for vascular morphogenesis in mice. *Genes and Development*, 14:1343–1352, 6 2000. ISSN 08909369. doi: 10.1101/gad.14.11.1343.
- G. Krenning, V. G. Barauna, J. E. Krieger, M. C. Harmsen, and J.-R. A. J. Moonen. Endothelial plasticity: Shifting phenotypes through force feedback. *Stem Cells International*, 2016, 2016. doi: 10.1155/2016/9762959.

- M. D. Laarman, G. Geeven, P. Barnett, G. J. Rinkel, W. de Laat, Y. M. Ruigrok, and J. Bakkers. Chromatin conformation links putative enhancers in intracranial aneurysm-associated regions to potential candidate genes. *Journal of the American Heart Association*, 8, 5 2019. ISSN 20479980. doi: 10.1161/JAHA.118.011201. URL <https://www.ahajournals.org/doi/abs/10.1161/JAHA.118.011201>.
- J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science (New York, N.Y.)*, 313:1929–35, 9 2006. doi: 10.1126/science.1132939. URL <http://www.ncbi.nlm.nih.gov/pubmed/17008526>.
- S. E. Lamhamedi, B. A. Menegaz, V. Ramamoorthy, R. A. Aiyer, R. L. Maywald, A. S. Buford, D. K. Doolittle, K. S. Culotta, J. E. O’Dorisio, and J. A. Ludwig. An oral formulation of yk-4-279: Preclinical efficacy and acquired resistance patterns in ewing sarcoma. *Molecular Cancer Therapeutics*, 14:1591–1604, 7 2015. ISSN 1535-7163. doi: 10.1158/1535-7163.MCT-14-0334. URL <https://mct.aacrjournals.org/content/14/7/1591><https://mct.aacrjournals.org/content/14/7/1591.abstract>.
- N. Lan, B. Massam, S. Kulkarni, and C. Lang. Pulmonary arterial hypertension: Pathophysiology and treatment. *Diseases*, 6:38, 5 2018. ISSN 2079-9721. doi: 10.3390/diseases6020038.
- E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. Fitzhugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. Levine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson,

- G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, M. L. Hong, J. Dubois, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. D. L. Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, A. Patrinos, M. J. Morgan, and I. H. G. S. Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409:860–921, 2 2001. ISSN 00280836. doi: 10.1038/35057062. URL <http://www.nature.com/articles/35057062>.
- K. B. Lane, R. D. Machado, M. W. Pauciulo, J. R. Thomson, J. A. Phillips, J. E. Loyd, W. C. Nichols, R. C. Trembath, M. Aldred, C. A. Brannon, P. M. Conneally, T. Foroud, N. Fretwell, R. Gaddipati, D. Koller, E. J. Loyd, N. Morgan, J. H. Newman, M. A. Prince, C. V. Güell, and L. Wheeler. Heterozygous germline mutations in *bmpr2*, encoding a *tgf-* receptor, cause familial primary pulmonary hypertension. *Nature Genetics*, 26:81–84, 9 2000. ISSN 10614036. doi: 10.1038/79226.
- A. W. Lange, A. R. Keiser, J. M. Wells, A. M. Zorn, and J. A. Whitsett. Sox17 promotes cell cycle progression and inhibits *tgf-b/smad3* signaling to initiate progenitor cell behavior in the respiratory epithelium. *Plos ONE*, 4, 5 2009. doi: 10.1371/journal.pone.0005711. URL www.plosone.org<http://www.ncbi.nlm.nih.gov/pubmed/19479035><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2682659>.
- A. W. Lange, H. M. Haitchi, T. D. LeCras, A. Sridharan, Y. Xu, S. E. Wert, J. James, N. Udell, P. J. Thurner, and J. A. Whitsett. Sox17 is required for normal pulmonary

- vascular morphogenesis. *Developmental Biology*, 387:109–120, 3 2014. ISSN 1095564X. doi: 10.1016/j.ydbio.2013.11.018.
- M. Laplante and D. M. Sabatini. mtor signaling in growth control and disease. *Cell*, 149:274, 4 2012. ISSN 10974172. doi: 10.1016/J.CELL.2012.03.017. URL [/pmc/articles/PMC3331679/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3331679/)[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3331679/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3331679/?report=abstract).
- M. H. Larson, L. A. Gilbert, X. Wang, W. A. Lim, J. S. Weissman, and L. S. Qi. Crispr interference (crispr) for sequence-specific control of gene expression. *Nature protocols*, 8:2180–96, 11 2013. ISSN 1750-2799. doi: 10.1038/nprot.2013.132. URL <http://www.ncbi.nlm.nih.gov/pubmed/24136345><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3922765>.
- E. M. Lau, Y. Tamura, M. D. McGoon, and O. Sitbon. The 2015 esc/ers guidelines for the diagnosis and treatment of pulmonary hypertension: A practical chronicle of progress. volume 46, pages 879–882. European Respiratory Society, 10 2015. ISBN 1522-9645. doi: 10.1183/13993003.01177-2015. URL <http://www.ncbi.nlm.nih.gov/pubmed/26318161>.
- I. P. Laumanns, L. Fink, J. Wilhelm, J. C. Wolff, R. Mitnacht-Kraus, S. Graef-Hoechst, M. M. Stein, R. M. Bohle, W. Klepetko, M. A. Hoda, R. T. Schermuly, F. Grimminger, W. Seeger, and R. Voswinckel. The noncanonical wnt pathway is operative in idiopathic pulmonary arterial hypertension. *American Journal of Respiratory Cell and Molecular Biology*, 40:683–691, 6 2009. ISSN 10441549. doi: 10.1165/rcmb.2008-0153OC.
- A. Lavelle, R. Sugrue, G. Lawler, N. Mulligan, B. Kelleher, D. M. Murphy, and S. P. Gaine. Sitaxentan-induced hepatic failure in two patients with pulmonary arterial hypertension, 9 2009. ISSN 09031936.
- N. D. Lawson, N. Scheer, V. N. Pham, C. H. Kim, A. B. Chitnis, J. A. Campos-Ortega, and B. M. Weinstein. Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development (Cambridge, England)*, 128:3675–3683, 2001. ISSN 0950-1991. doi: 10.1242/dev.039990. URL <http://dev.biologists.org/content/128/19/3675.long>[https://doi.org/10.1016/S0950-1991\(01\)00129-5](https://doi.org/10.1016/S0950-1991(01)00129-5)<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1281622/>.
- S. Lee, I. K. I. Kim, J. S. Ahn, D. C. Woo, S. T. Kim, S. Song, G. Y. Koh, H. S. Kim, B. H. Jeon, and I. K. I. Kim. Deficiency of endothelium-specific transcription factor sox17 induces intracranial aneurysm. *Circulation*, 131:995–1005, 3 2015. ISSN 15244539. doi: 10.1161/CIRCULATIONAHA.114.012568. URL <https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.114.012568><https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.114.012568>.

- S.-H. Lee, S. Lee, H. Yang, S. Song, K. Kim, T. L. Saunders, J. K. Yoon, G. Y. Koh, and I. Kim. Notch pathway targets proangiogenic regulator sox17 to restrict angiogenesis. *Circulation Research*, 115:215–226, 7 2014. ISSN 0009-7330. doi: 10.1161/CIRCRESAHA.115.303142. URL <https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.115.303142>.
- K. O. Leslie and M. R. Wick. *Practical pulmonary pathology : a diagnostic approach*. Elsevier, 3 edition, 2005. ISBN 9780323442886. URL <https://books.google.be/books?hl=en&lr=&id=70k7DwAAQBAJ&oi=fnd&pg=PP1&dq=Practical+pulmonary+pathology++a+diagnostic+approach+leslie+k&ots=CL5MXk309U&sig=fgIyI32Xx6iQoJZ9Qurkh1iRF44#v=onepage&q=Practical%20pulmonary%20pathology%20%3A%20a%20diagnostic%20approach%20leslie%20k&f=false>.
- J. Li, S. G. Kim, and J. Blenis. Rapamycin: one drug, many effects. *Cell metabolism*, 19:373, 3 2014. ISSN 15504131. doi: 10.1016/J.CMET.2014.01.001. URL [/pmc/articles/PMC3972801/](https://pmc/articles/PMC3972801/)<https://pmc/articles/PMC3972801/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3972801/>.
- W. Li, M. Petrimpol, K. D. Molle, M. N. Hall, E. J. Battegay, and R. Humar. Hypoxia-induced endothelial proliferation requires both mtorc1 and mtorc2. *Circulation Research*, 100:79–87, 1 2007. ISSN 00097330. doi: 10.1161/01.RES.0000253094.03023.3f. URL <https://www.ahajournals.org/doi/abs/10.1161/01.res.0000253094.03023.3f>.
- Y. M. Liao, Y. Song, Y. K. Li, J. H. Du, and Y. Zhou. Sox17, -catenin and cyclind1 expression in the endometrioid adenocarcinoma and influence of 5-aza on expression. *Cancer Gene Therapy 2019 27:3*, 27:256–263, 9 2019. ISSN 1476-5500. doi: 10.1038/s41417-019-0135-5. URL <https://www.nature.com/articles/s41417-019-0135-5>.
- N. Lim and P. Pavlidis. Evaluation of connectivity map shows limited reproducibility in drug repositioning. *bioRxiv*, page 845693, 11 2019. doi: 10.1101/845693. URL <https://www.biorxiv.org/content/10.1101/845693v1>.
- Y. Ling, M. K. Johnson, D. G. Kiely, R. Condliffe, C. A. Elliot, J. S. R. Gibbs, L. S. Howard, J. Pepke-Zaba, K. K. Sheares, P. A. Corris, A. J. Fisher, J. L. Lordan, S. Gaine, J. G. Coghlan, S. J. Wort, M. A. Gatzoulis, and A. J. Peacock. 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*, 186:790–796, 10 2012. ISSN 1073449X. doi: 10.1164/rccm.201203-0383OC.
- G. Y. Liu and D. M. Sabatini. mtor at the nexus of nutrition, growth, ageing and disease. *Nature Reviews Molecular Cell Biology 2020 21:4*, 21:183–203, 1 2020. ISSN 1471-

0080. doi: 10.1038/s41580-019-0199-y. URL <https://www.nature.com/articles/s41580-019-0199-y>.
- M. Liu, L. Zhang, G. Marsboom, A. Jambusaria, S. Xiong, P. T. Toth, E. V. Benevolenskaya, J. Rehman, and A. B. Malik. Sox17 is required for endothelial regeneration following inflammation-induced vascular injury. *Nature Communications*, 10: 2126, 5 2019. ISSN 20411723. doi: 10.1038/s41467-019-10134-y. URL <http://www.nature.com/articles/s41467-019-10134-y>.
- L. Long, M. L. Ormiston, X. Yang, M. Southwood, S. Gräf, R. D. Machado, M. Mueller, B. Kinzel, L. M. Yung, J. M. Wilkinson, S. D. Moore, K. M. Drake, M. A. Aldred, P. B. Yu, P. D. Upton, and N. W. Morrell. Selective enhancement of endothelial bmpr-ii with bmp9 reverses pulmonary arterial hypertension. *Nature Medicine*, 21:777–785, 7 2015. ISSN 1546170X. doi: 10.1038/nm.3877.
- A. P. Looney, R. Han, L. Stawski, G. Marden, M. Iwamoto, and M. Trojanowska. Synergistic role of endothelial erg and flil in mediating pulmonary vascular homeostasis. *American Journal of Respiratory Cell and Molecular Biology*, 57:121–131, 7 2017. ISSN 15354989. doi: 10.1165/RCMB.2016-0200OC/SUPPL_FILE/DISCLOSURES.PDF. URL [/pmc/articles/PMC5516275/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5516275/)[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5516275/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5516275/?report=abstract).
- L. Ma, D. Roman-Campos, E. D. Austin, M. Eyries, K. S. Sampson, F. Soubrier, M. Germain, D. A. Tregouët, A. Borczuk, E. B. Rosenzweig, B. Girerd, D. Montani, M. Humbert, J. E. Loyd, R. S. Kass, and W. K. Chung. A novel channelopathy in pulmonary arterial hypertension. *New England Journal of Medicine*, 369:351–361, 2013. ISSN 15334406. doi: 10.1056/NEJMoa1211097.
- K. Marinou, C. Christodoulides, C. Antoniadou, and M. Koutsilieris. Wnt signaling in cardiovascular physiology, 12 2012. ISSN 10432760.
- B. A. Maron, D. L. Bhatt, M. Nykiel, S. Kinlay, and A. B. Waxman. Protocol for vasoreactivity testing with epoprostenol in pulmonary hypertension. *Critical Pathways in Cardiology: A Journal of Evidence-Based Medicine*, 11:40–42, 3 2012. ISSN 1535-282X. doi: 10.1097/HPC.0b013e3182480725. URL <http://journals.lww.com/00132577-201203000-00007>.
- F. A. Masri, W. Xu, S. A. Comhair, K. Asosingh, M. Koo, A. Vasanji, J. Drazba, B. Anand-Apte, and S. C. Erzurum. Hyperproliferative apoptosis-resistant endothelial cells in idiopathic pulmonary arterial hypertension. *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 293:548–554, 9 2007. ISSN 10400605. doi: 10.1152/AJPLUNG.00428.2006/ASSET/IMAGES/LARGE/

- ZH50090749450004.JPEG. URL <https://journals.physiology.org/doi/abs/10.1152/ajplung.00428.2006>.
- Y. Mohamedi, T. Fontanil, S. Cal, T. Cobo, and Álvaro J. Obaya. Adamts-12: Functions and challenges for a complex metalloprotease. *Frontiers in Molecular Biosciences*, 8: 378, 4 2021. ISSN 2296889X. doi: 10.3389/FMOLB.2021.686763/BIBTEX.
- H. Morisaki, N. Nakanishi, S. Kyotani, A. Takashima, H. Tomoike, and T. Morisaki. Bmpr2 mutations found in japanese patients with familial and sporadic primary pulmonary hypertension. *Human Mutation*, 23:632–632, 6 2004. ISSN 1059-7794. doi: 10.1002/humu.9251. URL <http://doi.wiley.com/10.1002/humu.9251>.
- F. Murray, H. H. Patel, R. Y. S. Suda, S. Zhang, P. A. Thistlethwaite, J. X.-J. Yuan, and P. A. Insel. Expression and activity of camp phosphodiesterase isoforms in pulmonary artery smooth muscle cells from patients with pulmonary hypertension: role for pde1. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 292:L294–L303, 1 2007. ISSN 1040-0605. doi: 10.1152/ajplung.00190.2006. URL <https://www.physiology.org/doi/10.1152/ajplung.00190.2006>.
- A. Musa, L. S. Ghorai, S.-D. Zhang, G. Glazko, O. Yli-Harja, M. Dehmer, B. Haibe-Kains, and F. Emmert-Streib. A review of connectivity map and computational approaches in pharmacogenomics. *Briefings in Bioinformatics*, 19:506–523, 2018. doi: 10.1093/BIB. URL <http://m.bib.oxfordjournals.org/content/early/2017/01/14/bib.bbw134.long>.
- M. T. Nasim, T. Ogo, M. Ahmed, R. Randall, H. M. Chowdhury, K. M. Snape, T. Y. Bradshaw, L. Southgate, G. J. Lee, I. Jackson, G. M. Lord, J. S. R. Gibbs, M. R. Wilkins, K. Ohta-Ogo, K. Nakamura, B. Girerd, F. Coulet, F. Soubrier, M. Humbert, N. W. Morrell, R. C. Trembath, and R. D. Machado. Molecular genetic characterization of smad signaling molecules in pulmonary arterial hypertension. *Human Mutation*, 32: 1385–1389, 12 2011. ISSN 10597794. doi: 10.1002/humu.21605. URL <http://doi.wiley.com/10.1002/humu.21605>.
- K. K. Niakan, H. Ji, R. Maehr, S. A. Vokes, K. T. Rodolfa, R. I. Sherwood, M. Yamaki, J. T. Dimos, A. E. Chen, D. A. Melton, A. P. McMahon, and K. Eggan. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Development*, 24:312, 2 2010. ISSN 08909369. doi: 10.1101/GAD.1833510. URL [/pmc/articles/PMC2811832/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2811832/)[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2811832/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2811832/?report=abstract).
- A. Nishihara, T. Watabe, T. Imamura, and K. Miyazono. Functional heterogeneity of bone morphogenetic protein receptor-ii mutants found in patients with primary pulmonary

- hypertension. *Molecular Biology of the Cell*, 13:3055–3063, 9 2002. ISSN 10591524. doi: 10.1091/mbc.E02-02-0063.
- T. Nishimura, J. L. Faul, G. J. Berry, I. Veve, R. G. Pearl, and P. N. Kao. 40-o-(2-hydroxyethyl)-rapamycin attenuates pulmonary arterial hypertension and neointimal formation in rats. *American Journal of Respiratory and Critical Care Medicine*, 163: 498–502, 2001. ISSN 1073449X. doi: 10.1164/ajrccm.163.2.2006093.
- R. Nogueira-Ferreira, R. Ferreira, and T. Henriques-Coelho. Cellular interplay in pulmonary arterial hypertension: Implications for new therapies, 5 2014. ISSN 01674889.
- T. Néstor, H. Masckauchán, C. J. Shawber, Y. Funahashi, C.-M. Li, and J. Kitajewski. Wnt/b-catenin signaling induces proliferation, survival and interleukin-8 in human endothelial cells. *Angiogenesis*, 8:43–51, 2005. doi: 10.1007/s10456-005-5612-9.
- S. K. Oh, D. Kim, K. Kim, K. Boo, Y. S. Yu, I. S. Kim, Y. Jeon, S. K. Im, S. H. Lee, J. M. Lee, Y. Ko, H. Lee, D. Park, S. Fang, and S. H. Baek. Ror is crucial for attenuated inflammatory response to maintain intestinal homeostasis. *Proceedings of the National Academy of Sciences of the United States of America*, 116:21140–21149, 10 2019. ISSN 10916490. doi: 10.1073/PNAS.1907595116/SUPPL_FILE/PNAS.1907595116.SAPP.PDF. URL www.pnas.org/cgi/doi/10.1073/pnas.1907595116.
- E. C. Oliveira, A. L. P. Ribeiro, and C. F. S. Amaral. Adenosine for vasoreactivity testing in pulmonary hypertension: A head-to-head comparison with inhaled nitric oxide. *Respiratory Medicine*, 104:606–611, 4 2010. ISSN 09546111. doi: 10.1016/j.rmed.2009.11.010. URL <https://pubmed.ncbi.nlm.nih.gov/19962292/>.
- H. Olschewski, G. Simonneau, N. Galiè, T. Higenbottam, R. Naeije, L. J. Rubin, S. Nikkho, R. Speich, M. M. Hoeper, J. Behr, J. Winkler, O. Sitbon, W. Popov, H. A. Ghofrani, A. Manes, D. G. Kiely, R. Ewert, A. Meyer, P. A. Corris, M. Delcroix, M. Gomez-Sanchez, H. Siedentop, and W. Seeger. Inhaled iloprost for severe pulmonary hypertension. *New England Journal of Medicine*, 347:322–329, 8 2002. ISSN 00284793. doi: 10.1056/NEJMoa020204.
- M. L. Ormiston, R. S. Godoy, K. R. Chaudhary, and D. J. Stewart. The janus faces of bone morphogenetic protein 9 in pulmonary arterial hypertension. *Circulation research*, 124:822–824, 3 2019. ISSN 1524-4571. doi: 10.1161/CIRCRESAHA.119.314753. URL <http://www.ncbi.nlm.nih.gov/pubmed/30870124>.
- R. Paddenberger, P. Stieger, A. L. von Lilien, P. Faulhammer, A. Goldenberg, H. H. Tillmanns, W. Kummer, and R. C. Braun-Dullaeus. Rapamycin attenuates hypoxia-induced pulmonary vascular remodeling and right ventricular hypertrophy in mice. *Respiratory Research*, 8:15, 2 2007. ISSN 14659921. doi: 10.

- 1186/1465-9921-8-15. URL <http://respiratory-research.biomedcentral.com/articles/10.1186/1465-9921-8-15>.
- P. Palasingam, R. Jauch, C. K. L. Ng, and P. R. Kolatkar. The structure of sox17 bound to dna reveals a conserved bending topology but selective protein interaction platforms. *Journal of Molecular Biology*, 388:619–630, 5 2009. ISSN 0022-2836. doi: 10.1016/J.JMB.2009.03.055. URL <https://www.sciencedirect.com/science/article/pii/S0022283609003635?via%3Dihub>.
- A. I. Papaioannou, E. Zakyntinos, K. Kostikas, T. Kiropoulos, A. Koutsokera, A. Ziogas, A. Koutroumpas, L. Sakkas, K. I. Gourgoulisanis, and Z. D. Daniil. Serum vegf levels are related to the presence of pulmonary arterial hypertension in systemic sclerosis. *BMC Pulmonary Medicine*, 9:1–8, 5 2009. ISSN 14712466. doi: 10.1186/1471-2466-9-18.
- A. Parpaleix, V. Amsellem, A. Houssaini, S. Abid, M. Breau, E. Marcos, D. Sawaki, M. Delcroix, R. Quarck, A. Maillard, I. Couillin, B. Ryffel, and S. Adnot. Role of interleukin-1 receptor 1/myd88 signalling in the development and progression of pulmonary hypertension. *European Respiratory Journal*, 48:470–483, 8 2016. ISSN 13993003. doi: 10.1183/13993003.01448-2015. URL <http://ow.ly/1Fpe3008RLs>.
- R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14:417–419, 2017. ISSN 15487105. doi: 10.1038/nmeth.4197.
- A. J. Peacock, N. F. Murphy, J. J. V. McMurray, L. Caballero, and S. Stewart. An epidemiological study of pulmonary arterial hypertension. *The European respiratory journal*, 30:104–9, 7 2007. ISSN 0903-1936. doi: 10.1183/09031936.00092306. URL <http://www.ncbi.nlm.nih.gov/pubmed/17360728>.
- V. A. D. J. Perez, T. P. Alastalo, J. C. Wu, J. D. Axelrod, J. P. Cooke, M. Amieva, and M. Rabinovitch. Bone morphogenetic protein 2 induces pulmonary angiogenesis via wnt --catenin and wnt - rhoa - rac1 pathways. *Journal of Cell Biology*, 184:83–99, 1 2009. ISSN 00219525. doi: 10.1083/jcb.200806049.
- V. D. J. Perez, K. Yuan, T. P. Alastalo, E. Spiekerkoetter, and M. Rabinovitch. Targeting the wnt signaling pathways in pulmonary arterial hypertension, 2014. ISSN 18785832.
- S. Pott and J. D. Lieb. What are super-enhancers? *Nature Genetics* 2015 47:1, 47:8–12, 12 2014. ISSN 1546-1718. doi: 10.1038/ng.3167. URL <https://www.nature.com/articles/ng.3167>.
- G. Pousada, V. Lupo, S. Cástro-Sánchez, M. Álvarez Satta, A. Sánchez-Monteagudo, A. Balóira, C. Espinós, and D. Valverde. Molecular and functional characterization of

- the bmpr2 gene in pulmonary arterial hypertension. *Scientific Reports*, 7:1–19, 12 2017. ISSN 20452322. doi: 10.1038/s41598-017-02074-8.
- K. W. Prins, T. Thenappan, E. K. Weir, R. Kalra, M. Pritzker, and S. L. Archer. Repurposing medications for treatment of pulmonary arterial hypertension: What’s old is new again. *Journal of the American Heart Association*, 8:e011343, 1 2019. ISSN 2047-9980. doi: 10.1161/JAHA.118.011343. URL <http://www.ncbi.nlm.nih.gov/pubmed/30590974><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6405714>.
- P. Puigdevall, L. Piccari, I. Blanco, J. A. Barberà, D. Geiger, C. Badenas, M. Milà, R. Castelo, and I. Madrigal. Genetic linkage analysis of a large family identifies fign as a candidate modulator of reduced penetrance in heritable pulmonary arterial hypertension. *Journal of Medical Genetics*, 56:481–490, 7 2019. ISSN 14686244. doi: 10.1136/jmedgenet-2018-105669.
- L. Qiao, T. Nishimura, L. Shi, D. Sessions, A. Thrasher, J. R. Trudell, G. J. Berry, R. G. Pearl, and P. N. Kao. Endothelial fate mapping in mice with pulmonary hypertension. *Circulation*, 129:692–703, 2 2014. ISSN 00097322. doi: 10.1161/CIRCULATIONAHA.113.003734.
- X. A. Qu and D. K. Rajpal. Applications of connectivity map in drug discovery and development, 12 2012. ISSN 13596446.
- B. Ranchoux, F. Antigny, C. Rucker-Martin, A. Hautefort, C. Péchoux, H. J. Bogaard, P. Dorfmueller, S. Remy, F. Lecerf, S. Planté, S. Chat, E. Fadel, A. Houssaini, I. Anegon, S. Adnot, G. Simonneau, M. Humbert, S. Cohen-Kaminsky, and F. Perros. Endothelial-to-mesenchymal transition in pulmonary hypertension. *Circulation*, 131:1006–18, 3 2015. ISSN 1524-4539. doi: 10.1161/CIRCULATIONAHA.114.008750. URL <http://www.ncbi.nlm.nih.gov/pubmed/25593290>.
- G. Rawadi, B. Vayssière, F. Dunn, R. Baron, and S. Roman-Roman. Bmp-2 controls alkaline phosphatase expression and osteoblast mineralization by a wnt autocrine loop. *Journal of Bone and Mineral Research*, 18:1842–1853, 10 2003. ISSN 08840431. doi: 10.1359/jbmr.2003.18.10.1842.
- C. V. Remillard, D. D. Tigno, O. Platoshyn, E. D. Burg, E. E. Brevnova, D. Conger, A. Nicholson, B. K. Rana, R. N. Channick, L. J. Rubin, D. T. O’Connor, and J. X. Yuan. Function of kv1.5 channels and genetic variations of kcna5 in patients with idiopathic pulmonary arterial hypertension. *American Journal of Physiology - Cell Physiology*, 292, 5 2007. ISSN 03636143. doi: 10.1152/ajpcell.00405.2006.

- A. M. Reynolds, M. D. Holmes, S. M. Danilov, and P. N. Reynolds. Targeted gene delivery of bmpr2 attenuates pulmonary hypertension. *European Respiratory Journal*, 39:329–343, 2 2012. ISSN 09031936. doi: 10.1183/09031936.00187310.
- S. B. Rho, B. R. Kim, and S. Kang. A gene signature-based approach identifies thioridazine as an inhibitor of phosphatidylinositol-3-kinase (pi3k)/akt pathway in ovarian cancer cells. *Gynecologic Oncology*, 120:121–127, 1 2011. ISSN 00908258. doi: 10.1016/j.ygyno.2010.10.003.
- C. J. Rhodes, K. Batai, M. Bleda, M. Haimel, L. Southgate, M. Germain, M. W. Pauculo, C. Hadinnapola, J. Aman, B. Girerd, A. Arora, J. Knight, K. B. Hanscombe, J. H. Karnes, M. Kaakinen, H. Gall, A. Ulrich, L. Harbaum, I. Cebola, J. Ferrer, K. Lutz, E. M. Swietlik, F. Ahmad, P. Amouyel, S. L. Archer, R. Argula, E. D. Austin, D. Badesch, S. Bakshi, C. Barnett, R. Benza, N. Bhatt, H. J. Bogaard, C. D. Burger, M. Chakinala, C. Church, J. G. Coghlan, R. Condliffe, P. A. Corris, C. Danesino, S. Dettette, C. G. Elliott, J. Elwing, M. Eyries, T. Fortin, A. Franke, R. P. Frantz, A. Frost, J. G. N. Garcia, S. Ghio, H.-A. Ghofrani, J. S. R. Gibbs, J. Harley, H. He, N. S. Hill, R. Hirsch, A. C. Houweling, L. S. Howard, D. Ivy, D. G. Kiely, J. Klinger, G. Kovacs, T. Lahm, M. Laudes, R. D. Machado, R. V. M. Ross, K. Marsolo, L. J. Martin, S. Moledina, D. Montani, S. D. Nathan, M. Newnham, A. Olschewski, H. Olschewski, R. J. Oudiz, W. H. Ouwehand, A. J. Peacock, J. Pepke-Zaba, Z. Rehman, I. Robbins, D. M. Roden, E. B. Rosenzweig, G. Saydain, L. Scelsi, R. Schilz, W. Seeger, C. M. Shaffer, R. W. Simms, M. Simon, O. Sitbon, J. Suntharalingam, H. Tang, A. Y. Tcheurbanov, T. Thenappan, F. Torres, M. R. Toshner, C. M. Treacy, A. V. Noordegraaf, Q. Waisfisz, A. K. Walsworth, R. E. Walter, J. Wharton, R. J. White, J. Wilt, S. J. Wort, D. Yung, A. Lawrie, M. Humbert, F. Soubrier, D.-A. Tréguët, I. Prokopenko, R. Kittles, S. Gräf, W. C. Nichols, R. C. Trembath, A. A. Desai, N. W. Morrell, M. R. Wilkins, U. N. B. R. D. Consortium, U. P. C. S. Consortium, and U. P. B. Consortium. Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. *The Lancet. Respiratory medicine*, 7:227–238, 3 2019. ISSN 2213-2619. doi: 10.1016/S2213-2600(18)30409-0. URL <http://www.ncbi.nlm.nih.gov/pubmed/30527956><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6391516>.
- M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26: 139–140, 1 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp616.
- L. J. Rubin, P. Nicod, L. D. Hillis, and B. G. Firth. Treatment of primary pulmonary hypertension with nifedipine. a hemodynamic and scintigraphic evaluation. *Annals of Internal Medicine*, 99:433–438, 1983. ISSN 00034819. doi: 10.7326/0003-4819-99-4-433.

- L. J. Rubin, D. B. Badesch, R. J. Barst, N. Galiè, C. M. Black, A. Keogh, T. Pulido, A. Frost, S. Roux, I. Leconte, M. Landzberg, and G. Simonneau. Bosentan therapy for pulmonary arterial hypertension. *New England Journal of Medicine*, 346:896–903, 3 2002. ISSN 00284793. doi: 10.1056/NEJMoa012212.
- K. Saito, I. Nobuhisa, K. Harada, S. Takahashi, M. Anani, H. Lickert, M. Kanai-Azuma, Y. Kanai, and T. Taga. Maintenance of hematopoietic stem and progenitor cells in fetal intra-aortic hematopoietic clusters by the sox17-notch1-hes1 axis. *Experimental Cell Research*, 365:145–155, 4 2018. ISSN 10902422. doi: 10.1016/j.yexcr.2018.02.014. URL <https://www.sciencedirect.com/science/article/pii/S0014482718300880?via%3Dihub>.
- W. Schachterle, C. R. Badwe, B. Palikuqi, B. Kunar, M. Ginsberg, R. Lis, M. Yokoyama, O. Elemento, J. M. Scandura, and S. Rafii. Sox17 drives functional engraftment of endothelium converted from non-vascular cells. *Nature Communications* 2017 8:1, 8:1–12, 1 2017. ISSN 2041-1723. doi: 10.1038/ncomms13963. URL <https://www.nature.com/articles/ncomms13963>.
- M. Scharpfenecker, M. van Dinther, Z. Liu, R. L. van Bezooijen, Q. Zhao, L. Pukac, C. W. Löwik, and P. ten Dijke. Bmp-9 signals via alk1 and inhibits bfgf-induced endothelial cell proliferation and vegf-stimulated angiogenesis. *Journal of Cell Science*, 120:964–972, 3 2007. ISSN 00219533. doi: 10.1242/jcs.002949.
- R. T. Schermuly, H. A. Ghofrani, M. R. Wilkins, and F. Grimminger. No title, 8 2011. URL <http://www.nature.com/articles/nrcardio.2011.87>.
- N. Selimovic, C. H. Bergh, B. Andersson, E. Sakiniene, H. Carlsten, and B. Rundqvist. Growth factors and interleukin-6 across the lung circulation in pulmonary hypertension. *European Respiratory Journal*, 34:662–668, 9 2009. ISSN 09031936. doi: 10.1183/09031936.00174908.
- A. Q. Sheikh, A. Misra, I. O. Rosas, R. H. Adams, and D. M. Greif. Smooth muscle cell progenitors are primed to muscularize in pulmonary hypertension. *Science Translational Medicine*, 7:308ra159, 10 2015. ISSN 19466242. doi: 10.1126/scitranslmed.aaa9712.
- Y. Shi and J. Massagué. Mechanisms of tgf- signaling from cell membrane to the nucleus, 6 2003. ISSN 00928674.
- M. Shintani, H. Yagi, T. Nakayama, T. Saji, and R. Matsuoka. A new nonsense mutation of smad8 associated with pulmonary arterial hypertension. *Journal of Medical Genetics*, 46:331–337, 5 2009. ISSN 00222593. doi: 10.1136/jmg.2008.062703.
- J. M. Shulman, N. Perrimon, and J. D. Axelrod. Frizzled signaling and the developmental control of cell polarity, 11 1998. ISSN 01689525.

- M. Simon, M. Gombert-Maitland, R. Oudiz, R. Machado, F. Rischard, J. Elinoff, B. Grigorian, A. Schmid, S. Hou, N. Desai, and M. Gladwin. Patients with severe pulmonary arterial hypertension treated with abi-009, nab-sirolimus, an mtor inhibitor: Interim results from a phase 1 clinical trial. *Journal of the American College of Cardiology*, 73:1927, 3 2019. ISSN 07351097. doi: 10.1016/s0735-1097(19)32533-1.
- M. Simons, E. Gordon, and L. Claesson-Welsh. Mechanisms and regulation of endothelial vegf receptor signalling, 10 2016. ISSN 14710080.
- D. Sinner, S. Rankin, M. Lee, and A. M. Zorn. Sox17 and -catenin cooperate to regulate the transcription of endodermal genes. *Development*, 131:3069–3080, 7 2004. ISSN 09501991. doi: 10.1242/dev.01176.
- D. Sinner, J. J. Kordich, J. R. Spence, R. Opoka, S. Rankin, S.-C. J. Lin, D. Jonatan, A. M. Zorn, and J. M. Wells. Sox17 and sox4 differentially regulate beta-catenin/t-cell factor activity and proliferation of colon carcinoma cells. *Molecular and cellular biology*, 27:7802–15, 11 2007. ISSN 1098-5549. doi: 10.1128/MCB.02179-06. URL <http://www.ncbi.nlm.nih.gov/pubmed/17875931><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2169141>.
- O. Sitbon, M. Humbert, J. Jagot, O. Taravella, M. Fartoukh, F. Parent, P. Herve, and G. Simonneau. Inhaled nitric oxide as a screening agent for safely identifying responders to oral calcium-channel blockers in primary pulmonary hypertension. *European Respiratory Journal*, 12, 1998.
- O. Sitbon, M. Humbert, X. Jaïs, V. Ioos, A. M. Hamid, S. Provencher, G. Garcia, F. Parent, P. Hervé, and G. Simonneau. Long-term response to calcium channel blockers in idiopathic pulmonary arterial hypertension. *Circulation*, 111:3105–3111, 6 2005. ISSN 00097322. doi: 10.1161/CIRCULATIONAHA.104.488486. URL <https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.104.488486>.
- O. Sitbon, M. Delcroix, E. Bergot, A. B. Boonstra, J. Granton, D. Langleben, P. E. Subías, N. Galiè, T. Pfister, J. C. Lemarié, and G. Simonneau. Epite-2: An open-label study assessing the transition to a new formulation of intravenous epoprostenol in patients with pulmonary arterial hypertension. *American Heart Journal*, 167:210–217, 2 2014. ISSN 00028703. doi: 10.1016/j.ahj.2013.08.007.
- O. Sitbon, R. Channick, K. M. Chin, A. Frey, S. Gaine, N. Galiè, H.-A. Ghofrani, M. M. Hoeper, I. M. Lang, R. Preiss, L. J. Rubin, L. D. Scala, V. Tapson, I. Adzerikho, J. Liu, O. Moiseeva, X. Zeng, G. Simonneau, and V. V. McLaughlin. Selexipag for the treatment of pulmonary arterial hypertension. *New England Journal of Medicine*, 373:2522–2533, 12 2015. ISSN 0028-4793. doi: 10.1056/NEJMoa1503184. URL <http://www.nejm.org/doi/10.1056/NEJMoa1503184>.

- O. Sitbon, M. Gomberg-Maitland, J. Granton, M. I. Lewis, S. C. Mathai, M. Rainisio, N. L. Stockbridge, M. R. Wilkins, R. T. Zamanian, and L. J. Rubin. Clinical trial design and new therapies for pulmonary arterial hypertension. *The European respiratory journal*, 53, 1 2019. ISSN 13993003. doi: 10.1183/13993003.01908-2018.
- G. M. Sizemore, J. R. Pitarresi, S. Balakrishnan, and M. C. Ostrowski. The ets family of oncogenic transcription factors in solid tumours. *Nature Reviews Cancer* 2017 17:6, 17:337–351, 4 2017. ISSN 1474-1768. doi: 10.1038/nrc.2017.20. URL <https://www.nature.com/articles/nrc.2017.20>.
- N. Sommer, H. A. Ghofrani, O. Pak, S. Bonnet, S. Provencher, O. Sitbon, S. Rosenkranz, M. M. Hoeper, D. G. Kiely, A. Ghofrani, O. Pak, S. Bonnet, S. Provencher, O. Sitbon, S. Rosenkranz, M. M. Hoeper, and D. G. Kiely. Current and future treatments of pulmonary arterial hypertension, 2 2020. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/bph.15016>.
- C. Soneson, M. I. Love, and M. D. Robinson. Differential analyses for rna-seq: Transcript-level estimates improve gene-level inferences [version 2; referees: 2 approved]. *F1000Research*, 4, 2016. ISSN 1759796X. doi: 10.12688/F1000RESEARCH.7563.2.
- H. W. Song, K. L. Foreman, B. D. Gastfriend, J. S. Kuo, S. P. Palecek, and E. V. Shusta. Transcriptomic comparison of human and mouse brain microvessels. *Scientific Reports* 2020 10:1, 10:1–14, 7 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-69096-7. URL <https://www.nature.com/articles/s41598-020-69096-7>.
- F. Soubrier, W. K. Chung, R. Machado, E. Grünig, M. Aldred, M. Geraci, J. E. Loyd, C. G. Elliott, R. C. Trembath, J. H. Newman, and M. Humbert. Genetics and genomics of pulmonary arterial hypertension. volume 62, pages D13–D21. Elsevier, 12 2013. ISBN 1558-3597 (Electronic)0735-1097 (Linking). doi: 10.1016/j.jacc.2013.10.035. URL <https://www.sciencedirect.com/science/article/pii/S0735109713058786?via%3Dihub>.
- L. Southgate, R. D. Machado, S. Gräf, and N. W. Morrell. Molecular genetic framework underlying pulmonary arterial hypertension, 2 2020. ISSN 17595010.
- E. Spiekerkoetter, X. Tian, J. Cai, R. K. Hopper, D. Sudheendra, C. G. Li, N. El-Bizri, H. Sawada, R. Haghighat, R. Chan, L. Haghighat, V. D. J. Perez, L. Wang, S. Reddy, M. Zhao, D. Bernstein, D. E. Solow-Cordero, P. A. Beachy, T. J. Wandless, P. T. Dijke, and M. Rabinovitch. Fk506 activates bmpr2, rescues endothelial dysfunction, and reverses pulmonary hypertension. *The Journal of Clinical Investigation*, 123:3600, 8 2013. ISSN 00219738. doi: 10.1172/JCI65592. URL [/pmc/articles/PMC3726153/](https://pmc/articles/PMC3726153/)[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3726153/](https://pmc/articles/PMC3726153/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3726153/).

- E. Stacher, B. B. Graham, J. M. Hunt, A. Gandjeva, S. D. Groshong, V. V. McLaughlin, M. Jessup, W. E. Grizzle, M. A. Aldred, C. D. Cool, and R. M. Tudor. Modern age pathology of pulmonary arterial hypertension. *American Journal of Respiratory and Critical Care Medicine*, 186:261–272, 8 2012. ISSN 1073449X. doi: 10.1164/rccm.201201-0164OC.
- J. P. Stasch, E. M. Becker, C. Alonso-Alija, H. Apeler, K. Dembowski, A. Feurer, R. Gerzer, T. Minuth, E. Perzborn, U. Pleiß, H. Schröder, W. Schroeder, E. Stahl, W. Steinke, A. Straub, and M. Schramm. No-independent regulatory site on soluble guanylate cyclase. *Nature*, 410:212–215, 3 2001. ISSN 00280836. doi: 10.1038/35065611.
- J. P. Stasch, P. Pacher, and O. V. Evgenov. Soluble guanylate cyclase as an emerging therapeutic target in cardiopulmonary disease. *Circulation*, 123:2263–2273, 5 2011. ISSN 00097322. doi: 10.1161/CIRCULATIONAHA.110.981738.
- R. S. Stearman, Q. M. Bui, G. Speyer, A. Handen, A. R. Cornelius, B. B. Graham, S. Kim, E. A. Mickler, R. M. Tudor, S. Y. Chan, and M. W. Geraci. Systems analysis of the human pulmonary arterial hypertension lung transcriptome. *American Journal of Respiratory Cell and Molecular Biology*, 60:637–649, 6 2019. ISSN 15354989. doi: 10.1165/RCMB.2018-0368OC/SUPPL_FILE/DISCLOSURES.PDF. URL [/pmc/articles/PMC6543748/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6543748/)[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6543748/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6543748/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6543748/).
- L. C. Steffes, A. A. Froistad, A. Andruska, M. Boehm, M. McGlynn, F. Zhang, W. Zhang, D. Hou, X. Tian, L. Miquerol, K. Nadeau, R. J. Metzger, E. Spiekerkoetter, and M. E. Kumar. A notch3-marked subpopulation of vascular smooth muscle cells is the cell of origin for occlusive pulmonary vascular lesions. *Circulation*, 142:1545–1561, 10 2020. ISSN 15244539. doi: 10.1161/CIRCULATIONAHA.120.045750. URL <http://ahajournals.org>.
- K. R. Stenmark, N. Davie, M. Frid, E. Gerasimovskaya, and M. Das. Role of the adventitia in pulmonary vascular remodeling, 4 2006. ISSN 15489213. URL www.physiologyonline.org.
- A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, J. Gould, J. F. Davis, A. A. Tubelli, J. K. Asiedu, D. L. Lahr, J. E. Hirschman, Z. Liu, M. Donahue, B. Julian, M. Khan, D. Wadden, I. C. Smith, D. Lam, A. Liberzon, C. Toder, M. Bagul, M. Orzechowski, O. M. Enache, F. Piccioni, S. A. Johnson, N. J. Lyons, A. H. Berger, A. F. Shamji, A. N. Brooks, A. Vrcic, C. Flynn, J. Rosains, D. Y. Takeda, R. Hu, D. Davison, J. Lamb, K. Ardlie, L. Hogstrom, P. Greenside, N. S. Gray, P. A. Clemons, S. Silver, X. Wu, W. N. Zhao, W. Read-Button, X. Wu, S. J. Haggarty, L. V. Ronco, J. S. Boehm, S. L. Schreiber, J. G. Doench, J. A. Bittker, D. E.

- Root, B. Wong, and T. R. Golub. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, 171:1437–1452.e17, 11 2017. ISSN 10974172. doi: 10.1016/j.cell.2017.10.049.
- S. Suchting, C. Freitas, F. L. Noble, R. Benedito, C. Bréant, A. Duarte, and A. Eichmann. The notch ligand delta-like 4 negatively regulates endothelial tip cell formation and vessel branching. *Proceedings of the National Academy of Sciences of the United States of America*, 104:3225–3230, 2 2007. ISSN 00278424. doi: 10.1073/pnas.0611177104.
- Y. Sun, C. H. Liu, J. P. SanGiovanni, L. P. Evans, K. T. Tian, B. Zhang, A. Stahl, W. T. Pu, T. M. Kamenecka, L. A. Solt, and J. Chen. Nuclear receptor ror regulates pathologic retinal angiogenesis by modulating socs3-dependent inflammation. *Proceedings of the National Academy of Sciences of the United States of America*, 112:10401–10406, 8 2015. ISSN 10916490. doi: 10.1073/PNAS.1504387112/SUPPL_FILE/PNAS.1504387112.SAPP.PDF.
- T. Suzuki, E. J. Carrier, M. H. Talati, A. Rathinasabapathy, X. Chen, R. Nishimura, Y. Tada, K. Tatsumi, and J. West. Isolation and characterization of endothelial-to-mesenchymal transition cells in pulmonary arterial hypertension. *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 314:L118–L126, 1 2018. ISSN 15221504. doi: 10.1152/ajplung.00296.2017.
- M. R. Swift, V. N. Pham, D. Castranova, K. Bell, R. J. Poole, and B. M. Weinstein. Sox factors and notch regulate nr2f2 gene expression during venous differentiation in zebrafish. *Developmental biology*, 390:116–25, 6 2014. ISSN 1095-564X. doi: 10.1016/j.ydbio.2014.03.018. URL <http://www.ncbi.nlm.nih.gov/pubmed/24699544><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4104406>.
- Q. Szabo, F. Bantignies, and G. Cavalli. Principles of genome folding into topologically associating domains. *Science Advances*, 5, 2019. ISSN 23752548. doi: 10.1126/SCIADV.AAW1668. URL </pmc/articles/PMC6457944/>[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6457944/](/pmc/articles/PMC6457944/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6457944/).
- S. Takahashi, I. Nobuhisa, K. Saito, M. Gerel, A. Itabashi, K. Harada, M. Osawa, T. A. Endo, A. Iwama, and T. Taga. Sox17-mediated expression of adherent molecules is required for the maintenance of undifferentiated hematopoietic cluster formation in midgestation mouse embryos. *Differentiation; research in biological diversity*, 115:53–61, 9 2020. ISSN 1432-0436. doi: 10.1016/J.DIFF.2020.08.001. URL <https://pubmed.ncbi.nlm.nih.gov/32891959/>.
- D. S. Tan, M. Holzner, M. Weng, Y. Srivastava, and R. Jauch. Sox17 in cellular reprogramming and cancer. *Seminars in Cancer Biology*, 67:65–73, 12 2020. ISSN 1044-579X. doi: 10.1016/J.SEMCANCER.2019.08.008.

- R. C. Team. R: A language and environment for statistical computing., 2021. URL <https://www.r-project.org>.
- K. Teichert-Kuliszewski, M. J. Kutryk, M. A. Kuliszewski, G. Karoubi, D. W. Courtman, L. Zucco, J. Granton, and D. J. Stewart. Bone morphogenetic protein receptor-2 signaling promotes pulmonary arterial endothelial cell survival. *Circulation Research*, 98:209–217, 2 2006. ISSN 0009-7330. doi: 10.1161/01.RES.0000200180.01710.e6. URL <https://www.ahajournals.org/doi/10.1161/01.RES.0000200180.01710.e6>.
- T. Thenappan, S. J. Shah, S. Rich, and M. Gomberg-Maitland. A usa-based registry for pulmonary arterial hypertension: 1982-2006. *The European respiratory journal*, 30: 1103–10, 12 2007. ISSN 0903-1936. doi: 10.1183/09031936.00042107. URL <http://www.ncbi.nlm.nih.gov/pubmed/17804449>.
- P. J. Thul, L. Akesson, M. Wiking, D. Mahdessian, A. Geladaki, H. A. Blal, T. Alm, A. Asplund, L. Björk, L. M. Breckels, A. Bäckström, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Hober, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C. M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J. M. Schwenk, A. Sivertsson, E. Sjöstedt, M. Skogs, C. Stadler, D. P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoglu, F. Pontén, K. V. Feilitzén, K. S. Lilley, M. Uhlén, and E. Lundberg. A subcellular map of the human proteome. *Science*, 356, 5 2017. ISSN 10959203. doi: 10.1126/science.aal3321.
- G. Thurston, I. Noguera-Troise, I. B. Lobov, C. Daly, J. S. Rudge, N. W. Gale, S. J. Wiegand, and G. D. Yancopoulos. Delta-like ligand 4/notch pathway in tumor angiogenesis, 2008.
- M. I. Townsley. Structure and composition of pulmonary arteries, capillaries, and veins, 1 2012. URL <http://doi.wiley.com/10.1002/cphy.c100081>.
- R. C. Trembath, J. R. Thomson, R. D. Machado, N. V. Morgan, C. Atkinson, I. Winship, G. Simonneau, N. Galie, J. E. Loyd, M. Humbert, W. C. Nichols, J. Berg, A. Manes, J. McGaughan, M. Pauciulo, L. Wheeler, and N. W. Morrell. Clinical and molecular genetic features of pulmonary hypertension in patients with hereditary hemorrhagic telangiectasia. *New England Journal of Medicine*, 345:325–334, 8 2001. ISSN 0028-4793. doi: 10.1056/NEJM200108023450503. URL <http://www.nejm.org/doi/abs/10.1056/NEJM200108023450503>.
- L. Tu, A. Desroches-Castan, C. Mallet, L. Guyon, A. Cumont, C. Phan, F. Robert, R. Thuillet, J. Bordenave, A. Sekine, A. Huertas, O. Ritvos, L. Savale, J. J. Feige, M. Humbert, S. Bailly, and C. Guignabert. Selective bmp-9 inhibition partially protects against experimental pulmonary hypertension. *Circulation Research*, 124:846–855, 3

2019. ISSN 15244571. doi: 10.1161/CIRCRESAHA.118.313356. URL <https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.118.313356>.
- R. M. Tuder, B. Groves, D. B. Badesch, and N. F. Voelkel. Exuberant endothelial cell growth and elements of inflammation are present in plexiform lesions of pulmonary hypertension. *American Journal of Pathology*, 144:275–285, 2 1994. ISSN 00029440. URL [/pmc/articles/PMC1887146/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1887146/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1887146/?report=abstract).
- H. Uwamori, Y. Ono, T. Yamashita, K. Arai, and R. Sudo. Comparison of organ-specific endothelial cells in terms of microvascular formation and endothelial barrier functions. *Microvascular research*, 122:60, 3 2019. ISSN 10959319. doi: 10.1016/J.MVR.2018.11.007. URL [/pmc/articles/PMC6294313//pmc/articles/PMC6294313/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6294313/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6294313/?report=abstract).
- M. Vaillancourt, G. Ruffenach, J. Meloche, and S. Bonnet. Adaptation and remodelling of the pulmonary circulation in pulmonary hypertension, 4 2015. ISSN 0828282X. URL <https://www.sciencedirect.com/science/article/pii/S0828282X14014937?via%3Dihub>.
- C. J. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. G. Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Y. Wang, A. Wang, X. Wang, J. Wang, M. H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferreira, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Ko-

- duru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Rendon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. D. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. The sequence of the human genome. *Science*, 291:1304–1351, 2 2001. ISSN 0036-8075. doi: 10.1126/SCIENCE.1058040. URL <https://pubmed.ncbi.nlm.nih.gov/11181995/>.
- C. E. Ventetuolo, A. Praestgaard, H. I. Palevsky, J. R. Klinger, S. D. Halpern, and S. M. Kawut. Sex and haemodynamics in pulmonary arterial hypertension. *European Respiratory Journal*, 43:523–530, 2 2014. ISSN 09031936. doi: 10.1183/09031936.00027613.
- N. Villa, L. Walker, C. E. Lindsell, J. Gasson, M. L. Iruela-Arispe, and G. Weinmaster. Vascular expression of notch pathway receptors and ligands is restricted to arterial vessels. *Mechanisms of Development*, 108:161–164, 10 2001. ISSN 09254773. doi: 10.1016/S0925-4773(01)00469-5.
- D. Wang, F. Wang, K. H. Shi, H. Tao, Y. Li, R. Zhao, H. Lu, W. Duan, B. Qiao, S. M. Zhao, H. Wang, and J. Y. Zhao. Lower circulating folate induced by a fidgetin intronic variant is associated with reduced congenital heart disease susceptibility. *Circulation*, 135:1733–1748, 5 2017. ISSN 15244539. doi: 10.1161/CIRCULATIONAHA.116.025164. URL <https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.116.025164>.
- D. Wang, Z. Chen, A. W. K. Yeung, and A. G. Atanasov. Differences between common endothelial cell models (primary human aortic endothelial cells and ea.hy926 cells) revealed through transcriptomics, bioinformatics, and functional analysis. *Current Research in Biotechnology*, 3:135–145, 1 2021. ISSN 2590-2628. doi: 10.1016/J.CRBIOT.2021.05.001.

- G. Wang, R. Fan, R. Ji, W. Zou, D. J. Penny, N. P. Varghese, and Y. Fan. Novel homozygous bmp9 nonsense mutation causes pulmonary arterial hypertension: a case report. *BMC Pulmonary Medicine*, 16:17, 12 2016. ISSN 1471-2466. doi: 10.1186/s12890-016-0183-7. URL <http://www.biomedcentral.com/1471-2466/16/17>.
- R. C. Wang, F. M. Jiang, Q. L. Zheng, C. T. Li, X. Y. Peng, C. Y. He, J. Luo, and Z. A. Liang. Efficacy and safety of sildenafil treatment in pulmonary arterial hypertension: a systematic review. *Respiratory medicine*, 108:531–537, 3 2014. ISSN 1532-3064. doi: 10.1016/J.RMED.2014.01.003. URL <https://pubmed.ncbi.nlm.nih.gov/24462476/>.
- G. Wei, R. Srinivasan, C. Z. Cantemir-Stone, S. M. Sharma, R. Santhanam, M. Weinstein, N. Muthusamy, A. K. Man, R. G. Oshima, G. Leone, and M. C. Ostrowski. Ets1 and ets2 are required for endothelial cell survival during embryonic angiogenesis. *Blood*, 114:1123–1130, 2009. ISSN 00064971. doi: 10.1182/blood-2009-03-211391.
- A. Weiss, M. C. Neubauer, D. Yerabolu, B. Kojonazarov, B. C. Schlueter, L. Neubert, D. Jonigk, N. Baal, C. Ruppert, P. Dorfmueller, S. S. Pullamsetti, N. Weissmann, H. A. Ghofrani, F. Grimminger, W. Seeger, and R. T. Schermuly. Targeting cyclin-dependent kinases for the treatment of pulmonary arterial hypertension. *Nature Communications* 2019 10:1, 10:1–17, 5 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-10135-x. URL <https://www.nature.com/articles/s41467-019-10135-x>.
- R. J. White. Estrogen: Friend or foe in pulmonary hypertension? *American Journal of Respiratory and Critical Care Medicine*, 193:1084–1086, 5 2016. ISSN 15354970. doi: 10.1164/RCCM.201512-2511ED/SUPPL_FILE/DISCLOSURES.PDF. URL www.atsjournals.org.
- M. R. Wilkins, J. Aman, L. Harbaum, A. Ulrich, J. Wharton, and C. J. Rhodes. Recent advances in pulmonary arterial hypertension. *F1000Research*, 7:1128, 2018. ISSN 2046-1402. doi: 10.12688/f1000research.14984.1. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6058465/https://f1000research.com/articles/7-1128/v1>.
- Y. Wu, J. Wharton, R. Walters, E. Vasilaki, J. Aman, L. Zhao, M. R. Wilkins, and C. J. Rhodes. The pathophysiological role of novel pulmonary arterial hypertension gene sox17. *European Respiratory Journal*, page 2004172, 2 2021. ISSN 0903-1936. doi: 10.1183/13993003.04172-2020. URL <https://erj.ersjournals.com/content/early/2021/02/11/13993003.04172-2020https://erj.ersjournals.com/content/early/2021/02/11/13993003.04172-2020.abstract>.
- J. D. Wythe, L. T. Dang, W. P. Devine, E. Boudreau, S. T. Artap, D. He, W. Schachterle, D. Y. Stainier, P. Oettgen, B. L. Black, B. G. Bruneau, and J. E. Fish. Ets factors regulate vegf-dependent arterial specification. *Developmental Cell*, 26:45–58, 7 2013. ISSN 15345807. doi: 10.1016/j.devcel.2013.06.007.

- H. Yang, S. S. Lee, S. S. Lee, K. Kim, Y. Yang, J. H. Kim, R. H. Adams, J. M. Wells, S. J. Morrison, G. Y. Koh, and I. Kim. Sox17 promotes tumor angiogenesis and destabilizes tumor vessels in mice. *The Journal of Clinical Investigation*, 123:418, 1 2013. ISSN 1558-8238. doi: 10.1172/JCI64547. URL <http://www.ncbi.nlm.nih.gov/pubmed/23241958><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3533291>.
- J. Yang, R. J. Davies, M. Southwood, L. Long, X. Yang, A. Sobolewski, P. D. Upton, R. C. Trembath, and N. W. Morrell. Mutations in bone morphogenetic protein type ii receptor cause dysregulation of id gene expression in pulmonary artery smooth muscle cells: implications for familial pulmonary arterial hypertension. *Circulation research*, 102:1212–21, 5 2008. ISSN 1524-4571. doi: 10.1161/CIRCRESAHA.108.173567. URL <http://www.ncbi.nlm.nih.gov/pubmed/18436795>.
- Z. Yang, X. Xu, C. Gu, J. Li, Q. Wu, C. Ye, A. V. Nielsen, L. Mao, J. Ye, K. Bai, F. Guo, C. Tang, and Y. Zhao. Chemicals orchestrate reprogramming with hierarchical activation of master transcription factors primed by endogenous sox17 activation. *Communications Biology* 2020 3:1, 3:1–10, 10 2020. ISSN 2399-3642. doi: 10.1038/s42003-020-01346-w. URL <https://www.nature.com/articles/s42003-020-01346-w>.
- K. Yasuno, K. Bilguvar, P. Bijlenga, S. K. Low, B. Krischek, G. Auburger, M. Simon, D. Krex, Z. Arlier, N. Nayak, Y. M. Ruigrok, M. Niemelä, A. Tajima, M. V. U. Z. Fraunberg, T. Dóczi, F. Wirjatijasa, A. Hata, J. Blasco, A. Oszvald, H. Kasuya, G. Zilani, B. Schoch, P. Singh, C. Stüer, R. Risselada, J. Beck, T. Sola, F. Ricciardi, A. Aromaa, T. Illig, S. Schreiber, C. M. V. Duijn, L. H. V. D. Berg, C. Perret, C. Proust, C. Roder, A. K. Ozturk, E. Gaál, D. Berg, C. Geisen, C. M. Friedrich, P. Summers, A. F. Frangi, M. W. State, H. E. Wichmann, M. M. Breteler, C. Wijmenga, S. Mane, L. Peltonen, V. Elio, M. C. Sturkenboom, P. Lawford, J. Byrne, J. MacHo, E. I. Sandalcioğlu, B. Meyer, A. Raabe, H. Steinmetz, D. Rüfenacht, J. E. Jääskeläinen, J. Hernesniemi, G. J. Rinkel, H. Zembutsu, I. Inoue, A. Palotie, F. Cambien, Y. Nakamura, R. P. Lifton, and M. Günel. Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nature genetics*, 42:420–425, 5 2010. ISSN 1546-1718. doi: 10.1038/NG.563. URL <https://pubmed.ncbi.nlm.nih.gov/20364137/>.
- B. Yates, B. Braschi, K. A. Gray, R. L. Seal, S. Tweedie, and E. A. Bruford. Genenames.org: the hgnc and vgnc resources in 2017. *Nucleic acids research*, 45:D619–D625, 2017. ISSN 1362-4962. doi: 10.1093/nar/gkw1033. URL <http://www.ncbi.nlm.nih.gov/pubmed/27799471><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5210531>.

- E. S. Yi, H. Kim, H. Ahn, J. Strother, T. Morris, E. Masliah, L. A. Hansen, K. Park, and P. J. Friedman. Distribution of obstructive intimal lesions and their cellular phenotypes in chronic pulmonary hypertension: A morphometric and immunohistochemical study. *American Journal of Respiratory and Critical Care Medicine*, 162:1577–1586, 10 2000. ISSN 1073449X. doi: 10.1164/ajrccm.162.4.9912131. URL <http://www.atsjournals.org/doi/abs/10.1164/ajrccm.162.4.9912131>.
- V. Yodsurang, Y. Tang, Y. Takahashi, C. Tanikawa, Y. Kamatani, A. Takahashi, Y. Momozawa, N. Fuse, J. Sugawara, A. Shimizu, A. Fukushima, A. Hishida, N. Furusyo, M. Naito, K. Wakai, T. Yamaji, N. Sawadai, M. Iwasaki, S. Tsugane, M. Hirata, Y. Murakami, M. Kubo, and K. Matsuda. Genome-wide association study (gwas) of ovarian cancer in japanese predicted regulatory variants in 22q13.1. *PLOS ONE*, 13:e0209096, 12 2018. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0209096. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0209096>.
- T. Yuen, J. Iqbal, L. L. Zhu, L. Sun, A. Lin, H. Zhao, J. Liu, P. K. Mistry, and M. Zaidi. Disease-drug pairs revealed by computational genomic connectivity mapping on gba1 deficient, gaucher disease mice. *Biochemical and Biophysical Research Communications*, 422:573–577, 6 2012. ISSN 0006291X. doi: 10.1016/j.bbrc.2012.05.027.
- S. Zahler, J. Liebl, R. Fürst, and A. M. Vollmar. Anti-angiogenic potential of small molecular inhibitors of cyclin dependent kinases in vitro. *Angiogenesis*, 13:239–249, 9 2010. ISSN 09696970. doi: 10.1007/S10456-010-9181-1/TABLES/3. URL <https://link.springer.com/article/10.1007/s10456-010-9181-1>.
- Y. Zhang, W. Bao, K. Wang, W. Lu, H. Wang, H. Tong, and X. Wan. Sox17 is a tumor suppressor in endometrial cancer. *Oncotarget*, 7:76036–76046, 2016. ISSN 1949-2553. doi: 10.18632/ONCOTARGET.12582. URL <https://pubmed.ncbi.nlm.nih.gov/27738313/>.
- Y. D. Zhao, A. I. Campbell, M. Robb, D. Ng, and D. J. Stewart. Protective role of angiopoietin-1 in experimental pulmonary hypertension. *Circulation Research*, 92:984–991, 5 2003. ISSN 00097330. doi: 10.1161/01.RES.0000070587.79937.F0. URL <http://www.ncbi.nlm.nih.gov/pubmed/12690034>.
- J. Zhou, P. Su, L. Wang, J. Chen, M. Zimmermann, O. Genbacev, O. Afonja, M. C. Horne, T. Tanaka, E. Duan, S. J. Fisher, J. Liao, J. Chen, and F. Wang. mtor supports long-term self-renewal and suppresses mesoderm and endoderm activities of human embryonic stem cells. *Proceedings of the National Academy of Sciences*, 106:7840–7845, 5 2009. ISSN 0027-8424. doi: 10.1073/PNAS.0901854106. URL <https://www.pnas.org/content/106/19/7840https://www.pnas.org/content/106/19/7840.abstract>.

- W. Zhou, K. Wang, J. Wang, J. Qu, G. Du, and Y. Zhang. Sox17 inhibits tumor metastasis via wnt signaling in endometrial cancer. *OncoTargets and therapy*, 12:8275–8286, 2019. ISSN 1178-6930. doi: 10.2147/OTT.S220536. URL <https://pubmed.ncbi.nlm.nih.gov/31632077/>.
- Y. Zhou, J. Williams, P. M. Smallwood, and J. Nathans. Sox7, sox17, and sox18 cooperatively regulate vascular development in the mouse retina. *PloS one*, 10: e0143650, 12 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0143650. URL <http://www.ncbi.nlm.nih.gov/pubmed/26630461><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4667919><http://dx.plos.org/10.1371/journal.pone.0143650>.
- N. Zhu, C. L. Welch, J. Wang, P. M. Allen, C. Gonzaga-Jauregui, L. Ma, A. K. King, U. Krishnan, E. B. Rosenzweig, D. D. Ivy, E. D. Austin, R. Hamid, M. W. Pauciulo, K. A. Lutz, W. C. Nichols, J. G. Reid, J. D. Overton, A. Baras, F. E. Dewey, Y. Shen, and W. K. Chung. Rare variants in sox17 are associated with pulmonary arterial hypertension with congenital heart disease. *Genome Medicine*, 10:56, 2018. ISSN 1756994X. doi: 10.1186/s13073-018-0566-x. URL <http://www.ncbi.nlm.nih.gov/pubmed/30029678><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6054746>.

7 Supplemental Work

The following section contains all supplemental tables, and figures to accompany this thesis.

7.1 Supplemental Tables

Regent	Additional Information	Company	Catalogue code
Dulbecco's modified eagle medium (DMEM) – High glucose	4500 mg/L glucose, L-glutamine, and sodium bicarbonate, without sodium pyruvate, liquid, sterile-filtered.	SA	D5796-6X500ML
FBS	Low endotoxin and hemoglobin levels.	Tf	26140079
L-Glutamate	Without L-glutamine, liquid, sterile-filtered, BioReagent, suitable for cell culture	SA	G7513-100ML
Sodium pyruvate	100 mM, sterile-filtered, BioReagent.	SA	S8636-100ML
MEM non-essential amino acids	Without L-glutamine, liquid, sterile-filtered, BioReagent.	SA	M7145-100ML
Fibronectin	Sterile-filtered, BioReagent	SA	F1141-5MG
Endothelial basal medium with supplemental pack (EGM2)		Promocell	C-22211 C-39211
Trypsin solution	1X, sterile-filtered, BioReagent, 0.5 g porcine trypsin and 0.2 g EDTA. 4Na per litre of Hanks Balanced Salt Solution with phenol red	SA	T2601-500ML
NE-PER nuclear and cytoplasmic extraction reagents kit		Tf	78833
Amicon centrifugal filter (300Kd NMWL)		Merck	36100101
Pierce Coomassie protein kit		Tf	23200
TBE buffer (10x)	BioReagent, suitable for electrophoresis	SA	T4415-1L
DNA retardation gels (6% w/v)		Tf	EC63652
Biodyne B nylon membrane, 0.45um, 8cmx12cm		Tf	77016

Regent	Additional Information	Company	Catalogue code
EMSA 10X annealing buffer	TRIS [100 mM] EDTA [10 mM] NaCl [500mM]	All SA	T1503-1KG (from lightshift kit) S7653-250G
Lightshift chemiluminescent EMSA kit		Tf	20148
EMSA 10X binding buffer	Poly(dIdC) [50 ng/ul] MgCl ₂ [5 mM] ZnSO ₄ [5 mM]	All Tf, Mgcl ₂ SA	2014E M8266-100G
DNA loading dye (5x)		Qiagen	239901
Ampicillin		SA	10835242001
LB agar		SA	L2897-1KG
GenElute plasmid miniprep kit		SA	PLN350- 1KT
Pureyeild plasmid midiprep kit		Promega	A2492
PEIpro		VWR	115-010
Lenti-X concentrator		Takarabio	631232
Blasticidin		SA	15205- 25MG
RNeasy plus kit		Qiagen	74136
Multiscribe reverse transcriptase		Tf	4368814
Power SYBRGreen mastermix		Tf	4357659
siRNA-SOX17		Tf	s34626
siRNA-SOX7		Tf	s38029
siRNA-SOX18		Tf	s28927
siRNA-GAPDH		Tf	4390849
siRNA-Negative 1		Tf	4390843
Optimem	With L-Glutamine and phenol-red.	Tf	31985070
Lipofectamine RNAi MAX		Tf	13778100

Regent	Additional Information	Company	Catalogue code
Phosphate buffered saline (PBS)	Modified, without calcium chloride and magnesium chloride, liquid, sterile-filtered.	SA	D8537-500ML
RIPA buffer		SA	20-188
Protease inhibitor cocktail		Tf	78430
Phosphatase inhibitor cocktail		SA	P0044-5ML
Laemmli SDS sample buffer		Tf	J60015
NuPage, Nis-TRIS mini protein gel		Tf	NP0336BOX
NuPage MES-SDS buffer		Tf	NP0002
PVDF transfer pack		Biorad	1704156
Dried milk powder	Thermophile-free, spray dried powder.	Tf	LP0033B
TBST		Tf	28360
Immobilon crescendo western HRP substrate		Merck	WBLUR0500
Spectr multicolor broad range protein ladder		Tf	26623
Recombinant Human VEGF (165)		PeproTech	100.20
MTT		SA	M2128-100MG
MTT detergent	Isopropanol 94.65% (v/v) NP-40 5% (v/v) HCL 0.35% (v/v)	FisherSci S SA	14-650-208 74385 258148
Caspase-Glo 3/7 kit		Promega	G8090
Lipopolysaccharide		SA	L4391-1MG
TNF-alpha	HEK293-derived.	R&D	10291-TA

Regent	Additional Information	Company	Catalogue code
CellTitre-Glo kit		Promega	G7570
FITC-Dextran	40kDa	SA	FD40S-100MG
DiMethyl Sulfoxide (DMSO)		SA	D2650-100ML
Aminopurvalanol-a		APExBIO	B6935 1mg
YK-4279		APExBIO	A3946 10mM
Sirolimus		APExBIO	A8167-APE-25mg
ddH2O	Sterile filtered, for molecular biology	SA	95284-6 X100ML

Table 7.1: Materials and reagents. SA, Sigma Aldrich. Tf, Thermofisher.

Condition	Up-regulated Gene List	Gene Names	Down-regulated Gene List	Gene Names
SOX17-Promoter Repression	98	EIF3CL, TRG-AS1, MEG8, MINDY4, RAB3C, GALNT12, POLE2, APCDD1L, ELFN2, ESPN, FANCD2, BARD1, KIF15, FER1L4, IFITM1, CHCHD5, ZNF471, KIF19, LAMA2, FERMT1, FGL2, ATP2A3, LRRRC4, ATG9B, ZEB2, ZNF10, ANO2, TMEM156, ZFHX4-AS1, FANCC, APIP, RCAN3, SPC24, ZNF653, SCN3B, DNA2, APOLD1, STC1, DGCR11, PPM1L, SKA3, MAPK13, NME1-NME2, MCEE, HELLS, CYB5RL, TRBC2, PLVAP, CBS, CENPI, CDKN3, THAP10, PRIM1, ZNF559, AH11, PCBD2, SULT1A1, ORC1, ACKR1, MMS22L, BBS5, GZF1, SPATA13, PBK, FAM111B, THAP6, ZNF283, ZFYVE16, CPXM1, MTRF1, RAD51, CDCA5, C15orf41, DHH, CACFD1, PTC1, DEPDC1B, DNAJC25-GNG10, DEPDC5, CENPF, KIF2C, CLBA1, KIF20A, CD68, PARBP, CCNB2, PIMREG, RRM2, PLEKHM3, ARHGAP11A, CXCL6, SLC25A16, BICRA, LMNB1, LINS1, RNASEH2A, TOX, ZWINT, BIRC5, ORC6, C1orf112, SHC3, TMEM97, SPIN4, TNFRSF6B, DLGAP5, RGS5, MRC1, DEPDC1, CENPW, IGFBP3, MGARP, CXCL1, MYBL2, TNRC6B, CEP55, CDCA2, SLC17A9, TTK, MAD2L1, SIPA1L1, PCLAF, KNTC1, SULT1A4, CCNA2, MTO1, PRC1, MCUR1, TRIP13, PTC2, KLHL8, SCLT1, ITGAE, COL13A1, TK1, SPAAR, UBE2E2, SDC1, PHF19, PTPN2, CDC20, PTTG1, NDC80, IL1RL1, AP5S1, PLCL1, FGFR1OP, DENND4A, FAM76B, FEN1.	106	IKBKGP1, CDRT4, TMEM132B, TLX1, UPK3BL1, DOCK8, NEBL, ZC3H6, C2CD4A, MAN1B1-DT, HYPK, TRPS1, KCNK3, POU2F1, LGALS1, CAPS, TECTA, FGFR4, TBXA2R, ABCB1, FAM156B, PFKFB2, SLC25A35, C1orf50, TRIM45, MITF, PLA2G6, DNAH14, POU6F1, HSD11B1L, INHBA, BTBD11, SGIP1, NEUROD2, SHROOM1, IGF2, GLT8D2, SATB2, DACH1, AXIN2, HLX, GOLGA8B, HSPA2, ZNF862, TRIM62, PKD1P1, WDR91, ARG2, CHRFAM7A, ALDH1A1, IQGAP2, FASN, PKD1, PLCG2, ITPR1, TRO, POR, NIPAL2, CD34, PLA2G4C, IKZF5, SELE, CYB5R4, CCL2, SLC25A42, NPR2, CASKIN2, ZNF446, SLC9A7, ZBTB7B, SPTBN5, PGGT1B, MKS1, ACSM3, FRS2, GAB1, SLC26A6, ICOSLG, UNKL, ZNF780A, STC2, PLXNA3, RNF217, CTXN1, SBF2, PHF8, ZHX2, ADPRH, TRAF5, ADARB1, RBMS2, PHC3, KLHL17, CRTCL1, PGF, CRACR2B, PIBF1, PIAS4, BCL9, CHST1, SETD1A, ZFX, ARL5A, ICAM1, THBS1, DEPP1, MAPK8, HERC2P9, PITPNM1, ARGLU1, BMP4, DLL4, MT-ND5, PER1, GNAI1, SLC2A6, CPNE5, SORT1, STIM1, UBE2Q1, TRIM69, MT-CYB, EIF4ENIF1, HERC2P3, SART3, USP20, SLC25A23, IPO13, PGGHG, JUP, MGP, PIGQ, RGS19, C12orf65, MAD1L1, PIK3CD, VASH1, AGO4, FABP5, RELB, MYCBP2, RHOB, LAMB2, LAMC1, MCOLN1, SMAD1, INSR, APPBP2, MRTFA, PLXNB1.

Table 7.2: CMap *SOX17* promoter repression gene lists

Condition	Up-regulated Gene List	Gene Names	Down-regulated Gene List	Gene Names
SOX17-Promoter Activation	150	INO80B-WBP1, RIMS1, RPS4Y1, GJA5, STAC, SEMA3G, KDM5D, USP9Y, IL1A, CD69, SOX17, TGFB1, NPTX2, SLC16A9, DOCK7, CXCL3, PPP1R12B, PTPN7, CXCL1, GJA4, CXCL2, DNA2, TCF15, ZNF519, DIO2, MYRF, LRP4, UNC5B, SELE, HEY1, TCTEX1D1, CCNA1, FST, POLR2J4, VCAM1, TMEM79, ADAM12, NOSTRIN, DCLK1, SDR42E1, KBTBD3, ARHGAP28, CDC6, JAM2, LIPG, APOL4, INHBA, JMJD7, CCDC15, PHGDH, PPARGC1B, INSR, LFNG, PLLP, ZNF585A, CCL2, RFTN2, DTL, TUSC3, CKAP2L, HSPA2, PLK4, ELFN2, ALG14, LYPD1, ACKR3, RND1, DEPDC1, MMS22L, ARHGEF25, ZBTB14, EFNB2, PODN, TTK, PAPLN, DTNA, NRN1, PEG10, GPRIN3, PSRC1, TTC4, CHEK2, IL17RA, COL3A1, ZNF229, TOX, XRRA1, BUB1B, NUF2, NR5A2, GPHN, ZNF383, ACSM3, MYLK, RUNX1, TTC26, KIF20A, CXCL6, COL4A1, IL1RAP, MYBL2, BLM, AURKA, TNFAIP2, ITM2A, MCM5, E2F1, MATN3, TAPT1, SPAG5, RAD51AP1, BIRC3, F2RL1, SPC25, HJURP, NFKBIZ, KSR2, CHST1, TSPAN13, TULP4, MIS18A, ADGRG6, E2F8, CCNB2, RIPK2, ZWINT, GORAB, UBE2S, SORBS2, KIF26A, ZWILCH, SKA3, SLC7A11, CENPE, SERPINE2, AURKB, IL3RA, ITGB8, ANLN, NOTCH4, NOCT, GINS1, ABL2, ADAMTS6, DLGAP5, ADAM23, SOX13, IKBKE, GSTCD, RNASE1.	80	CDC14B, KCNMA1, MSMP, KCND1, SAMD12, IL17RE, ACP5, RAG1, RGPDP2, H4C9, FAM74A4, FAM27E3, GRIK5, ZNF490, ZNF571, LRRC34, WASH7P, NOTCH3, FAM221A, ACTR3B, GTF2IRD2, ZNF653, ITGB4, PCDHA10, GALNT8, NEK10, CCL14, RARRES1, ZNF607, NTM, VGLL3, FOXO6, PHLPP1, SESN3, H2AC6, PTGIR, TM7SF2, C1orf56, PDK4, PDE2A, GALNT15, GALNTL5, SLX1B-SULT1A4, SEMA4F, SLC4A8, SLC40A1, MREG, ZNF767P, CHD2, RCCD1, ADAMTS18, RNASET2, STAG3L5P, MMAA, USP49, PHC3, OCLN, DEPDC5, HYAL3, GABPB2, GABPB1, GSTM2, FAM78A, HID1, NPR1, MYEOV, MTURN, FAM217B, TIAM2, HEATR6, EPHA4, ARID4A, WDR90, SCN1B, CDKN2B, IL17D, ZNF286A, SNED1, THBD, ITGA10.

Table 7.3: CMap *SOX17* promoter activation gene lists

Condition	Up-regulated Gene List	Gene Names	Down-regulated Gene List	Gene Names
SOX17-Signal 1 Repression	150	BMPR1A, MEGF10, NPIPA7, LINC01002, RAB3C, CPA4, PPT2-EGFL8, MSMP, ZIC1, ZBTB43, ADCYAP1R1, SH2D2A, THAP10, CD36, GABRE, DRAXIN, SERHL2, LLGL2, ACP5, PSG4, ZNF774, TRBC2, GNG2, APCDD1L, CDO1, PTCH1, RELN, KRT15, ZNF26, FSD1L, LTB4R, TEDC2, CDT1, BAALC, CPNE7, CCBE1, IMMP2L, IL1RL1, PIK3CG, ZNF283, MAPK13, POLE2, MFSD2A, LYVE1, BDH1, DENND1B, ZNF324B, STC1, SHQ1, PTPN22, KRT19, COQ8B, MAP3K6, DMXL2, AKAP9, ADIRF, CCDC68, TNFRSF6B, TREX1, LETM2, NAA16, ADAMTS18, BBS5, TMEM273, ZNF616, CENPJ, MGARP, SMYD4, CRCP, STRIP2, VPS13B, EGFR, ULK3, SLC4A8, TBC1D2, PDE2A, CENPI, EPC1, COL4A6, DDHD1, ADAMTS1, ZNF91, NEGR1, GFPT2, CSPP1, HELLS, MFSD14C, RAB8A, DAGLB, GAS8, ITGAE, NQO1, NBPF14, ATG16L2, NCR3LG1, ZNF280D, FAH, DIS3L.	150	BACH2, SEMA3G, SOX17, RGPDI, AFAP1L2, SCAPER, LARGE1, MATN2, TBC1D32, TMEM117, C2CD4A, RBPMS2, CX3CL1, CCL2, ZNF850, MGAT5B, SLC15A3, RASA4B, POU6F1, C16orf95, HDAC9, CTSS, OPLAH, HSPA2, PKHD1L1, GJA5, SEMA4D, RAI2, COL3, BMF, GDFGP1, SHISA3, NIBAN1, IL3RA, BTBD11, CTSK, APBA1, MXD3, SEMA7A, PITPNM1, SHROOM1, EPHB1, LYPD1, ATXN7L1, CPA3, PAPLN, HOXA-AS2, CHN2, GTF2IRD2, CD34, ATP7B, TBL1XR1, ITGA1, FBXO32, PREX2, MCTP1, FILIP1, NAT1, SLC6A2, PKD1P1, A2M, CEP162, HERC2P3, TCIM, ERCC4, VWA1, CXCL2, PLLP, SEMA4F, EHD3, NTN4, MAMLD1, COMMD3-BMI1, ZFX, ALS2CL, TNFSF10, ACKR3, LAMP3, POMT2, ZNF75A, ATPAF2, EDN1, FANCG, ALDH1A1, AIF1L, ABLIM1, RGS3, COLEC12, C15orf54, KLRG1, WARS1, ZSCAN18, GBP4, SLC2A8, COL8A1, FABP4, TMC6, NYNRIN, PDGFB, LZTS3, SPNS2, BMP4, PRR5L, FAM117A, MMP2, LIG1.

Table 7.4: CMap *SOX17*-signal 1 repression gene lists

Gene Name	Drug	Condition	FWD Primer (5'-3')	REV Primer (5'-3')
<i>ITGAE</i>	Sirolimus	SOX17-promoter repression	-GAGGTCATCTGCTCATGTTTCAGT-	-TTGAAGGGTGTGGGTCTCGT-
<i>RAB8A</i>	Sirolimus	SOX17-promoter repression	-ACGGTTTCGGACGATCACAA-	-TCCAGTTCGGATGTTGTGCG-
<i>FAH</i>	Sirolimus	SOX17-promoter repression	-TTCACGGTGAGACCAAAAGTCAG-	-GCTTGATGATGCTGAGGTCCA-
<i>PKD1P1</i>	Sirolimus	SOX17-promoter repression	-CGACAGCTGCCTGGACTCAT-	-TAGAGCCCTCACCTCAGCGT-
<i>MXD3</i>	Sirolimus	SOX17-promoter repression	-GAGCATGGTTATGCGTCCCT-	-CCTGCGCTTCTCCAGTTCAT-
<i>PITPNM1</i>	Sirolimus	SOX17-promoter repression	-CTTTGATGCCCACGAAGGCT-	-TTTAGCTGCCTCGGCTCCAG-
<i>MAMLD1</i>	Sirolimus	SOX17-promoter repression	-CAGGGGGTGATCTGCAACAG-	-TGACTCTCCAAGGCGTTTCC-
<i>IL1RL1</i>	Sirolimus	SOX17-promoter repression	-AGAGGAAAACAAACCCACAAGG-	-CTGGCCGGTGACATTACAGAT-
<i>BMPR1A</i>	Sirolimus	SOX17-promoter repression	-GAAAAAGTGGCGGTGAAAGT-	-TAGAGCTGAGTCCAGGAACC-
	Aminopurvalanol-a	SOX17-promoter repression		
<i>CENPI</i>	Aminopurvalanol-a	SOX17-promoter repression	-GCCTCATCTCCAGGCTTTGT-	-CCGGTTTCTTTGCTTCACGG-
<i>NQO1</i>	Aminopurvalanol-a	SOX17-promoter repression	-AGGCTGGTTTGAGCGAGTGT-	-ATGTCCCCGTGGATCCCTTG-
<i>SHQ1</i>	Aminopurvalanol-a	SOX17-promoter repression	-TCAGGAAAATGACCCAGCGT-	-AGCTGGGCCTTTGTAAGGGA-
<i>POLE2</i>	Aminopurvalanol-a	SOX17-promoter repression	-TCACCAGCACCTCCAACCTG-	-CCTCTGGACCAGGTACAAACACA-
	Aminopurvalanol-a	Signal-1 repression		
<i>ACKR3</i>	Aminopurvalanol-a	SOX17-promoter repression	-ATTTGATTGCCCGCCTCAGA-	-GACGCTTTTGTGGGCATGT-
<i>NYNRIN</i>	Aminopurvalanol-a	SOX17-promoter repression	-GCAGGAATGGTTCATGGTGC-	-GCACAGGCCCTTCAGGTATT-
<i>LAMP3</i>	Aminopurvalanol-a	SOX17-promoter repression	-CAAACATGCGGTGGTGATGT-	-GTGTAGTCAGACGAGCACTCA-
<i>AURKA</i>	YK-4279	SOX17-promoter activation	-GTGGCGGAGCGTCAAGT-	-AAATATCCCCGCACTCTGGC-
<i>CNEPE</i>	YK-4279	SOX17-promoter activation	-ACACGGATGCTGGTGACCTC-	-TGCCAAGGCACCAAGTAACTC-
<i>DEPDC1</i>	YK-4279	SOX17-promoter activation	-TGCCATGAAGTGCCTAGCAA-	-ACTGTGATGTAGCCACAAACAACC-
<i>KIF20A</i>	YK-4279	SOX17-promoter activation	-GCTTCGGCGACTAGGTGTGA-	-TTGCGTACCACAGACCCCAA-
	Aminopurvalanol-a	Signal-1 repression		
<i>GABPB1</i>	YK-4279	SOX17-promoter activation	-GACCTGGAGGGGTGGTGAAC-	-TCTGTGGCCACTACTGGAGTT-
<i>SEMA4F</i>	YK-4279	SOX17-promoter activation	-TCGCTTTATCCCTGCCCTTC-	-TGGACACATCCTCTTTCTTGCC-
<i>CDC20</i>	Aminopurvalanol-a	Signal-1 repression	-CAAGGAGCTCATCTCAGGCCA-	-GTCCAAC TCAAACAGCGCCA-
<i>CCNA2</i>	Aminopurvalanol-a	Signal-1 repression	-GCGCTGGCGGTACTGAAAGTC-	-TGCTTTCCAAGGAGGAACGGT-
<i>SORT1</i>	Aminopurvalanol-a	Signal-1 repression	-CCGGGTCCGGGACTTC-	-GCTATCTCCAACCCAGGACAAG-
<i>PLXNA3</i>	Aminopurvalanol-a	Signal-1 repression	-CTCCATGGCTGGTGTCAATTGTG-	-GCGTGTCTGAGGGGATCTTGA-
<i>LAMB2</i>	Aminopurvalanol-a	Signal-1 repression	-CAGACCCCTACAGCTCACGG-	-TGCCACGTACAACCAGCTCA-

Table 7.5: Gene, drug, condition and primer information for connectivity map qPCR. FWD, forward primer. REV, reverse primer.

7.2 Supplemental Figures

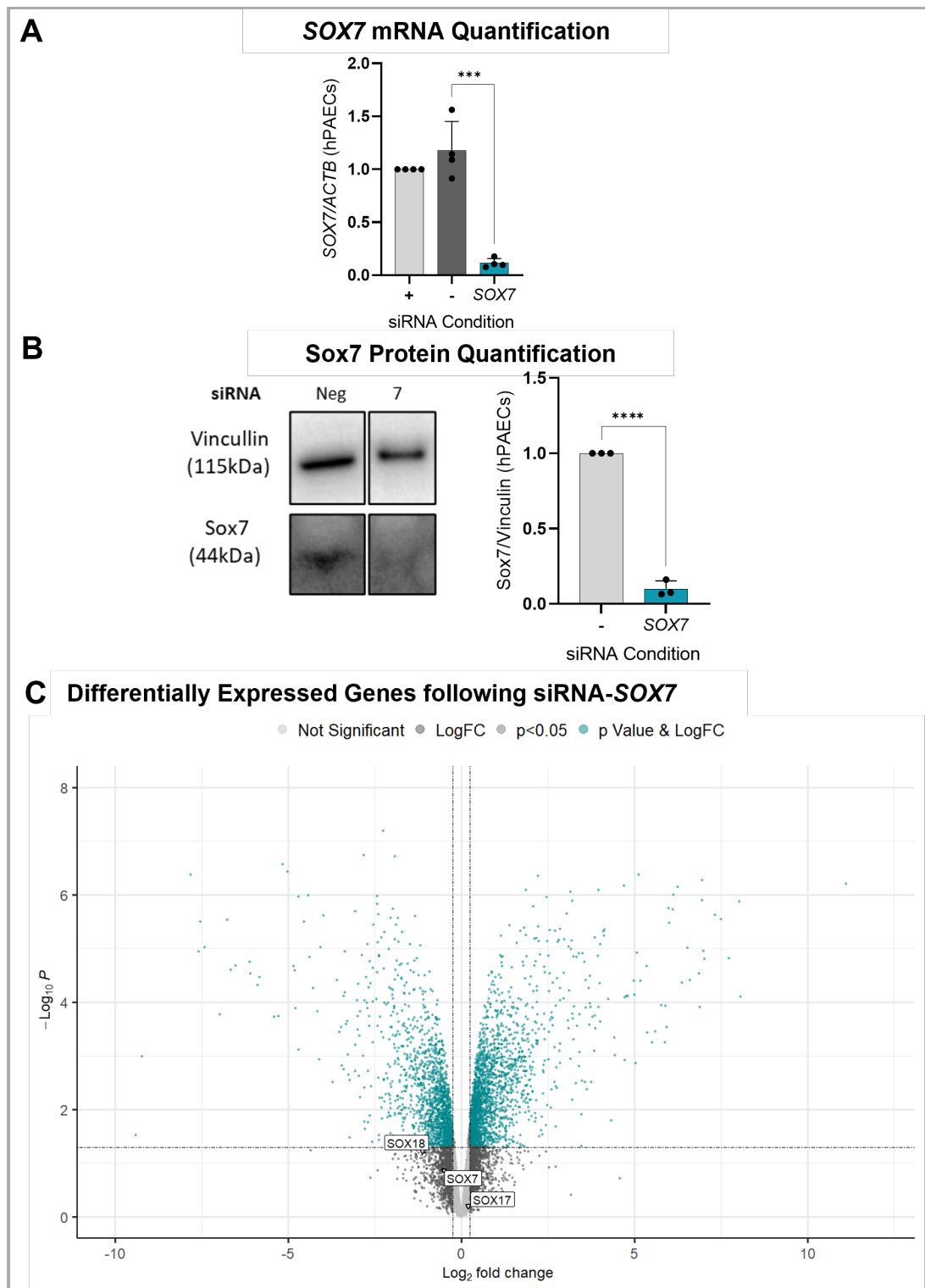


Figure 7.1

Knockdown of *SOX7* following siRNA. **A)** Relative mRNA quantification of *SOX7* by qPCR. **B)** Representative image of a western blot of Sox7 and vinculin protein in siRNA-*SOX7* and siRNA-negative samples. Quantification of western blotting of Sox7. All statistical tests are paired, one-way, student's t test. *** - $p < 0.0005$, **** - $p < 0.0001$. Points are averages of either three or four separate experiments ($n=3/4$). 12251 variables. All in hPAECs following *SOX17*-siRNA treatment for 48 hours.

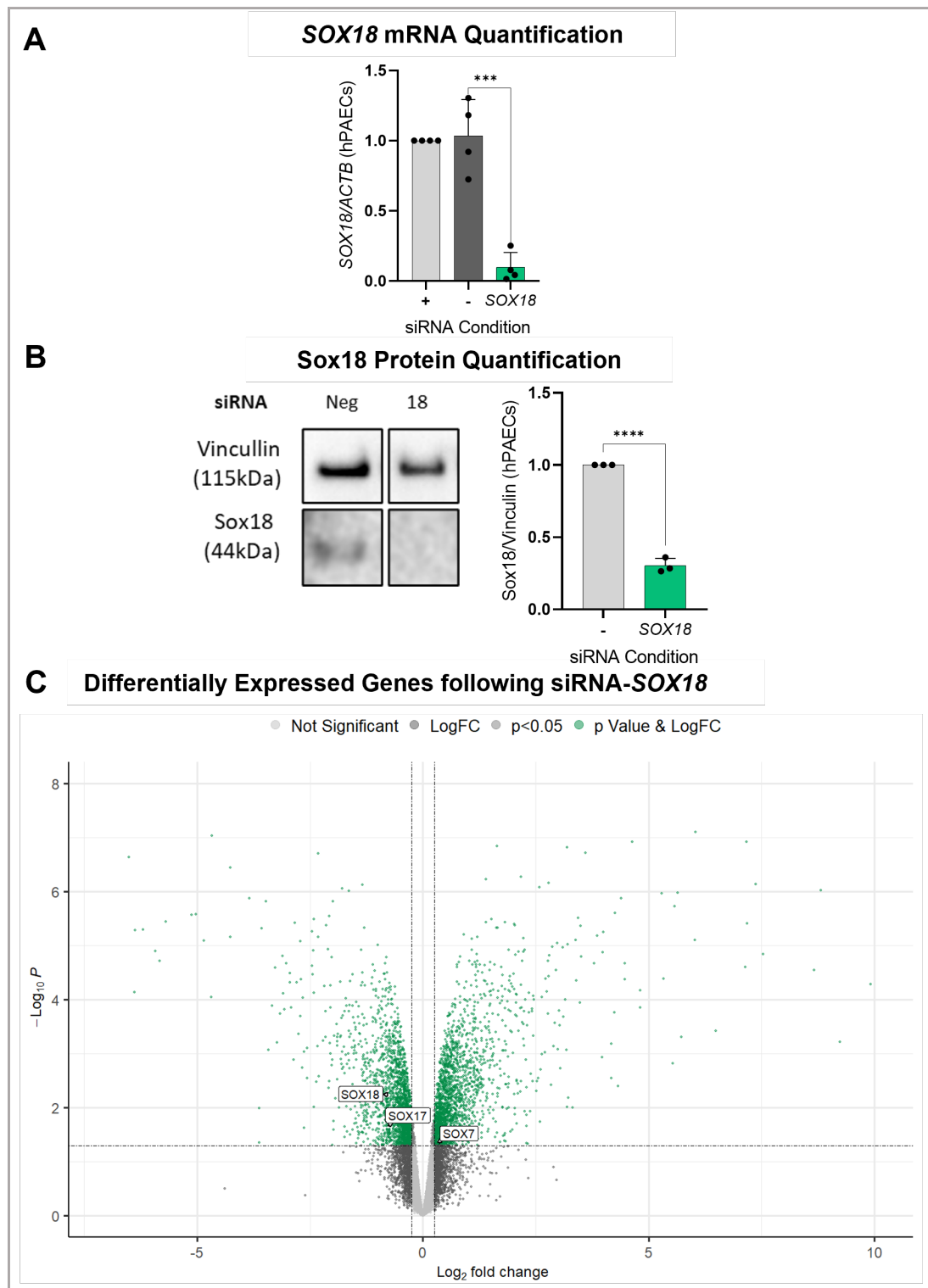


Figure 7.2

Knockdown of *SOX18* following siRNA. **A)** Relative mRNA quantification of *SOX18* by qPCR. **B)** Representative image of a western blot of Sox18 and vinculin protein in siRNA-*SOX18* and siRNA-negative samples. Quantification of western blotting of Sox18. All statistical tests are paired, one-way, student's t test. *** - $p < 0.0005$, **** - $p < 0.0001$. Points are averages of either three or four separate experiments ($n=3/4$). 12205 variables. All in hPAECs following *SOX17*-siRNA treatment for 48 hours.

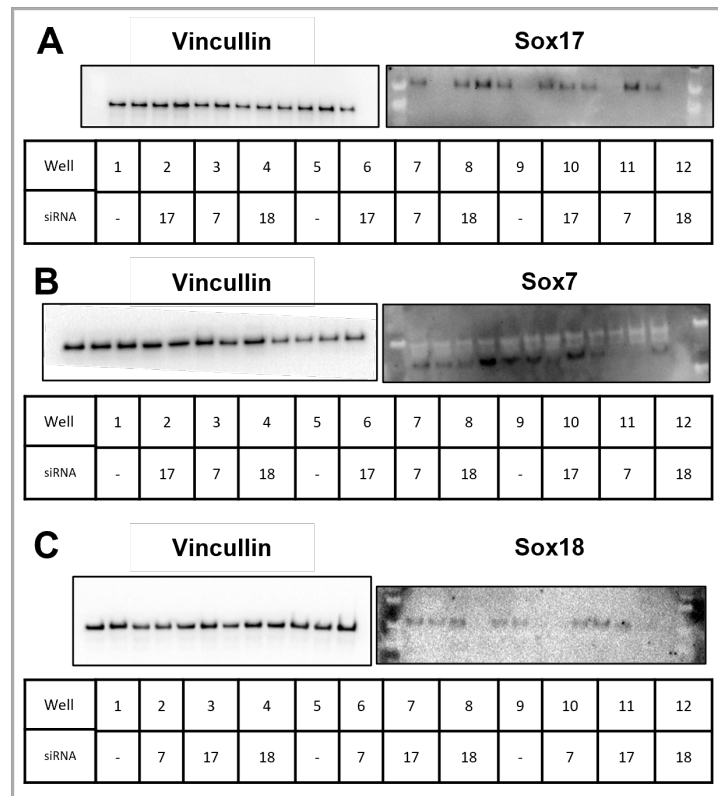


Figure 7.3

Full images of western blots shown in this thesis. A) Vincullin and Sox17 bands from western blot shown in figures 4.1 and 4.8A. B) Vincullin and Sox7 bands from western blot shown in figure 4.8B and figure 7.1. C) Vincullin and Sox18 bands from western blot shown in figure 4.8C and figure 7.2.