Transcriptional regulation of endothelial *KLF2* in vascular development



Lucija Fleisinger

Green Templeton College

University of Oxford

Supervisors: Dr Sarah De Val and Dr Alice Neal

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

Abstract

Transcriptional regulation of endothelial *KLF2* in vascular development

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY, TRINITY TERM 2022
LUCIJA FLEISINGER, GREEN TEMPLETON COLLEGE, UNIVERSITY OF OXFORD

The development of a functional vascular system is critically dependent on shear stress, a hemodynamic force exerted onto the vascular endothelium by flowing blood. Krüppel-like factor 2 (KLF2) is the central transcriptional regulator of the endothelial shear stress response, yet the mechanisms by which shear stress signalling controls *KLF2* expression remain controversial. Activation of *KLF2* by shear stress was previously hypothesised to occur through a MEK5-ERK5-MEF2 pathway directly interacting with the *KLF2* promoter, thereby inducing *KLF2* expression. However, *KLF2* is expressed in many different cell types in response to a variety of different upstream signals, suggesting cell type-specific *KLF2* expression might be regulated by yet undescribed distal enhancers. Further, MEF2 factors not only control differentiation of many non-endothelial cell types, but also regulate the expression of endothelial genes independent of shear stress signalling, suggesting additional transcriptional regulators must interact with MEF2 to induce context-specific *KLF2* expression. In this thesis, I identify and characterise two *KLF2* enhancers able to drive reporter gene expression specifically in endothelial cells (ECs) and endocardial cushions (ECCs) independently of the KLF2 promoter, which is by itself unable to drive any reporter expression in development. Pharmacological inhibition of cardiac contractions during development resulted in a reduction of reporter expression driven by both enhancers, suggesting they might be regulated by shear stress-responsive transcription factors (TFs). Validated phylogenetic foot-printing combined with mutational analysis of both *KLF2* enhancers implicated roles for MEF2 binding alongside a number of other TFs. Crucially, a previously unidentified MEF2-HBOX double motif was found to be required for the endothelial activity of both KLF2 enhancers, and was sufficient to identify a novel EC-specific enhancer for *KLF4*, a gene that compensates for *KLF2 in vivo*. This work demonstrates previously unappreciated complexity in the transcriptional regulation of *KLF2*, and provides an opportunity to better understand transcriptional dynamics at the locus of this disease-relevant gene.

Declaration

This thesis is submitted to the Department of Physiology, Anatomy, and Genetics, University of Oxford, in fulfilment of the requirements for the degree of Doctor of Philosophy. This thesis is entirely my own work and describes my own research, except where otherwise stated. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Any data that are not my own work are acknowledged and referenced appropriately. This work has not been submitted for another degree in this or any other University.

Lucija Fleisinger

Chapter 3

- Zebrafish embryo microinjections were carried out by Dr Svanhild Nornes
- Zebrafish endothelial ATAC-seq dataset was generated by Dr Kira Chouliaras
- *Klf2* RNA Scope® experiments were carried out in collaboration with Dr Helena Rodriguez-Caro
- F0 transient transgenic mouse embryos were generated by Cyagen Inc.
- Stable transgenic lines were generated by Dr Ben Davies (The Wellcome Centre for Human Genetics, University of Oxford)
- Embryos were sectioned by Dr Sarah De Val and Indrika Ratnayaka (Ludwig Institute of Cancer Research, University of Oxford)

Chapter 4

- F0 transient transgenic mouse embryos were generated by Cyagen Inc.
- Embryos were sectioned by Dr Sarah De Val and Indrika Ratnayaka (Ludwig Institute of Cancer Research, University of Oxford)
- GATA2 EMSA analysis was carried out by Dr Sarah De Val
- MEF2A/C ChIP-seq datasets were provided by Prof Mukesh Jain (Case Western Reserve University)
- MEF2-HBOX double motif analysis using MEFA/C ChIP-seq datasets was carried out by Dr Susann Bruche

Chapter 5

- Embryos were sectioned by Dr Sarah De Val and Indrika Ratnayaka (Ludwig Institute of Cancer Research, University of Oxford)

Word count: 47,632

Acknowledgements

To Sarah,

I feel immensely privileged and grateful for having been able to start my science journey under your supervision. Your fierce intelligence, brilliant ideas, and an unparalleled sense of humour make science seem like the best occupation in the world. Thank you for trusting me with a demanding project I was genuinely interested in, and for providing me with countless precious opportunities that helped me develop as a scientist. My only hope for other DPhil students is to find a supervisor like you.

~

To Alice,

I consider myself very fortunate for being taught how to do science, and how to be a scientist, by you. You taught me how to think, how to have ideas, how to question everything, and how to trust my own judgement. I will always look up to you, and I will always cherish the time we spent in the lab together.

~

To Sophie, Svanhild, Dorota, Susann, Helena, and Salwa,

You are all incredibly kind and knowledgeable people, and I had the absolute best time hanging out with you inside and outside the lab, talking about science and life. Each and every one of you has been a source of mentorship, friendship, motivation, and support, and I am so grateful to have worked with all of you.

~

To Prof Duncan Sparrow, Prof Nicola Smart, Prof Kim Dora, and Prof David Greaves,

Thank you for your generous mentorship, your dedication to students, and your considered advice in times of confusion. I truly appreciated your academic support throughout this journey.

~

To Joe,

By far the best part of my DPhil was meeting you! Your cheerful disposition and a sense of perspective have carried me through boring troubleshooting spells, stressful presentations, and lockdown blues. You have been the greatest source of joy, kindness, and support over the last few years. The balance you bring to my life makes me a better person, and a better scientist.

~

Moji družini,

Hvala za vašo neskončno ljubezen in podporo, ki me spremljata na vsakem koraku!

Abbreviations

A	Artery
AMPK	AMP-activated protein kinase
ATAC	Assay for transposase-accessible chromatin using sequencing
AV	Atrio-ventricular
AVC	Atrio-ventricular cushion
AVV	Atrio-ventricular valve
BAEC	Bovine aortic endothelial cells
BBB	Blood brain barrier
BMP	Bone morphogenetic protein
BP	Base pair
BRE	B recognition element
ChIP	Chromatin immunoprecipitation
CCM	Cerebral cavernous malformations
CREB	cAMP response element-binding protein
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
CXCL12	CXC chemokine ligand 12
CXCR4	CXC chemokine receptor type 4
DA	Dorsal aorta
DBD	DNA-binding domain
DCE	Downstream core element
DLL4	Delta-like protein 4
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
DPE	Downstream promoter element
DRE	DNA replication-related element
DTT	Dithiothreitol
E	Embryonic day
EC	Endothelial cell
ECC	Endocardial cushion
ECM	Extra-cellular matrix
EMCN	Endomucin
EMSA	Electrophoretic mobility shift assay
En	Endocardium
EMT	Endothelial to mesenchymal transition
EP300	E1A binding protein P300
ERG	Ets-related gene
ERK5	Extracellular signal regulated kinase 5
ETS	The E26 transformation-specific
ETV2	ETS variant transcription factor 2
FBS	Foetal bovine serum
FGF2	Fibroblast growth factor 2
FLI1	Friend leukaemia integration 1
FOX	Forkhead box

GFP	Green fluorescent protein
GPCR	G protein-coupled receptor
GTF	General transcription factor
GTP	Guanosine-5'-triphosphate
H	Histone
HAEC	Human aortic endothelial cells
HBOX	Homeobox
HDAC	Histone deacetylase
HEY1	Hes-related family BHLH transcription factor with YPRW motif 1
HF	High fidelity
HIF1	Hypoxia-inducible factor 1
HLX	H2.0-like homeobox transcription factor
HMVEC	Human microvascular endothelial cells
HRP	Horseradish peroxidase
HUVEC	Human umbilical vein endothelial cells
ICA	Internal carotid artery
IP	Immunoprecipitation
IUPAC	International union of pure and applied chemistry
KLF	Krüppel-like factor
KO	Knockout
LCC	Left common carotid
LEC	Lymphatic endothelial cells
MAPK	Mitogen-activated protein kinase
MEF2	Myocyte enhancer factor-2
MEK5	Mitogen/extracellular signal-regulated kinase-5
MEKK3	Mitogen-activated protein kinase kinase 3
MMP	Matrix metalloprotease
MT	Mutant
MTE	Motif ten element
NICD	Notch intracellular domain
NCX1	Na^{+}/Ca^{2+} exchange protein
NE	Nuclear envelope
NELF	Negative elongation factor
NET	Nuclear envelope transmembrane protein
NFAT	Nuclear factor of activated T-cells
NO	Nitric oxide
NOS	Nitric oxide synthase
NR2F2	Nuclear receptor subfamily 2 group F member 2
NTP	Nucleoside triphosphate
OCR	Open chromatin region
OT	Outflow tract
P	Post-natal day
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDGFB	Platelet-derived growth factor subunit B
PECAM	Platelet endothelial cell adhesion molecule

PFA	Paraformaldehyde
PIC	Pre-initiation complex
PT	Pulmonary trunk
PTM	Post-translational modification
RET	Rearranged during transcription
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPBJ	Recombination signal binding protein for immunoglobulin kappa J region
RT	Room temperature
RUNX1	Runt-related transcription factor 1
S1PR	Sphingosine-1-phosphate receptor
SAGA	Spt-Ada-Gcn5 acetyltransferase
SCL	Stem cell leukaemia
SILAC	Stable isotope labelling by amino acids in cell culture
SKMC	Human skeletal muscle cells
SMC	Structural maintenance of chromosomes
SNP	Single nucleotide polymorphism
SOX	SRY-related HMG-box
SV	Sinus venosus
SV40	Simian vacuolating virus 40
TAD	Topologically-associated domain
TAF	TBP-associated factor
TAL1	T-cell acute lymphocytic leukaemia protein 1
TBE	Tris-borate-EDTA
TBP	TATA-binding protein
TBS	Tris-buffered saline
TEAD	TEA domain
TEF	Thyrotrope embryonic factor
TF	Transcription factor
TFBS	Transcription factor binding site
TGF	Transforming growth factor
TNF	Tumour necrosis factor
TRP	Transient receptor potential
TSS	Transcription start site
V	Vein
VEGF	Vascular endothelial growth factor
VEGFR2	Vascular endothelial growth factor receptor
VSMC	Vascular smooth muscle cell
WT	Wild-type

Table of contents

Abstract II

Declaration III

Acknowledgements IV

Abbreviations V

Table of contents VIII

1 Introduction 1

1.1 Transcriptional regulation of gene expression 2

1.1.1 Structural organisation of the vertebrate genome 2

1.1.2 Gene transcription 5

1.1.2.1 Transcription initiation 5

1.1.2.2 Promoter pausing, elongation, and termination 7

1.1.2.3. Promoters 8

1.1.2.4 Enhancers 9

1.1.2.5 Enhancer-promoter interactions in nuclear space 11

1.1.2.6 Mechanisms of enhancer-promoter interactions 14

1.1.2.7 Transcription factors 15

1.2 Vascular development 18

1.2.1 Vasculogenesis 18

1.2.2 Angiogenesis 19

1.2.2.1 Sprouting angiogenesis 20

1.2.2.2 Intussusceptive angiogenesis 21

1.2.3 Vessel maturation and remodelling 21

1.2.4 Arterio-venous differentiation 22

1.2.5 Haemodynamic forces in vascular development 24

1.2.5.1 Shear stress 24

1.2.5.2 Physiological roles of shear stress in vascular development 25

1.2.5.3 Signalling pathways induced by shear stress 27

1.3 Transcriptional regulation downstream of shear stress signalling 29

1.3.1 Myocyte enhancer factor 2 family 29

1.3.2 Krüppel-like factor family 31

1.3.3 KLF2 regulation by shear stress 32

1.4 Thesis aims and hypotheses 36

2 Materials and methods 38
2.1 Transgenic mice 39
2.1.1 Generation of transient transgenic mouse embryos 39
2.1.2 Generation and breeding of stable transgenic mouse lines 39
2.1.3 Genotyping 40
2.1.4 Dofetilide studies 41
2.2 Transgenic zebrafish 42
2.2.1 Husbandry 42
2.2.2 Tol2 transposon transgenesis 42
2.2.3 Image acquisition and analysis 42
2.3 Histology 43
2.3.1 Fixation and staining of embryos 43
2.3.2 Paraffin embedding and sectioning 45
2.3.3 Klf2 RNAScope® on E11.5 sections 45
2.3.4 Immunostaining following RNAScope® 46
2.4 Enhancer prediction using the UCSC genome browser 48
2.5 Cloning of enhancer constructs 49
2.5.1 Gateway cloning 49
2.5.2 Cut-paste cloning 51
2.5.3 Mutagenesis 52
2.5.4 Enhancer sequences 52
2.5.4.1 Human KLF2 putative enhancer sequences 52
2.5.4.2 Zebrafish putative *klf2a* enhancers 54
2.5.4.3 Mutated KLF2-41 human sequences (tested in mice) 56
2.5.4.4 Mutated KLF2-66 human sequences (tested in mice) 58
2.5.4.5 Phylogenetically foot printed enhancer sequences 58
2.6 Electrophoretic mobility shift assay 59
2.6.1 Plasmids for generation of recombinant proteins 59
2.6.2 Probe labelling and gel shifts 59
3 Identification of endothelial-specific enhancers for *KLF2* 66
3.1 Rationale 67
3.1.1 Utilising enhancer analysis to understand regulatory pathways controlling endothelial responses to shear stress 67
3.1.2 Identification of endothelial-specific enhancers driving KLF2 expression in development 68
3.2 KLF2 locus analysis and enhancer prediction 69
3.2.1 An In silico screen of the human KLF2 locus to identify putative KLF2 enhancers 69

3.3 Analysis of putative KLF2 enhancers in transient transgenic zebrafish embryos 72
3.3.1 Putative KLF2 enhancers in transient transgenic zebrafish embryos 72
3.4 In silico screen of the zebrafish klf2a locus for identification of putative klf2a enhancers 74
3.4.1 Putative klf2a enhancers in transient transgenic zebrafish embryos 77
3.4.1.1 Rationale 77
3.4.1.2. Putative klf2a enhancers in transient transgenic zebrafish embryos 77
3.4.2 Detailed analysis of the klf2a-6 putative enhancer in transient transgenic zebrafish embryos 78
3.4.2.1 Rationale 78
3.4.2.2 Analysis of the klf2a-6 putative enhancer in transient transgenic zebrafish embryos. 79
3.5 Analysis of putative KLF2 enhancer:LacZ expression in transient transgenic mouse embryos .. 80
3.5.1 Rationale 80
3.5.2 Endogenous Klf2 mRNA expression in wild type mouse embryos 81
3.5.2.1 Rationale 81
3.5.2.2 Analysis of *Klf2* mRNA expression in the E12.5 WT mouse embryo 82
3.5.3 Putative KLF2 enhancers in transient transgenic mouse embryos 84
3.5.3.1 KLF2-41 enhancer:*LacZ* expression in transient transgenic mouse embryos 84
3.5.3.2 KLF2-66 enhancer:*LacZ* expression in transient transgenic mouse embryos 86
3.5.3.3 KLF2-8 enhancer*:LacZ* expression in transient transgenic mouse embryos 89
3.5.3.4 KLF2-34:*LacZ* and KLF2+23:*LacZ* expression in transient transgenic mouse embryos . 90
3.6 Analysis of KLF2-41:LacZ transgene expression in the developing mouse vasculature 91
3.6.1 Rationale 91
3.6.2 KLF2-41:LacZ stable transgenic line analysis in whole mount mouse embryos 92
3.6.3 Time-course analysis of KLF2-41:LacZ expression in the systemic vasculature during development 93
3.6.4 KLF2-41:LacZ transgene expression in the organ-specific vasculature 96
3.6.4.1 KLF2-41:*LacZ* transgene expression in the embryonic and post-natal heart 96
3.6.4.2 KLF2-41:*LacZ* transgene expression analysis in post-natal organs 98
3.7 Analysis of KLF2-66:LacZ transgene expression in the developing mouse vasculature 100
3.7.1 Analysis of KLF2-66:LacZ activity during mouse development 100
3.7.2 Time-course analysis of KLF2-66:LacZ transgene expression during development 100
3.7.3 KLF2-66:LacZ transgene expression in the organ-specific vasculature 104
3.7.3.1 KLF2-66:*LacZ* transgene expression in the embryonic and post-natal heart 104
3.7.3.2 KLF2-66:*LacZ* transgene expression analysis in postnatal tissues 105
3.8 Putative KLF2 promoter:LacZ expression in transient transgenic mouse embryos 106
3.8.1 Rationale 106

3.8.2 KLF2 promoter activity in transient transgenic mouse embryos 109
3.9 Discussion 110
3.9.1 KLF2 and klf2a putative enhancers 110
3.9.2 Human KLF2-66 and KLF2-41 enhancers drive reporter gene expression in the endothelium throughout development and post-natally 112
3.9.3 Human KLF2 promoter region is unable to drive LacZ expression at E12.5 116
3.9.4 Developmental dynamics of the KLF2 gene locus 116
3.10 Conclusion 117
4 Delineating direct transcriptional regulators of KLF2 and KLF4 enhancers 119
4.1 Rationale 120
4.2 Analysis of evolutionarily conserved KLF2-41 sequence in transgenic mouse embryos 120
4.2.1 KLF2-41 phylogenetic foot-printing 120
4.2.2 Klf2-6 enhancer:LacZ expression in transient transgenic mouse embryos 122
4.2.3 pKlf2-41 enhancer:LacZ expression in transient transgenic mouse embryos 125
4.3 Identification of transcription factors interacting with the KLF2-41 enhancer 129
4.3.1 ETS transcription factors 131
4.3.2 RBPJ transcription factor 135
4.3.3 NFAT factors 141
4.3.4 MEF2 transcription factors 143
4.4 Identification of TFs interacting with the KLF2-66 enhancer 147
4.4.1 Validated phylogenic footprinting of the KLF2-66 enhancer 147
4.4.2 ETS transcription factors 150
4.4.3 RBPJ transcription factors 151
4.4.4 GATA factors bind the KLF-66 enhancer in vitro 153
4.4.5 MEF2 transcription factors 155
4.5 The MEF2-HBOX motif is an essential element conserved in both KLF2 enhancers 156
4.5.1 A requirement for the HBOX site in KLF2-41:LacZ transgene expression 157
4.5.2 A requirement for the HBOX site in KLF2-66:LacZ transgene expression 159
4.6 The MEF2-HBOX is also found within the KLF2 promoter 160
4.7 The MEF2-HBOX motif is not commonly found in the genome but is shared by an upstream KLF4 endothelial enhancer 163
4.8 Identification and validation of the KLF4-148 enhancer 164
4.8.1 The KLF4-148 element is an endothelial-specific enhancer 165
4.8.2 Putative transcriptional regulators of the KLF4-148 enhancer 167
4.8.3 ETS factors 169
4.8.4 GATA factors 170

4.8.5 NFAT factors 171
4.8.6 MEF2 factors 172
4.9 What binds the HBOX motif? 173
4.9.1 Is the MEF2-HBOX site bound by a MEF2 dimer? 174
4.9.2 FOX factors interact with the MEF2, but not the HBOX motif 177
4.10 Discussion 181
4.10.1 KLF2/4 regulatory regions are transcriptionally regulated by ETS and MEF2 factors 181
4.10.1.1 ETS factors 181
4.10.1.2 MEF2 factors 182
4.10.2 RBPJ, GATA, and NFAT factors and regulation of EC subtype-specific expression 184
4.10.2.1 RBPJ 184
4.10.2.2 KLF2-66 enhancer contains a functional GATA binding motif 185
4.10.2.3 KLF2-41 and KLF4-148 enhancer contain functional NFAT binding motifs 187
4.10.3 A previously unidentified HBOX motif found in KLF2/4 regulatory regions is required for KLF2/4 enhancer:reporter transgene expression 187
4.10.4 Identifying the transcriptional regulators of the MEF2-HBOX double motif 188
4.10.5 A model of transcriptional regulation at the KLF2 gene locus 189
4.11 Conclusion 190
5 KLF2 enhancer activity in response to shear stress 191
5.1 Rationale 192
5.1.1 Aim 195
5.2 KLF2 enhancer:LacZ transgene expression in the dorsal aorta and adult organs 195
5.2.1 KLF2-41:LacZ expression in the post-natal dorsal aorta 195
5.2.2 KLF2-66:LacZ expression in the post-natal dorsal aorta 197
5.2.3 KLF2-41:LacZ transgene expression in the adult dorsal aorta 198
5.2.4 KLF2-66:LacZ transgene expression in the adult dorsal aorta 200
5.2.5 KLF2-41:LacZ expression in adult organs 200
5.2.6 KLF2-66:LacZ expression in adult organs 202
5.3 KLF2-41:LacZ and KLF2-66:LacZ expression following pharmacological manipulation of blood flow 203
5.3.1 Rationale 203
5.3.2 KLF2-41:LacZ expression in Dofetilide-treated embryos 206
5.3.3 KLF2-66:LacZ expression in Dofetilide-treated embryos 212
5.4 Discussion 224
5.4.1 KLF2-66:LacZ and KLF2-41:LacZ transgene expression corresponds to shear stress magnitudes in the post-natal dorsal aorta 224

5.4.2 KLF2-66:LacZ and KLF2-41:LacZ transgene expression is absent from the adult dorsal aorta 225
5.4.3 KLF2-66:LacZ and KLF2-41:LacZ transgene expression is reduced following Dofetilide treatment 228
5.5 Conclusion 229
6 Conclusion 230
6.1 Overall conclusions 231
6.2 Limitations 232
6.2.1 Statistical analysis 232
6.2.2 Blinding 233
6.2.3 Zebrafish transgenic studies 234
6.2.4 Mouse transgenic studies 234
6.2.5 Image annotation 235
6.3 Future directions 235
7 References 237

1 Introduction

1.1 Transcriptional regulation of gene expression

During embryonic development of a complex multicellular organism, a single cell can give rise to an astonishing variety of different cell types. Whilst most cells contain largely identical copies of their genetic material, different cell types perform highly specific functional roles within their resident tissue. To achieve this, cells utilise only a certain set of genes, in a highly dynamic yet tightly regulated process of gene expression which defines cell identity and function. Consequently, understanding both the mechanisms of gene expression genome-wide, and the precise transcriptional regulation of single genes central to specific cellular processes, remain key endeavours of modern biology.

While genes are the primary functional units of the coding DNA, their expression relies on interactions with their cognate gene regulatory elements, which generally reside in the non-coding genome. Enhancers are the most prominent class of gene regulatory elements due to their unique ability to recruit tissue-specific transcription factors (TFs) in response to developmental and environmental cues. Enhancers bound by TFs will interact with the promoters of their cognate genes to achieve precise spatio-temporal control of tissue-specific gene expression. In this section, I will describe how eukaryotic genomes are organised, how transcription is initiated, and how enhancers interact with TFs and their cognate promoters to regulate gene expression.

1.1.1 Structural organisation of the vertebrate genome

The configuration of the eukaryotic genome is controlled by the structural properties of DNA molecules and their associated proteins. A single cell's worth of DNA molecules that together span two meters are able to fit into the confines of a six micrometre cell nucleus by way of compaction, a process that maintains both gene accessibility and structural flexibility necessary to mediate highly dynamic processes such as gene transcription. In eukaryotic cells, DNA is packed into chromatin, a molecular conglomerate composed of DNA wrapped around histone proteins. The basic unit of

chromatin is a nucleosome, an octamer of histone molecules wrapped into 146 base pairs (bp) of DNA 1.65 times[1–3]. Within the nucleosome, histones are organised in a cylindrical shape composed of two heterodimers of each histone H3 and H4 flanked by two heterodimers of histones H2A and H2B[4–6]. The nucleosomes are connected by linker DNA bound to the nucleosome core by a class of linker histones (H1)[7,8] that play important roles in higher order chromatin organisation[9].

While the nucleosome is the basic structural unit of chromatin, histone-bound DNA is largely inaccessible to TF binding, and therefore transcriptionally silent. Local depletion in nucleosomes defines transcriptionally active genomic regions, referred to as open chromatin regions (OCR). In line with this, chromatin exists in two higher order structures primarily defined by nucleosome density[10,11]. Heterochromatin is a regularly structured, condensed form of chromatin where DNA is inaccessible to TFs, with the exception of pioneer factors, a class of TFs able to productively interact with histone-bound DNA[12,13]. Euchromatin is a looser chromatin structure conducive to TF binding and is characteristic of transcriptionally active regions. However, the extent of nucleosome packaging is subject to significant local variation and is dynamically regulated even within euchromatin[14,15].

In addition to nucleosome density, histone post-translational modifications (PTMs) are also a crucial determinant of transcriptional activity[16]. Each PTM serves as a tag for a specific chromatin state, which in turn controls both access to DNA and recruitment of effector proteins that can bind to PTMs and render a genomic region active or inactive. Each core histone has an amino- and a carboxy- histone tail, both amenable to PTMs including methylation, acetylation, phosphorylation, and others. PTMs are mediated by enzymes (histone acetyl transferases and histone deacetylases, histone methyl and demethyl transferases, histone kinases, histone phosphatases, and others). Both chromatin accessibility and histone marks have been mapped across whole genomes of many different species and cell types using approaches including histone PTM chromatin immunoprecipitation sequencing (ChIP-seq), assay for transposase-accessible chromatin sequencing (ATAC-seq), and DNAse I

hypersensitivity assays[17,18]. These publicly available datasets are invaluable tools for predicting cell type-specific gene regulatory regions, which generally lie within OCRs and are flanked by histones with specific PTMs.

Chromatin molecules are further condensed into chromosomes. Upon entering interphase, each chromosome is tightly condensed into a characteristic rod-like shape. In non-dividing cells, where chromosomes exist in a de-condensed state, each chromosome still occupies a specific position within the nucleus referred to as a chromosome territory[19,20]. Many chromosome territories are evolutionarily conserved, and repositioning of chromosomes from their natural territories and the resulting dysregulation of gene expression has been associated with disease states[21,22]. Chromatin territories segregate primarily depending on their gene density and transcriptional activity. They are nominally divided into territories A and B, with territory A corresponding to euchromatic, transcriptionally active regions, and territory B corresponding to heterochromatic, transcriptionally inactive regions. Gene-poor and inactive chromosomes are more likely to localise to the nuclear periphery, whereas gene-dense, transcriptionally active chromosomes tend to be located towards the centre of the nucleus[23]. Relocating human chromosomes to change their native position within the nucleus was found to affect the expression of many, though not all, genes[24].

The arrangement of the nuclear compartments plays an important role in genome organisation. The nuclear envelope (NE) is studded with nuclear pores, which are dense with nuclear envelope transmembrane proteins (NETs). NETs and nuclear pore complexes are both able to bind chromatin-binding proteins, including histones, and in this way tether often inactive and gene-poor chromosome regions to maintain their position within the nucleus[25]. Nuclear lamin proteins have been shown to anchor heterochromatin, and lamin A (*Lmna*) knockout (KO) mice lack peripheral heterochromatin localisation[26]. Larger scale genome organisation, such as chromosome segregation with respect to transcriptional activity, is therefore further reinforced by the nuclear architecture.

1.1.2 Gene transcription

According to the central dogma of molecular biology, the entirety of an organism's genetic information is encoded within the cell's DNA. The process of transcription generates messenger RNA (mRNA) molecules that carry specific pieces of information encoded by the DNA to the cytoplasm, where this information is translated into functional protein. In addition to the production of mRNA, non-coding RNA molecules, such as long non-coding RNAs (lncRNAs), micro RNAs (mRNAs), and enhancer (eRNAs) may be generated during transcription, although the biological function of these molecules remains an active area of investigation[27,28]. The process of transcription from DNA to mRNA can be divided into four steps: initiation, proximal promoter pausing, elongation, and termination.

1.1.2.1 Transcription initiation

In eukaryotes, protein-coding genes are transcribed by RNA polymerase II (Pol II). While purified Pol II is sufficient for RNA synthesis in cell culture systems, it is by itself unable to recognise a promoter sequence[29]. Aside from specifically interacting with a gene promoter, Pol II must also be recruited in response to specific cues, thereby initiating context-dependent gene transcription at short notice. A number of protein complexes in addition to Pol II are crucial both for timely Pol II recruitment and integration of transcriptional information from distal regulatory regions[30,31].

The recruitment of Pol II to the promoter relies on initiation factors, TFs able to recognise and interact with core promoter elements to form a pre-initiation complex (PIC). This complex contains Pol II, six general transcription factors (GTFs), the Mediator complex, and transcriptional co-regulators (**Figure 1.1a**). The TATA-binding protein (TBP), a component of the general transcription factor IID (TFIID), was found to be required for PIC assembly[32,33]. TBP interacts with one of the core promoter elements, the TATA box, and thereby induces a clamp-like conformational change that immobilises the DNA and enables the rest of GTFs to assemble in a step-wise manner (**Figure 1.1b**)[34,35]. While PIC assembly has

been initially described at a TATA box-containing promoter, it is now clear that not all promoters contain a TATA box[36–38]. In other promoter types, PIC assembly is thought to be mediated by TBP-associated factor (TAF) subunits of the TFIID complex interacting with other promoter elements, for example TAF1 and TAF2 bind the Initiator (Inr) element[39], whereas TAF6 and TAF9 bind the downstream promoter element (DPE)[40,41]. These are further described in Section 1.1.2.3. Different elements of PIC are evolutionarily conserved from human to yeast, suggesting that initiation of transcription by Pol II is an ancient process universal to eukaryotic cells, despite variability in core promoter elements[42].

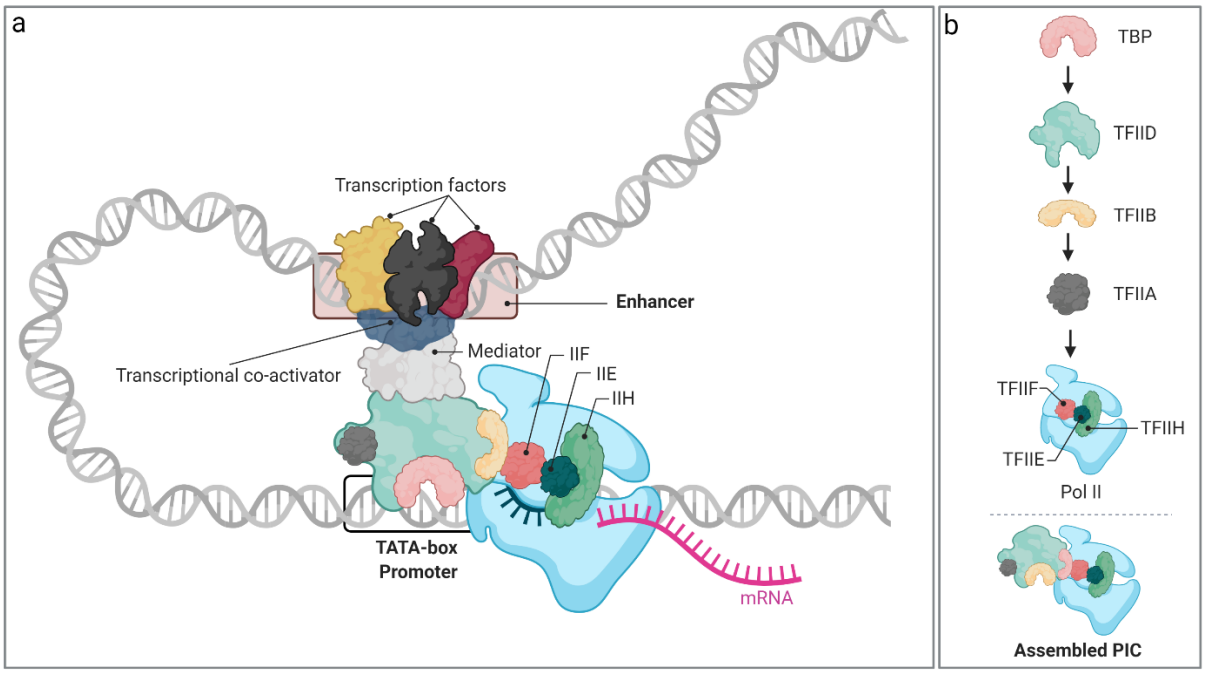


Figure 1.1: Molecular mechanism of transcription initiation. (a) Enhancers recruit sequence-specific TFs, which interact with transcriptional co-regulators and the Mediator complex that allow the TF-bound enhancer to interact with the transcriptional machinery recruited to the cognate core promoter. **(b)** GTFs are recruited to the TATA-box motif in the promoter region, starting with TBP able to recognise the TATA-box sequence, followed by TFIID, TFIIB, and TFIIA GTFs. TFIID interacts with the mediator to stabilise the enhancer-promoter interaction. TFIIF-bound Pol II is recruited to the complex, followed by TFIIE and TFIIH binding. The above figure was adapted from "Eukaryotic gene regulation-transcription initiation", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates

1.1.2.2 Promoter pausing, elongation, and termination

Promoter pausing is a rate-limiting process during which Pol II, already recruited to the core promoter, only transcribes 20-120 nucleotides before its arrest downstream of the transcription start site (TSS)[43]. The pausing process is thought to poise Pol II for transcription while it awaits the release signal, allowing it to initiate elongation at short notice having already achieved time and energy-consuming steps of recruitment and initiation. Promoter-pausing is especially common at the promoters of developmentally regulated and stimulus-controlled genes where rapid activation of transcription is required, such as heat shock protein (*hsp*) genes in fly[44] or immune-responsive genes in mammals[45]. Pol II pausing at promoters is mediated by CG-rich DNA motifs, such as the fly DPE, also known as the "pause button"[46]. In addition to core promoter motifs and their associated TFs, negative elongation factor (NELF) and DRB-sensitivity-inducing factor (DSIF) are thought to contribute to paused Pol II stabilisation[47]. It has been shown that NELF deletion results in re-positioning of nucleosomes into nucleosome-free regions at the promoter, reducing gene expression, suggesting pausing functions to maintain gene accessibility[48]. This is consistent with the fact that enhancer-promoter looping, an indicator of gene activity, has been associated with paused Pol II in the fly genome[49]. Pol II release is the definitive start of transcription elongation and is mediated by positive transcription elongation factor-b (P-TEFb) complex, the inhibition of which not only prevents Pol II release, but also abolishes virtually all transcription[50]. In addition to the P-TEFb complex, TFs and co-regulators also have roles both in stimulating Pol II pausing and in Pol II pause release, suggesting different regulatory factors, on balance, regulate Pol II pause entry and exit[51].

The rate of transcription elongation increases throughout the transcription process and is variable between different genes[52]. The factors that influence the elongation rate include the presence of histone marks, histone chaperons, and nucleosome remodellers[53]. The nature of the DNA sequence can stall the process, for example CG-rich sequences take longer to transcribe due to their complex DNA topology[54]. Transcription termination occurs after Pol II passes through the poly(A) signal[55].

1.1.2.3. Promoters

Promoters are gene regulatory sequences that define the location and direction of transcription initiation by serving as docks for PIC assembly[31]. The term "core promoter" describes the minimal sequence required to initiate transcription and is thought to extend 50-100bp either upstream or downstream of the TSS. In addition to the TSS, core promoters also contain GTF binding motifs required for transcription initiation. The upstream promoter-proximal region is often dense with additional TF binding motifs that integrate signals from distal enhancers[56]. While core promoters are sufficient for transcription initiation, their basal activity is generally low, as measured using reporter assays in cell systems[57] and *in vivo*[58].

The assembly of PIC for transcription initiation requires specific core promoter motifs[59]. The first and best characterised core promoter motif is the TATA box, commonly located 25-50bp upstream of a single dominant TSS with a consensus binding motif of $TATA^{A}{}_{T}A^{A}{}_{T}$[60]. This consensus motif is bound by TBP (see 1.1.2.1). While the TATA box motif is evolutionarily conserved from yeast to human, the consensus TATA box motif is only present in between 5%-15% of all human core promoters[61].

Since the discovery and characterisation of the TATA box, other core promoter motifs have been described. The Inr motif is present in 46% of all human core promoters and is located precisely at the TSS. The Inr, often accompanied by the DPE, is commonly present in core promoters that do not contain a TATA box[61]. Several TFIID subunits are able to interact with the DPE, and strict Inr-DPE spacing rules apply to enable TFIID binding[38]. Additional core promoter motifs include TFIIB recognition elements (BREs) and downstream core elements (DCEs) in humans, and the motif ten element (MTE) and DRE-binding factor (DREF) in fly[62]. While many core promoter motifs are conserved, there is a significant proportion that differ between species, both in consensus motif and function[63]. For example, Inr is present in all metazoan genomes, yet its consensus binding motif is species-specific[64].

In addition to the presence of specific binding motifs, the overall nucleotide composition of core promoters also relates to their function. The feature that remains most predictive of vertebrate promoters is the overlap with regions of high CG content and high CpG dinucleotide density, called CpG islands (CGIs), found in 50-60% of core promoters[65]. Consistent with their prevalence, CGI promoters correspond to OCRs and are frequently marked by H3K4me3 characteristic of promoter regions. Promoters with CGIs have been associated with the vast majority of housekeeping genes, and about half of tissue-specific genes in mouse and human genomes[66,67].

An important remaining question is how different promoter properties, such as differences in core promoter motifs and nucleotide composition, relate to promoter function. It has been shown that CGI promoters generally lack core promoter motifs, consistent with the idea of dispersed transcription initiation from multiple TSSs during housekeeping gene activation. The presence of core promoter elements is thought to underlie focused transcription initiation from a single TSS characteristic of specific enhancer-promoter interactions in developmental gene regulation[68]. A notable exception is the α-globin promoter which contains both CGIs and a TATA box[69]. In non-mammalian genomes that do not contain CGIs, different core promoter motifs have been associated with the regulation of functionally distinct genes. For example, fly core promoter TATA box, DPE, and Inr motifs were all associated with regulating developmental genes, whereas the DREF motif was associated with the housekeeping transcriptional programme[70]. While often transcriptionally silent by themselves, promoter sequences encode important transcriptional information that commonly requires an activating transcriptional switch: an enhancer.

1.1.2.4 Enhancers

Enhancers are non-coding DNA sequences dense with DNA binding motifs that recruit sequence-specific TFs, then interact with gene promoters to regulate gene expression. Active enhancers lie within OCRs where they are accessible to TF binding and are flanked by nucleosomes with histone

modifications characteristic of enhancers: H3K4me1 (enhancer-specific) and H3K27ac (also marking active promoters)[71]. While promoters are fixed in their position directly upstream of the TSS, enhancers can be anywhere from one to tens of thousands of kilobases (kb) upstream or downstream with respect to the promoter[72,73], and sometimes reside within inter-genic regions of their cognate or an entirely different gene[74,75]. Consistent with variability in position, enhancers are able to amplify transcription regardless of their relative distance and orientation[57]. The estimated number of enhancers in the human genome is counted in millions, vastly outnumbering the 21,000 protein-coding genes[17,76], although only a small percentage of putative enhancer regions have been fully validated[77].

Enhancers can exist in a silenced, poised, or an active state. Silenced enhancers are inaccessible to TFs and reside within genomic regions devoid of enhancer-specific histone modifications. However, they can be poised by pioneer factors able to direct local nucleosome depletion, thereby increasing enhancer accessibility to TF binding[78]. Although positioned within an OCR, H3K4me1-marked poised enhancers are unable to regulate gene expression, and only become active by acquiring H3K27Ac signatures which enables Pol II recruitment[79]. Enhancer poising prevents premature gene activation and is thought to be especially important for spatio-temporally precise gene expression. Enhancers are active when they are regulating their cognate gene expression, as indicated by the presence of H3K27Ac histone modification[80] which positively correlates with the expression of the most proximal gene[81–83].

To regulate transcription, enhancers need to recruit co-regulators in addition to TFs. While not required for basal transcription from promoters, co-regulators are known to modify the chromatin landscape to enable enhancer-promoter interactions. They are widely expressed and can be recruited to chromatin by a range of TFs, but they do not by themselves exhibit sequence-specific binding. Examples of co-activator proteins include histone acetyltransferases (for example p300, SAGA

complex), histone methyltransferases (MLL3/4, CARM1), chromatin remodelling factors (Brg1) and the Mediator complex which facilitates enhancer-promoter interactions[84,85]. ChIP-seq mapping of transcriptional co-activator binding genome-wide can be used to predict active enhancers[86].

Developmental enhancers are classically viewed as modular and additive, meaning multiple enhancers can regulate a single gene, each controlling specific aspects of endogenous gene expression[87]. Consistent with this notion, true enhancers in reporter assays typically replicate specific aspects of the total endogenous gene expression pattern (**Figure 1.2a**)[88–90]. A single gene can also be regulated by multiple enhancers that show overlapping patterns of reporter expression, and often act in a redundant fashion. The primary enhancer, often the first to be discovered but not necessarily the first to evolve, is accompanied by "shadow" enhancers thought to ensure robustness of gene expression against mutations or environmental variability throughout evolution (**Figure 1.2b**)[91–93].

1.1.2.5 Enhancer-promoter interactions in nuclear space

Topologically associating domains (TADs) are defined as genomic regions that more frequently interact within themselves than with each other. Each TAD is an architectural feature that envelops enhancers and promoters in a local regulatory hub that enables productive enhancer-promoter interactions[94]. An average TAD is 185kb in size, but TAD domains up to 1 MB in size have been reported[95,96]. TADs are composed of multiple sub-TADs, which encapsulate regulatory regions that interact with each other under specific conditions, for instance a context-specific enhancer-promoter interaction[97].

TAD boundaries are enriched in TSSs and correspond to the distribution of activating and repressing histone marks[98]. During each cell division, the chromosomes decondense and TADs are lost, then faithfully re-assembled, suggesting that TADs are architecturally rigid, consistent with the fact that TAD boundaries are often evolutionarily conserved between cell types and across species[99,100]. In

contrast, sub-TADs vary considerably between different cell types[101], consistent with the nature of specific enhancer-promoter interactions known to be both transient and heavily context-dependent.

Interactions within TADs are mediated by two types of protein complexes: cohesin and CCCTC binding factor (CTCF). Both are enriched at TAD boundaries, but even more enriched within TADs[94]. CTCF binding is required for TAD formation, as CTCF depletion leads to dysregulation of transcription and disorganised contacts of genomic regions within the CTCF-depleted region[102–104]. Cohesin depletion resulted in decrease in interaction between regions within a single TAD, but did not affect TAD location or organisation, consistent with the idea that cohesin regulates smaller scale sub-TAD interactions[105].

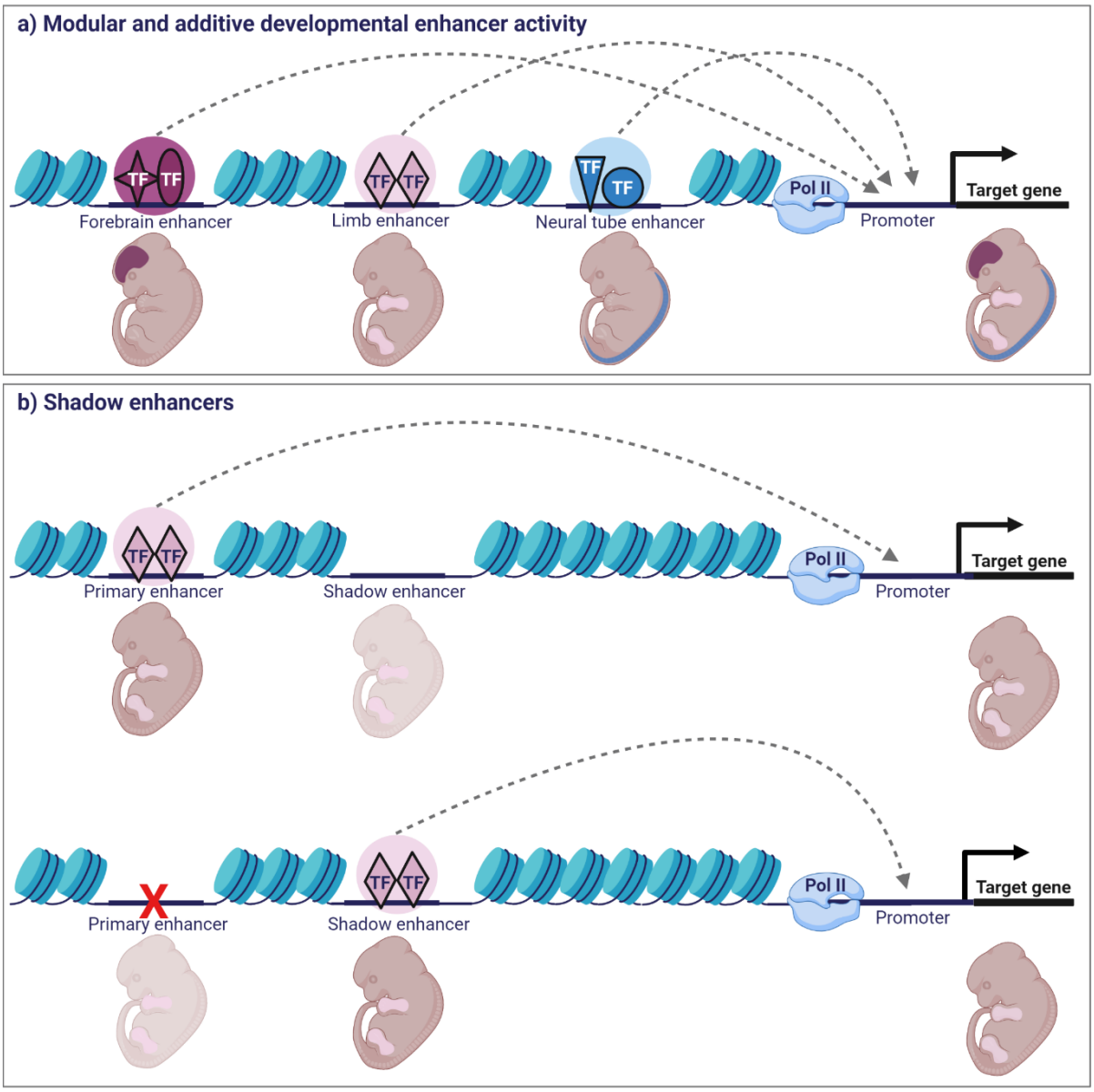


Figure 1.2: A classical view of developmental enhancer action. (a) Representation of modular and additive enhancer activity. The example gene is expressed in the forebrain, the limbs, and the neural tube. Distinct enhancer elements control the expression of the gene in each of these compartments through recruitment of specific TFs. The sum of the expression patterns controlled by distinct enhancer sequences equals to the overall expression pattern of the gene. **(b)** Representation of the shadow enhancer concept. The primary enhancer regulates gene expression by recruiting specific TFs. If the primary enhancer is unable to regulate gene expression, for example due to a mutation within the primary enhancer sequence, the shadow enhancer can recruit the same TFs to drive gene expression. In an artificial transgenic system, both the primary and shadow enhancers might be able to direct reporter expression to overlapping compartments within the embryo. This figure was created using BioRender.com (2022).

1.1.2.6 Mechanisms of enhancer-promoter interactions

In order for gene expression to occur, an enhancer and a promoter need to interact with each other. While segregation of the genome into TADs and sub-TADs considerably reduces the otherwise vast genomic space that an enhancer must sample in order to find its cognate promoter, the enhancer can still reside hundreds of kbs away. We now have models explaining the mechanism underlying enhancer-promoter interactions, including the tracking model, the looping model, and the loop extrusion model (**Figure 1.3**)[106,107]. Here, I describe key ideas from all these non-mutually-exclusive models. The loop extrusion model is an updated description of the mechanistic basis describing cohesin and CTCF-mediated TAD and subTAD formation. Cohesin is a ring-like protein able to simultaneously envelop two DNA strands, and at the same time move across the chromatin, tracking the DNA until it encounters a CTCF-bound boundary element on both strands. The idea of tracking is central to the tracking model, which suggests that a TF and Pol II-bound enhancer would uni-directionally track the chromatin until it encountered a promoter, resulting in an enhancer-promoter interaction. However, DNA tracking is unlikely to support highly dynamic gene regulation processes, especially in settings where multiple nearby genes are simultaneously activated by distinct enhancer-promoter interactions. The looping model suggests that in response to TF and co-regulator binding, an enhancer would loop onto its cognate promoter to induce gene expression. However, it does not explain how specificity of enhancer-promoter interactions is achieved to avoid erroneous gene activation, especially during long-range enhancer-promoter interactions[108]. In the context of the loop extrusion model, cohesin mediates bi-directional chromatin tracking until it encounters CTCF binding sites on both DNA strands, thereby determining the dimensions of the loop equivalent to a sub-TAD. Cohesin stalling at CTCF binding sites results in loop extrusion, creating a domain within which promoters and enhancers are brought into close proximity in an area of high TF and co-regulator density, facilitating their interactions[109,110].

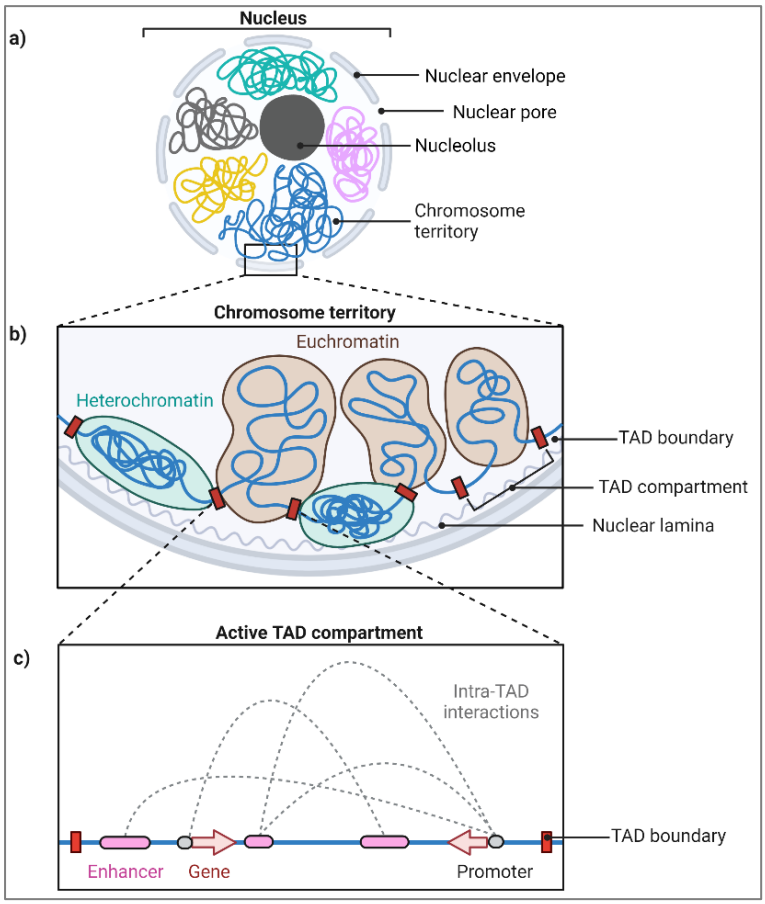


Figure 1.3: Structural organisation of the nucleus determines enhancer-promoter interaction dynamics. (a) Within the eukaryotic nucleus, each chromosome occupies a specific territory. **(b)** Within the chromosome territory, transcriptionally inactive heterochromatin is tethered to the nuclear lamina at the nuclear periphery. In contrast, transcriptionally active euchromatin localises towards the centre of the nucleus. TAD compartments are defined by TAD boundaries enriched for cohesin and CTCF proteins. **(c)** Within an active TAD compartment, enhancers interact with their cognate promoters. A specific enhancer-promoter interaction is referred to as a sub-TAD. This figure was adapted from "Chromosome organization in the nucleus: TADs", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates

1.1.2.7 Transcription factors

Transcription factors are DNA-binding proteins that regulate gene transcription. The human genome encodes between 1400 and 1600 TFs, and each cell is estimated to express ~200 TFs at any point in time[111]. Each TF contains a sequence-specific DNA binding domain (DBD), which binds to TF binding motif sequences generally between 6-12bp in length, and activation or repression domains responsible for interactions with transcriptional co-factors required for transcription initiation at the core promoter[112].

Sequence-specific TFs bind enhancers in order to regulate gene expression in a spatio-temporally specific manner. In addition to regulating other genes in response to a variety of cues, TFs are often able to regulate their own expression to create a feedback loop which activates or represses the signalling pathways downstream[112]. TF binding to regulatory elements that regulate their own

expression, their regulation of other TF-encoding and non-TF genes creates a TF network. While a TF network might not be active in its entirety at all times, understanding TF networks in specific cell types helps us understand global transcriptional programmes controlling cell identity and function[113].

Enhancer activation generally requires binding of multiple different TFs, including lineage-specifying factors, and factors that bind in response to environmental cues. In addition to binding motif identity, TF binding is also regulated by relative motif order, orientation, and spacing. These parameters are thought to regulate the number, type, and affinity of TF binding motifs for TF binding, and define combinatorial binding capacity and stability of an enhancer[87,114,115]. The TF binding "rules" derived from observations of these parameters are together referred to as enhancer grammar, or motif syntax. It has been suggested that even enhancers with low affinity binding sites can induce robust tissue-specific gene expression when their syntax is optimally organised[116].

Motif syntax forms the basis of models describing the mechanisms of enhancer activation by TF binding. There are currently three main models describing this: the enhanceosome model, the billboard model, and the TF collective model. The enhanceosome model suggests that TF binding sites (TFBSs) must be precisely arranged in a rigid linear fashion to allow TF binding. This type of model rests on the assumption of deep sequence conservation that precisely determines orientation and spacing of different TFBSs[87]. The most frequently cited example of an enhanceosome model of enhancer binding is the 55bp interferon-β (*IFNB1*) enhancer, for which DNA binding domains of eight bound TFs have been constructed[117]. Mutating any of the eight TF binding motifs altered the binding potential of all eight factors, implying direct cooperativity between them and confirming the strict motif syntax. While a few other examples of enhanceosome enhancers exist[118,119], such rigid TFBS organisation is thought to be relatively rare[120]. In contrast to the enhanceosome model, the billboard model suggests there are no constraints on TFBS arrangement, so long as all of the TFs are able to interact with their TFBS[121]. Consistent with the billboard model, both binding site permutation and *in vivo* enhancer

studies show there is often flexibility in TF binding order[122–125]. Finally, the TF collective model suggest that both TF binding to DNA, as well as additional TF-TF interactions are important for enhancer activation. This model implies a degree of flexibility in motif organisation and even motif structure and puts emphasis on the presence and interaction between a defined group of TFs[126,127]. Mechanisms of TF binding remain an active area of research. While analysis of specific enhancers might suggest requirements either for extreme motif syntax, or the lack of any, the syntax probably depends on the enhancer role and function[120]. Recently there have been efforts to employ deep learning to understand motif syntax rules in fly S2 cells. The emerging models were not only able to predict active enhancers but could be used to design *de novo* enhancers with desired properties[128]. Such models will likely be cell type-specific and differ significantly between different species but might be useful to predict cell type-specific active enhancers and understand their mode of TF binding in the future.

TFs interact with DNA, each other, co-regulators, and the surrounding nucleosomes to stabilise the DNA-TF interaction and productively engage with the promoter. TFs are in constant competition with histones for access to DNA and can actively modulate nucleosomes to prevent their repositioning[129]. Direct cooperativity between two neighbouring TFs reinforces the occupancy of the binding motifs for both TFs. Further, binding of additional TFs to a gene regulatory region counteracts nucleosome repositioning and leads to a net increase in TF binding also referred to as assisted loading or collaborative competition[130,131].

1.2 Vascular development

In a mammalian embryo, the cardiovascular system is the first to develop, consistent with its primary purpose of supplying oxygen and nutrients to all other organ systems. Before the onset of cardiac contractions, endothelial cells (EC) arise from mesoderm-derived progenitors called angioblasts and assemble into a primitive vascular plexus. This *de novo* assembly of simple blood vessels from single angioblasts is known as vasculogenesis. After initiation of blood flow following the first cardiac contractions, the primitive vascular plexus expands into a mature vascular network, a complex hierarchy of stereotypically organised arteries, veins, and capillaries. This process of new blood vessel formation from an existing endothelial network is known as angiogenesis[132,133]. Within the primitive vascular plexus, some ECs already assume their endothelial subtype identity, for example by expressing arterial or venous specifying genes and kickstarting the process of arterio-venous differentiation[134]. While genetic specification of different endothelial subtypes is of great importance in endothelial function, additional environmental cues, such as hemodynamic forces exerted onto ECs by flowing blood, play a major role in regulating subtype-specific endothelial identity and function[135]. In this part of the introduction, I will first describe different stages of vascular development, then focus on the role of hemodynamic forces in these processes, and the molecular mechanisms that underlie them.

1.2.1 Vasculogenesis

In zebrafish embryos, mesodermal progenitors emerge either from the anterior lateral plate mesoderm to form the head vasculature, or from the posterior lateral plate mesoderm to form a precursor vascular cord at the midline that gives rise to the dorsal aorta and the posterior cardinal vein[136]. In mouse embryos, vasculogenesis begins in the yolk sac at embryonic day 7.5 (E7.5), where the first ECs emerge from within blood islands: structures of hematopoietic cells surrounded by ECs. Blood islands form a ring around the visceral yolk sac which coalesces to form the vitelline vessels[134].

An independent population of angioblasts forms the embryonic allantois which gives rise to the umbilical vessels[137]. Almost simultaneously, the heart development is initiated with the formation of a linear heart tube[138], and the dorsal aortae and the cardinal veins develop from within the aortic primordia, a layer of mesoderm-derived cells positioned lateral to the midline. Multiple mesodermal sources contribute angioblasts to different parts of the embryo as it develops[139].

The majority of angioblasts differentiate through progressive restriction of the mesodermal lineage and transition into the endothelial lineage, whereas the rest emerge from haemangioblasts: bipotent precursors of both EC and haematopoietic cells[134,140]. Conversely, a population of differentiated ECs referred to as haemogenic endothelium can give rise to haematopoietic cells[141]. Angioblast differentiation is controlled by several tissue-derived factors, including fibroblast growth factor 2 (FGF2), bone morphogenetic protein 4 (BMP4), and vascular endothelial growth factor (VEGF), that play roles in mesoderm specification, EC differentiation and EC survival[142]. The emergence of the primitive vascular plexus is transcriptionally controlled by E26 transformation specific (ETS) family member ETS variant transcription factor 2 (ETV2), which has been shown to interact with Forkhead box protein C2 (FOXC2) to bind the FOX:ETS binding motif found in a number of validated early endothelial enhancers[143]. Additional transcriptional regulators, including T-cell acute lymphocytic leukaemia protein 1 (TAL1), GATA factors, other ETS factors, and FOX factors, have also been implicated in regulation of vasculogenesis[136].

1.2.2 Angiogenesis

Angiogenesis is initiated when a pro-angiogenic stimulus such as hypoxia or nutrient deprivation, induces endothelial activation, sprout formation, lumenisation, and maturation of the newly formed vessel through blood perfusion[144]. Most major vessels in an embryo form by vasculogenesis and subsequently expand and mature through angiogenesis. However, certain vascular networks, such as intersegmental vessels and organ-specific vascular beds including the retinal, coronary, and brain

vascular networks, form entirely by angiogenesis[145–147]. Vascular growth in adult organisms, for instance during wound healing and tumour vascularisation, also proceeds by angiogenesis[148,149].

1.2.2.1 Sprouting angiogenesis

Early blood vessels are supported by a basement membrane composed of protein-rich extracellular matrix (ECM) which prevents ECs from leaving their positions within the vessel wall. Angiogenic sprouting is initiated in response to a VEGF-A signal, which activates matrix metalloproteases (MMPs), such as MMP1, that locally dissolve the ECM. In order to diffuse the perivascular cells surrounding the forming sprout, MMPs also release sequestered pro-angiogenic growth factors, such as angipoietin-2 (ANG2), which mediate perivascular cell detachment[150]. Compromising the integrity of the ECM and the perivascular cell layer allows the ECs to invade the extra-vascular space.

Within the angiogenic sprout, ECs can assume two distinct phenotypes: a tip cell or a stalk cell identity. The single tip cell projects filopodia at the forefront of the angiogenic sprout, whereas stalk cells support the tip cell by proliferating, elongating, and lumenising the forming sprout. In response to VEGF-A, tip cells initiate migration, whereas stalk cells proliferate. The same VEGF-A signal interacts with VEGFR2 receptor on tip cells to stimulate filopodia formation and further reinforce tip cell selection. Simultaneously, VEGF-A induces the expression of *Dll4*, a Notch pathway ligand, in tip cells. DLL4 secreted by tip cells activates the Notch pathway in surrounding stalk cells, downregulating the VEGFR2 pathway and thereby suppressing tip cell behaviour in stalk cells[148]. The VEGF-Dll4/Notch signalling pathway therefore regulates tip cell selection, as the first cell to produce high levels of DLL4 in response to VEGF-A will assume the tip cell phenotype and suppress the same phenotype arising in surrounding cells [151]. However, the tip cells at the leading edge of the sprout are replaced in a matter of hours in cell culture[152], and it has been shown that *Dll4* expression is dynamically regulated through the interaction of TEL/CtBP repressor complex at the *Dll4* promoter[153]. *Dll4* expression during sprouting angiogenesis is additionally mediated by MEF2 factors through an intronic *Dll4* enhancer[154].

1.2.2.2 Intussusceptive angiogenesis

Intussusception is a process in which existing blood vessels are separated, thereby forming two new vessels. In contrast to sprouting angiogenesis, where new sprouts extend outwards, during intussusception, the endothelial processes protrude inwards into the vessel lumen from opposing vascular walls. Once the two processes meet, they form a transluminal pillar, to which perivascular cells and fibroblasts are recruited, leading to ECM deposition. As the pillar strengthens, the original vessel is split into two new vessels. Intussusception is not well understood but is thought to be especially relevant in the later stages of development, for example for vessel branching, and can be induced in response to an increase in blood flow[155].

1.2.3 Vessel maturation and remodelling

Vessel remodelling and maturation include lumen formation, perivascular cell recruitment, development of the ECM, and acquisition of organ-specific cell properties such as EC fenestrations[148]. A crucial step in vascular network maturation is perfusion by blood which strengthens the integrity of the existing connections and leads to regression of non-perfused vessels[156]. For perfusion to occur, newly formed vessels need to be lumenised. In intersomitic vessels, a lumen is formed by cell hollowing, characterised by coalescence of intracellular vacuoles between neighbouring EC[157]. In large diameter vessels, a lumen is formed by cord hollowing, during which EC are thought to change their morphology and relocate in order to form a hollow vessel[158]. This process is thought to be mediated by GTPase RHO, the genetic loss of which results in no lumen formation[159].

In addition to perfusion, the deposition of ECM, recruitment of perivascular cells, and upregulation of EC-EC junction formation also contribute to the process of vascular maturation. Pericytes are perivascular cells that establish contacts with EC of capillaries and immature blood vessels, whereas mature arteries and veins are enveloped by vascular smooth muscle cells (VSMCs)[134]. Vessel maturation is mediated by the Transforming Growth Factor-β (TGF-β) signalling pathway, the loss of

which leads to poor vessel integrity due to loss of perivascular cell recruitment[160,161], and Platelet Derived Growth Factor Subunit B (PDGFB) pathway, the loss of which leads to pericyte deficiency and a number of associated vascular defects[162]. In addition, Sphingosine-1-phosphate receptor (S1PR) signalling, ANG-1 signalling, EphrinB2, and Notch signalling pathways are all required for vascular maturation[160].

Vascular remodelling is the process during which the network branches and prunes to achieve its final stabilised state. This dynamic process is thought to be mediated by ECM proteases and their inhibitors which regulate the local release of pro and anti-angiogenic growth factors also sequestered within the ECM. It is thought that local branching pattern of the vascular network is controlled by spatio-temporally dynamic concentration gradients of these factors, which also influence the proliferation and apoptosis of both EC and perivascular cells[148,149].

1.2.4 Arterio-venous differentiation

Blood vessels can be broadly divided into arteries and veins. The primary functional difference between the two is that arteries carry oxygenated blood away from the heart towards other organs, whereas veins bring deoxygenated blood back to the heart. Structurally, both arteries and veins have an inner endothelial lining. However, arteries contain thicker layers of tunica media, adventitia, and VSMCs, allowing for transport of blood through small diameters under high pressure. Veins have thinner extra-endothelial layers and a larger diameter able to accommodate greater blood volume to compensate for the lack in pressure. In order to successfully return the blood to the heart under low pressure, veins contain valves to counter retrograde flow. Arterio-venous differentiation is the developmental process during which differences between arteries and veins are first established[163,164].

Early arterial and venous EC are thought originate from different angioblast subpopulations, suggesting that arterial and venous fates can be genetically pre-determined from early in

vasculogenesis[136,142,165,166]. In the zebrafish embryo, it was initially thought that both the dorsal aorta and the cardinal vein arise from the common cord thought to initially expresses both *EfnB2* and *EphB4* [167,168]. However, it is now believed that zebrafish venous and arterial EC arise from different angioblast populations and differentiate under the control of VEGF and Hedgehog (Hh) morphogens[169]. In the mouse embryo, the cardinal vein is formed through sprouting of *EphB4*$^{+}$ ECs from the dorsal aorta which forms first[170]. Expression of *EphrinB2* and *EphB4*, and Neuropilin-1 (*Nrp1*) and Neuropilin-2 (*Nrp2*) in arterial and venous ECs, respectively, has been observed prior to onset of cardiac contractions[165,171], suggesting the establishment of arterial and venous identities is initially independent of blood flow.

The Notch signalling pathway is required for the establishment and maintenance of arterial identity[172,173]. There are four transmembrane Notch receptors (Notch 1-4) and five Notch ligands (Dll1, Dll3, Dll4, Jagged1, and Jagged2). Upon ligand receptor interaction, Notch intracellular domain (NCID) of the receptor is cleaved, resulting in its translocation to the nucleus, where it is thought to activate arterial-specific target genes, including *Hes* and *Hey* gene families which act as transcriptional repressors. The repressive function of *Hey1* and *Hey2* appears to be required for arterial identity, as mice with combined *Hey1/2* deletion fail to remodel their vasculature and show reduced expression of arterial markers including *EfnB2*[174,175]. However, recent work has also implicated Notch signalling in regulation of cell cycle, suggesting that this, rather than direct transcriptional gene regulation, is the primary role for Notch signalling in arterial differentiation[176]. Additional transcriptional regulators of arterial identity include SOXF and FOX TF families, both of which are required for correct arterial identity[74,177–180]. Venous EC markers include Nr2f2, EphB4, and Nrp-2. In the absence of Nr2f2, arterial markers are expressed in venous EC, suggesting Nr2f2 both actively induces venous fate and represses arterial fate[181]. Venous identity upstream of *EphB4* and *Nr2f2* genes is regulated by the SMAD TF family downstream of BMP- Alk3 receptor signalling[73].

1.2.5 Haemodynamic forces in vascular development

Due to their position on the inner lining of all blood vessels, ECs are constantly exposed to blood and the haemodynamic forces it exerts onto the vessel wall. ECs are not only able to sense the force exerted by blood, but convert its components, including the rate, pulsatility, and pattern of blood flow, into intracellular signals that ultimately induce gene expression programs that control EC phenotypes.

1.2.5.1 Shear stress

There are two major types of physical forces exerted onto vascular endothelium by flowing blood. The first is circumferential stretch, the force of blood pressure stretching the vessel wall radially and outwards. The resulting change in vessel diameter is a physiological adaptation to the pulsatility of flowing blood that ensures the match in blood supply to tissue demand on a short-term basis. The second is fluid shear stress, a frictional force exerted onto ECs parallel to the direction of flowing blood that affects almost every aspect of EC function[182]. Shear stress is required for vascular development, as mouse models in which cardiac contractions, and consequently blood flow, have been inhibited do not survive, and their blood vessels never remodel into a stereotypical vascular network[183].

Shear stress is defined as the frictional force of blood per unit area of an endothelial layer within the blood vessel. It is most commonly measured in dynes (dyn) per cm^2, is directly proportional to flow velocity and blood viscosity, and inversely proportional to vessel diameter, as indicated in the equation below.

$$\sigma(shear) = \frac{4\eta Q}{\pi R^3}$$

η...blood viscosity
Q...flow rate
R...vessel radius

Assuming constant blood flow velocity and blood viscosity within the vascular system, differences between shear stress levels in different blood vessels are largely attributable to differences in vessel

diameters[184]. The variability in shear stress levels exists not only between different vascular beds, but between different blood vessels of the same vascular bed, and even within structurally different areas of large blood vessels such as the dorsal aorta. In adult small diameter arteries shear stress can be as high as 50 dyn/cm^2 but is only about 10 dyn/cm^2 in the straight part of the large diameter aorta. Similarly shear stress can be as high as 20 dyn/cm^2 in small diameter veins, and as low as 1 dyn/cm^2 in the large diameter cardinal vein[185,186]. While the shear stress magnitudes experienced by embryonic blood vessels are less well defined, they have been measured to be 0 dyn/cm^2 and 5.5 dyn/cm^2 in E8.5 and E10.5 mouse embryos, respectively[187].

In addition to variation in shear stress rates, distinct patterns of shear stress are generated depending on vessel geometry. ECs located in straight blood vessels are exposed to laminar shear, a uniform pattern of blood flow that is also pulsatile due to the nature of cardiac contractions. On the other side, ECs located in vascular bifurcations or curvatures, such as the inner curvature of the aortic arch, experience disturbed shear stress patterns characterised by their non-uniform, oscillatory nature. Different shear stress patterns lead to activation of different mechanotransduction pathways and give rise to vastly different EC gene expression profiles[182,188,189].

1.2.5.2 Physiological roles of shear stress in vascular development

When the chick embryo heart was removed prior to the onset of the first cardiac contraction the chick yolk sac vasculature did not remodel[190]. Later studies in mouse "no heartbeat" models, where genes encoding proteins required for embryonic cardiac contractions, including the *Mlca*$^{-/-}$ and *Ncx1*$^{-/-}$, were deleted, led to similar results. *Mlc2a*$^{-/-}$ embryos developed an abnormal, developmentally delayed circulation resembling the primitive vascular plexus, and died by E9.5. Further studies based on sequestering erythroblasts to reduce blood viscosity showed a similar phenotype. This was rescued upon restoring blood viscosity, demonstrating it is the force of shear stress and not other properties

of blood, such as nutrient or oxygenation status, that are required for correct vascular development[191]. The lack of blood flow was additionally shown to affect the remodelling of the cardiac outflow tract (OT), and cardiac valve development, which are both areas exposed to high shear stress magnitudes[192,193].

Blood flow inhibition is embryonically lethal because it is required for many important developmental processes, including arterio-venous differentiation, angiogenesis, regression, and vascular remodelling[156,183,194]. High shear stress has been shown to maintain arterial identity, as the arteriovenous differentiation failed in *Ncx*$^{-/-}$ embryos, which was demonstrated by the abnormal connections between the dorsal aorta and cardinal vein, and both the lack of arterial markers and ectopic expression of venous markers in the dorsal aorta of these embryos[171,195]. Similar evidence comes from chick embryos, where high shear stress was found to induce the expression of arterial specific genes, whereas venous and lymphatic genes were downregulated[196]. In chick and mouse embryos, high pulsatile shear stress was shown to induce Gap Junction Protein Alpha 5 (*Gja5*) expression, with *Gja5* ablation leading to reduced arteriogenesis in peripheral occlusive disease models, suggesting *Gja5* might play a role in flow-regulated arteriogenesis[197]. This is consistent with evidence from adult ECs which suggest that exposure to high shear stress imparts a quiescent phenotype through Notch signalling and regulation of cell cycle in arterial ECs[198]. Arterio-venous specification in the brain vasculature of zebrafish embryos is also regulated by haemodynamic forces through modulation of chemokine signalling[199]. In addition to the rate of shear stress, pattern of shear stress also contributes to arterio-venous differentiation, with pulsatile flow found to activate arterial genes, and steady flow activating venous-specific genes in adult ECs[197]. This is consistent with the elastic aorta experiencing high pulsatility from cardiac contractions, whereas veins experience less pulsatility due to the lack of a substantial VSMC layer. While blood flow is dispensable for the initial specification of EC subtypes and for the formation of the dorsal aorta and the cardinal vein, it is later required for maintenance of arterio-venous specification.

Shear stress also affects angiogenesis. Studies manipulating blood viscosity to alter shear stress levels have found that high shear stress levels inhibit angiogenesis both in quail and mouse embryos[200], as well as in human ECs *in vitro*[201], although this effect only applies to embryonic veins, whereas sprouting from arteries remained unaffected. In stabilised, perfused vessels, low levels of shear stress are thought to act as a stimulus for angiogenesis, similar to hypoxia or nutrient deprivation. However, in a vascular network where shear stress levels are low overall, or exhibit an oscillatory pattern, new sprouts are predisposed to branch regression[202]. Shear stress therefore controls vascular remodelling in a highly context-dependent manner.

1.2.5.3 Signalling pathways induced by shear stress

The force of shear stress is sensed by ECs through highly specialised cellular structures or single molecules called mechanosensors. Mechanosensors sense different aspects of shear stress-encoded information, including pattern, rate, and pulsatility, then pass this information onto adaptor molecules able to transduce it into activation of intracellular signalling cascades. This ultimately leads to changes in gene expression, which in turn influence EC behaviour.

Mechanosensors are located on either the basal, junctional, or apical EC surface. Apical mechanosensors include the primary cilia, particularly prominent in mechanotransduction during development[203–205], the glycocalyx, a glycoproteins-rich structure that interacts with components of blood; G- Protein-Coupled Receptors (GPCRs), such as S1P1 which undergo conformational changes in response to shear stress; and ion channels, for example Piezo1, which respond to differences in EC membrane potential in response to shear stress[186]. Junctional mechano-sensors include platelet endothelial cell adhesion molecule-1 (PECAM-1) that interacts with VE-cadherin and VEGFR2 to form a junctional mechanosensory complex required for shear sensing in all ECs[206]. Integrins are positioned at the basal EC surface and bind ECM ligands including fibronectin and collagen, activating signalling

cascades through focal adhesions sites and cytoskeletal proteins. The context-specific hierarchy of mechanosensors and the cross-talk between them is an active area of investigation.

The application of shear stress onto cultured ECs was shown to stimulate the release of K^+ and Ca^{2+} ions, prostacyclin, and nitric oxide (NO), production of reactive oxygen species (ROS), and activation of many kinases, such as MAP kinases, and GTPases such as RhoA and Rac which go on to activate TFs, leading to gene expression changes[186,207].

1.3 Transcriptional regulation downstream of shear stress signalling

The ultimate consequence of mechanotransduction is an altered gene expression profile[208,209]. Application of shear stress to cultured ECs has been estimated to alter the expression of up to 3000 genes, with distinct shear stress rates and distinct shear stress patterns inducing distinct groups of genes in different EC lines[210,211]. Despite the evidently dynamic transcriptional regulation of the EC subtype and context-specific shear stress response, our understanding of specific gene regulatory networks that orchestrate it remains limited. While many studies have been carried out in cultured ECs in response to shear stress application, our knowledge in particular of genes that lie downstream of mechanotransduction pathways in developmental and adult *in vivo* settings, and how these genes are transcriptionally regulated, is lacking. In this part of the introduction, I focus on the two best characterized groups of TFs known to regulate shear stress signalling in ECs: Myocyte enhancer factor 2 (MEF2) and Krüppel-like factor (KLF) families. I then focus on *KLF2*, the gene encoding KLF2 which has been extensively studied in the context of transcriptional regulation by shear stress.

1.3.1 Myocyte enhancer factor 2 family

Myocyte enhancer factor 2 (MEF2) belongs to the MADS family of TFs expressed in smooth muscle, skeletal muscle, cardiac muscle, neural crest, neurons, lymphocytes, and ECs. All factors from the family contain a conserved MADS-box domain and an adjacent MEF2 domain, which controls DNA-binding, MEF2 dimerisation, and its interactions with transcriptional co-factors. The C-terminal region functions as a transcriptional activation domain and is divergent among MEF2 family members. They all bind the $^{C}_{T}TA^{T}_{A}T_{A}T_{A}TA^{A}_{G}$ consensus motif as homo- or heterodimers, often interacting with other TFs to activate target gene expression in different settings. While MEF2 factors function as transcriptional activators, their repressive function is mediated by class IIa histone deacetylases (HDACs) able to associate with the MADS domain to form a repressive complex on MEF2-regulated genes[212].

The vertebrate genome encodes four MEF2 factors: MEF2A, B, C, and D, from which MEF2A, MEF2C, and MEF2D are expressed in ECs, as well as skeletal and cardiac muscle. Analysis of an EC-specific Mef2C enhancer expressed as early as E8.5 revealed a role for vascular ETS factors in regulating *Mef2c* activity[213]. Global *Mef2C* null mice die from cardiac looping defects and show additional vascular defects[214,215]. However, EC-specific *Mef2C* deletion showed no endothelial phenotype, suggesting the vascular defects in global KO embryos are secondary to heart defects[216]. However, it has been suggested that the lack of vascular defects in these mice may be due to compensation by other MEF2 factors in the absence of *Mef2C*. Global *Mef2A* null embryos are viable but die post-natally due to cardiac defects[217], while *Mef2D* loss-of-function embryos are viable but exhibit an impaired response to cardiac stress signals[218]. In each case, potential compensation from other MEF2 factors has been proposed. This is supported by studies in skeletal muscle, where deletion of MEF2A, C and D results in a total loss of regenerative capacity that is rescued by any single *Mef2* allele[219].

Within the EC, MEF2 factors are thought to maintain vascular integrity and cell survival during development[212]. The mitogen-activated protein (MAP) kinase ERK5 is required for *Mef2* expression, and is itself required for vascular integrity, as *Erk5* null embryos die between E9.5 and E10 and exhibit defects in vascular maturation. However, cardiac looping defects were also observed in these embryos, which could be the underlying cause of compromised vascular integrity phenotype[220]. Further, HDAC7 is known to repress MEF2 signalling and is itself required for the maintenance of vascular integrity[221,222]. MEF2 factors are required for the regulation of sprouting angiogenesis directly downstream of VEGF-A, as they regulate multiple genes expressed during sprouting angiogenesis, including *Dll4* and *Hlx* through sprout-specific enhancers[154]. The deletion of *Mef2A*, *C*, and *D* specifically in ECs of adult mice led to rapid lethality with systemic inflammation, haemorrhage, thrombocytopenia, and leukocytosis, and abrogated expression of two most prominently flow-responsive genes known to regulate these processes: *Klf2* and *Klf4*[223].

1.3.2 Krüppel-like factor family

Krüppel-like factors (KLFs), named after the orthologue founding factor deleterious for fly development[224], are a 17-member zinc finger TF family that share homology in their DNA-binding zinc-finger region but are divergent in their non-DNA binding activation and repression domains[225–227]. Three members of the family, KLF2, KLF4, and KLF6, are robustly expressed in ECs[228]. The best studied member of the family, *Klf2*, is expressed in the developing ECs and ECCs as early as E9.5, more robustly in arterial than venous ECs[229]. The zebrafish *Klf2* orthologue *klf2a* is expressed at 36 hours post fertilisation (hpf) in the dorsal aorta and intersegmental vessels[230].

Klf2 is required for vascular development in mouse embryos. Systemic deletion of *Klf2* was found to induce severe intra-embryonic and intra-amniotic haemorrhaging associated with a reduction in tunica media thickness, followed by embryonic death between E12.5 and E14.5 likely due to defects in vessel wall stabilisation[231]. Targeted *Tek-Cre* mediated deletion of *Klf2* in endothelial and hematopoietic lineages during embryonic development recapitulated the global KO phenotype, resulting in lethality between E12.5 and E14.5 due to vascular maturation defects that lead to embryonic heart failure[229]. Further investigations into this phenotype on a different genetic background revealed a further role for *Klf2* in EMT during cardiac development[232]. Interestingly, compound global deletion of *Klf2* and *Klf4* led to embryonic death by E10.5, earlier than the single global deletion of either of the two genes, suggesting there is a degree of genetic compensation between the two factors when one is absent. Embryonic lethality at this early point is thought to be due to loss of vascular integrity, as endothelial breaks are observed in the primary head vein, and bulbous EC phenotype. Expression of *Nos3*, *Vegfr2*, and *Ocln* genes were dramatically reduced in E9.5 compound KO embryos compared to WT and *KLF2* KO embryos[233]. Genetic compensation between KLF2 and KLF4 is also apparent in the adult setting, where a single copy of either gene is sufficient for survival[234]. Similar to MEF2 factors, it is likely that compensation from other KLF factors plays a significant role in limiting phenotypes in the single null mice. For example, in contrast to systemic *Klf2*

deletion, systemic *Klf4* KO embryos were born at Mendelian ratios, but died shortly after birth due to loss of skin barrier function, with no vascular abnormalities reported in these mice[235]. Understanding both the overlapping and divergent functions of KLF2 and KLF4 in developing and adult ECs remains an active area of investigation.

The KLFs are thought to negatively regulate EC proliferation, and are preferentially expressed in growth-arrested states, like the quiescent endothelium[182]. In response to laminar shear stress in mature ECs, both KLF2 and KLF4 have been shown to inhibit interleukin-1β (IL-1β) and tumour necrosis factor-α (TNF-α)-induced EC activation and inflammation through inhibition of the Nuclear Factor kappa-light-chain-enhancer of activated B cells (NF-kB) signalling pathway[236–238], and maintain an anti-thrombotic endothelial state[239,240]. Both KLF2 and KLF4 have been shown to downregulate Endothelin-1 (*Edn1*) expression and induce *Nos3* expression, leading to general endothelial health[241,242]. However, challenges in generating reliable antibodies against KLF2 have limited identification and understanding of direct KLF2 targets.

1.3.3 KLF2 regulation by shear stress

Multiple studies of *KLF2* induction by shear stress in EC culture culminated in a generally accepted model of *KLF2* activation. In this model, laminar shear stress interacts with mechanosensors on the EC surface to induce MEK5, a kinase phosphorylates ERK5, another kinase, which phosphorylates MEF2, leading to its translocation into the EC nucleus. MEF2 then interacts directly with the *KLF2* promoter to induce *KLF2* expression (**Figure 1.4a**)[225,243,244]. Although the regulatory capacity of the *KLF2* promoter has never been assessed *in vivo*, there is substantial evidence in support of other components of this pathway. Genetic perturbation of *Erk5* leads to embryonic death, and *Erk5* hemizygous embryos do not express *Klf2* mRNA[245]. Both dominant negative MEK5 and MEF2 mutants prevent flow-mediated induction of *KLF2* expression in EC culture[246]. Repression of *KLF2* expression was shown to proceed through the oscillatory shear stress-induced Tumour Necrosis Factor Alpha

(TNF-α)-mediated inflammatory pathway which leads to the HDAC4 and p65-mediated inhibition of MEF2 binding at the KLF2 promoter (**Figure 1.4b**)[247].

The expression pattern of *Klf2* mRNA is consistent with the idea that shear stress regulates *Klf2* expression. In both mouse and zebrafish embryos, *klf2* mRNA is robustly expressed after the onset of shear stress. In addition to onset of expression which corresponds to blood flow initiation, *Klf2* mRNA expression levels are higher in arterial ECs and cardiac regions exposed to high levels of shear, such as the OT of the heart and the ECCs which will later develop into valves[248]. In contrast, only limited *Klf2* mRNA expression is detected in venous ECs, especially after the onset of shear. Robust *Klf2* mRNA expression can also be detected in capillary ECs[248].

In the adult dorsal aorta, EC responses to shear stress are defined by the type of shear stress EC are exposed to, which depends on their location within the dorsal aorta. In the descending aorta, where ECs are exposed to high laminar shear, they are quiescent, align with the direction of blood flow, are largely impermeable and express anti-inflammatory genes. In contrast, in the inner curvature of the aortic arch and vessel bifurcations, where ECs are exposed to disturbed shear stress patterns, they exhibit high cell turnover, poor alignment with the direction of flow, high permeability, and upregulate pro-inflammatory gene expression[182,188,189]. In line with this distinction in phenotypes that are the consequence of shear stress types, *Klf2* mRNA expression also follows this pattern, and is higher in the descending aorta than the inner curvature of the aortic arch in the human dorsal aorta, and the mouse artery collar model, where *Klf2 mRNA* expression was higher in the area of high shear stress[249].

Disturbed shear stress patterns in the inner curvature of the aortic arch (e.g. low flow, flow separation, flow reversal, and flow turbulence) have been associated with increased EC inflammation and formation of atherosclerotic plaques[250,251]. Atherosclerosis is an inflammatory vascular disease characterised by accumulation of lipids and immune cells that lead to plaque formation[182,252]. Despite

the systemic nature of risk factors associated with the development of atherosclerosis, plaque formation is highly focal, corresponding to areas of low and disturbed shear stress. Analysis of adult *Klf2* and *Klf4* KO models in the context of atherosclerosis suggests both contribute to the protective EC phenotypes that are lost where their expression is reduced, for example in areas of disturbed shear stress. Deleting one copy of *Klf2* increased atherosclerotic plaque burden on an *ApoE*$^{-/-}$ model[253], and endothelial-specific *Klf4* deletion using constitutively active *Cdh5-Cre* showed significantly increased atherosclerotic plaque burden and increased inflammatory cell infiltration on an *ApoE*$^{-/-}$ background, predisposing to pro-thrombotic and pro-adhesive EC phenotypes[254]. *Klf2* KO mice exhibited significantly larger strokes and impairments in blood-brain barrier (BBB) function, whereas *Klf2* overexpressing mice were protected against ischaemic stroke and maintained BBB function[255]. Taken together, these studies suggest that *Klf2* and *Klf4* maintain EC phenotypes that lead to vascular protective functions in pathological settings. Understanding how these two genes are transcriptionally regulated thus has important implications for understanding vascular disease.

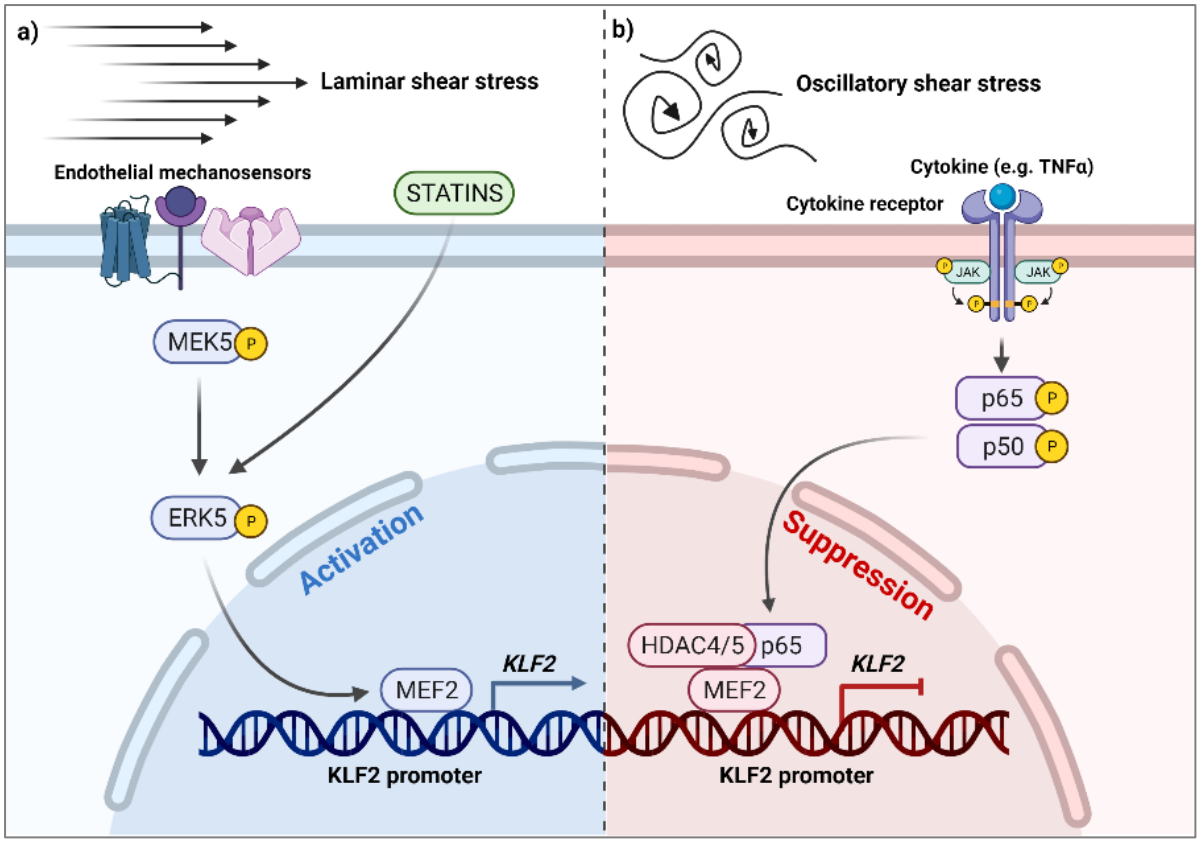


Figure 1.4: Pathways activating and inhibiting *KLF2* expression in cultured ECs. (a) High laminar shear activates mechano-transduction complexes on the EC surface, which result in activation of the MEK5-ERK5-MEF2 signalling pathway. Interaction of MEF2 factors with the *KLF2* promoter induced *KLF2* expression. **(b)** Oscillatory shear stress activates cytokine signalling, leading to recruitment of HDAC4/5 which interact with p65, a transcriptional effector of the pro-inflammatory NF-κB signalling pathway, to inhibit MEF2-induced *KLF2* expression via epigenetic modification of the surrounding chromatin. This figure was adapted from "Blank comparison pathway", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates

1.4 Thesis aims and hypotheses

Despite extensive investigations into regulation of *KLF2* by the MEKK3-MEK5-ERK5 signalling cascade that leads to MEF2 binding at the *KLF2* promoter to induce *KLF2* expression, many questions about transcriptional regulation of *KLF2* remain unanswered. The *KLF2* promoter is conserved and open across many different cell types yet flow-sensitive expression is restricted to ECs. Additionally, MEF2 factors are themselves not specific to ECs, instead responsible for regulating differentiation of many non-endothelial cell types including skeletal and cardiac muscle, neural crest, and SMCs[212]. Lastly, MEF2 factors are not specifically expressed in response to flow: they are found in all EC subtypes and prominently feature in the transcriptional regulation of signalling pathways that are not regulated or affected by shear stress. For example, MEF2 factors are required for the activity of an HLX-3 enhancer expressed specifically at the angiogenic sprout during development, yet inhibition of blood flow using the *tnnt2* morpholino affected neither enhancer activity, nor the expression of *hlx1* mRNA in zebrafish embryos[154]. It is therefore unlikely that MEF2 factors alone control the expression of a gene with complex subtype-specific and context-dependent expression pattern both during development and in response to a variety of environmental cues, such as shear stress and inflammatory stimuli, all through a single gene regulatory element ubiquitous to all of these scenarios. It is however possible that additional transcriptional regulators work with MEF2 factors to regulate *KLF2* expression, or alternatively that specificity of signalling is achieved through another layer of regulation, for example through post-translational modification.

The control of context-dependent and spatio-temporally precise developmental gene expression often relies on enhancers, yet no enhancers for *KLF2* have ever been described. In addition, the *KLF2* activation pathways have been studied exclusively in cultured ECs so far, and data on *KLF2* regulation during development is still lacking. Understanding how genes central to specific signalling pathways are regulated is crucial to the understanding of how they are activated by upstream signalling

pathways. When studying disease-relevant genes such as *KLF2*, understanding their behaviour in seemingly unrelated settings can give us important information about signalling pathways that could be targeted, and ones that should not, maximising the chances of developing therapeutic strategies.

The aim of this thesis is to investigate how endothelial *KLF2* is transcriptionally regulated in development. I hypothesised that (1) transcriptional control of endothelial *KLF2* expression would be regulated by enhancers in addition to the promoter, (2) context-specific *KLF2* expression downstream of different signalling pathways would be regulated by direct transcriptional regulators in addition to MEF2, and (3) *KLF2* enhancers and direct transcriptional regulators would be regulated by shear stress.

This thesis has the following aims:

- **a)** Identification, validation, and characterisation of the *KLF2* enhancers (Chapter 3)
- **b)** Analysis of direct transcriptional regulators of identified *KLF2* enhancers (Chapter 4)
- **c)** Investigation of the effects of shear stress on *KLF2* enhancer activity (Chapter 5)

2 Materials and methods

2.1 Transgenic mice

2.1.1 Generation of transient transgenic mouse embryos

Transient (F0) transgenic mouse embryos were generated by pronuclear injection outsourced to Cyagen Inc (https://www.cyagen.com/us/en/). Each putative enhancer construct was injected into fertilised mouse oocytes on the C57BL/6J genetic background. Injected oocytes were transplanted into pseudo-pregnant surrogate mothers to obtain embryos that were harvested at E12.5. All embryos were genotyped by PCR to confirm the presence of the desired transgene and stained with X-Gal to identify *LacZ* expressing embryos. Embryos were fixed in 4% paraformaldehyde (PFA), then shipped to the Ludwig Institute for Cancer Research, University of Oxford, where they were further dissected and imaged using a binocular dissecting microscope (M165C, Leica) supplied with an accessory ProgRes CF Scan camera. The pattern of reporter gene expression was recorded qualitatively and designated to fall under one of the following categories: vascular, ectopic, or no expression. Images were captured in ProgRes CapturePro software (Jenoptik). Following whole-mount imaging, embryos were processed for histological analysis as described in Section 2.3.2. All transgene enhancer sequences are provided in Section 2.4.4.

2.1.2 Generation and breeding of stable transgenic mouse lines

All animal procedures were approved by local ethical review and licensed by the UK Home Office in accordance with the Animals (Scientific Procedures) Act 1986 (PP1224162 and PP7967296). Mice were housed and bred at the Functional Genomics Facility at the Wellcome Trust Centre for Human Genetics (WTCHG). All stable transgenic mouse lines were generated on a C57BL/6J background by pro-nuclear injection into the fertilised mouse oocytes by Dr Ben Davies and his research group (WTCHG). Prior to injection, each putative enhancer construct was linearised by restriction digestion using XhoI and NotI enzymes and removed from the backbone by agarose gel electrophoresis. The band containing the enhancer construct was extracted from the gel and purified using QIAquick Gel Extraction kit (Cat.#

28706X4, Qiagen). The oocytes injected with linearised constructs were implanted into pseudo-pregnant surrogate mothers who gave rise to the F0 generation. After birth, these F0 pups were genotyped for the presence of *LacZ* to identify founder mice. Founder mice were then bred with WT C57BL/6 counterparts and their embryos analysed at specified developmental stages to obtain a time-course of enhancer-directed reporter gene expression throughout development. Whole-mount embryonic analysis was carried out as described in 2.1.1.

2.1.3 Genotyping

All samples were subjected to crude lysis (Jackson Lab Manuals) and genotyped by PCR followed by agarose gel electrophoresis. Each tissue sample (embryonic yolk sac or adult ear biopsy) was digested in 100µL of 25mM NaOH/0.2mM EDTA lysis buffer at 98°C for 1h and vortexed prior to the addition of 100µL of 40mM Tris-HCl (pH 5.5). For each PCR reaction, 2µL of the crude lysate was added to GoTaq G2 Green master mix (Cat. #M7122, Promega) and LacZ primers (**Tables 2.1,2.2, and 2.3**). PCR products were run on 1.5% agarose gel at 180V and the genotype was determined according to the presence or absence of a LacZ product band.

DNA lysate	1-2µL
GTG mix	10µL
Primer (F)	2µL
Primer (R)	2µL
H_2O	4-5µL
Total volume	20µL

Table 2.1: Genotyping PCR protocol.

Temperature	**Time**	**No. of cycles**
94°C	3min	1
94°C	30s	30
55°C	30s	
72°C	30s	
72°C	5min	1

Table 2.2: Genotyping PCR cycle protocol.

Primer name	**Forward sequence**	**Reverse sequence**
hsp68_LacZ	GTCGTTTTACAACGTCGTGAC	GATGGGCGCATCGTAACCGTG

Table 2.3: Genotyping primers.

2.1.4 Dofetilide studies

Dofetilide experiments were carried out under PP7967296. Mice were housed and bred at the Functional Genomics Facility at the Wellcome Trust Centre for Human Genetics (WTCHG). To obtain embryos at specific developmental stages, vaginal plugs were used as an indication for the onset of gestation.

Dofetilide (Cat.# PZ0016-5MG, Sigma-Aldrich) was dissolved in H_2O acidified with HCl (pH 2.7) and kept at a stock solution of 1mg/mL at 4°C for 1 week before a fresh solution was prepared. Pregnant females were weighed a day before first Dofetilide administration, and the 2mg/kg dose of Dofetilide administered by oral gavage in a volume of 100µL was adjusted according to the weight of each individual animal. Acidified H_2O (pH 2.7, 100µL) was administered by gavage to vehicle-treated control pregnant females on a schedule that matched the schedule of Dofetilide administration in experimental groups. Transgenic embryos were collected between 12h and 16h after the last Dofetilide or vehicle administration and dissected from the uteri of pregnant mice culled by cervical dislocation, then processed and stained as described in Section 2.3.1. For the pilot experiment where 2mg/kg Dofetilide was administered once daily between E6 and E10, one vehicle-treated and one Dofetilide-treated litter was analysed per stable transgenic line. For further experiments where 2mg/kg Dofetilide was administered twice daily between E7 and E10, three experimental litters and one vehicle-control litter were analysed.

2.2 Transgenic zebrafish

2.2.1 Husbandry

All animal procedures were approved by local ethical review and licensed by the UK Home Office in accordance with the Animals (Scientific Procedures) Act 1986 (PP1224162). Adult zebrafish were maintained at 28.5°C in system water at the Biomedical Services housed in the Weatherall Institute of Molecular Medicine (WIMM).

2.2.2 Tol2 transposon transgenesis

F0 transient mosaic transgenic zebrafish embryos were generated by Dr Svanhild Nornes using Tol2-mediated integration. Wild-type fertilised zebrafish oocytes were obtained by timed spawning of WT adult zebrafish. One cell stage fertilised oocytes were injected with 0.5nl of 50ng/µL Tol2 transposase and 60ng/µL enhancer-containing construct using the FemtoJet microinjector (Eppendorf). Upon harvest, embryos were kept in E3 medium (5mM NaCl, 0.17mM KCl, 0.33mM CaCl2, 0.33mM MgSO4 and 10^{-5}% Methylene Blue) at 28.5°C.

2.2.3 Image acquisition and analysis

Prior to imaging, embryos were dechorionated using fine forceps and anaesthetised with 0.1% tricaine mesylate (Cat.#5040, Sigma Aldrich). For each expression screen, individual embryos were transferred into wells of a flat bottom 96-well plate. *GFP* reporter gene expression was detected with a Zeiss LSM 710 confocal microscope at 24hpf and 48hpf. Whole fish were imaged using the tile scan command and combined with Z-stack collection at 488nm excitation and 509nm emission. For each putative enhancer, the number of fish with any detectable reporter gene expression per total number of injected fish was recorded. The pattern of reporter gene expression was recorded qualitatively and designated to fall under one of the following categories: vascular, ectopic, or no expression.

2.3 Histology

2.3.1 Fixation and staining of embryos

To obtain embryos at specific developmental stages, vaginal plugs were used as an indication for the onset of gestation. The mating was assumed to have taken place at midnight preceding the detection of the plug. Transgenic embryos were collected at each time point and dissected from the uteri of pregnant mice culled by cervical dislocation. Embryos were rinsed and dissected in ice cold PBS. One yolk sac tissue sample per embryo was frozen to be used for genotyping. After dissection, embryos were fixed at 4°C in a solution of 2% PFA and 0.2% glutaraldehyde dissolved in PBS. Depending on age, embryos were fixed for varying amounts of time (10 min for E7.5-8.5, 20min for E9.0, 30min for E9.5, 60min for E10.0-E11.5, 1.5h for E12.5, 2h for E13.5, 2.5h for E14.5, 3h for E15.5). Chest cavities of embryos older than E12.5 were opened to allow for more efficient infiltration of fixative and X-Gal into internal organs. Following fixation, embryos were rinsed in PBS, then incubated either in PBS (E8.5-E11.5) or rinse solution (0.1% sodium deoxycholate, 0.2% Nonidet P-40, 2mM MgCl2 in PBS) if E12.5 or older, twice for 30min at 4°C. Depending on age, embryos were then stained between 6-8h (most transgenic lines) or overnight at 37°C (KLF2-66:*LacZ* stable transgenic line 2) with 1mg/mL X-Gal/DMF dissolved in stain solution (5mM K4Fe(CN)6, 5mM K3Fe(CN)6, 0.1% sodium deoxycholate, 0.2% Nonidet P-40, 2mM MgCl2 in PBS). After staining, the embryos were rinsed in PBS, then fixed overnight in 4% PFA at 4°C. Some embryos at E13 or older had their skin, hearts, lungs, and/ or hindbrain removed prior to imaging. Embryos were analysed for expression of *LacZ* as described in 2.1.1.

For the analysis of transgene expression in the post-natal and the adult dorsal aorta, the dorsal aorta was first dissected from the mouse pups or adult mice. The dissection was aided by the use of micro-pins and a Sylgard™ (Cat.# SIAL761036-5EA, Merck)-coated petri dish. Following dissection, the dorsal aortae were pre-fixed in a solution of 2% PFA and 0.2% glutaraldehyde dissolved in PBS, either for

30min (if post-natal) or 1h (if adult). The aortae were then incubated in rinse solution (0.1% sodium deoxycholate, 0.2% Nonidet P-40, 2 mM MgCl2 in PBS) twice for 30min at 4°C, and stained with 1mg/mL X-Gal/DMF dissolved in stain solution (5mM K4Fe(CN)6, 5mM K3Fe(CN)6, 0.1% sodium deoxycholate, 0.2% Nonidet P-40, 2mM MgCl2 in PBS) for 2h at RT (post-natal dorsal aorta analysis of the KLF2-41:*LacZ* stable transgenic line), 8h at RT (post-natal dorsal aorta analysis of the KLF2-66:*LacZ* stable transgenic line), or 24h at 37°C (adult dorsal aorta analysis of the KLF2-41:*LacZ* and KLF2-66:*LacZ* stable transgenic lines).

For the analysis of transgene expression in post-natal and adult organs, the different organs were dissected, then pre-fixed in a solution of 2% PFA and 0.2% glutaraldehyde dissolved in PBS, either for 3h (if post-natal) or 4h (if adult). The organs were then incubated in rinse solution (0.1% sodium deoxycholate, 0.2% Nonidet P-40, 2mM MgCl2 in PBS) twice for 30min at 4°C, and stained with 1mg/mL X-Gal/DMF dissolved in stain solution (5mM K4Fe(CN)6, 5mM K3Fe(CN)6, 0.1% sodium deoxycholate, 0.2% Nonidet P-40, 2mM MgCl2 in PBS) for 2h at RT (post-natal organ analysis of the KLF2-41:*LacZ* stable transgenic line), 8h at RT (post-natal organ analysis of the KLF2-66:*LacZ* stable transgenic line), or 24h at 37°C (adult organ analysis of the KLF2-41:*LacZ* and KLF2-66:*LacZ* stable transgenic lines).

For the analysis of Dofetilide and vehicle treated embryos, the embryos were dissected, then pre-fixed in a solution of 2% PFA and 0.2% glutaraldehyde dissolved in PBS, between 20min and 1h, depending on stage of developmental delay observed following Dofetilide administration. The embryos were then incubated in PBS once for 30min at 4°C, and stained with 1mg/mL X-Gal/DMF dissolved in stain solution (5mM K4Fe(CN)6, 5mM K3Fe(CN)6, 0.1% sodium deoxycholate, 0.2% Nonidet P-40, 2mM MgCl2 in PBS) for 8h at RT (analysis of the KLF2-41:*LacZ* stable transgenic line), and for 6-8h at 37°C (analysis of the KLF2-66:*LacZ* stable transgenic line). Vehicle and Dofetilide-treated embryos for the same stable transgenic line were always incubated for the same length of time to allow for direct

comparison in transgene expression observed between vehicle control and Dofetilide-treated embryos.

2.3.2 Paraffin embedding and sectioning

Embryos were processed by dehydration through a series of ethanol washes, cleared with xylene and infiltrated with liquid paraffin, all using a tissue processor (Excelsior AS, Cat.# A82300001, Thermo Scientific). Embryos were then embedded in steel moulds using the HistoStar Embedding Workstation (Cat.# A81000002, Thermo Scientific). Sectioning was carried out by Indrika Ratnayaka and Dr Sarah De Val (Ludwig Institute for Cancer Research, University of Oxford) using HM 355S Automatic Microtome (Cat.# 905200, Thermo Scientific). Sections (6µm) were mounted onto glass slides and left to dry for 48-72h at RT or 2-3h at 37°C. Once dry, slides selected for imaging were de-waxed through a series of Histoclear II (2x 5 min) and ethanol (2x100% EtOH for 3min each, 1x 90% EtOH for 1min, 1x 70% EtOH for 30s) washes, rehydrated in dH_2O (1min), counterstained with nuclear fast red (30-60s), then dehydrated (1x 70% EtOH for 30s, 1x 90% EtOH for 30s, 2x 100% EtOH for 2min each, 3x Histoclear II for 5min each). Slides were mounted with VectaMount permanent mounting medium (Vector Labs) and left to dry overnight. The slides were counter-stained with nuclear fast red dye (Vector Labs), then imaged using a slide scanner (Hamamatsu NanoZoomer S210, C13239 series).

2.3.3 Klf2 RNAScope® on E11.5 sections

For the RNA scope experiment, the Mm-Klf2-O1 RNAScope® probe (Cat. #510671, BioTechne, UK) was used on E11.5 paraffin sections, using the RNAScope® Multiplex Fluorescent Reagent Kit v2 (Cat.#323100, BioTechne, UK). Sections were cleared through three 5min histoclear incubations, then dehydrated in 100% ethanol for 2min, and finally dried for 5min at 60°C. Sections were then incubated with RNAScope® Hydrogen Peroxide solution for 10min at RT, then washed with agitation in ultrapure H_2O five times, with each wash lasting 1-2min. For antigen retrieval, the slides were acclimatised in ultrapure H_2O for 10s at 100°C, then submerged in RNAScope® 1x Target Retrieval Reagent and

incubated for a further 15min at 100°C using a steamer. Following target retrieval, slides were rinsed in ultrapure H_2O for 15s, then dehydrated in 100% ethanol for 2min, and dried for 5min at 60°C. A wax pen was used to draw a hydrophobic barrier around each section. The sections were then treated with the RNAScope® Protease Plus reagent and incubated for 30min at RT. Following the protease treatment, slides were washed with agitation in ultrapure H_2O twice for 1min each time. Hybridisation probes were diluted to 1X, then applied to sections in drops until the whole section was covered. The slides were incubated for 2h at 40°C in the hybridisation oven. Following hybridisation, the slides were washed with the RNAScope® 1X Wash Buffer for 2min at RT, with agitation. Signal amplification was carried out using the RNAScope® Multiplex FL V2 amplification reagents 1, 2 and 3, which were applied sequentially with 30min incubation at 40°C between each step. Following the amplification process, the slides were washed twice with agitation in 1X Wash Buffer, for 2min per wash. To develop the HRP signal, the Opal™ Dyes were reconstituted 1:750 in RNAScope® TSA buffer. The *Klf2* signal was developed using Opal™ Dye 570 at 1:3000 dilution, and the CD31 signal was developed using Opal™ Dye 750. After development, the slides were further processed for immunohistochemistry (Section 2.3.4).

2.3.4 Immunostaining following RNAScope®

Following RNAScope® for *Klf2* and *CD31*, the slides were immuno-stained for Endomucin (EMCN). Slides were washed in 1 x PBS twice for 2 min each time. The slides were then carefully dried with a tissue to remove excess PBS and placed in a humidified chamber. Sections were incubated with blocking solution (4% FBS, 10% donkey serum, 0.2% Triton X-100 in PBS) for 1h at RT. Following blocking, the primary antibody solution (anti-EMCN raised in rat, Lot.#H1512, Santa Cruz Biotechnology, dissolved 1:50 in blocking solution) was applied to section in a humidified chamber. The sections were incubated with the primary anti-body overnight at 4°C. The following morning, the slides were rinsed three times in 1 x PBS for 2min per wash. The excess PBS was removed using tissue. Secondary anti-body (Alexa488, dissolved 1:500 in blocking solution) was applied to sections before

incubation for 1h at RT. Following secondary anti-body staining, the slides were rinsed in 1 x PBS three times, then incubated for 30s with DAPI solution. Following multiple washes with TBS-T, the slides were mounted and dried overnight in the dark. The slides were stored at 4°C until imaging, which was carried out using the Zeiss LSM980 microscope and the Zen Blue 3.3 software.

2.4 Enhancer prediction using the UCSC genome browser

The details of *in silico* enhancer prediction are described in Section 3.2. The KLF2 locus analysis included the analysis of DNAse I hypersensitivity in 46 non-endothelial and six EC lines. The list of cell lines used for this analysis is shown in **Table 2.4**.

Cell Line[1]	views[2]	Track Name[3]
Treg Wb78495824	Signal	Treg Wb78495824 DNaseI DGF Per-base Signal from ENCODE/UW
Th17	Signal	Th17 DNaseI DGF Per-base Signal from ENCODE/UW
Th2 Wb54553204	Signal	Th2 Wb54553204 DNaseI DGF Per-base Signal from ENCODE/UW
Th2	Signal	Th2 DNaseI DGF Per-base Signal from ENCODE/UW
Th1 Wb33676984	Signal	Th1 Wb33676984 DNaseI DGF Per-base Signal from ENCODE/UW
Th1	Signal	Th1 DNaseI DGF Per-base Signal Rep 2 from ENCODE/UW
T-47D	Signal	T-47D DNaseI DGF Per-base Signal from ENCODE/UW
SkMC	Signal	SKMC DNaseI DGF Per-base Signal from ENCODE/UW
SK-N-SH RA	Signal	SK-N-SH_RA DNaseI DGF Per-base Signal from ENCODE/UW
SAEC	Signal	SAEC DNaseI DGF Per-base Signal from ENCODE/UW
RPMI-7951	Signal	RPMI-7951 DNaseI DGF Per-base Signal from ENCODE/UW
NHLF	Signal	NHLF DNaseI DGF Per-base Signal from ENCODE/UW
NHDF-neo	Signal	NHDF-neo DNaseI DGF Per-base Signal from ENCODE/UW
NHDF-Ad	Signal	NHDF-Ad DNaseI DGF Per-base Signal from ENCODE/UW
NH-A	Signal	NH-A DNaseI DGF Per-base Signal from ENCODE/UW
NB4	Signal	NB4 DNaseI DGF Per-base Signal from ENCODE/UW
M059J	Signal	M059J DNaseI DGF Per-base Signal from ENCODE/UW
HVMF	Signal	HVMF DNaseI DGF Per-base Signal from ENCODE/UW
HSMM	Signal	HSMM DNaseI DGF Per-base Signal from ENCODE/UW
HRCEpiC	Signal	HRCEpiC DNaseI DGF Per-base Signal from ENCODE/UW
HPF	Signal	HPF DNaseI DGF Per-base Signal from ENCODE/UW
HPdLF	Signal	HPdLF DNaseI DGF Per-base Signal from ENCODE/UW
HPAF	Signal	HPAF DNaseI DGF Per-base Signal from ENCODE/UW
HMVEC-LLy	Signal	HMVEC-LLy DNaseI DGF Per-base Signal from ENCODE/UW
HMVEC-LBl	Signal	HMVEC-LBl DNaseI DGF Per-base Signal from ENCODE/UW
HMVEC-dLy-Neo	Signal	HMVEC-dLy-Neo DNaseI DGF Per-base Signal from ENCODE/UW
HMVEC-dBl-Neo	Signal	HMVEC-dBl-Neo DNaseI DGF Per-base Signal from ENCODE/UW
HMVEC-dBl-Ad	Signal	HMVEC-dBl-Ad DNaseI DGF Per-base Signal from ENCODE/UW
HMF	Signal	HMF DNaseI DGF Per-base Signal from ENCODE/UW
HIPEpiC	Signal	HIPEpiC DNaseI DGF Per-base Signal from ENCODE/UW
HGF	Signal	HGF DNaseI DGF Per-base Signal from ENCODE/UW
HFF	Signal	HFF DNaseI DGF Per-base Signal from ENCODE/UW
HEEpiC	Signal	HEEpiC DNaseI DGF Per-base Signal from ENCODE/UW
HCPEpiC	Signal	HCPEpiC DNaseI DGF Per-base Signal from ENCODE/UW
HCM	Signal	HCM DNaseI DGF Per-base Signal from ENCODE/UW
HCFaa	Signal	HCFaa DNaseI DGF Per-base Signal from ENCODE/UW
HCF	Signal	HCF DNaseI DGF Per-base Signal from ENCODE/UW
HAEpiC	Signal	HAEpiC DNaseI DGF Per-base Signal from ENCODE/UW
HA-sp	Signal	HA-sp DNaseI DGF Per-base Signal from ENCODE/UW
HA-h	Signal	HA-h DNaseI DGF Per-base Signal from ENCODE/UW
H7-hESC	Signal	H7-hESC DNaseI DGF Per-base Signal from ENCODE/UW
GM12865	Signal	GM12865 DNaseI DGF Per-base Signal from ENCODE/UW
GM06990	Signal	GM06990 DNaseI DGF Per-base Signal from ENCODE/UW
CD4+ Naive Wb11970640	Signal	CD4+ Naive Wb11970640 DNaseI DGF Per-base Signal from ENCODE/UW
AoAF	Signal	AoAF DNaseI DGF Per-base Signal from ENCODE/UW
AG10803	Signal	AG10803 DNaseI DGF Per-base Signal from ENCODE/UW
LHCN-M2	Signal	LHCN-M2 DNaseI DGF Per-base Signal from ENCODE/UW
HUVEC	Signal	HUVEC DNaseI DGF Per-base Signal from ENCODE/UW
HepG2	Signal	HepG2 DNaseI DGF Per-base Signal from ENCODE/UW
B cells CD20+ RO01778	Signal	B cells CD20+ RO01778 DNaseI DGF Per-base Signal from ENCODE/UW
A549	Signal	A549 DNaseI DGF Per-base Signal from ENCODE/UW
K562	Signal	K562 DNaseI DGF Per-base Signal from ENCODE/UW

Table 2.4: The list of cell lines used for DNAse I hypersensitivity analysis. Endothelial cell lines are marked by black squares.

2.5 Cloning of enhancer constructs

2.5.1 Gateway cloning

Putative human *KLF2* and zebrafish *klf2a* enhancer sequences were predicted *in silico* using the UCSC Genome Browser (Section 3.2.1). An attempt was made to amplify each putative 600bp to 1000bp long enhancer piece by PCR from genomic DNA (see **Tables 2.5** and **2.6** for PCR amplification protocols and **Table 2.7** for sequence-specific primers used). KLF2+23 could not be successfully amplified by PCR and was therefore generated as a synthetic gene (GeneSynthesis, Life Technologies). Various pieces of the klf2a-6 enhancer and conserved enhancer versions from various species (zebrafish, platypus, Tasmanian devil) were generated synthetically as linear double-stranded DNA fragments (GeneArt strings, Life Technologies) as genomic DNA of those species was not readily available. Putative enhancer sequences generated as GeneArt strings had their polyA tails synthesised immediately prior to cloning. Putative enhancer generated as a synthetic gene was amplified from the synthetic plasmid backbone by PCR using enhancer sequence-specific primers (Table 5).

	Taq PCR mix (µL)	**Phusion HF Taq PCR mix (µL)**
H_2O	17.9	27.5
10x PCR buffer	3	10
50µM Mg^{+2}	1	0
10µM dNTPs	0.6	1
1µM primer (F)	3	5
1µM primer (R)	3	5
100ng/µL DNA template	1	1
Enzyme	0.5	0.5
Total volume	30	50

Table 2.5: Protocols for enhancer amplification through PCR.

Temperature	Time	No. of cycles
94°C	5min	1
94°C	30s	30
58°C	45s	
72°C	1min	1

Table 2.6: Enhancer amplification PCR protocol.

Primer name	Forward sequence	Reverse sequence
KLF2-66	GATCGCAACATTGCACTCCAG	CAAGGCTAGTCTCGAATGCC
KLF2-41	GGTTGCACAACACTGCATG	GAAACTGAGTCCCAGCGTTAG
KLF2-34	GCATGCTCTTATGCCTGGC	GCCAAGATTGCACCACTAC
KLF2-8	CTCAGCTCCAAGCATTGCAG	CAGCCTCGATGGTGATTATC
KLF2+23	CTGGCCAACATGGTGAAAC	CATATTGCCCAGGCTGGTC
klf2a-83	CAGCATTCAAGTTGCTTAACC	CCCTTGGATGTTAATGTCATTG
klf2a-60	CAGTCCACAAACACGCCCTC	CACTCTATGTGGGCGGAG
klf2a-41	GTAACAAGCTGTGGATGCTAAC	GCTAGCGAGTCATGAGGAC
klf2a-6	GACATCCCTTTTTGATGGCA	CCATTCACTTTCCCTGTCGT
klf2a+23	TTTGTGTGCATGCTCTCTGT	CTGCAGCGTTTCCTCAGAG
KLF2-0.4 (pMK)	CTAGGCGCGCCATGAGCTC	GATCCAAGCGGCCGCTACTC
KLF2-41 mut Mef2	GGTTGCACAACACTGCATG	GAAACTGAGTCCCAGCGTTAG
KLF2-41 mut Tata	GGTTGCACAACACTGCATG	GAAACTGAGTCCCAGCGTTAG
KLF2-41 mut Rbpj	GGTTGCACAACACTGCATG	GAAACTGAGTCCCAGCGTTAG
KLF2-41mutEtsCDE	GGTTGCACAACACTGCATG	GAAACTGAGTCCCAGCGTTAG
KLF4-148	GAGTCGGGCAAAAGCGAAATC	GCTTTTTTGGTGGCACCTTG

Table 2.7: Primers used to amplify enhancers from genomic DNA or Gene Synthesis vector backbone.

Regardless of the initial amplification strategy, each putative enhancer was cloned, using the Gateway technology (Invitrogen), into an entry vector (PCR8 TOPO-TA) according to the manufacturer's instructions. The resulting construct was transformed into chemically competent *E. coli* cells and plated onto spectinomycin-coated agar plates to select for successfully transformed bacteria carrying the pCR8 vector. Plates were incubated overnight at 37°C. The following day, 6-10 spectinomycin resistant colonies were each used to inoculate an aliquot of LB broth with added spectinomycin to ensure cultivation of colonies carrying the construct. Each liquid culture was incubated overnight, then used to miniprep DNA following the manufacturer's protocol (Qiagen). DNA concentration of each sample was determined using a Nanodrop spectrophotometer. Samples were then digested using at least two distinct restriction enzyme strategies per sample and ran on an agarose gel to confirm

insertion and correct insert orientation (5'-3'). Finally, samples were Sanger sequenced (Eurofins) to ensure sequence identity to the initial putative enhancer sequences.

Next, each putative enhancer was transferred from the entry vector (pCR8) either into a zebrafish (pE1b-containing) or a mouse (hsp68-containing) compatible expression vector using the Gateway LR clonase II enzyme (Invitrogen) according to the manufacturer's protocol. The resulting constructs were then transformed, incubated, inoculated, miniprepped, digested, and sequenced as described above. For generation of transiently transgenic embryos by pro-nuclear injection, each construct was midi-prepped according to manufacturer's protocol (Cat.#12143, Qiagen) and re-sequenced prior to posting to Cyagen.

2.5.2 Cut-paste cloning

The *KLF2* core promoter (KLF2-0.4) was generated as a synthetic gene (GeneArt Synthesis, Life Technologies) with accessory restriction enzyme sequences flanking both 5' (NotI, KpnI, ECoRI) and 3' (SpeI, HindIII, ECoRI) ends. To clone KLF2-0.4 into the pAUG expression vector for transient expression in mouse embryos, KLF2-0.4 was released from the vector backbone using a combination of NotI (Cat.#R3189L, NEB) and SpeI (Cat.#R3133L, NEB), then purified by agarose gel electrophoresis followed by gel extraction (Cat.#28706X4, QIAquick gel extraction, Qiagen). The pAUG vector was cut with the same restriction enzymes, then treated with Antarctic phosphatase enzyme (Cat.#M0289S, NEB) according to the manufacturer's protocol to dephosphorylate 5' and 3' ends preventing self-ligation. The linearised pAUG was then column purified (Cat.#28104, Qiagen) to remove residual enzymes. The insert and destination vector were ligated in a 1:4 vector to insert ratio using T4 ligase (Cat.#M0202S, NEB) following the manufacturer's protocol. The resulting construct was transformed, inoculated, miniprepped, sequenced, and midi-prepped as described in 2.4.1.1.

2.5.3 Mutagenesis

Mutant enhancer sequences were generated either as custom linear double-stranded DNA fragments (GeneArt strings, Life Technologies) or as synthetic genes (GeneSynthesis, Life Technologies) designed to contain the desired mutations. For mutagenesis, the core motifs for each putative TF were determined through JASPAR analysis (database of TF-DNA binding preferences, http://jaspar.genereg.net/) and general consensus in the field. Nucleotides contained within the core motif were mutated through random scrambling. Mutated motifs were entered into TRANSFAC™, a data base of eukaryotic TFs, to ensure that scrambling did not increase the probability of TF binding in the context of current knowledge. Sequences generated as synthetic genes were amplified by PCR using enhancer-specific primers, then cloned upstream of hsp68 and *LacZ* for testing in transgenic mouse embryos as described in 2.4.1.

2.5.4 Enhancer sequences

2.5.4.1 Human KLF2 putative enhancer sequences

KLF2-66 (human sequence tested in mice and zebrafish)

GATCGCAACATTGCACTCCAGCCTGGGCAAGAAGAGCAAAACTCCATCTCGAAATAAAAATAAAAATAAAAA
ATAAATGAAATGGAGACCCCAGTACCCCCCGTTGATAGATAAGAGACTGAGATTTCTAGAGGGGCTCAGCTTC
CCTGGGGGACAGCTGTTTAGACCCAGAGCCAATTTTCAACTCTTGGGCAGCCAGACTGCAAGTCCAGTGCTCC
TTCCAAAACTTGTGTGTATGGAGAACCCACCACCTGTGCCCTGACCCCCTCCTGGTTCCCCACAGCCCCTGCCC
AGGCCAGGACACCCTGTCCCCTGGGTGTCCCTGAGGTCACAGATAAGGCCGCTGGGCCCGGCTATCGGCAGG
AAACATCTTGGCAGGACAGGAAGCGCGGCATGGGTCAGCCGCGCCTGCACCCGGCTCAGCCTATCAGCGGCC
AGATATTTTGGGAGCCACTATATAAACCCCAACTCGGCTCCAAGCTGTTCCCCGGGGAAGCCTCAGGCTTCAT
GTTCTGCGGCAGCCCTGAGGCTGCTGGAACACGGGGCCCGGGAGGAGCTCAGGAGCTGAGCGGGACAGAG
TGACAGGGGTGGACTCCGTCCGGGCATCACTGCAGGCCAGAGGGTGCTGGCTCTGAATCGAGCTAGCCCCTA
TTCAGTCCTGGTGCAGACACTCTCACTTGCTGGGCAAGTGGCTTCACCGCCCTGTGCCTCAGTTTCCCCCTCCC
CACATCCCTCCTTTGATCCTGTCCCAGGGAAAAACACACCTCCAGGTCCTGGAGCCTTTTCTGAAGTTTCTGGG
CCGACATCTGAGGGCCAACCCACTTTCTCCCAATGCAAAGCAGGGCCCAGGAATCTGCATCTTAAAACCTGCC
TTGGCCGGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCATGAGGTC
AGGCATTCGAGACTAGCCTTG

KLF2-41 (human sequence tested in mice and zebrafish)

GGTTGCACAACACTGCATGTATAAAATGCCACTACTTTAAAAATGGTTAATTTCATGTGAATTGCACTTTAATA
AAAATAAACAAGTATTTAAAAATCCACCAATGCAAACACTTCTCTGTTCCTCTGTGGTGGAGGAAGGGACACG
CGCTTTTTTTTCCGACCTTAGGAAGGAACAAGGGAGCCGGGGTCCCCTCCCAGCCTGGGAGCCCTGGGCACA
GTCCCGGCTCATTTGTCAGAGCTATCGGAGCCGTCCTCGGGCTGGTGGGAGTTCAGGGCTCTGAAAGGTTTTC

TGTCAAGGCTTGAAAGGGGGCCAGGTTTTTTTCCCCCCGGAGCCGCGCAGTCTCGGGGCTGTTGTTCTCAGCA
ATCGCAGGGCCTCGTGTTAGCAGGAAGCACAGCCAAGTAGGGTTTCCTGCGTGTTGGAGAGAGGAAGCTCCG
TAATGTTCTGGGAGGCGATGGTTAAAAATAACTCCGGTATATAAAGACAGCGGAGGGTCCCCTTGTTCGCTCA
CTCGGGCGCCGGCCGGCTGGACGCAGGGCCGAGCAGGTGGTTTGGGGCCTCGGGAAGGCCAAACCCCCGCC
TCTGGGCCCCTGGCTGGGGAAGACACCAGCCAAGTTCAGAGCCCCAAGTCGGCCTCACTTCCACAACTCAGCG
TCAGGGACACCGTGGGCGTTTCTGTTTCAAAACGCTTTTCTCCAGCAAAGAACGTAACCTCAAGCTGCTGTCA
GGGTAGAGGAATCCCTGCCCCCCGCCGCCACCCAGATAGGTTGGGCCGTGATTTTGGCTGTGCCTGGCAGTG
ACGATGTAGCGCACCCAGGCTGGGGGGCGCTGGCAGCCAGGGTTCAAATCCCGGCTCTGCCACCTAACGCTG
GGACTCAGTTTC

KLF2-34 (human sequence tested in mice and zebrafish)

GCATGCTCTTATGCCTGGCTAATTTTTGTATTTTTTTGTAGAGATGAGCTCTTGCTATGTTGCCCACCTGCTGGG
ACTCAATCCACCAAGTAGATGTCTCCAATTCTCCACCATTCCCCAAGACAAGTCTGGTGACAAAGAATTGGTTC
AGAACCTGGGAAGTCTTTGAAGGTCCACTTTTGTCAGATGACACATGATCTAAATTGGTCCCCAACCCCAGGC
CGCCGGAAAGGCTAGTGAGCATGGGGTGAGCATAGAATCACCGCTCTATGAAAACATGCCTCCCTCAGCCCT
GCCTCCTGGGCCACCGAGGTCTGATTCACGCTTCCTCGTATTCAAACAGAGTCGGTGCAGCTGGAAGCAACCC
TGCCAGGGGTAAGTCAGCCAGTGCAGGTAACATTTCTCCCCCTCGATTTTTCCCAGAACCCGGCTCCGGGCCC
AACACAGGATGCCTGGGGGATGTCAGAGGATTCTAAGTGAGGAAACCCCTGCCGTCCATTGCTGTTCTGTTCC
CCCAGGGACGCTTTGTCCAGGGCCCACCGATCCATCACACATACAAGGCCCAGGGCCCTGGAATGTTTAAGA
GCCCACAAAAAAATAGCCTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTCGCCCAGGCTGGAGGGTAGTG
GTGCAATCTTGGC

KLF2-8 (human sequence tested in mice and zebrafish)

CTCAGCTCCAAGCATTGCAGCGACAAGGTGGTGACCCGAGATAAGAAGCGCACTTGCAGAGGGTGCCTCTGC
CCGTATCTCTCCGCATCCTCAAGAAACACAGAACTATCAAACATTTTTGCAACATGAGGAGCCAAGTTTTGCAA
CATACTGAAAAAGGGTTGTGCAAACTTCAGCAGAAACACCCCAATTGCCCCAGCTGGGAGTCATCATCCAGTT
GCTGCTGTAATGTGATTGTAATTCCTCCAGTGTTCATAATTTGTCTTGGAAGCTCAGCTGTATTTTACATGACGT
GTTTTCTGTAGAATTCCTGAGTACTTGCTCTTGGGATCTGAGCCAAGTGCACACAATAGCTATTTGGGGCAAAA
CAGCTGGTAAAGAGAAAGGCAGCTTAGACCCTGAAGAAAAGTGGAAAGAATTCAAGGTTACTAGAAATGTCC
AGGGCCCAGTAAAAGAAGATTGTAATCTCTCTATGTGCAGAAACAGGCCATATACTTCATCAGTTTCCTCTCCA
TGCTCAGCTCAGCATGGAGCGGAAGCTCAGTATTTATTTGTGGCTCAGACGAAAAACTTCTTGTTTGCAAGTTT
CCTGAGCCCTTGAAGACAGCCTGGCAGGAGACAACAAATGCAAAGCTGCACACGTACCACTTATCCAATCCCC
TGTTAGTGACAGACATCACCAGTCAATCAGTCAATCACAGACATGACCAATCCATCACAGCACTGGGCCTGGC
TACAACCTCAGAATCTGCACCAGCACAGTGCTAAAAGCAACTATTATCAACAGATTTGAGGAAAATATTCTGCC
TCCTTGGTCTACATTTATAATGTTGAGTTGTTAGGAGAAGATAAGGGGCAGCAAGAAAGAGATGACAATAAG
ATAATCACCATCGAGGCTG

KLF2+23 (human sequence tested in mice and zebrafish)

CTGGCCAACATGGTGAAACCCCATCTTCACTAAAAATAAACAAAATTAGCCCGGCGTGGTGGTGGGTGCCTAT
AATTTCAGCTACTCGGGGAGCTGAGGCAGAATTGCTTGAACCTGGGAGGTGGAGGTTGCAGGGAGCTGAGA
TCACACCATTGCACCCCAGCCTAGGCACCAAGAGCATAACTGTCTCAAAATAATAATAATAATAATAATAATAA
TGATAATAATAATAATATAATAAAAATAAAGACCACTCAACGTAAACACGACTCTACCCAGGCAAGCCCACAT
CACGTCATCTCTTTGGAATAAGCTTGCGATAACCATCATATTTCCATTTCCTTTTATATTTTTCCAAAAACAGCTC
ATGATACAATTTGTACAGGTCCTAGTAACTTTGGGGAAAATTTTGGAAAAAAAAAAGTAGCATGTTTCAAAGG
AACTTGATCTGATAAGAGGCAGAAATATTGACTCCATTATTTCAAGAGGCAGTTTCTTTGTTTCCCAGACGCTT
CCTGACAGTTTTAAAAATAGCCTTCCACAACTGTTCCAGGATGAGAATGTCCTTTTACATAAAATGGGCTTTGA

AGATGTGATTTCATTTACTCAGAGGAGGCTGTAGATCATTCGTCATCCTTTTCCAAAACATGGCTTGGACCGGG
CAATGTGGGCCCTTGAATTCCAGTTGAACACCAGTCCAAACTGGGTTGTTGGATGCGGAGAAGATAAAGGGT
ACCCACAAAGAGGACACGGCTCGTTACAATCTGGGAGACAAAACGCAATGGGGTTGCAGTCAAAAGTTGCCT
CTTGCGACCAGGTGTGGTGGCTCACGCCTGTAATCCCAGCACTTCGGGAGGCTGAGGCAGGTGGATCACCTG
AGGTCAGGAGTTCGAGACCAGCCTGGGCAATATG

KLF2-0.4 (human sequence tested in mice)

GCGGCCGCTCGAATTCCAGGTACCCGGGATGCTCGCGCAGGCGGGAGCCCGGGGCGCCCGGGGGATCCCTG
AGCGTCACGCCGCTGTTGTGGAGCGCGTGTTGACAACGTCGCCGGGGAGACGGGCGGGGGCGGGGCCCGG
GAGAGGGGGAGGCGCGGCCCTGGCGGCGCGCGAGGGGCCGGGCTGTCAGCGCAAGGCCCAGGCCGCCGCA
GTGGCCACGGCCGCTGCCGCCCGCCGGCTTATATACCGCGGCTAAATTTAGGCTGCGCCCGGAGCTCGTCCCC
ATCCGGGACGCGTTTCCGCCGCCGCCGCTTTGGCCCGGCCCCCGCGCGCGCCGCGCCTATAAGGCTTGGGCG
GGCCCGGCCGCGGCCCACAGAGCCGTCCCCGCCACTAGTCGAAGCTT

KLF4-148 (human sequence tested in mice)

GGGCGAGTGCCCCGAGGGAGTCGGGCAAAAGCGAAATCCCGGCTTCCTACCCGGGACAGCGGGAGAGCCG
AGTTTTCCGCAGCGGCCGCGGGCTCGGAGCTTTTAAGGGTTTCTAATCCTGGAAGGCTGTCTGACATCACCCC
GTTTCTTGTCGGCTGATGTTTCGTACAAGCCTCTCATTTCCTCAGTGTTTTCAGAGCCACCAATTACTGCAACAG
TCTCGTGTTTATGTTTGCGCGCGCTCTCCTGCCTCGCTGGCCCTCCGCCTGGGAGGCTGCGCTTCCCTCCGACG
CGCGGAGCCCAGAGGGGGTGGGTCGGGGGAGCGGGGAGGGGGGAATCTGGATCAGGCATTGAAGAAAAA
GGGAGGGGTCCCCCTTAATTTTTTTTCATTGACTTCAGCACCATGTGATCAGGAAGTCTGGCCTCCCTCCATTTC
CCCTCCCGACTAAAGGGAAACATTGTGTAGCAGCCGCCTGGGCCACCGGTGGGATGGCCTTCGCTGCCTGAC
GTAGGGAGATAAAAATAACCGGCATATTTAAGGCAGGATCAGGAATCCCGGCGCTCACACGCGGCCTGGTCA
GTTCCCGAGGCCGCCCGGCAGGCAGCGCAGCCTGCGGGGAGGCCCCGCGCTCGACTGTGCGCGCCCCACCG
CAGGCAGTGCTAGGGCGTACGACCGGACCCCGACCCTGACCCCGACCTTGACCTCGGGACCCAGGCCGACTC
AGGCCTGGCTGAGCCTCCCAGATTCCCCAGGGAGGCCGAGCAAGACCGGCCCGAGGCAGTCTGGCTCAGGCT
CGCTGGGCTGGGGCTGGGCCTGGGTTCTCTGTCCGGCCTCCCGGGAGACGAGAAGGAAGGATGCCGTGGCA
AAGGGAGGGTAGTGCGAAGTTTGATGCCGAGGGGGGTTACAGAGCCTCTGTCTCAAGAGAAGGCGCACAAG
GTGCCACCAAAAAAGC

2.5.4.2 Zebrafish putative *klf2a* enhancers

klf2a-83 (zebrafish sequence tested in zebrafish)

CAGCATTCAAGTTGCTTAACCTACATCTTGTAATCGTATTATATATAATCCTATGAATCTGATCACTTCCATTATA
AGTGCCTCACCATAAATTACAAATTTGCTTTATTTTCTAAAAAAGATGGGTAGGTGGAAATGCTTTTAACATTT
GTCAGTGCTGTCGATTGAGCTTATCCCAGAATATACAAGGAAAAATCAATCAATGGCACACCTTGACCTGTTAT
AAATCTGTCACATGGTCCAGCTCAGCGACTGCCATCACATTCATTTCAGGTGGCATTAGTTTGGGTTAGAGAGT
GCCAGGGCCAGAGGTGCCCATATCTTCTGATGGTTTTCTATCAGCTGATTACACATATCCTGTGATGACGGTTC
ATTCCCTCTCCGGGAGCCTGAGGCTGGTATACAAAATAAAGAGGAAATGGAGACACTTTACTGTGGACTGAG
CCGAGGCTGAAGTTCCTGTATGCTGTGATGGACAGGAAATAAAGAGTCTCGTCATGCAAGTGATTCAGTAATA
TATAGACACGCCATCATACAATCCTGCTGCTGTAAGACGCTGGGTGGAAGCTGTAATGAACATAAAGATTGAC
AATTATAACATGCAATTATTTAATATTCAGAGAGTCAATAAGATACAAATCAACTACAAACAATAAAGACCAAT
ACATGGATCAACATCTATGTAACTGCAGTAAGAAATGTATGTAGTATTGCATTGTTGTAAACATTTGCCCCCAA
AGTAGTCTGAGACCCCTGTAAAAAACTCAATGACATTAACATCCAAGGG

klf2a-60 (zebrafish sequence tested in zebrafish)

CAGTCCACAAACACGCCCTCAGTCCATGTTAATGCACAAAAGGAATAAGCGTTTCAGCTCCAACACAATAAAC
CTTACGTTTTATGGCCTATAACTAGACCTCAAAGGCTTTTGCGTAGCGACGGCCCCGAAAATTGTTGACTTGGC
ACAAGCCTAGCACTGTAATTTAGGGCAAAAACAGCGTTTACAGTTGATGACGGATGCAGCCAGACAAAAGTG
TATCCTTCGCATACTCCATTAAACTAGCATGCACGGTTAAGCTCGCAATACGGTCATCTTCTGTTGTTGATGAG
TAGGTAAACACGTGCAGCCCACCAGGCTCCTGCCCTCACCGTGACCTTCACACTTCCGCTAGTCCTTTAGATCT
GTCAGAGACGGTCTCAGTTAAATACAAAGACCCTAAACGCCCTTTCCAGGGCACATGGGTGATTCCTAGGCTG
TGACAATGAATGCCAGGATATTTGACAGAGAGGAATGCTGTTTTTTACTTAAGGCAGGAGTGAAAATAAGTTT
GTCTTTGACTGGGATAGTCTAAGGACAAGTAGAAGTGCTTTTGCCTTCTTCAAAGTTTAATAGAGAATGTTAGC
TCACTCTTTAAATCATTACTACTACTTCAATGAGAGTTGACATGAAGCTAAATTCAACAATGTTTAAAAAGGGA
GGATCAAGATAAATCTGATTGGCTGTTAGCCGTGCGATATTCTGCAATAACAGTGCTTGTACAGCCTCTTCACC
CTTGTGTATTACTCCGCCCACATAGAGTG

klf2a-41 (zebrafish sequence tested in zebrafish)

GGTAACAAGCTGTGGATGCTAACGCACCAAGTTTACACAAACATCCCAAATGAAAGCATCAGATTTTCTGACT
ACACCCTTGCGTAGTCGATCCCTGCCGGGAGACTCTTCGGAAAATTTGAGCAAGTTTGGGAATCGAGATACGC
TGCTGTTAAAAACAAATAATTCACTCAGATGCTCCAGTGTATACAAGCACACGCTGTTTAGGCCTTGTAATCCT
TTGCTAATGCATTACTAGCTCAAATAAAAAAGGCCTTGATGTATCATTAGTGTCATAGCGCGCCACCGGGAAG
CCAGAGTAGACACATGTGCTTACAGTCGAGGCATCTCATCAGCGCGTGTGTTTACACTCGGACCAAAACATTG
GCTCTCACGCAGCTGACGGGGCCATGTGTTCCTGCACACGGCACGCCGCTATCACCGACAGTGTTTCTACAGT
CCTCACAAAGGTCAGCAAAACATGCCAATGTCTTGTCCCTCGGCGCCTGGCTATCAACAGGAAACATCTTTCGC
AGAACAGGAAGTGCTACTCTCTAATTGGCCCAAAGTGATTCAGTGAGAACCTCTTCGTCCACCAATGAGAGCG
CCGCTATTTTGGGAGAGGCTATATAAGTAGACTCACAGAGCTTCAGTGCGCCAGAGCTCACATGAGGTAGTTG
GTAGTTGAGGTTTCCTGGTGAACTGGGAACTGAGTAATGAAAGTGTTGTTTTTATTCTTGCTGAAAGTGTACA
AATGACAAATCAGTCAAACACTTGTGTTGCCCAAGACATTTTAGTCCTCATGACTCGCTAGC

klf2a-6 (zebrafish sequence tested in zebrafish)

GACATCCCTTTTTGATGGCATAATTATATTCATTTTTGTGAATAATGTGTGACTGAAAATGCAATGTCAACAACT
CAAATAACAGTTCATATTGAGGAACAGATACTGCAACAAAATCATCGGGAAACACTGTTTAAGAAGAAATACA
CACTGTTAGCGGCAGCATGACATACAAATGGTTTCATTTTGAGCTCATCGTTAATCTTAAGGGCTAAAAAAAA
AACGTGAGCCACACGTCAAACTCCCTTTCAGACAATCACTTGCTCCGCTAAAATCTAATCACACAAATATGCAA
CATGATTTTAACTCTTTTTCCTTTTGAGTCAGTGCCTGGTCATCACAAACGCCATCCTTTCCTGATTGGCTGCGA
ACGGTCCGGCGTGCAGGCAGGAAGCTGAGCCAGGCTTTGCTTCCTGAGCGTCCGAGAGAGGAAGCCCTGTG
ATAGGCGAGAGGGAGTGGGGTTAAAAATAGTGGCGTATATAAAGGCTCGCTCAGAATCCCATTCCGTCTCAA
ACTCTCACTCGCACACTTTCTGGCAGGACTCTCGGTTTCCTTCTCTCTCGAAAGCAGGGCCCAGGGCAGCCGCA
ACGCCGTCCTGCTGTGAAGAAAGAGCAGGTGTTTCCCAGGGTTTGACTCAAGCAGCAAAATATGCATTACGTT
GGATAACCAAAGCAAGAAAAGAAAAAAGAAAACCTTTTTCCAAATAAATGCTCAGTGCAGTTTACTTGAGCTC
GGTTCACACGCAAAGCATGATACCTTTTCCAAAACGACAGGGAAAGTGAATGG

klf2a-6/600 bp (zebrafish sequence tested in zebrafish)

ATACAAATGGTTTCATTTTGAGCTCATCGTTAATCTTAAGGGCTAAAAAAAAAACGTGAGCCACACGTCAAACT
CCCTTTCAGACAATCACTTGCTCCGCTAAAATCTAATCACACAAATATGCAACATGATTTTAACTCTTTTTCCTTT
TGAGTCAGTGCCTGGTCATCACAAACGCCATCCTTTCCTGATTGGCTGCGAACGGTCCGGCGTGCAGGCAGGA
AGCTGAGCCAGGCTTTGCTTCCTGAGCGTCCGAGAGAGGAAGCCCTGTGATAGGCGAGAGGGAGTGGGGTT
AAAAATAGTGGCGTATATAAAGGCTCGCTCAGAATCCCATTCCGTCTCAAACTCTCACTCGCACACTTTCTGGC
AGGACTCTCGGTTTCCTTCTCTCTCGAAAGCAGGGCCCAGGGCAGCCGCAACGCCGTCCTGCTGTGAAGAAAG
AGCAGGTGTTTCCCAGGGTTTGACTCAAGCAGCAAAATATGCATTACGTTGGATAACCAAAGCAAGAAAAGA

AAAAAGAAAACCTTTTTCCAAATAAATGCTCAGTGCAGTTTACTTGAGCTCGGTTCACACGCAAAGCATGATAC
CTTTTCCAAAACG

klf2a-6/419 bp (zebrafish sequence tested in zebrafish)

CAAACTCCCTTTCAGACAATCACTTGCTCCGCTAAAATCTAATCACACAAATATGCAACATGATTTTAACTCTTT
TTCCTTTTGAGTCAGTGCCTGGTCATCACAAACGCCATCCTTTCCTGATTGGCTGCGAACGGTCCGGCGTGCAG
GCAGGAAGCTGAGCCAGGCTTTGCTTCCTGAGCGTCCGAGAGAGGAAGCCCTGTGATAGGCGAGAGGGAGT
GGGGTTAAAAATAGTGGCGTATATAAAGGCTCGCTCAGAATCCCATTCCGTCTCAAACTCTCACTCGCACACTT
TCTGGCAGGACTCTCGGTTTCCTTCTCTCTCGAAAGCAGGGCCCAGGGCAGCCGCAACGCCGTCCTGCTGTGA
AGAAAGAGCAGGTGTTTCCCAGGGTTTGACTCAAGCAGCAAAATATGCATTA

klf2a-6/200 bp (zebrafish sequence tested in zebrafish)

CTTTCCTGATTGGCTGCGAACGGTCCGGCGTGCAGGCAGGAAGCTGAGCCAGGCTTTGCTTCCTGAGCGTCC
GAGAGAGGAAGCCCTGTGATAGGCGAGAGGGAGTGGGGTTAAAAATAGTGGCGTATATAAAGGCTCGCTCA
GAATCCCATTCCGTCTCAAACTCTCACTCGCACACTTTCTGGCAGGACTCTCGGTTT

klf2a+23 (zebrafish sequence tested in zebrafish)

TTTGTGTGCATGCTCTCTGTCTGTGGCTGATTCAAAGTGGGGACTAGGAGATCTAACATGGCATTACTGCCAA
AAGTTCGGTCAAATATTGCACAGTAATGAGCAAAATTAGCTTTAACTTGGCAAATGACTACAGAGCAAACGGG
ATGATCAACTTCCTGGTTCACTTCCTGGTGACAACCTCTGCTTCACATTGTGCTGCCTCCATGCCGTGCATTTAA
AAGCCTTAAACACAAGTCTCTTATTTACATCACTTTAACTTTACATTTTAACTTTACATTAAACTGACATTTTTTTA
AAAAGTTATACATACTGTTTTATGTTAAATGAGTTATAATTAACATAAAAACTTATGTTGTCTTACTGATCACTA
TATAATTTAACAGTTACAATGTTTGAAATGTGTTTTAAGAGAAGAACACAAGTAAATGTGAATTATTCAGGGT
GAAAAGTTGAGGCATCTAGGCCACACTACCCACTACATGTGTAAACACACACACGCACACATACAGGAAGTG
GGGAGGGGTGTGTATAGATCAGTGCTTGATGCTGGCAGTGCAGGCTAACCACAGATCATCGAAGCATGTTAG
CAATACCACCCTAAAAACACCTAACAAACCTCCTTCTTTAACCTCCATACCAAAAATTAAACCTCAGTTACACCA
GGAAAGCGCTCACTTGTTTTCTAGGATCTAAAAAAAAAACATAAACATGCTTCTGCAGGAAAAAAGAGTGAGT
TGCACTGATTCTCCTTTTCCTCTAAATGGCGACAGCTCTGAGGAAACGCTGCAG

2.5.4.3 Mutated KLF2-41 human sequences (tested in mice)

KLF2-41 mut MEF2

GGTTGCACAACACTGCATGTATAAAATGCCACTACTTTAAAAATGGTTAATTTCATGTGAATTGCACTTTAATA
AAAATAAACAAGTATTTAAAAATCCACCAATGCAAACACTTCTCTGTTCCTCTGTGGTGGAGGAAGGGACACG
CGCTTTTTTTTCCGACCTTAGGAAGGAACAAGGGAGCCGGGGTCCCCTCCCAGCCTGGGAGCCCTGGGCACA
GTCCCGGCTCATTTGTCAGAGCTATCGGAGCCGTCCTCGGGCTGGTGGGAGTTCAGGGCTCTGAAAGGTTTTC
TGTCAAGGCTTGAAAGGGGGCCAGGTTTTTTTCCCCCCGGAGCCGCGCAGTCTCGGGGCTGTTGTTCTCAGCA
ATCGCAGGGCCTCGTGTTAGCAGGAAGCACAGCCAAGTAGGGTTTCCTGCGTGTTGGAGAGAGGAAGCTCCG
TAATGTTCTGGGAGGCGATG**GTGCAAAAC**GACTCCGGTATATAAAGACAGCGGAGGGTCCCCTTGTTCGCTC
ACTCGGGCGCCGGCCGGCTGGACGCAGGGCCGAGCAGGTGGTTTGGGGCCTCGGGAAGGCCAAACCCCCGC
CTCTGGGCCCCTGGCTGGGGAAGACACCAGCCAAGTTCAGAGCCCCAAGTCGGCCTCACTTCCACAACTCAGC
GTCAGGGACACCGTGGGCGTTTCTGTTTCAAAACGCTTTTCTCCAGCAAAGAACGTAACCTCAAGCTGCTGTC
AGGGTAGAGGAATCCCTGCCCCCCGCCGCCACCCAGATAGGTTGGGCCGTGATTTTGGCTGTGCCTGGCAGT
GACGATGTAGCGCACCCAGGCTGGGGGGCGCTGGCAGCCAGGGTTCAAATCCCGGCTCTGCCACCTAACGCT
GGGACTCAGTTTC

KLF2-41 mut HBOX

GGTTGCACAACACTGCATGTATAAAATGCCACTACTTTAAAAATGGTTAATTTCATGTGAATTGCACTTTAATA
AAAATAAACAAGTATTTAAAAATCCACCAATGCAAACACTTCTCTGTTCCTCTGTGGTGGAGGAAGGGACACG
CGCTTTTTTTTCCGACCTTAGGAAGGAACAAGGGAGCCGGGGTCCCCTCCCAGCCTGGGAGCCCTGGGCACA
GTCCCGGCTCATTTGTCAGAGCTATCGGAGCCGTCCTCGGGCTGGTGGGAGTTCAGGGCTCTGAAAGGTTTTC
TGTCAAGGCTTGAAAGGGGGCCAGGTTTTTTTCCCCCCGGAGCCGCGCAGTCTCGGGGCTGTTGTTCTCAGCA
ATCGCAGGGCCTCGTGTTAGCAGGAAGCACAGCCAAGTAGGGTTTCCTGCGTGTTGGAGAGAGGAAGCTCCG
TAATGTTCTGGGAGGCGATGGTTAAAAATAACTCCGG**GCGCGCCC**GACAGCGGAGGGTCCCCTTGTTCGCTC
ACTCGGGCGCCGGCCGGCTGGACGCAGGGCCGAGCAGGTGGTTTGGGGCCTCGGGAAGGCCAAACCCCCGC
CTCTGGGCCCCTGGCTGGGGAAGACACCAGCCAAGTTCAGAGCCCCAAGTCGGCCTCACTTCCACAACTCAGC
GTCAGGGACACCGTGGGCGTTTCTGTTTCAAAACGCTTTTCTCCAGCAAAGAACGTAACCTCAAGCTGCTGTC
AGGGTAGAGGAATCCCTGCCCCCCGCCGCCACCCAGATAGGTTGGGCCGTGATTTTGGCTGTGCCTGGCAGT
GACGATGTAGCGCACCCAGGCTGGGGGGCGCTGGCAGCCAGGGTTCAAATCCCGGCTCTGCCACCTAACGCT
GGGACTCAGTTTC

KLF2-41 mut RBPJ

CACAACACTGCATGTATAAAATGCCACTACTTTAAAAATGGTTAATTTCAACCAAATTGCACTTTAATAAAAAT
AAACAAGTATTTAAAAATCCACCAATGCAAACACTTCTCTGTTCCTCTGTGGTGGAGGAAGGGACACGCGCTT
TTTTTTCCGACCTTAGGAAGGAACAAGGGAGCCGGGGTCCCCTCCCAGCCTGGGAGCCCTGGGCACAGTCCC
GGCTCATTTGTCAGAGCTATCGGAGCCGTCCTCGGGCTGGTGGGAGTTCAGGGCTCTGAAAGGTTTTCTGTCA
AGGCTTGAAAGGGGGCCAGGTTTTTTTCCCCCCGGAGCCGCGCAGTCTCGGGGCTGTTG**TCCTTT**GCAATCGC
AGGGCCTCGTGTTAGCAGGAAGCACAGCCAAGTAGGGTTTCCTGCGTGTTGGAGAGAGGAAGCTCCGTAATG
TTC**AGAGGG**GCGATGGTTAAAAATAACTCCGGTATATAAAGACAGCGGAGGGTCCCCTTGTTCGCTCACTCG
GGCGCCGGCCGGCTGGACGCAGGGCCGAGCAGGTGGTTTGGGGCCTCGGGAAGGCCAAACCCCCGCCTCTG
GGCCCCTGGCTGGGGAAGACACCAGCCAAGTTCAGAGCCCCAAGTCGGCCTCACTTCCACAACTCAGCGTCA
GGGACACCGTGGGCGTTTCTGTTTCAAAACGCTTTTCTCCAGCAAAGAACGTAACCTCAAGCTGCTGTCAGGG
TAGAGGAATCCCTGCCCCCCGCCGCCACCCAGATAGGTTGGGCCGTGATTTTGGCTGTGCCTGGCAGTGACGA
TGTAGCGCACCCAGGCTGGGGGGCGCTGGCAGCCAGGGTTCAAATCCCGGCTCTGCCACCTAACGCTGGGAC
TCAG

KLF2-41 mut ETS cde

CACAACACTGCATGTATAAAATGCCACTACTTTAAAAATGGTTAATTTCATGTGAATTGCACTTTAATAAAAAT
AAACAAGTATTTAAAAATCCACCAATGCAAACACTTCTCTGTTCCTCTGTGGTGGAGGAAGGGACACGCGCTT
TTTTTTCCGACCTTAGGAAGGAACAAGGGAGCCGGGGTCCCCTCCCAGCCTGGGAGCCCTGGGCACAGTCCC
GGCTCATTTGTCAGAGCTATCGGAGCCGTCCTCGGGCTGGTGGGAGTTCAGGGCTCTGAAAGGTTTTCTGTCA
AGGCTTGAAAGGGGGCCAGGTTTTTTTCCCCCCGGAGCCGCGCAGTCTCGGGGCTGTTGTTCTCAGCAATCGC
AGGGCCTCGTGTTAGCA**GAGA**GCACAGCCAAGTAGGG**TCTC**CTGCGTGTTGGAGA**GAGA**GAGCTCCGTAAT
GTTCTGGGAGGCGATGGTTAAAAATAACTCCGGTATATAAAGACAGCGGAGGGTCCCCTTGTTCGCTCACTCG
GGCGCCGGCCGGCTGGACGCAGGGCCGAGCAGGTGGTTTGGGGCCTCGGGAAGGCCAAACCCCCGCCTCTG
GGCCCCTGGCTGGGGAAGACACCAGCCAAGTTCAGAGCCCCAAGTCGGCCTCACTTCCACAACTCAGCGTCA
GGGACACCGTGGGCGTTTCTGTTTCAAAACGCTTTTCTCCAGCAAAGAACGTAACCTCAAGCTGCTGTCAGGG
TAGAGGAATCCCTGCCCCCCGCCGCCACCCAGATAGGTTGGGCCGTGATTTTGGCTGTGCCTGGCAGTGACGA
TGTAGCGCACCCAGGCTGGGGGGCGCTGGCAGCCAGGGTTCAAATCCCGGCTCTGCCACCTAACGCTGGGAC
TCAG

2.5.4.4 Mutated KLF2-66 human sequences (tested in mice)

KLF2-66 mutHBOX

TAAAAAATAAATGAAATGGAGACCCCAGTACCCCCCGTTGATAGATAAGAGACTGAGATTTCTAGAGGGGCT
CAGCTTCCCTGGGGGACAGCTGTTTAGACCCAGAGCCAATTTTCAACTCTTGGGCAGCCAGACTGCAAGTCCA
GTGCTCCTTCCAAAACTTGTGTGTATGGAGAACCCACCACCTGTGCCCTGACCCCCTCCTGGTTCCCCACAGCC
CCTGCCCAGGCCAGGACACCCTGTCCCCTGGGTGTCCCTGAGGTCACAGATAAGGCCGCTGGGCCCGGCTATC
GGCAGGAAACATCTTGGCAGGACAGGAAGCGCGGCATGGGTCAGCCGCGCCTGCACCCGGCTCAGCCTATC
AGCGGCCAGATATTTTGGGAGCCAC**CTATCGCG**CCCCAACTCGGCTCCAAGCTGTTCCCCGGGGAAGCCTCAG
GCTTCATGTTCTGCGGCAGCCCTGAGGCTGCTGGAACACGGGGCCCGGGAGGAGCTCAGGAGCTGAGCGGG
ACAGAGTGACAGGGGTGGACTCCGTCCGGGCATCACTGCAGGCCAGAGGGTGCTGGCTCTGAATCGAGCTA
GCCCCTATTCAGTCCTGGTGCAGACACTCTCACTTGCTGGGCAAGTGGCTTCACCGCCCTGTGCCTCAGTTTCC
CCCTCCCCACATCCCTCCTTTGATCCTGTCCCAGGGAAAAACACACCTCCAGGTCCTGGAGCCTTTTCTGAAGTT
TCTGGGCCGACATCTGAGGGCCAACCCACTTTCTCCCAATGCAAAGCAGGGCCCAGGAATCTGCATCTTAAAA
CCTGCCTTGGCCGGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCAT
GAGGTCAGGCATTCGAGACTAGCCTTG

2.5.4.5 Phylogenetically foot printed enhancer sequences

tdKLF2-66 (orthologous Tasmanian devil sequence tested in mice)

AGGTACCACCACTATCATGCTACCCGCCCTGATACACTGGGTTCTCTTCTTCTCCCTCTCTGACTCCCGAGCTGT
GAGGTCACATTCTAGAGTTCTGTTTCTCCAGTTCAGCAGCCCTCTTCCTCCCCGGGCAGCTTGACAGCCAGCCG
TAGCAAGCAGGGGGGCAGACAGCCCGGAGTTTATTTTTTTTTCCCTGCTTGAGTTGTTTCACCAGCTCTGGCAT
GGAGAGGCACGTGGACCTGAGGGTGCTGCGAACTTTCCGATGGCGAAGCTCTGGCAGGGGTTACAGGGGTC
AGTGATAACCCTGGGAGTTCAGAGCAACGATGTGGCCCATGCTATCAACAGGAAACATCTTGGAGCAGCAGC
ACAGGAACTGAGGGATTGTCATTGGCCCGGGCCAGAAGCAGGACCACGGCTGCTCGGCCTATCAGCGGGCA
GATATTTTGGGAGCTGCTATATAAGCTCCTCCAAGCACACAGTGTCTCTTGGACACAGGAAAACAAGCTCATG
TGGAAAACTCAGCCGGGTCCTGGGGCCGCAGCCTTTTACTGCCCTATGAATACTCAGTCCGGCGTGATGGTTC
CTCCGTACGGCTAAGGGGTGGGAGGGAGGGTGTAAATATTAACCTGGGGGGAGGGAGGTGCTTAGATTTTG
GGCCGGTCATGCACTTGGATAGAAAAATGACATCGTTAATTTCAATAAAGTTAGTTTTCTTTGTAATCCTATGT
AGTTTATTTTATGCATTTAAACACAGATTTTTAGACGAGGGTCCCTAGGCTTCACCAGACAGTTTCTTCCACAG
AAGGTTAAGAACCCCTGATTTAGGGGGTGGGAGTAATTTAGGACTGTAAATCTAAAGCTAGAAGGGACACCA
GCAAACATCTGGGCCACTTTCCTCATTTTATAGAGATGGAA

klf2a-6 (orthologous zebrafish sequence tested in mice)

TGAGCTCATCGTTAATCTTAAGGGCTAAAAAAAAAACGTGAGCCACACGTCAAACTCCCTTTCAGACAATCACT
TGCTCCGCTAAAATCTAATCACACAAATATGCAACATGATTTTAACTCTTTTTCCTTTTGAGTCAGTGCCTGGTC
ATCACAAACGCCATCCTTTCCTGATTGGCTGCGAACGGTCCGGCGTGCAGGCAGGAAGCTGAGCCAGGCTTT
GCTTCCTGAGCGTCCGAGAGAGGAAGCCCTGTGATAGGCGAGAGGGAGTGGGGTTAAAAATAGTGGCGTAT
ATAAAGGCTCGCTCAGAATCCCATTCCGTCTCAAACTCTCACTCGCACACTTTCTGGCAGGACTCTCGGTTTCCT
TCTCTCTCGAAAGCAGGGCCCAGGGCAGCCGCAACGCCGTCCTGCTGTGAAGAAAGAGCAGGTGTTTCCCAG
GGTTTGACTCAAGCAGCAAAATATGCATTACGTTGGATAACCAAAGCAAGAAAAGAAAAAA

pKLF2-41 (orthologous platypus sequence tested in mice)

GTGGAAGTGGTATAAATAATATTTTCTACTATAAAAAAATGTAAACGTAGAAAAGGGCTATTTACATGGGAGC
ATTTGGTGCCGTTTCAAGCTAATTTGTCACTGATATAGGAGCCTCATCCGCTAGCGGGAGAGAATCCAGAGCT

TTCAAAGGTTTTCTGTCAGGATTTGAAAAGGTGCCAGGTTTTTTTTAATGGAGCCAAGCAGTCTCAGGCCTGTT
GTTCTCAGAGATTTTGTGAGCCAGGGCCGCATGTTAGCAGGAAACAGAGCCAAGTGGACTTCCTGAGAGTCT
GAGAGAGGAAACACTGTAATAAGCATGGAGGCTGTGGTTAAAAATAACTCTGGTATATAAAGGCTGCAAAGG
GTCCCCTCTCACTCTCACACACGCCGCTGCCTCTTGGCCGCTTCAGGCTGACGGCCGGAGCGCCGAGACGGAG
AGCAGGTGGTTTGGGAGCGAAAGGAGGCAGAAGACATTTCTCTGGACAGTGAAACCTCAGGATTAAAGGCT
CCCTGGTTTCCAGAAATCAGAGCCCCGAGGGGTTTGGGGTTTTCATCGTGGAGAGCGACCCTGTTTTCAAGTG
TTCAAGGGAGCTGGACGTCTTAGCATCTAAACTG

2.6 Electrophoretic mobility shift assay

2.6.1 Plasmids for generation of recombinant proteins

Recombinant ETS1-DBD, MEF2A, MEF2C, MEF2D, RBPJ, NFATC1, FOXC1, FOXC2, and GATA2 proteins for EMSA were *in vitro* transcribed and translated from pCITE2 (ETS1 DBD, MEF2C, NFATc1, and GATA2) or pcDNA3 (RBPJ, MEF2A, MEF2D, FOXC1, FOXC2) vectors using the TNT Quick Coupled Transcription/Translation System (Cat.# L1170, Promega) using the T7 polymerase.

2.6.2 Probe labelling and gel shifts

Double-stranded oligonucleotides (**Tables 2.7, 2.8, 2.9, 2.10,** and **2.11**) were annealed, then labelled with ^{32}P-dCTP, using DNA Polymerase I Large (Klenow) Fragment (Cat.# M2201, Promega) to fill in overhanging 5' ends, and purified on a non-denaturing polyacrylamide-TBE gel (10%). Probe concentration was measured and diluted to 20-40000 counts per minute/µL in 40mM KCl.

For the gel shift, binding reactions containing recombinant protein and probe were pre-incubated at room temperature in 1x binding buffer (40mM KCl, 15mM HEPES pH 7.9, 1mM EDTA, 0.5mM DTT, 5% glycerol), 2 µLpoly dI-dC (Cat.# 20148E, Thermo Fisher Scientific), and 100 x excess competitor DNA where indicated for 10min before the addition of radiolabelled probe. Reactions were incubated for an additional 25min at room temperature after probe addition and electrophoresed on a 10% (ETS) or 6% (MEF2,RBPJ, GATA2, NFATC1, FOXC1, FOXC2) non-denaturing polyacrylamide gel for 2h. Gels were dried for 1 h and 45min at 80°C and visualised by autoradiography.

Sequence name	Forward sequence	Reverse sequence
Rbpj_control site	ctagTTATTTTTGGCG**TGGGA**ACG CGGGGAGCAC	ctagGTGCTCCCCGCGT**TCCCA**CGCCAAAAATAA
Mef2_control site	ctagGATCCTCATCTT**TTAAAAATAA**CTTTTCAAA AG	ctagCTTTTGAAAAG**TTATTTTTTAA**AAGATGAGG ATC
Ets1_control site	ctagGGAGTTACTCTC**TTCC**TGTTATG	ctagTGTCATAACA**GGAA**GAGAGTAA
Gata2_control site	ctagGCTAGGT**TTATCA**CTGCCTC	ctagGAGGCAG**TGATAA**ACCTAGC
Nfat_control site	ctagCTTCATTT**TTTCC**ATTTACTGCAGA	ctagTCTGCAGTAAAT**GGAAA**AAATGAAG
FoxC1/2_control site	ctagGGAGGAGCAGC**CTGTTTGTTTGC**CAGATCT GTGC	ctagGCACAGATCTG**GCAAACAAACAG**GCTGCT CCTCC

Table 2.8: Control oligonucleotides.

Sequence name	Forward sequence	Reverse sequence
KLF2-41 Rbpj_ab WT	ctagGGGTCCCC**TCCCA**GCC**TGGGA**GCCCTGGGC	ctagGCCCAGGGC**TCCCA**GGC**TGGGA**GGGGACC C
KLF2-41 Rbpj_ab MT	ctagGGGTCCCC**TACAG**GCCG**TGAAG**CCCTGGG C	ctagGCCCAGGGC**TTCAC**GGC**CTGTA**GGGGACC C
KLF2-41 Rbpj_c WT	ctagCTCGGGCTGG**TGGGA**GTTCAGGGC	ctagGCCCTGAAC**TCCCA**CCAGCCCGAG
KLF2-41 Rbpj_c MT	ctagCTCGGGCTGG**GAGTA**GTTCAGGGC	ctagGCCCTGAAC**TACTC**CCAGCCCGAG
KLF2-41 Rbpj_d WT	ctagGGCTGTTGT**TCTCA**GCAATCGCAG	ctagCTGCGATTGC**TGAGA**ACAACAGCC
KLF2-41 Rbpj_d MT	ctagGGCTGTTGT**TATAG**GCAATCGCAG	ctagCTGCGATTGC**CTATA**ACAACAGCC
KLF2-41 Rbpj_e WT	ctagCGTAATGTTC**TGGGA**GGCGATGGT	ctagACCATCGCC**TCCCA**GAACATTACG
KLF2-41 Rbpj_e MT	ctagCGTAATGTTC**CAGTA**GGCGATGGT	ctagACCATCGCC**TACTG**GAACATTACG
KLF2-41 Mef2_a WT	ctagGGCGATGG**TTAAAAATAA**CTCCGGT	ctagACCGGAG**TTATTTTTAA**CCATCGCC
KLF2-41 Mef2_a MT	ctagGGCGATGG**TGCAAAACGA**CTCCGGT	ctagACCGGAG**TCGTTTTGCA**CCATCGCC
KLF2-41 Mef2_b WT	ctagTAACTCCG**GTATATAAAG**ACAGCGGA	ctagTCCGCTTGT**CTTTATATAC**CGGAGTTA
KLF2-41 Mef2_b MT	ctagTAACTCCG**GCCGCACACA**CAAGCGGA	ctagTCCGCTTGT**GTGTGCGGC**CGGAGTTA
KLF2-41 MEF2-HBOX WT	ctagTGG**TTAAAAATAA**CTCCG**GTATATAAAG**A	ctagT**CTTTATATAC**CGGAG**TTATTTTTAA**CCA
KLF2-41 mutMEF2-HBOX	ctagTGG**TGCAAGAGCA**CTCCG**GTATATAAAG**A	ctagT**CTTTATATAC**CGGAGTGCTCTTGCACCA
KLF2-41 MEF2-mutHBOX	ctagTGG**TTAAAAATAA**CTCCG**GCAACTGGTG**A	ctagT**CACCAGTTGC**CGGAG**TTATTTTTAA**CCA
KLF2-41 mutMEF2-mutHBOX	ctagTGG**TGCAAGAGCA**CTCCG**GCAACTGGTG**A	ctagT**CACCAGTTGC**CGGAG**TGCTCTTGCA**CCA

KLF2-41 ETS_a WT	ctagTTTTTT**TTCC**GACCTTA	ctagTAAGGTC**GGAA**AAAAAA
KLF2-41 ETS_a MT	ctagTTTTTT**TCTC**GACCTTA	ctagTAAGGTC**GAGA**AAAAAA
KLF2-41 ETS_b WT	ctagGACCTTA**GGAAGGAA**CAAGGGA	ctagTCCCTTG**TTCCTTCC**TAAGGTC
KLF2-41 ETS_b MT	ctagGACCTTA**GAGAGAGA**CAAGGGA	ctagTCCCTTG**TCTCTCTC**TAAGGTC
KLF2-41 ETS_c WT	ctagAGGTTTTT**TTCC**CCCCGGAG	ctagCTCCGGGG**GGAA**AAAAACCT
KLF2-41 ETS_c MT	ctagAGGTTTTT**GGAC**CCCCGGAG	ctagCTCCGGGG**GTCC**AAAAACCT
KLF2-41 ETS_d WT	ctagTGTTAGCA**GGAA**GCACAGCC	ctagGGCTGTGC**TTCC**TGCTAACA
KLF2-41 ETS_d MT	ctagTGTTAGCA**TTCA**GCACAGCC	ctagGGCTGTGC**TGAA**TGCTAACA
KLF2-41 ETS_e WT	ctagAGTAGGGT**TTCC**TGCGTGTT	ctagAACACGCA**GGAA**ACCCTACT
KLF2-41 ETS_e MT	ctagAGTAGGGT**GGAC**TGCGTGTT	ctagAACACGCA**GTCC**ACCCTACT
KLF2-41 ETS_f WT	ctagTGGAGAGA**GGAA**GCTCCGTA	ctagTACGGAGC**TTCC**TCTCTCCA
KLF2-41 ETS_f MT	ctagTGGAGAGA**TTCA**GCTCCGTA	ctagTACGGAGC**TGAA**TCTCTCCA
KLF2-41 ETS_g WT	ctagGGGCCTCG**GGAA**GGCCAAAC	ctagGTTTGGCC**TTCC**CGAGGCCC
KLF2-41 ETS_g MT	ctagGGGCCTCG**TTCA**GGCCAAAC	ctagGTTTGGCC**TGAA**CGAGGCCC
KLF2-41 ETS_h WT	ctagCTGGCTGG**GGAA**GACACCAG	ctagCTGGTGTC**TTCC**CCAGCCAG
KLF2-41 ETS_h MT	ctagCTGGCTGG**TTCA**GACACCAG	ctagCTGGTGTC**TGAA**CCAGCCAG
KLF2-41 NFAT_a WT	ctagGCTTTTT**TTTCC**GACCTTAG	ctagCTAAGGTC**GGAAA**AAAAAGC
KLF2-41 NFAT_a MT	ctagGCTTTTT**CAAAT**GACCTTAG	ctagCTAAGGTC**ATTTG**AAAAAGC
KLF2-41 NFAT_b WT	ctagGGTTTT**TTTCC**CCCCGGAG	ctagCTCCGGGG**GGAAA**AAAACC
KLF2-41 NFAT_b MT	ctagGGTTTT**GATAG**CCCCGGAG	ctagCTCCGGGG**CTATC**AAAACC
KLF2-41 NFAT_c WT	ctagGTAGGG**TTTCC**TGCGTGTT	ctagAACACGCA**GGAAA**CCCTAC
KLF2-41 NFAT_c MT	ctagGTAGGG**ACACG**TGCGTGTT	ctagAACACGCA**CGTGT**CCCTAC

Table 2.9: KLF2-41 enhancer oligonucleotides.

Sequence name	Forward sequence	Reverse sequence
KLF2-66 RBPJ_a WT	ctagAAGAGAC**TGAGA**TTTCTAGA	ctagTCTAGAAA**TCTCA**GTCTCTT
KLF2-66 RBPJ_a MT	ctagAAGAGAC**TTTTT**TTTCTAGA	ctagTCTAGAAA**AAAAA**GTCTCTT
KLF2-66 RBPJ_bc WT	ctagCCACGTG**TCCCT**GAGG**TCACA**GA	ctagTC**TGTGA**CCTC**AGGGA**CACGTGG
KLF2-66 RBPJ_bc MT	ctagCCACGTG**GAAAG**GAGG**TCACA**GA	ctagTC**TGTGA**CCTC**CTTTC**CACGTGG
KLF2-66 RBPJ_d WT	ctagATATTT**TGGGA**GCCACTA	ctagTAGTGGC**TCCCA**AAATAT
KLF2-66 RBPJ_d MT	ctagATATTT**CAAGA**GCCACTA	ctagTAGTGGC**TAAAC**AAAATAT
KLF2-66 RBPJ_e WT	ctagGCAGACAC**TCTCA**CTTGCTG	ctagCAGCAAG**TGAGA**GTGTCTGC
KLF2-66 RBPJ_e MT	ctagGCAGACAC**AGCGA**CTTGCTG	ctagCAGCAAG**TCGCT**GTGTCTGC
KLF2-66 RBPJ_f WT	ctagGATCCTG**TCCCA**GGGAAA	ctagTTTCCC**TGGGA**CAGGATC
KLF2-66 RBPJ_f MT	ctagGATCCTG**GGGGG**GGGAAA	ctagTTTCCC**CCCCC**CAGGATC
KLF2-66 MEF2_a WT	ctagCGGCCAG**ATATTTTCTG**CGGC	ctagGCCG**CAGAAAATAT**CTGGCCG
KLF2-66 MEF2_a MT	ctagCGGCCAG**ATAGCTTTT**CTGCGGC	ctagGCCGCAG**AAAAGCTAT**CTGGCCG
KLF2-66 MEF2_b WT	ctagAGCCA**CTATATAAAC**CCCAA	ctagTTGGG**GTTTATATAG**TGGCT
KLF2-66 MEF2_b MT	ctagAGCCA**CGAGAGAAAC**CCCAA	ctagTTGGG**GTTTCTCTCG**TGGCT
KLF2-66 ETS_a WT	ctagCTCAGC**TTCC**CTGGGGGAC	ctagGTCCCCCAG**GGAA**GCTGAG
KLF2-66 ETS_a MT	ctagCTCAGC**TCTC**CTGGGGGAC	ctagGTCCCCCAG**GAGA**GCTGAG
KLF2-66 ETS_b WT	ctagAGTGCTCC**TTCC**AAAACTT	ctagAAGTTTT**GGAA**GGAGCACT
KLF2-66 ETS_b MT	ctagAGTGCTCC**TGAC**AAAACTT	ctagAAGTTTT**GTCA**GGAGCACT
KLF2-66 ETS_c WT	ctagTCCTGG**TTCC**CCACAG	ctagCTGTGG**GGAA**CCAGGA
KLF2-66 ETS_c MT	ctagTCCTGGT**GACC**CACAG	ctagCTGTGGG**TTAC**CAGGA
KLF2-66 ETS_d WT	ctagATCGGCA**GGAA**ACATCTTGGC	ctagGCCAAGATGT**TTCC**TGCCGAT
KLF2-66 ETS_d MT	ctagATCGGCA**GGGG**ACATCTTGGC	ctagGCCAAGATGT**CCCC**TGCCGAT
KLF2-66 ETS_e WT	ctagAGGACA**GGAA**GCGCGG	ctagCCGCGC**TTCC**TGTCCT
KLF2-66 ETS_e MT	ctagAGGACA**GCCA**GCGCGG	ctagCCGCGC**TGGC**TGTCCT

KLF2-66 ETS_fg WT	ctagGCTG**TTCC**CCGG**GGAA**GCCTC	ctagGAGGC**TTCC**CCGG**GGAA**CAGC
KLF2-66 ETS_fg MT	ctagGCTG**CCCC**CCGG**GGGG**GCCTC	ctagGAGGC**CCCC**CCGG**GGGG**CAGC
KLF2-66 ETS_h WT	ctagGCTGCT**GGAA**CACGGG	ctagCCCGTG**TTCC**AGCAG
KLF2-66 ETS_h MT	ctagGCTGCT**GCCA**CACGGG	ctagCCCGTG**TAAC**AGCAG
KLF2-66 ETS_i WT	ctagCTCAGT**TTCC**CCCTCC	ctagGGAGGG**GGAA**ACTGAG
KLF2-66 ETS_i MT	ctagCTCAGT**TAAC**CCCTCC	ctagGGAGGG**GCCA**ACTGAG
KLF2-66 ETS_j WT	ctagCTCCCCAC**ATCC**CTCCTTTG	ctagCAAAGGAG**GGAT**GTGGGGAG
KLF2-66 ETS_j MT	ctagCTCCCCAC**ACTC**CTCCTTTG	ctagCAAAGGAG**GAGT**GTGGGGAG
KLF2-66 ETS_k WT	ctagCTCCTTTGA**ATCC**TGTCCCAG	ctagCTGGGACA**GGAT**TCAAAGGAG
KLF2-66 ETS_k MT	ctagCTCCTTTGA**ACTC**TGTCCCAG	ctagCTGGGACA**GAGT**TCAAAGGAG
KLF2-66 ETS_l WT	ctagTGTCCCAG**GGAA**AAACAC	ctagGTGTTT**TTCC**CTGGGACA
KLF2-66 ETS_l MT	ctagGTCCCAG**GAGA**AAACAC	ctagGTGTTT**TCTC**CTGGGAC
KLF2-66 GATA_aa WT	ctagCCCCCCGT**TGATAGATAA**GAGAC	ctagGTCTCTTAT**CTATCA**ACGGGGGG
KLF2-66 GATA_aa MT	ctagCCCCCCGT**TGAGAGACAA**AGAGAC	ctagGTCTCTTTGT**CTCTCA**ACGGGGGG
KLF2-66 GATA_a WT	ctagGAGGTCAC**AGATAA**GGCCGCTG	ctagCAGCGGCC**TTATCT**GTGACCTC
KLF2-66 GATA_a MT	ctagGAGGTCAC**TTTCCC**GGCCGCTG	ctagCAGCGGCC**GGGAAA**GTGACCTC
KLF2-66 GATA_b WT	ctagCAGCGGCC**AGATAT**TTTGGGAG	ctagCTCCCAAA**ATATCT**GGCCGCTG
KLF2-66 GATA_b MT	ctagCAGCGGCC**TTTCCC**TTTGGGAG	ctagCTCCCAAA**GGGAAA**GGCCGCTG
KLF2-66 GATA_c WT	ctagGGCCCG**GCTATC**GGCAGGAA	ctagTTCCTGC**CGATAG**CCGGGCC
KLF2-66 GATA_c MT	ctagGGCCCG**GGAGAG**AGCAGGAA	ctagTTCCTGC**TCTCTC**CCGGGCC
KLF2-66 GATA_d WT	ctagGCTCAGC**CTATCA**GCGGCCAG	ctagCTGGCCGC**TGATAG**GCTGAGC
KLF2-66 GATA_d MT	ctagGCTCAGC**CACACA**GCGGCCAG	ctagCTGGCCGC**TGTGTG**GCTGAGC

Table 2.10: KLF2-66 enhancer oligonucleotides.

Sequence name	Forward sequence	Reverse sequence
KLF2-0.4 MEF2_a WT	ctagCGCAGC**CTAAATTTAG**CCGCG	ctagCGCAGC**CTAAATTTAG**CCGCG
KLF2-0.4 MEF2_a MT	ctagCGCAGC**CAAAAAAAAG**CCGCG	ctagCGCGG**CTTTTTTTTG**GCTGCG
KLF2-0.4 MEF2_b WT	ctagCCGCG**GTATATAAGC**CGGCGG	ctagCCGCCG**GCTTATATAC**CGCGG
KLF2-0.4 MEF2_b MT	ctagCCGCG**GTTTTTTTGC**CGGCGG	ctagCCGCCG**GCAAAAAAAC**CGCGG

Table 2.11: KLF2-0.4 promoter oligonucleotides.

Sequence name	Forward sequence	Reverse sequence
KLF4-148 ETS_ab WT	ctagGTTTCTA**ATCC**T**GGAA**GGCTGTC	ctagGACAGCC**TTCC**A**GGAT**TAGAAAC
KLF4-148 ETS_c WT	ctagCTCTCAT**TTCC**TCAGTGTTTT	ctagAAAACACTGA**GGAA**ATGAGAG
KLF4-148 ETS_c MT	ctagCTCTCAT**TCTC**TCAGTGTTTT	ctagAAAACACTGA**GAGA**ATGAGAG
KLF4-148 ETS_d WT	ctagGCTGCGC**TTCC**CTCCGAC	ctagGTCGGAG**GGAA**GCGCAGC
KLF4-148 ETS_ef WT	ctagGAGGGG**GGAA**TCT**GGAT**CAG	ctagCTG**ATCC**AGA**TTCC**CCCCTC
KLF4-148 ETS_g WT	ctagGTGATCA**GGAA**GTCTGGC	ctagGCCAGAC**TTCC**TGATCAC
KLF4-148 ETS_g MT	ctagGTGATCA**GAGA**GTCTGGC	ctagGCCAGAC**TCTC**TGATCAC
KLF4-148 ETS_h WT	ctagCCTCCAT**TTCC**CCTCCC	ctagGGGAGG**GGAA**ATGGAGG
KLF4-148 ETS_i WT	ctagTAAAG**GGAA**ACATTG	ctagCAATGT**TTCC**CTTTA
KLF4-148 ETS_j WT	ctagCACCGGTG**GGAT**GGCCTCC	ctagGGAGGCC**ATCC**CACCGGTG
KLF4-148 ETS_klm WT	ctagGGCA**GGAT**CA**GGAATCC**CGGC	ctagGCCG**GGATTCC**TG**ATCC**TGCC
KLF4-148 ETS_n WT	ctagGTCAG**TTCC**CGAGGCCG	ctagCGGCCTCG**GGAA**CTGAC
KLF4-148 MEF2_a WT	ctagGGGA**GATAAAAATAA**CCGGC	ctagGCCGG**TTATTTTTAT**CTCCC
KLF4-148 MEF2_a MT	ctagGGGA**GAGAGGAACTT**CCGGC	ctagGCCGG**AAGTTCCTCT**CTCCC
KLF4-148 HBOX_a WT	ctagCCGG**CATATTTAAG**GCAG	ctagCTGC**CTTAAATATG**CCGG
KLF4-148 HBOX_a MT	ctagGCCG**GCCAGCGAGT**GGCAG	ctagCTGC**CACTCGCTGG**CCGGC
KLF4-148 RBPJ_a WT	ctagCAAGCC**TCTCA**TTTCCTCAG	ctagCTGAGGAAA**TGAGA**GGCTTG

KLF4-148 RBPJ_b WT	ctagCCGC**CTGGGAG**GCTGCG	ctagCGCAGC**CTCCCAG**GCGG
KLF4-148 RBPJ_c WT	ctagGCACC**ATGTGAT**CAGG	ctagCCTGA**TCACAT**GGTGC
KLF4-148 RBPJ_d WT	ctagCACCG**GTGGGAT**GGCCTTC	ctagGAAGGCC**ATCCCAC**CGGTG
KLF4-148 RBPJ_e WT	ctagCGGCG**CTCACAC**GCGGCC	ctagGGCCGC**GTGTGAG**CGCCG
KLF4-148 GATA_a WT	ctagACGTAGGA**GGATAA**AAATAAC	ctagGTTATTT**TTATCC**TCCTACGT
KLF4-148 GATA_a MT	ctagACGTAGGA**TTTCCC**AAATAAC	ctagGTTATTT**GGGAA**ATCCTACGT
KLF4-148 NFAT_a WT	ctagCTCTCA**TTTCC**TCAGTG	ctagCACTGA**GGAAA**TGAGAG
KLF4-148 NFAT_b WT	ctagCCTCCA**TTTCC**CCTCCC	ctagGGGAGG**GGAAA**TGGAGG
KLF4-148 NFAT_c WT	ctagCTAAAG**GGAAA**CATTGTG	ctagCACAATG**TTTCC**CTTTAG
KLF4-148 NFAT_c MT	ctagCTAAAG**AAAAA**CATTGTG	ctagCACAATG**TTTTT**CTTTAG

Table 2.12: KLF24-148 enhancer oligonucleotides.

3 Identification of endothelial-specific enhancers for *KLF2*

3.1 Rationale

3.1.1 Utilising enhancer analysis to understand regulatory pathways controlling endothelial responses to shear stress

Transcriptional changes in response to shear stress have already been studied using EC culture and unbiased systems biology approaches[194,256–263]. However, cell lines are rarely representative of the dynamic developmental processes that occur *in vivo*, and -omics analyses are often carried out at specific time points after shear stress application, resulting merely in snapshot profiles of altered gene expression[210]. Consequently, the role and relevance of many genes thought to be induced by shear stress *in vitro* is yet to be established in continuous and *in vivo* settings. Some endothelial genes that emerged from studies of hallmark endothelial shear stress responses *in vitro* have been analysed using endothelial-specific knockout mouse models[229,264,265]. However, the loss of these genes often results in early lethality, muddling the determination of their role in shear stress responses and preventing investigations into adulthood[229,266]. The use of inducible endothelial-specific knockout technologies has allowed for the circumvention of this problem, but many genes influenced by shear stress encode TFs that belong to large protein families with compensatory potential[241,267–269], further complicating identification of TFs required, and characterisation of their precise role in shear stress-induced transcriptional regulatory pathways. Due to the limitations of these studies, our understanding of transcriptional networks governing endothelial responses to shear stress remains incomplete[182,194,257,270].

While mammalian core promoters have been shown to encode some TF binding information and serve as docks for the assembly of transcriptional machinery, it is widely accepted that, in mammals, promoters are not the primary regulators of complex patterns of gene expression[271]. Enhancers are composed of densely aggregated groupings of transcription factor binding motifs, and thereby allow for recruitment of TFs responsible for directing the spatio-temporal expression pattern of their cognate gene. So far, transcriptional analysis of *KLF2* has been limited to studies of the *KLF2* promoter

in cell line assays[272]. Given the requirement for *KLF2* expression for vascular maturation, a complex developmental process that requires tight transcriptional regulation, it is therefore highly unlikely that the *Klf2* promoter alone contains adequate information to direct the transcriptional regulation of *KLF2*, particularly since *Klf2* is expressed in other cell types as well as ECs[273,274]. Other endothelial genes with complex expression patterns are already known to be regulated by multiple distal enhancers in addition to their promoter regions. Enhancers for some of these genes required for vascular development have been repeatedly shown to drive robust reporter gene expression in the absence of their endogenous promoters *in vivo*[73,74,154,275]. Further, the promoter of a key developmental arterial-specifying gene *Dll4* has been shown to be entirely dispensable for *Dll4* enhancer-driven reporter gene expression in transgenic mouse assays[58]. Despite the crucial role of *KLF2* in vascular development, the current model of its transcriptional regulation by shear stress is promoter-centric, and potential regulation of *KLF2* expression by enhancer elements remains uninvestigated.

Enhancer-centric approaches for identification of direct transcriptional regulators and signalling pathways governing vascular developmental processes have been deployed by our group to understand arterio-venous differentiation[73,74,276], regulation of sprouting angiogenesis[154], and development of organ-specific vascular beds[277]. These studies demonstrated that enhancer analysis of EC subtype-specific genes can be successfully used to identify transcriptional programs governing cell identity and developmental processes. Here, a similar enhancer-centric approach was applied to investigate transcriptional regulation of endothelial responses to shear stress in the context of vascular development by analysing the gene locus of *KLF2*, a central shear-stress responsive TF.

3.1.2 Identification of endothelial-specific enhancers driving KLF2 expression in development

Despite a clear and crucial role for *KLF2* in vascular development, current knowledge of its transcriptional regulation rests on a promoter-centric model never tested in an *in vivo* setting. This chapter addresses my aim to find and characterise all endothelial-specific *KLF2* enhancers and to

evaluate the relative contributions of *KLF2* promoter and enhancer activity to *KLF2* expression in the developing endothelium. It was hypothesised that *in silico*, *in vitro* and *in vivo* analyses of the *KLF2* locus will identify one or more novel *KLF2* enhancers. Analysis and characterisation of these enhancer sequences will locate essential TF binding motifs, which will guide the identification of direct transcriptional regulators of *KLF2* during development. This knowledge will then be used to re-assess the promoter-centric model of *KLF2* regulation, either to validate or expand the current pathways thought to induce *KLF2* expression in response to shear stress. Potential identification of additional transcriptional regulators of *KLF2* in ECs would allow for analysis of previously undescribed signalling pathways that constitute the shear stress response in ECs.

3.2 KLF2 locus analysis and enhancer prediction

To understand the transcriptional regulation of *KLF2* in ECs, the first aim was to identify KLF2 enhancers within the human *KLF2* gene locus that are able to function as endothelial specific enhancers, here defined as having the ability to drive endothelial cell-specific reporter gene expression during vascular development. Once identified, the sequences of active *KLF2* enhancers could be used as a basis for identification of direct transcriptional regulators of *KLF2*, which would in turn allow for characterisation of signalling pathways that drive endothelial *KLF2* expression during vascular development.

3.2.1* An *In silico screen of the human KLF2 locus to identify putative KLF2 enhancers

An *in silico* enhancer screen of the human *KLF2* locus was carried out using the UCSC genome browser (https://genome.ucsc.edu/). An enhancer for any given gene can be positioned anywhere in the non-coding genome, including intergenic space and gene introns, far upstream or downstream of the gene in question[278]. Despite this significant variation in enhancer length and its distance from the promoter of the gene it regulates, most enhancers are thought to be 100bp-1000bp in length and lie within a

100kb of the promoter[77]. Therefore, an assumption was made that a *KLF2* enhancer would be found within a pre-defined *KLF2* locus.

"Gene locus" is a term used to describe a specified position of a gene on a chromosome. In order to focus the search for putative enhancers, the *KLF2* gene locus was therefore defined as the genomic landscape surrounding the *KLF2* gene, flanked on each side by the first neighbouring gene with no known expression in the developing endothelium (*AP1M1* on 5' and *EPS15L1* on 3' end). The resulting locus spanned 120 kb, 90 kb upstream and 30 kb downstream of the *KLF2* gene (**Figure 3.1**).

Within the 120kb region, putative enhancer sequences were identified based on a set of criteria constructed prior to locus analysis. Enhancers can be most readily distinguished from other cis-regulatory elements, such as promoters, by their epigenetic signature[279,280]. The first condition that all putative enhancers had to satisfy was therefore the presence of endothelial-specific peaks denoting histone marks characteristic of enhancer elements. The presence of specific histone marks indicative of promoters and enhancers has been mapped for different cell types, including human umbilical cord endothelial cells (HUVEC), within the scope of Encyclopaedia of DNA Elements (ENCODE) project[17,76,281]. In order to define putative enhancer sequences, three distinct epigenetic marks within the HUVEC track were considered. H3K4Me1 modification marks both active and inactive enhancers as well as promoters, whereas H3K27Ac modification marks active enhancers exclusively. Both modifications were considered positive indicators of enhancer regions. The presence of H3K4Me3 modification characteristic of promoter elements only was used to rule out promoter regions.

The second general feature of genomic regions with cis-regulatory potential is their location within areas depleted of nucleosomes characteristic of OCRs[282]. Each sequence marked with H3K4Me1 and H3K27Ac, but not H3K4Me3, modifications in the HUVEC track therefore also had to fall within an OCR detected as DNase I hypersensitivity. In order to identify endothelial specific enhancers, DNase I

hypersensitivity heat map was analysed for six endothelial cell lines, including human microvascular endothelial cells (HMVEC) and HUVEC, and 46 non-endothelial cell lines representative of other tissues, some of which were previously reported to express *KLF2*[242,283]. The details of endothelial and non-endothelial cell lines used in this analysis are given in Section 2.4. Only putative enhancers open specifically in endothelial, but not non-endothelial cell lines, were considered for further analysis.

Evolutionary conservation of the putative enhancer sequences was considered as the third and final criterion. Phylogenetic foot printing is a classic approach for predicting regulatory regions, and systematic analysis of highly conserved human non-coding DNA sequences in transgenic mouse reporter assays revealed that about half were enhancers[284]. Each putative enhancer satisfying the first two conditions was phylogenetically foot-printed to infer sequence conservation between vertebrate species including mouse, platypus, Tasmanian devil, chicken, frog, and zebrafish.

Considering the above criteria, *in silico* analysis yielded five putative *KLF2* enhancers, four upstream and one downstream of the *KLF2* gene. Each enhancer was denoted by its relative distance in kb from the TSS (KLF2-66, KLF2-41, KLF2-34, KLF2-8, and KLF2+23). All predicted enhancer sequences that were marked by endothelial-specific enhancer marks, but not promoter-specific histone marks; were located within an endothelial-specific OCR; and were evolutionary conserved at least from human to mouse. Each putative enhancer sequence was conserved further to varying degrees: for example, to egg-laying mammals (KLF2-66, 44.0% sequence identity; KLF2-8, 54.5% sequence identity; KLF2+23, 46.2% sequence identity) or birds (KLF2-34, 44.8% sequence identity). Only one putative enhancer, KLF2-41, was conserved to the zebrafish *klf2a* gene locus, although this is not visible in **Figure 3.1** due to a low degree of sequence conservation (45.7% sequence identity).

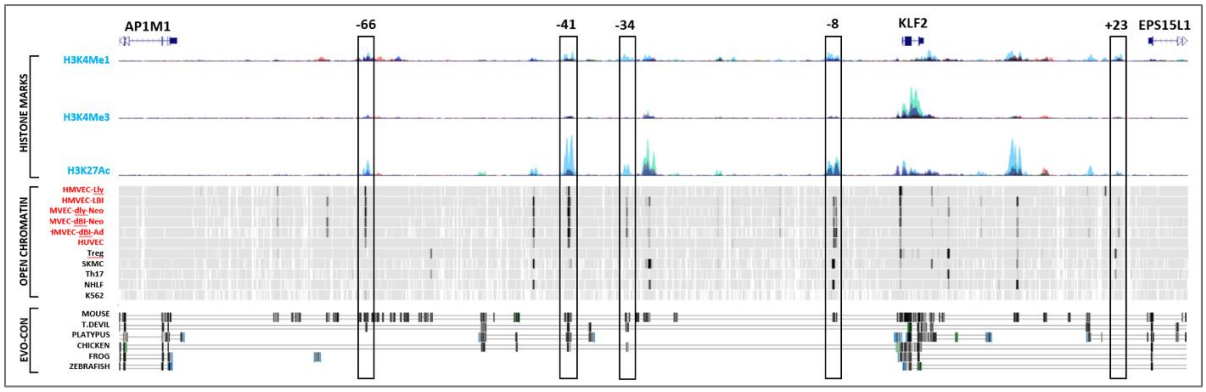


Figure 3.1: An *in silico* enhancer screen of the human *KLF2* gene locus predicts five previously uninvestigated putative enhancer sequences. The *KLF2* locus was defined as the genomic space surrounding the *KLF2* gene, flanked by one neighbouring gene on each side (*AP1M1* at 5' and *EPS15L1* at 3' end). Each putative enhancer region was denoted according to its distance from the TSS in kb (vertical black boxes). In order to qualify as a putative enhancer, each sequence had to satisfy the following criteria: 1) the presence of histone marks characteristic of enhancer (H3K4Me1 and H3K27Ac) but not promoter (H3K4Me3) elements, 2) residence within an area of DNaseI hypersensitivity reflective of open chromatin in EC (six cell lines) but not in non-EC (46 cell lines) cell types (only 5 non-endothelial cell lines are shown here), and 3) be evolutionary conserved at least from human to mouse. Conservation beyond mouse (shown for platypus, Tasmanian devil, chicken, *X. tropicalis*) was considered in order to maintain relevance to all experimental models used, but no sequence was excluded based on lack of conservation alone. The conservation of the KLF2-41 enhancer to the zebrafish *klf2a* locus is not visible in this figure due to the low degree of sequence conservation between the human and orthologous zebrafish sequences (45.7% sequence identity).

3.3 Analysis of putative KLF2 enhancers in transient transgenic zebrafish embryos

In order to establish whether human putative *KLF2* enhancers predicted *in silico* are functional during vascular development, their ability to drive *GFP* reporter gene expression in transgenic zebrafish was tested first.

3.3.1 Putative KLF2 enhancers in transient transgenic zebrafish embryos

3.3.1.1 Rationale

The ultimate aim of this project was to understand transcriptional regulation of *KLF2* in development of the mammalian endothelium. However, zebrafish have various advantages over mice as a model organism for developmental studies, including *ex vivo* embryogenesis, optical transparency of embryos and larvae, rapid embryonic development that can be tracked in real-time, low costs of maintenance, and relative ease of manipulation[285,286]. Due to these features, zebrafish have been successfully used in rapid and cost-effective developmental enhancer screens[287–289], which often

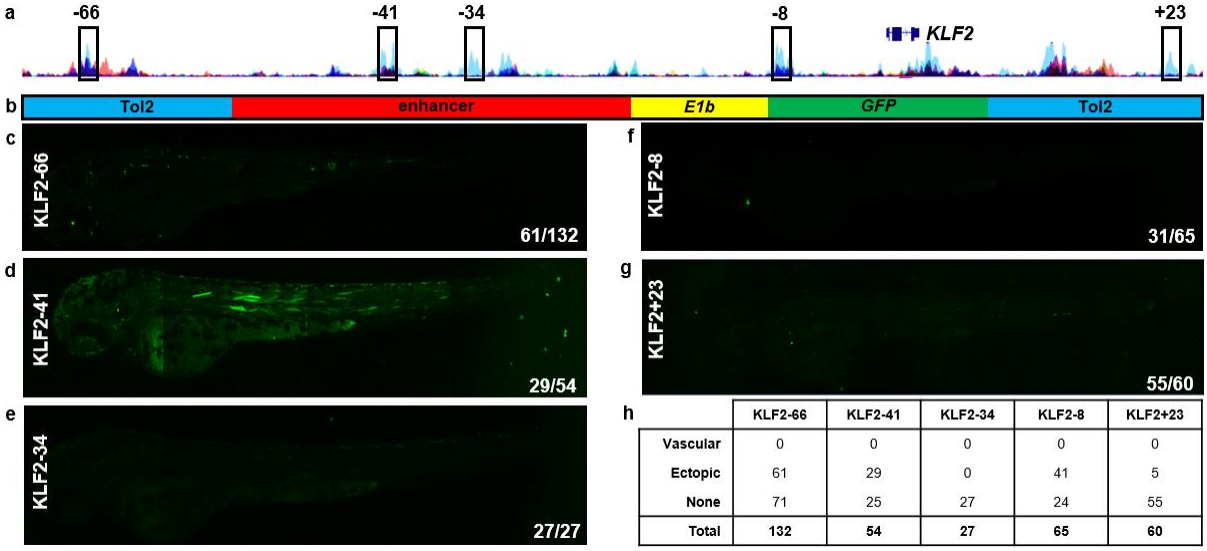
comprise the initial step in testing putative mammalian enhancer sequences predicted *in silico* for the validation of their expression *in vivo*. Working on an enhancer that is downstream of an evolutionary conserved pathway active in both zebrafish and mice would therefore allow for the use of zebrafish as the primary experimental model, with transgenic mice used to validate zebrafish findings in a mammalian system.

To test whether putative *KLF2* enhancers are functional *in vivo*, their ability to drive reporter gene expression specifically in the vasculature of F0 transgenic zebrafish embryos was investigated first. In zebrafish, a robust blood flow is established between 24 and 26 hpf[290]. As the aim was to identify shear stress-responsive regulatory elements, analysis of whether each putative enhancer could drive *GFP* expression in the developing zebrafish pre- (24hpf) and post- (48hpf) the initiation of blood flow was the starting point for *KLF2* enhancer analysis *in vivo*. In order to be tested in this zebrafish screen, each putative enhancer was cloned upstream of the E1b minimal promoter, which is by itself unable to drive reporter expression in ECs[291], and the *GFP* reporter gene into a Tol2 vector. The putative enhancer-containing Tol2 vector and Tol2 transposase mRNA were then co-injected into one cell stage zebrafish embryos, resulting in highly efficient Tol2-mediated introduction of transgene into the zebrafish germline as described previously[292]. The one cell stage zebrafish embryo microinjections were carried out by Dr Svanhild Nornes. The resulting mosaic zebrafish embryos transgenic for the enhancer:*GFP* constructs were then analysed for *GFP* expression.

3.3.1.2 Activity of human *KLF2* enhancers in transient transgenic zebrafish embryos

Putative *KLF2* enhancers predicted *in silico* were each tested in a transgenic zebrafish embryo screen in order to evaluate their ability to drive *GFP* expression in the zebrafish vasculature at 24 hpf and 48 hpf **(Figure 3.2)** Images are shown at 48 hpf, when robust blood flow should be established in all zebrafish embryos analysed. None of the human *KLF2* putative enhancer sequences were able to drive vascular *GFP* expression at any of the time points analysed, including pre- (24 hpf) and post- (48 hpf)

blood flow initiation in zebrafish. For each putative enhancer:*GFP* transgene, between 27 and 132 zebrafish were analysed, with each zebrafish representing an independent transgene integration event (**Figure 3.2h**). Ectopic *GFP* expression was detected in all but one human putative enhancer transgene (KLF2-34: *GFP*), most commonly in the yolk sac, the myocardium, and the skeletal muscle, and was likely a consequence of transgene integration into a transcriptionally permissive genomic region. This unbiased, hypothesis-free *in silico KLF2* putative enhancer screen was therefore unable to identify an enhancer element able to drive endothelium-specific reporter gene expression in zebrafish embryos.



h

	KLF2-66	KLF2-41	KLF2-34	KLF2-8	KLF2+23
Vascular	0	0	0	0	0
Ectopic	61	29	0	41	5
None	71	25	27	24	55
Total	132	54	27	65	60

Figure 3.2: Human putative *KLF2* enhancers are unable to drive endothelium-specific reporter gene expression when overexpressed in zebrafish embryos. **(a)** The human *KLF2* locus with putative enhancer regions within black boxes. **(b)** A diagram of the destination vector injected into one cell stage zebrafish embryos with each putative enhancer cloned upstream of the *E1b* minimal promoter and *GFP* reporter gene flanked by Tol2 sites. **(c-g)** Transgenic zebrafish embryos overexpressing either **(c)** the KLF2-66:*GFP* transgene, **(d)** the KLF2-41:*GFP* transgene, **(e)** the KLF2-34:*GFP* transgene, **(f)** the KLF2-8:*GFP* transgene, or **(g)** the KLF2+23:*GFP* transgene, showed no vascular *GFP* expression at 48 hpf. **(h)** A summary table showing the number of zebrafish that expressed each enhancer:*GFP* transgene analysed at 48 hpf, with *GFP* expression pattern classified as "vascular", "ectopic", or "none".

3.4 In silico screen of the zebrafish klf2a locus for identification of putative klf2a enhancers

Due to the inability of all putative KLF2 enhancers analysed to drive vascular *GFP* expression in transgenic zebrafish embryos at 48hpf, an alternative approach to testing putative KLF2 enhancers in zebrafish had to be considered. The sequence conservation of human KLF2 enhancers to zebrafish was

poor, with a single enhancer (KLF2-41) exhibiting some sequence identity (45.7%) with its zebrafish orthologue (Section 3.2.1, **Figure 3.1**). To ensure meaningful use of zebrafish as a model organism for the KLF2 enhancer screen, the possibility of a zebrafish genome-specific set of *klf2* enhancers had to be considered. The zebrafish *KLF2* orthologue, *klf2a*, is expressed after the onset of blood flow in the developing zebrafish endothelium and is thought to be regulated by shear stress in a manner similar to its regulation by shear stress in mammals[293–295]. Further, zebrafish enhancers with poor conservation to mammals have been previously shown to drive similar or identical patterns of gene expression in both fish and mammalian model systems[296,297]. It was therefore hypothesised that the human and zebrafish genomes contain distinct sets of enhancers for *KLF2/klf2*, which might not exhibit a high degree of sequence conservation but might nonetheless integrate evolutionarily-conserved transcriptional pathways that would regulate *KLF2/klf2* expression.

Due to an evolutionary whole genome duplication event in the common ancestor of teleost fish[298], zebrafish often carry two versions of the same gene with distinct sequence and therefore distinct gene regulatory properties[299]. In line with this, the zebrafish genome encodes two *KLF2* orthologues, *klf2a* and *klf2b*. Previously published *klf2a* and *klf2b* expression studies have shown robust expression of *klf2a* mRNA in the developing zebrafish vasculature after the onset of blood flow[230,300]. On the contrary, *klf2b* mRNA expression was absent from embryonic blood vessels. Therefore, only *klf2a* locus was analysed for putative zebrafish *klf2a* enhancers.

Contrary to the human genome, histone marks and DNase I hypersensitivity indicative of regulatory elements specifically in ECs have not been mapped for the zebrafish genome. *In silico* analysis parallel to that of human *KLF2* locus was therefore not possible. To predict candidate zebrafish *klf2a* enhancers, the putative enhancer-defining criteria had to be adjusted. As an alternative to analysing DNaseI hypersensitivity, a zebrafish ATAC-seq dataset (generated by Dr Kira Chouliaras, a previous PhD student in the lab) was used as an indicator of OCRs characteristic of enhancers in pan-endothelial

versus non-endothelial cell types (the datasets were provided by Dr Rui Monteiro and Dr Tomasz Dobrzycki from the WIMM, University of Oxford). Each putative enhancer predicted from the ATAC-seq was then phylogenetically foot-printed to other organisms and also to other species of fish to strengthen the rationale for testing it as a putative enhancer (data not shown).

As with the human locus, the *klf2a* zebrafish locus was defined as the genomic landscape between the first 5' and 3' neighbouring genes (*mrpl54* and *stil*) with no known expression in the developing zebrafish endothelium. The resulting locus spanned 250kb, 180kb upstream and 70kb downstream of the *klf2a* transcription start site (**Figure 3.3**). In order to qualify as a putative enhancer in this dataset, a sequence had to be located within an OCR in pan-endothelial but not non-endothelial cell types. Within the pre-defined locus, five putative zebrafish *klf2a* enhancer fitting this criterion were identified, four upstream and one downstream of the *klf2a* gene. Each enhancer was denoted by its relative distance in kb from the TSS (klf2a-83, klf2a-60, klf2a-41, klf2a-6, klf2a+23).

Consistent with the human *KLF2* enhancer prediction, only a single *klf2a* putative enhancer conserved to the human genome was identified. Phylogenetic foot-printing confirmed that the zebrafish klf2a-6 sequence mapped directly onto the KLF2-41 enhancer in the human genome, previously noted to be conserved to zebrafish (45.7% sequence identity). Consequently, the zebrafish klf2a-6 putative enhancer was from here onwards referred to as the zebrafish orthologue of the human KLF2-41 putative enhancer sequence.

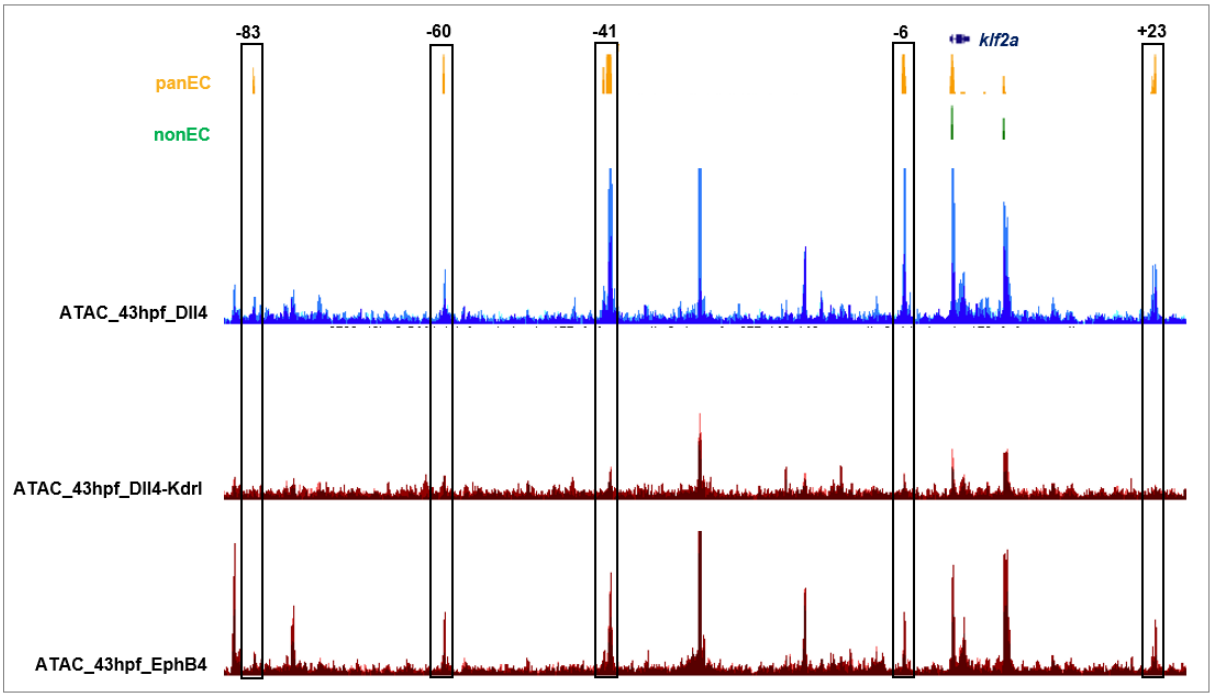


Figure 3.3: **An unbiased enhancer screen of the zebrafish ATAC-Seq dataset of open chromatin regions in different types of ECs vs non-EC predicted five putative zebrafish *klf2a* enhancer sequences.** Each putative enhancer region was named according to its distance from the TSS in kb (vertical black boxes). In order to qualify, each sequence must have been positioned within a pan-endothelial but not a non-endothelial region of open chromatin.

3.4.1 Putative klf2a enhancers in transient transgenic zebrafish embryos

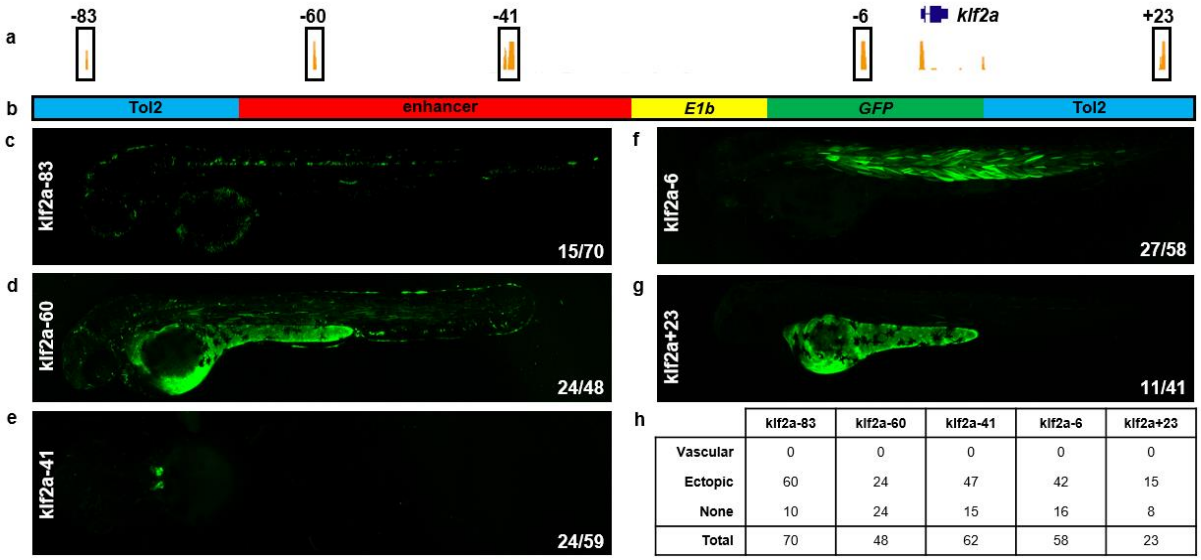
3.4.1.1 Rationale

Given none of the putative human *KLF2* putative enhancer were able to drive *GFP* expression in zebrafish embryos, the possibility of a zebrafish species-specific set of *klf2a* enhancers was tested in an *in vivo* zebrafish enhancer screen.

3.4.1.2. Putative klf2a enhancers in transient transgenic zebrafish embryos

As with the human putative enhancers, each zebrafish enhancer was cloned upstream of the E1b minimal promoter and GFP reporter gene, resulting in a construct that was injected into one cell stage zebrafish embryos. The resulting mosaic transgenic fish were analysed for GFP expression at 24 hpf and 48 hpf. For each putative enhancer:*GFP* transgene, between 23 and 70 zebrafish were analysed, with each zebrafish representing an independent transgene integration event (**Figure 3.4h**). None of

the zebrafish *klf2a* putative enhancer sequences were able to drive endothelium-specific *GFP* expression at any of the time points considered, including pre- (24 hpf) and post- (48 hpf) blood flow initiation (**Figure 3.4c-g**). Much like in the human putative *KLF2* enhancer screen in zebrafish, ectopic *GFP* expression was observed in all zebrafish putative enhancer sequences, most commonly in the yolk sac, the myocardium, and the skeletal muscle (**Figure 3.4h**).



h

	klf2a-83	klf2a-60	klf2a-41	klf2a-6	klf2a+23
Vascular	0	0	0	0	0
Ectopic	60	24	47	42	15
None	10	24	15	16	8
Total	70	48	62	58	23

Figure 3.4: Zebrafish putative *klf2a* enhancers were unable to drive endothelium-specific reporter gene expression when overexpressed in zebrafish embryos. (a) The zebrafish *klf2a* locus with putative enhancer regions within black boxes. **(b)** A diagram of the destination vector injected into one cell stage zebrafish embryos with each putative enhancer cloned upstream of the *E1b* minimal promoter and *GFP* reporter gene flanked by Tol2 sites. **(c-g)** Transgenic zebrafish embryos overexpressing either **(c)** the klf2a-83:*GFP* transgene, **(d)** the klf2a-60:*GFP* transgene, **(e)** the klf2a-41:*GFP* transgene, **(f)** the Klf2a-6:*GFP* transgene, or (g) the klf2a+23:*GFP* transgene, showed no vascular *GFP* expression at 48 hpf. **(h)** A summary table showing the number of zebrafish that expressed each enhancer:*GFP* transgene analysed at 48 hpf, with *GFP* expression pattern classified as "vascular", "ectopic", or "none".

3.4.2 Detailed analysis of the klf2a-6 putative enhancer in transient transgenic zebrafish embryos

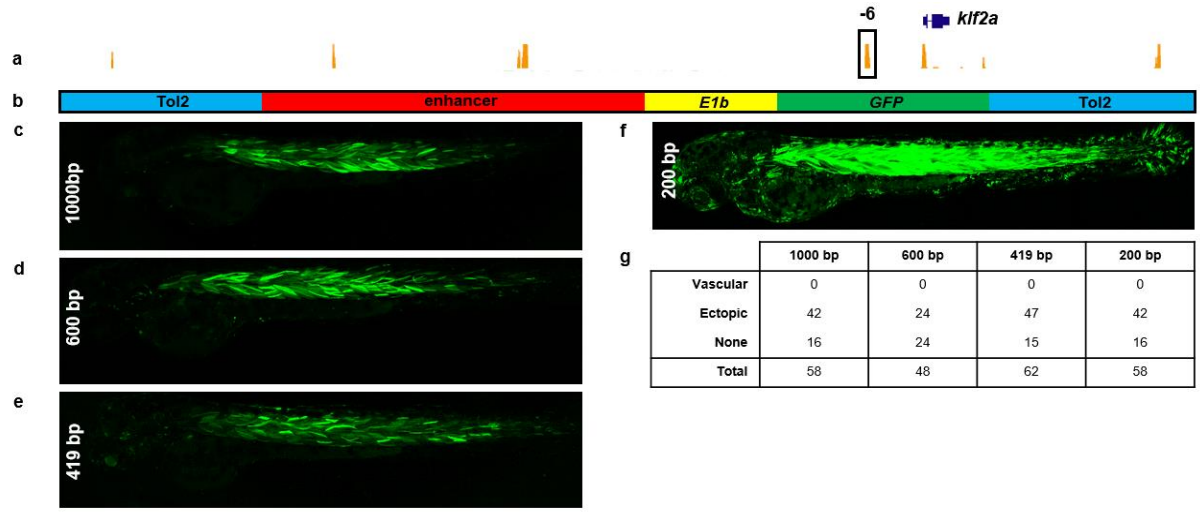
3.4.2.1 Rationale

Initial *KLF2* and *klf2a* putative enhancer screens in zebrafish failed to identify any enhancers capable of driving *GFP* expression in the vascular compartment. Although all *KLF2* and *klf2a* enhancers were cloned as 600-1000 bp long pieces, it is widely accepted that active enhancer elements vary in length from as few as 10 to as many as 1000 bp[279,301]. Due to the large variation in functional enhancer length,

it is difficult to predict the length of a putative enhancer. Extensive prior experience in enhancer analysis in zebrafish in our lab indicates that accessory sequence surrounding the core enhancer can influence enhancer-driven reporter gene expression, for instance increasing expression in ectopic areas. In order to test the possibility of surplus sequence masking the ability of *KLF2* and *klf2a* enhancers to drive *GFP* expression in transgenic zebrafish, the klf2a-6 zebrafish enhancer, which was evolutionary conserved to its human orthologous sequence, KLF2-41, was the focus of this analysis. The original 786 bp long piece of the zebrafish klf2a-6 putative enhancer was cut from each side, resulting in progressively shorter core enhancer piece with an increasing proportion of evolutionary conserved sequence. Enhancer sequences 600 bp (klf2a-6/600), 419 bp (klf2a-6/419), and 200bp (klf2a-6/600) in length were tested in transgenic zebrafish embryos and compared to GFP expression directed by the original 786 bp klf2a-6 putative enhancer piece.

3.4.2.2 Analysis of the klf2a-6 putative enhancer in transient transgenic zebrafish embryos

For each version of the putative klf2a-6:*GFP* transgene, between 48 and 62 zebrafish were analysed at 24 hpf and 48 hpf, with each zebrafish representing an independent transgene integration event (**Figure 3.5g**). Manipulating the length of the klf2a-6 enhancer strengthened *GFP* expression in ectopic areas, most notably in the skeletal muscle. The intensity of *GFP* expression increased in inverse proportion to sequence length, reaching highest intensity with the shortest piece (**Figure 3.5**). Under the control of the klf2a-6/200 enhancer piece, *GFP* expression was observed in many organs, including intersegmental vessels. However, the expression in intersegmental vessels was weak and not specific to the vasculature. Instead of leading to a more confined reporter gene expression, shortening the klf2a-6 enhancer sequence resulted in a more permissive induction of *GFP* expression. Much like the original 786 bp enhancer piece, all shorter versions of the same enhancer were also unable to drive *GFP* expression specifically in the developing zebrafish vasculature.



g

	1000 bp	600 bp	419 bp	200 bp
Vascular	0	0	0	0
Ectopic	42	24	47	42
None	16	24	15	16
Total	58	48	62	58

Figure 3.5: Shorter versions of the klf2a-6 zebrafish putative enhancer orthologous to the human KLF2-41 enhancer were all unable to drive endothelium-specific reporter gene expression when overexpressed in zebrafish embryos. (a) The location of the klf2a-6 enhancer in the zebrafish klf2a locus (black box). **(b)** A diagram of the destination vector injected into one cell stage zebrafish embryos with each putative enhancer cloned upstream of the *E1b* minimal promoter and *GFP* reporter gene flanked by Tol2 sites. **(c-g)** Transgenic zebrafish embryos overexpressing **(c)** the klf2a-6/1000bp:*GFP* transgene, **(d)** the klf2a-6/600bp:*GFP* transgene, **(e)** the klf2a-6/419bp:*GFP* transgene, or **(f)** the klf2a-6/200bp:*GFP* transgene, showed no vascular *GFP* expression at 48 hpf. **(g)** A summary table showing the number of zebrafish that expressed each enhancer:*GFP* transgene analysed at 48 hpf, with *GFP* expression pattern classified as "vascular", "ectopic", or "none".

In conclusion, the analysis of putative human and zebrafish KLF2/klf2a enhancers in transgenic zebrafish embryos did not identify any KLF2/klf2a enhancers able to drive reporter expression in the vasculature of zebrafish embryos before or after blood flow initiation.

3.5 Analysis of putative KLF2 enhancer:LacZ expression in transient transgenic mouse embryos

3.5.1 Rationale

Putative human and zebrafish *KLF2/ klf2a* enhancers were all unable to drive reporter gene expression in the developing zebrafish vasculature. While evolutionary conservation of enhancer sequences between species is a good predictor of developmental enhancers that are functional *in vivo*[302], the degree of conservation between different species varies. Predictably it tends to be higher between evolutionary closely related species and decreases as species diverged from each other throughout evolutionary history. In line with this, only one *KLF2* enhancer (KLF2-41) was conserved to zebrafish, whereas all of them were conserved to mouse. Due to a high degree of evolutionary conservation of

all predicted human *KLF2* enhancers to mouse, and to allow for the possibility of differential transcriptional regulation of *KLF2* and *klf2a* genes in mammals and fish, respectively, *KLF2* putative enhancers were next tested in transgenic mouse embryos.

Unlike zebrafish embryos that easily develop *ex vivo*, mouse embryo culture is not only more complicated, but also not feasible past E9.5[303]. Choosing the correct time point for embryo harvest is thus of crucial importance to enhancer:reporter transgene analysis, especially when it comes to generation of one-time transient embryos. For the purpose of analysing putative *KLF2* enhancers in transient transgenic mouse embryos, it was important to choose a time point after the establishment of a fully functional circulation. Embryos were harvested at E12.5, a time point well beyond the controversial onset of blood flow in mouse embryos.

3.5.2 Endogenous Klf2 mRNA expression in wild type mouse embryos

3.5.2.1 Rationale

Prior to testing putative *KLF2* enhancers, it was necessary to establish the pattern of endogenous *Klf2* mRNA expression in wild type mouse embryos at the same timepoint. Previous studies of endogenous *Klf2* mRNA expression[229,245] found it was robustly expressed in the endothelium of all major blood vessels, smaller angiogenic vessels, the OT, the ECCs, and the remainder of the endocardium of the developing heart at E9.5; at E11.5 the expression was observed in all these structures but was of much higher intensity, with stronger expression observed in arterial compared to venous ECs. At E14.5, *Klf2* mRNA was detected in endothelial cells of the umbilical artery and to a lesser extent the endothelium of the umbilical vein and was localised to the inner leaflets of mitral and aortic valves.

As endogenous *Klf2* expression at E12.5 has not been established, and this was precisely the time point at which the putative *KLF2* enhancers were to be tested, the *Klf2* mRNA expression was analysed in paraffin sections of a wild-type (WT) E12.5 embryo using RNAScope®. For this experiment, sections

with structures that were previously shown to express *Klf2* at different time points were chosen, focusing on the heart and surrounding vessels, as well as head vessels to look at *Klf2* mRNA expression in small diameter vessels, which are exposed to highest shear stress magnitudes, and the angiogenic hindbrain. In addition to the *Klf2* mRNA detection probe (Opal 570), a CD31 detection probe (Opal 750), which marks all ECs, was used also. This was combined with immunostaining for endomucin (Emcn, Alexa 488) which is expressed in most ECs, but is enriched in venous ECs and the endocardium of the heart at E12.5[304].

3.5.2.2 Analysis of *Klf2* mRNA expression in the E12.5 WT mouse embryo

In all sections analysed, *Klf2* mRNA expression largely co-localised with *CD31* and Emcn expression in ECs. High intensity *Klf2* mRNA expression was detected in the nuclei of arterial ECs, and other structures of arterial origin, such as the aortic sac, the pulmonary trunk, and the OT of the heart. Occasional *Klf2*-positive ECs were detected within the endothelium of large veins. Finally, *Klf2* mRNA expression was detected within small angiogenic vessels in the hindbrain, the branchial arches, and the developing limb buds (data not shown). In the head section of the E12.5 WT embryo, *Klf2* mRNA expression was most prominent in the internal carotid arteries and angiogenic capillaries within the hindbrain. In both these compartments, *Klf2* mRNA expression co-localised with *CD31* and Emcn mRNA expression. Limited *Klf2* mRNA expression was observed in single ECs within primary head veins and other smaller vessels where *CD31* and Emcn were robustly expressed (**Figure 3.6a**). Within the top cardiac section (**Figure 3.6b**), high intensity *Klf2* mRNA expression was detected in the aortic sac (origin of the thoracic aorta), branchial arch arteries, internal carotid arteries, and hindbrain capillaries. Individual *Klf2*-positive ECs were located within the cardinal veins. Robust *Klf2* mRNA expression was also observed within the endocardium of the OT (**Figure 3.6c and d**), with some *Klf2*-positive ECs observed in the ECC tissue and the remainder of the endocardium (**Figure 3.6d**). The overall *Klf*2 mRNA expression pattern in the E12.5 WT embryo was largely consistent with previously

reported *Klf2* expression patterns, although less prominent expression was detected in the venous endothelium compared to what was previously shown at E11.5[229].

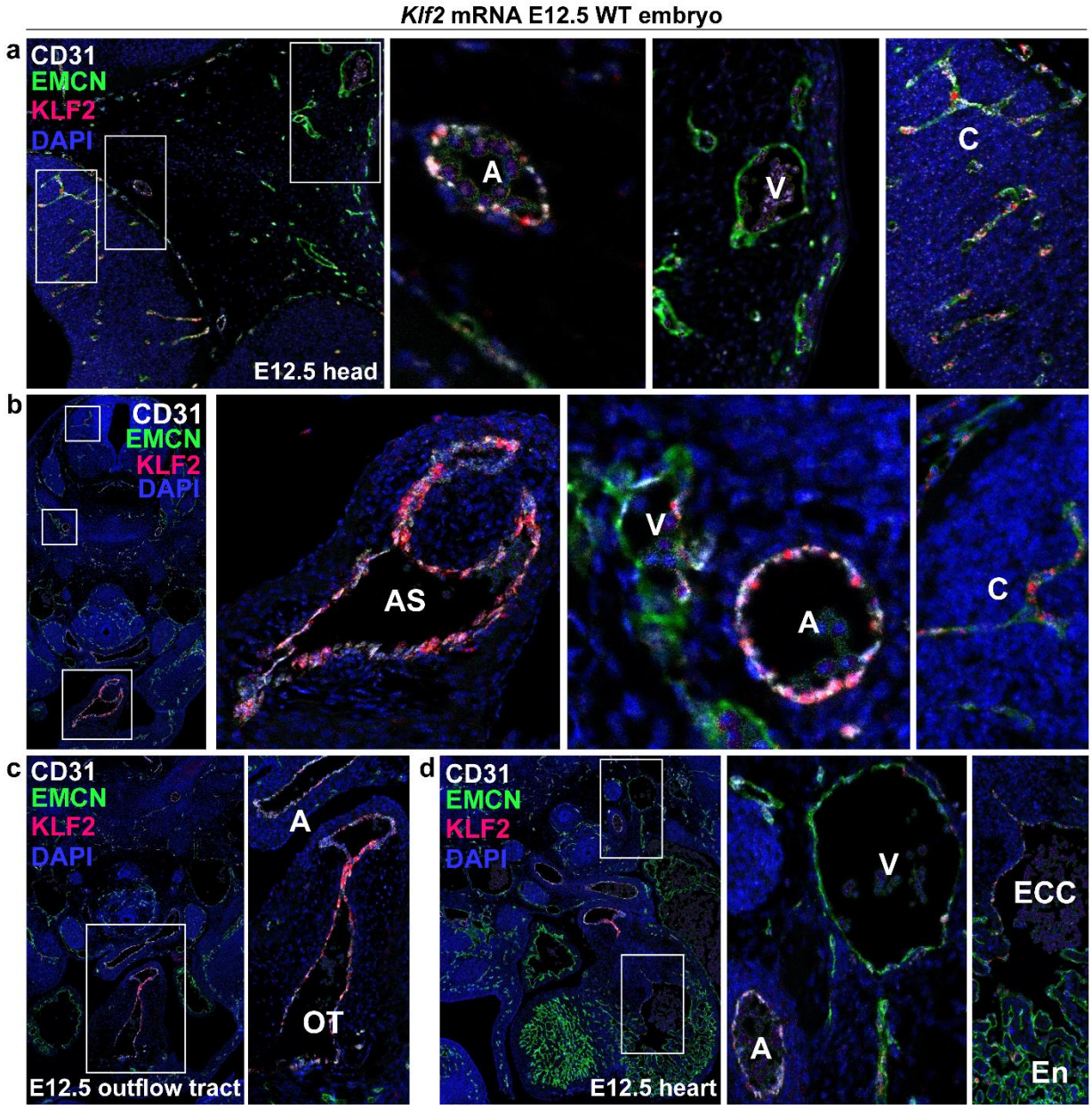


Figure 3.6: Endogenous *Klf2* mRNA expression in E12.5 WT mouse embryo sections. (a) Head section (left) with close up images of *CD31*+, *Klf2*+, Emcn+ carotid cartery, Emcn+ primary head vein, and *CD31*+, *Klf2*+, Emcn+ hindbrain capillaries. **(b)** Cardiac section (left) with close up images of *CD31*+, *Klf2*+ Emcn- aortic sac, *CD31*+, *Klf2*+ right internal carotid artery and *CD31*+, *Klf2*+, Emcn+ hind brain capillary. **(c)** Cardiac section (left) with close up images of the *CD31*+, *Klf2*+ OT and *CD31*+, *Klf2*+ pulmonary trunk. **(d)** Cardiac section (left) with close up images of *CD31*+, *Klf2*+ left doral aorta and *CD31*+, Emcn+ left cardinal vein, and *CD31*+, *Klf2*+ ECC and Emcn+, *Klf2*+ endocardium. **A= artery, V=vein, C=capillary, AS=aortic sac, OT=outflow tract, ECC=endocardial cushion, En=endocardium.** These experiments were carried out in collaboration with Dr Helena Rodriguez-Caro.

3.5.3 Putative KLF2 enhancers in transient transgenic mouse embryos

Putative *KLF2* enhancers predicted *in silico* were each tested in a transgenic mouse embryo screen in order to evaluate their ability to drive reporter gene expression in the mouse vasculature. Each enhancer sequence was cloned upstream of an hsp68 minimal promoter which is alone unable to drive reporter gene expression in ECs[305,306], and a *LacZ* reporter gene. *LacZ* activity was used as the read-out for enhancer activity within the endothelium of F0 transient transgenic mouse embryos.

Reporter *LacZ* expression, observed as blue staining following incubation of embryos with X-Gal, was analysed qualitatively, with *LacZ* expression determined to fall under one of the following categories: vascular, ectopic, or none. More detailed analysis of transgene expression was carried out after sectioning. Transient transgenic embryos were paraffin-embedded, then cut into 5µm transverse sections using the microtome (sectioning was carried out by Indrika Ratnayaka or Dr Sarah De Val). For each sectioned embryo, every 10th section was selected for analysis (in duplicate), with 50-100 sections analysed per embryo. The selected sections were then counter-stained with Nuclear Fast Red, then mounted and imaged using a slide scanner.

3.5.3.1 KLF2-41 enhancer:*LacZ* expression in transient transgenic mouse embryos

The generation of transient transgenic mouse embryos was outsourced to Cyagen, a provider of genetically-modified mouse models (the details of how transgenic embryos were generated are described in Chapter 2, Section 2.1.1). We received 13 PCR positive embryos transgenic for the KLF2-41:*LacZ* enhancer transgene (**Table 3.1**). Each of the 13 transgenic embryos represented an independent experiment as the site of transgene integration and the number of copies integrated were both random. Out of 13 embryos analysed, seven showed *LacZ* expression (**Figure 3.7**). Analysis of whole-mount embryos indicated that six out of seven embryos express *LacZ* in the vasculature. In one out of seven embryos (#41), KLF2-41:*LacZ* transgene expression was detected ectopically in the jaw region, but appeared absent from all vascular and non-vascular tissues. The pattern of vascular

transgene expression was largely consistent between transiently transgenic embryos analysed, which was especially evident following section analysis.

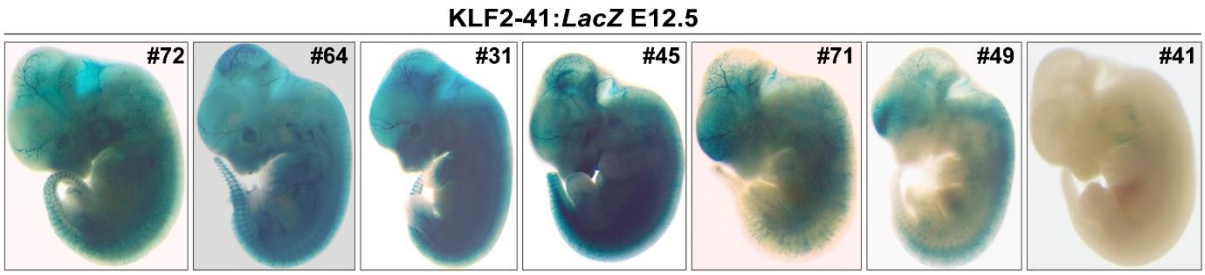


Figure 3.7: Transient transgenic KLF2-41:*LacZ* embryos harvested at E12.5 show robust vascular transgene expression. All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

In order to characterise the transgene expression pattern in KLF2-41:*LacZ* transient embryos, five out of six embryos with vascular transgene expression (#72, #64, #31, #45, #71) were sectioned. Transgene expression was largely consistent between the five embryos sectioned, with some differences between them likely attributable to poorer infiltration of X-Gal into the heart and surrounding vessels deep within the embryo. Some embryos (#45, #49) exhibited more ectopic expression in the neural tube, previously noted with the use of hsp68 minimal promoter upstream of *LacZ* in transgenic mouse embryos[307]. Section of all five embryos were analysed in the same way, but section analysis of one representative embryo (#31) is shown here (**Figure 3.8**).

In blood vessels throughout the embryo, KLF2-41:*LacZ* transgene expression was detected specifically in arterial and angiogenic ECs. Robust transgene expression was observed in the endothelium of all major arteries of the E12.5 embryo, including the internal carotid arteries, the paired dorsal aortae, branchial arch arteries, the umbilical arteries, superior mesenteric artery, and the abdominal portion of the aorta. Transgene expression was absent from all major veins, including the cardinal, subcardinal, and umbilical veins. In the heart of the embryo, low levels of transgene expression were detected throughout the endocardium, with highest levels of transgene expression detected in the

aortic sac (origin of the thoracic aorta), the pulmonary trunk, throughout the OT of the heart, and the endothelial lining of AVC tissue (**Figure 3.8d, e, f,** and **g**). Finally, high levels of transgene expression were detected in small capillaries and at the angiogenic front within the hindbrain (**Figure 3.8b** and **c**), the branchial arches, and the developing limb buds (**Figure 3.8h**).

Taken together, transgene expression observed in all but one transient KLF2-41:*LacZ* embryo recapitulated most aspects of endogenous *Klf2* expression at E12.5 (Section 3.5.2.2, **Figure 3.6**), excluding expression in a small subset of venous ECs. In two out of seven L*acZ* expressing embryos (#71, #49), slightly fainter endothelial-specific transgene expression was detected, yet it was still arterial EC-specific and largely corresponded to the transgene expression pattern observed in the other five *LacZ* positive embryos that drove expression in blood vessels.

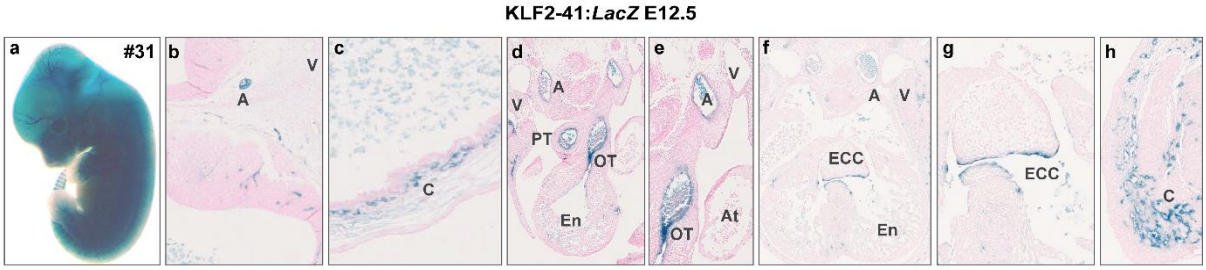


Figure 3.8: Analysis of representative KLF2-41:*LacZ* embryo sections at E12.5 shows robust endothelial cell and ECC-specific *LacZ* expression. (a) The whole-mount KLF2-41:*LacZ* E12.5 transient transgenic embryo #31. **(b)** *LacZ* expression was strongest within the endothelium of small arteries and capillaries of the head, but absent from venous ECs. **(c)** Robust *LacZ* expression was detected in the angiogenic head vessels. **(d)** KLF2-41:*LacZ* transgene was detected in the OT of the heart, the carotid arteries, the pulmonary trunk, and to a lesser extent in the remainder of the endocardium. It was absent from the anterior cardinal veins. **(e)** Close-up of *LacZ* expression in in the endothelium of left carotid artery and left anterior cardinal vein (no expression). **(f)** The KLF2-41:*LacZ* transgene was robustly expressed in the ECCs (here shown in AVC tissue) and to a much lesser extent in the remainder of the endocardium. **(g)** Close-up of *LacZ*-expressing AVC tissue. **(h)** KLF2-41:*LacZ* transgene was robustly expressed in the angiogenic vessels of the developing right forelimb bud. **A=artery, V=vein, C=capillary, PT=pulmonary trunk, OT=outflow tract, En=endocardium, At=atrium, ECC=endocardial cushion.**

3.5.3.2 KLF2-66 enhancer:*LacZ* expression in transient transgenic mouse embryos

We received nine PCR positive embryos transgenic for the KLF2-66:*LacZ* transgene (**Table 3.1**). Each of the nine transgenic embryos represented an independent experiment as the site of transgene integration and the number of copies integrated were both random. Five out of nine transiently

transgenic embryos had detectable *LacZ* expression as measured by X-Gal staining (**Figure 3.9**). In three out of five *LacZ* -expressing embryos (#54, #56, #91) robust vascular-specific expression was detected in the whole-mount embryos. In two out of five *LacZ* expressing embryos (#67, #48), faint vascular transgene expression was observed in surface capillaries, but was also seen in non-vascular tissues. Significantly more variation in transgene expression was observed in KLF2-66:*LacZ* transgenic embryos compared to KLF2-41:*LacZ* transgenic embryos, however the total number of transiently expressing embryos analysed with the latter was higher, possibly accounting for the more consistent expression pattern observed.

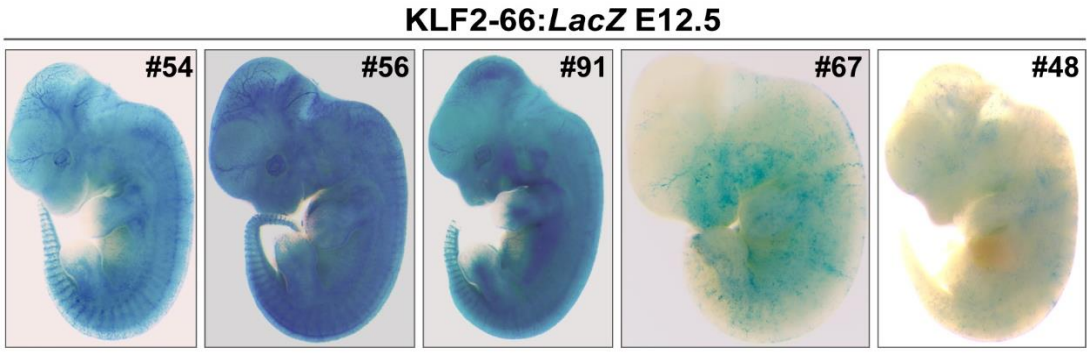


Figure 3.9: Transient transgenic KLF2-66:*LacZ* embryos harvested at E12.5 show robust vascular transgene expression. All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

In order to characterise the transgene expression pattern in KLF2-66:*LacZ* transient embryos, three out of five embryos with vascular transgene expression (#54, #56, and #91) were sectioned. Analysis of transgene expression (blue X-gal staining) against the nuclear fast red counter-stain in sections allowed for a more detailed analysis of transgene expression in these embryos. Sections of all three embryos were analysed in the same way, and transgene expression was consistent between the embryos analysed. The analysis of one representative embryo (#54) is shown here (**Figure 3.10**).

In blood vessels throughout the embryo, KLF2-66:*LacZ* transgene expression was detected in arterial, venous, and angiogenic ECs. However, transgene expression detected in arterial and angiogenic ECs

was more consistent, of higher intensity, and more continuous throughout the endothelial lining of all major arteries compared to weaker and more punctate expression detected in all major veins. Robust transgene expression was observed in the endothelium of all major arteries of the E12.5 embryo, including the internal carotid arteries, the paired dorsal aortae, branchial arch arteries, the umbilical arteries, superior mesenteric artery, and the abdominal portion of the aorta (**Figure 3.10b, e,** and **h**). Transgene expression, although weaker, was also present in all major veins, including the cardinal, subcardinal, and umbilical veins. In the heart of the embryo, low levels of transgene expression were detected throughout the endocardium, with highest levels of transgene expression detected in the aortic sac (origin of the thoracic aorta), the pulmonary trunk, throughout the OT of the heart, and the endothelial lining of the AVC tissue (**Figure 3.10d, f,** and **g**). Low levels of transgene expression were also detected in the remainder of the endocardium (**Figure 3.10f**). Finally, high levels of transgene expression were detected in small capillaries and at the angiogenic front within the hindbrain, the branchial arches, and the developing limb buds (**Figure 3.10b** and **c**).

Taken together, transgene expression observed in three transient KLF2-66:*LacZ* embryos recapitulated most aspects of endogenous *Klf2* expression at E12.5, including expression in arterial ECs, weaker expression in a subset of venous ECs, expression in the ECC tissue of the heart, and expression at the angiogenic front.

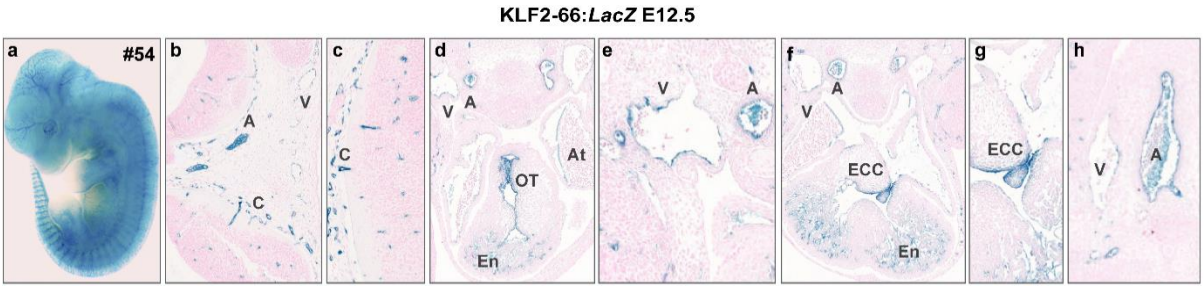


Figure 3.10: Analysis of representative KLF2-66:*LacZ* embryo sections at E12.5 shows robust endothelial cell and ECC-specific *LacZ* expression. (a) The whole-mount KLF2-66:*LacZ* E12.5 transient transgenic embryo #54. **(b)** *LacZ* expression was strongest in the EC of small arteries and capillaries of the head, and weaker in head veins. **(c)** Robust *LacZ* expression was detected in the angiogenic hind brain. **(d)** KLF2-66:*LacZ* transgene was detected in the OT of the heart, the carotid arteries, and to a lesser extent in anterior cardinal veins. **(e)** Close-up of *LacZ* expression in in the right carotid artery and right anterior cardinal vein. **(f)** KLF2-66:*LacZ* transgene

was robustly expressed in the ECCs (here shown in AVC tissue) and the remainder of the endocardium. **(g)** Close-up of *LacZ*-expressing AVC tissue. **(h)** KLF2-66:*LacZ* transgene was expressed in the endothelium of the abdominal aorta, and to a lesser extent in the endothelium of right subcardinal vein. **A=artery, V=vein, C=capillary, OT=outflow tract, En=endocardium, At=atrium, ECC=endocardial cushion.**

3.5.3.3 KLF2-8 enhancer:*LacZ* expression in transient transgenic mouse embryos

KLF2-8:*LacZ* enhancer transgene expression was analysed in a total of ten embryos PCR positive for the transgene (**Table 3.1**). Out of ten embryos analysed, eight showed some *LacZ* expression (**Figure 3.11**). In all eight *LacZ*-expressing embryos (#30, #81, #82, #86, #107, #115, #90, #93), some ectopic *LacZ* expression was detected, for example in the neural tube (#30, #93), the forebrain (#81, #86), the eye (#93), or the muscle (#107).

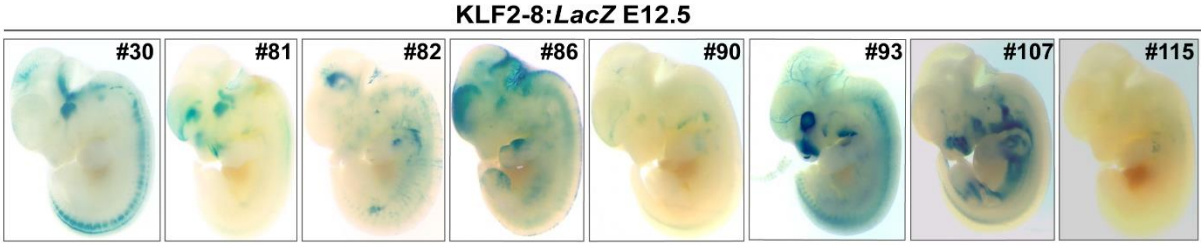


Figure 3.11: Transient transgenic KLF2-8:*LacZ* embryos harvested at E12.5 show robust vascular transgene expression. All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

From the eight *LacZ*-expressing embryos, two of them (#90, #93) exhibited low intensity *LacZ* expression in some blood vessels. After sectioning of these two embryos, barely any *LacZ* expression was observed in blood vessels of embryo #90. Embryo #93 on the other hand showed some *LacZ* expression in a subset of arterial ECs and VSMCs (**Figure 3.12b, d, e, g,** and **h**). Low intensity *LacZ* expression was also detected in the ECCs and the epicardium of this embryo (**Figure 3.12c, e,** and **f**). However, robust *LacZ* expression was also detected in non-vascular tissues of this embryo, including the eye, the jaw, the epicardium, and the neural tube (**Figure 3.12a**). Embryo #93 was thus unable to drive vascular-specific *LacZ* expression, possibly suggesting that in the case of this specific embryo, the transgene integrated into a transcriptionally permissive genomic region. As only one out of eight

analysed KLF2-8:*LacZ* transiently transgenic embryos was able to drive some vascular *LacZ* expression which was not specific to the vasculature, this enhancer was excluded from further analysis.

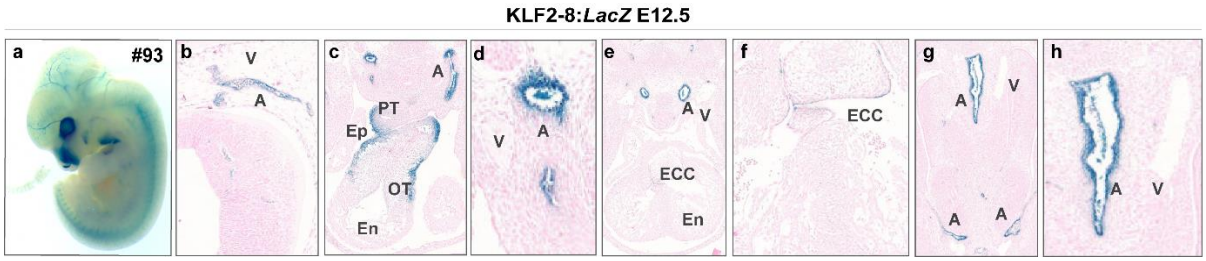


Figure 3.12: Analysis of the only KLF2-8:*LacZ* embryo sections with some vascular *LacZ* expression at E12.5 shows endothelial, vascular smooth muscle cell, and some ECC-specific *LacZ* expression. (a) The whole-mount KLF2-8:*LacZ* E12.5 transient transgenic embryo #93. **(b)** *LacZ* expression was strongest in endothelial cells of small arteries of the head, but absent from venous and capillary endothelial cells. **(c)** Robust *LacZ* expression was detected in the epicardium of the heart; the endothelial and VSMCs of the carotid arteries. It was absent from endothelial and vascular smooth muscle cells of the anterior cardinal veins. **(d)** Close-up of *LacZ* expression in the endothelial and vascular smooth muscle cells of the right carotid artery and its absence in the right anterior cardinal vein. **(e)** Low levels of KLF2-8:*LacZ* transgene expression were detected in the AVC tissue of this embryo. **(f)** Close-up of *LacZ*-expressing AVC tissue. **(g)** KLF2-8:*LacZ* transgene was expressed in the endothelium and vascular smooth muscle tissue of the abdominal aorta, and absent from the endothelium of left subcardinal vein. **(h)** Close-up of *LacZ* expression in the endothelium and vascular smooth muscle tissue of the abdominal aorta and its absence from left subclavian vein. **A=artery, V=vein, Ep=epicardium, PT=pulmonary trunk, OT=outflow tract, En=endocardium, ECC=endocardial cushion.**

3.5.3.4 KLF2-34:*LacZ* and KLF2+23:*LacZ* expression in transient transgenic mouse embryos

KLF2-34:*LacZ* enhancer transgene expression was analysed in a total of six embryos PCR positive for *LacZ*. Out of the six *LacZ*-positive embryos, three embryos had no detectable *LacZ* expression following X-Gal staining (**Table 3.1**). In the other three embryos (#31, #51, #66) only ectopic KLF2-34:*LacZ* transgene expression was detected (**Figure 3.13**).

KLF2+23:*LacZ* expression was analysed in a total of eight embryos PCR positive for *LacZ* (**Table 3.1**). In four out of the eight *LacZ* positive embryos, no KLF+23:*LacZ* transgene expression was detected. In the other four *LacZ* positive embryos, ectopic transgene expression was detected in all embryos (**Figure 3.13**). Embryos carrying either the KLF2-34:*LacZ* or the KLF2+23:*LacZ* transgene were not sectioned as no vascular *LacZ* expression was detected during analysis of the whole-mount embryos.

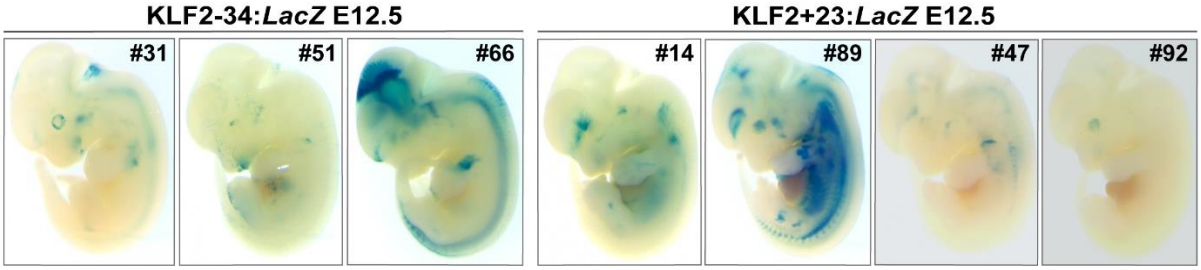


Figure 3.13: Transient transgenic KLF2-34:*LacZ* and KLF2+23:*LacZ* embryos harvested at E12.5 show no vascular transgene expression. All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

	KLF2-66:*LacZ*	KLF2-41:*LacZ*	KLF2-34:*LacZ*	KLF2-8:*LacZ*	KLF2+23:*LacZ*
Vascular	5	6	0	2	0
Ectopic	0	1	3	6	4
None	4	7	3	2	4
Total	**9**	**13**	**6**	**10**	**8**

Table 3.1: Summary of KLF2-66:*LacZ*, KLF2-41:*LacZ*, KLF2-34:*LacZ*, KLF2-8:*LacZ*, and KLF2+23:*LacZ* transgene expression detected in E12.5 mouse embryos. The expression patterns observed was classified as one of the three categories: vascular, ectopic, or no expression.

3.6 Analysis of KLF2-41:LacZ transgene expression in the developing mouse vasculature

3.6.1 Rationale

While transient transgenesis is an informative way of testing whether an enhancer is able to drive tissue-specific reporter gene expression, it limits the study of transgene expression to a single time point. Additionally, due to random integration of the transgene into the host genome and variability in number of copies integrated, the resulting variation in mosaic transgene expression complicates comparison between animals expressing different transgenes, especially when the number of transient embryos analysed per experiment is low. In order to gain insight into transgene expression throughout development and adulthood, stable KLF2-66:*LacZ* and KLF2-41:*LacZ* transgenic mouse lines were generated. Embryos were then harvested at different time points throughout embryonic development.

For detailed expression analysis, one to two embryos representative of the expression pattern observed in the whole litter were paraffin-embedded, then cut into 5µm transverse sections using the

microtome (sectioning was carried out by Indrika Ratnayaka or Dr Sarah De Val). For each sectioned embryo, every fifth or tenth section was selected for analysis, with 50-100 sections analysed per embryo depending on its stage of development. The selected sections were then counter-stained with Nuclear Fast Red, then mounted and imaged using a slide scanner.

3.6.2 KLF2-41:LacZ stable transgenic line analysis in whole mount mouse embryos

To analyse the spatio-temporal expression pattern of KLF2-41:*LacZ* enhancer activity in mice, six stable mouse lines transgenic for the KLF2-41:*LacZ* transgene (tested transiently in section 3.5.3.1) were generated (**Figure 3.14**). The founder animal from one of the six lines died prior to analysis. The remaining five lines were analysed for KLF2-41:*LacZ* expression, initially at E12.5-E13.5. Two of the five lines analysed (Lines 1 and 5) showed no *LacZ* expression at E12.5. Line 2 showed weak arterial-specific vascular expression of the transgene, with weak ectopic expression in the developing neural tube at E12.5 (data not shown). Lines 3 and 4 showed robust endothelial expression of the transgene, with Line 4 showing the highest intensity of KLF2-41:*LacZ* expression after staining with the least ectopic *LacZ* expression detected in the neural tube. From here onwards, Line 4 was used in KLF2-41:*LacZ* expression analyses throughout this thesis. The differences in transgene expression pattern observed between all five stable transgenic lines could be attributed to the random transgene integration site and transgene copy number variation between the individual stable transgenic lines.

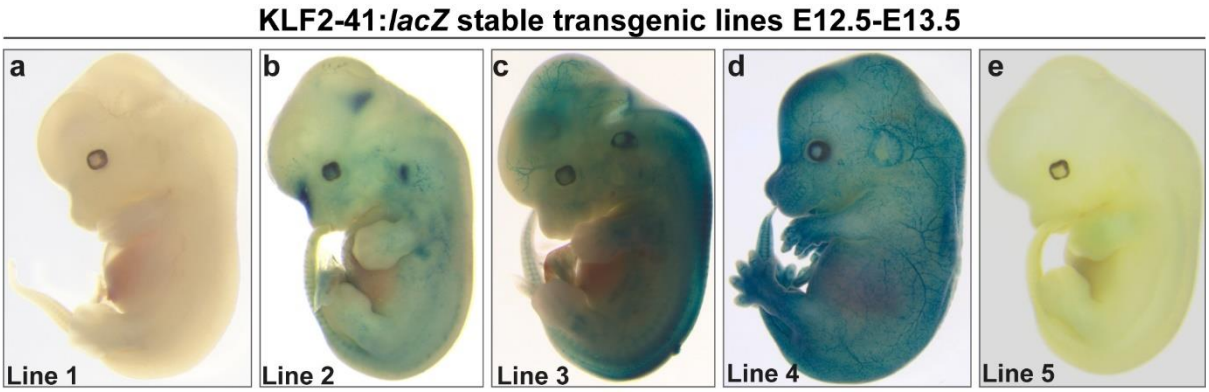


Figure 3.14: Representative whoe-mount X-Gal stained embryos from five independent stable KLF2-41:*LacZ* transgenic mouse lines at E12.5-E13.5. (a) KLF2-41:*LacZ* stable line 1 showed no KLF2-41:*LacZ* transgene

expression at E12.5. **(b)** KLF2-41:*LacZ* stable line 2 showed some arterial-specific KLF2-41:*LacZ* transgene expression at E12.5. **(c)** KLF2-41:*LacZ* stable line 3 showed arterial-specific KLF2-41:*LacZ* transgene expression at E12.5, with ectopic expression in the neural tube. **(d)** KLF2-41:*LacZ* stable line 4 showed robust arterial-specific KLF2-41:*LacZ* transgene expression at E12.5, with minimal ectopic expression within the neural tube. **(e)** KLF2-41:*LacZ* stable line 5 showed no endothelial KLF2-41:*LacZ* transgene expression at E12.5. One representative embryo for each stable line is shown above.

3.6.3 Time-course analysis of KLF2-41:LacZ expression in the systemic vasculature during development

Time-course analysis of KLF2-41:*LacZ* Line 4 expression was carried out between E8.5 and E15.5 to determine transgene expression onset and pattern throughout vascular development (**Figure 3.15**). The early time points of analysis (between E8.5 and E10.5) were investigated at shorter (0.5 day) intervals, as blood flow is thought to be initiated towards the end of this developmental window, although precise timing of blood flow initiation remains controversial (see Section 3.9.2)[308,309]. An important consideration was β-Gal half-life, estimated to be on the order of 20 hours[310]. The loss of or reduction in enhancer:reporter transgene expression would therefore be observed as reduced X-Gal staining up to one day later.

The type and identity of blood vessels were determined by anatomical location and were additionally distinguished by morphology and size. The Atlas of Mouse Development[311] and the eMouse Atlas[312] (the Kaufmann annotation) were used as aids for identification of different anatomical structures in mouse embryos throughout the time course analysis.

The KLF2-41:*LacZ* transgene was expressed specifically in ECs as early as E8.5 (**Figure 3.15a, b** and **c**). At this time-point expression in the whole-mount embryo appeared to be pan-endothelial, including robust expression not only in vitelline arteries and veins, but also in the allantois and the yolk sac endothelium of the embryo. At E9.5, robust transgene expression was detected throughout the paired dorsal aortae, the vitelline arteries; and first, second, and third branchial arch arteries. In the heart of the embryo, low levels of transgene expression were detected throughout the endocardium, with

highest levels of transgene expression detected in the aortic sac (origin of the thoracic aorta), the pulmonary trunk, throughout the OT of the heart, and the endothelial lining of atrio-ventricular and semilunar ECC tissue. Within the ECCs, the transgene expression was limited to the ECC cells, but not the mesenchymal cushion cells derived from invading endothelial cells. Punctate transgene expression was also observed in the primary head vein and the vitelline veins in the tail region, but the expression was much weaker there compared to arterial ECs. Throughout the embryo, the transgene was also robustly expressed at the angiogenic front, especially in small capillaries in the head and trunk regions (**Figure 3.15d** and **e**). Robust transgene expression was also observed within the endothelium of intersomitic vessels (**Figure 3.15c** and **d**).

At E10.5, the KLF2-41:*LacZ* transgene remained robustly expressed in all arterial structures that were already *LacZ* positive at E9.5. Additional transgene expression was observed in the angiogenic branchial arches, and within the perioptic vascular plexus. Weaker punctate transgene expression was still observed in some sections of the primary head and vitelline veins (**Figure 3.15f** and **g**). At E11.5, KLF2-41:*LacZ* reporter activity was entirely arterial-specific, with all punctate expression absent from the primary head and vitelline veins at this time point. In addition to continued expression in all arterial structures already described at earlier time points, especially strong transgene expression was observed in the angiogenic hindbrain, forelimb buds, and branchial arch capillaries, and the ECC tissue in the atrio-ventricular canal. Additional expression was observed in the internal carotid arteries, the aortic sac, pulmonary trunk, and the OT of the heart (**Figure 3.15h, i** and **j**).

At E12.5 and E13.5, transgene expression retained its specificity both to arteries and other structures of arterial origin, as well as the angiogenic front (hindbrain, branchial arches, developing limb buds). Transgene expression pattern observed in the E12.5 embryo recapitulated the expression pattern seen in transiently transgenic embryos (**Figure 3.15k, l,** and **m**). At E13.5 robust transgene expression was observed in leaflets of now fully developed left and right atrio-ventricular (mitral and tri-cuspid,

respectively) valves (**Figure 3.15n** and **o**), consistent with previous reports of *Klf2* expression in the developing valves[192]. At E15.5, the transgene retained its arterial specificity (**Figure 3.15p** and **q**). Transgene expression in the embryonic yolk sac remained arterial-specific between 9.5 and 15.5.

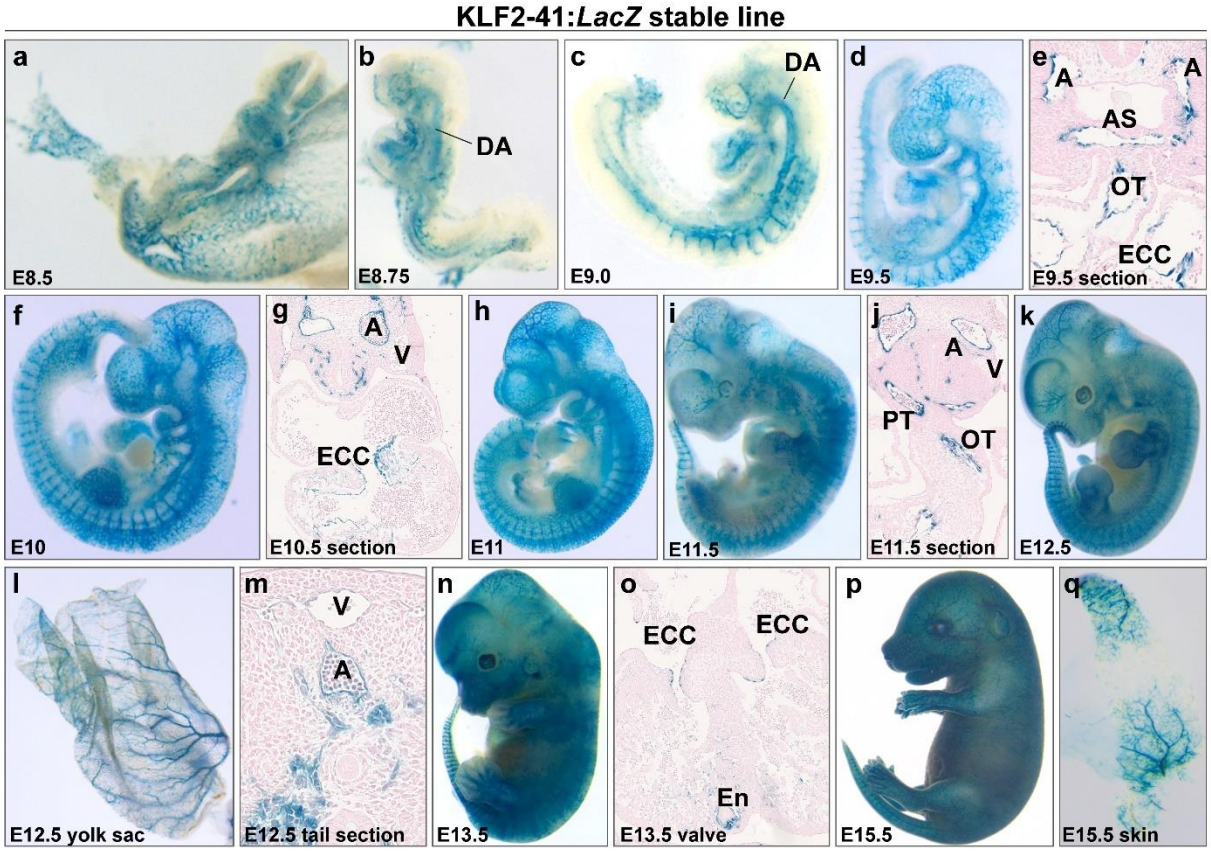


Figure 3.15: Time course analysis of KLF2-41:*LacZ* stable line 4 to assess transgene expression during embryonic development. (a) The KLF2-41:*LacZ* transgene was robustly expressed specifically in the endothelium of the E8.5 embryo. **(b)** At E8.75, the KLF2-41:*LacZ* transgene expression was clearly visible in the dorsal aorta. **(c)** At E9, the KLF2-41:*LacZ* transgene was still expressed in the dorsal aorta and the intersegmental vessels. **(d), (e)** At E9.5, the KLF2-41:*LacZ* transgene was expressed in the dorsal aorta, the intersegmental vessels, and at the angiogenic front. Section analysis of the E9.5 embryo showed KLF2-41:*LacZ* transgene expression within the paired dorsal aortae and the aortic sac, as well as the OT and AVC cushions of the heart. **(f), (g)** At E10, the KLF2-41:*LacZ* transgene remained robustly expressed in arterial ECs and at the angiogenic front. Section analysis of the E10.5 embryo showed KLF2-41:*LacZ* transgene expression within the endothelium of the dorsal aortae, the ECCs, and the remaining endocardium of the heart. **(h)** At E11, the KLF2-41:*LacZ* transgene remained robustly expressed in arterial ECs and at the angiogenic front. **(i), (j)** At E11.5, the KLF2-41:*LacZ* transgene remained robustly expressed in arterial ECs and at the angiogenic front. Section analysis of the E11.5 embryo showed KLF2-41:*LacZ* transgene expression within the endothelium of the dorsal aortae, the pulmonary trunk, the OT and AVC of the heart. **(k), (l), (m)** At E12.5, the KLF2-41:*LacZ* transgene remained robustly expressed in arterial ECs and at the angiogenic front. Within the E12.5 yolk sac, the KLF2-41:*LacZ* transgene remained arterial specific. A section of the E12.5 embryo tail region showed KLF2-41:*LacZ* expression remained specific to the endothelium of the tail artery, but not the tail vein. **(n), (o)** At E13.5, the KLF2-41:*LacZ* transgene remained robustly expressed in arterial ECs and at the angiogenic front. A section through the heart of the E13.5 embryo showed KLF2-41:*LacZ* transgene expression in ECCs and the remainder of the endocardium. **(p), (q)** The KLF2-41:*LacZ* transgene remained endothelial-specific at E15.5. Arterial-specific KLF2-41:*LacZ* expression was observed in the skin tissue

of the E15.5 embryo. **DA=dorsal aorta, A=artery, AS=aortic sac, OT=outflow tract, ECC=endocardial cushion, V=vein, PT=pulmonary trunk, En=endocardium.**

In conclusion, the time-course analysis of transgene expression between E8 and E15.5 demonstrated that KLF2-41:*LacZ* expression was pan-endothelial early in development, but became largely restricted to arterial and angiogenic ECs by E9.5. Further, the KLF2-41:*LacZ* transgene was robustly expressed in the AVC and OT cushions, and to a lesser extrent in the remainder of the endocardium throuhgout the timecourse. The KLF2-41:*LacZ* transgene expression therefore largely recapitulated the endogenous pattern of *Klf2* mRNA expression throughout development.

3.6.4 KLF2-41:LacZ transgene expression in the organ-specific vasculature

3.6.4.1 KLF2-41:*LacZ* transgene expression in the embryonic and post-natal heart

In addition to the arterial and angiogenic ECs, the KLF2-41:*LacZ* transgene was persistently expressed in the OT and AVC cushions of the developing heart, in line with previously published *Klf2* mRNA expression data[192]. Due to infiltration issues with the staining of larger embryos, the analysis of the KLF2-41:*LacZ* transgene in the ECCs became limited in the whole-mount embryo. To characterise the KLF2-41:*LacZ* transgene in the ECCs and cardiac valves in more detail, a timecourse of KLF2-41:*LacZ* transgene expression in whole-mount hearts was carried out between E11.5 and P8.

In addition to the analysis of the KLF2-41:*LacZ* transgene in the ECCs, whole-mount heart analysis represented an additional opportunity to characterise KLF2-41:*LacZ* expression in the coronary vasculature. The coronary vasculature is formed entirely by angiogenic sprouting of ECs from multiple independent origins. A large proportion of coronary arteries, capillaries, and veins on dorsal and lateral sides of the heart originate from the sinus venosus (SV), a cavity-like structure of venous origin residing between the paired vanae cavae on the dorsal side of the mammalian embryonic heart. Coronary vessels on the ventral side of the heart originate from the endocardium, together with the vasculature of the ventricular septum which is of the same origin. Finally, a small number of coronary

vessels sprout from the proepicardium, a transitory extracardiac embryonic structure which will give rise to the epicardium and contribute cells to other cardiac tissues, including the coronary vessels. In the early postnatal heart, SV-derived vessels can be found in the outer myocardial wall, whereas the coronaries of the inner myocardial wall derive from the endocardium[277].

To investigate the KLF2-41:*LacZ* expression in ECCs and coronary vessels, embryonic hearts were first dissected from the embryos or pups, then stained individually to avoid X-Gal infiltration issues encountered when staining late stage (E13.5 or older) embryos. After staining, a representative embryonic heart for each timepoint underwent sectioning to enable analysis not just of coronary vessels of the outer myocardial wall, but also of the inner myocardial wall vasculature and ECCs (later cardiac valves).

In E11.5 and E12.5 hearts, robust KLF2-41:*LacZ* transgene expression was observed in the OT of the heart visible on the dorsal side (**Figure 3.16a** and **b**). At E12.5, additional transgene expression was observed throughout the endocardium of the heart, and in two small populations of ECs, one located in the middle on the ventral surface, likely to be early endocardium-derived ECs; and the second on the dorsal face of the heart, likely to be SV-derived ECs (**Figure 3.16e** and **f**). At E13.5, there was still robust transgene expression in the OT, and in the visibly expanded population of SV-derived ECs on the dorsal side of the heart (**Figure 3.16i** and **j**). Endocardial transgene expression at 13.5 was limited to the endocardial lining the OT. At E15.5, the KLF2-41:*LacZ* transgene was expressed in the coronary microvasculature covering the whole surface of the heart, likely a combination of ECs derived from both the endocardium and the SV (**Figure 3.16l** and **m**). Transgene expression in early post-natal hearts (P0, P6, and P8) was still present in the coronary microvasculature on both sides of the heart, with further transgene expression observed in coronary veins and the SV especially at P6 and P8 (**Figure 3.16t** and **w**). Section analysis of a representative heart at each timepoint highlighted the previously

observed robust transgene expression within the OT, the ECCs, and cardiac valves throughout the timecourse (**Figure 3.16c, d, g, h, k, n, o, r, u, x**).

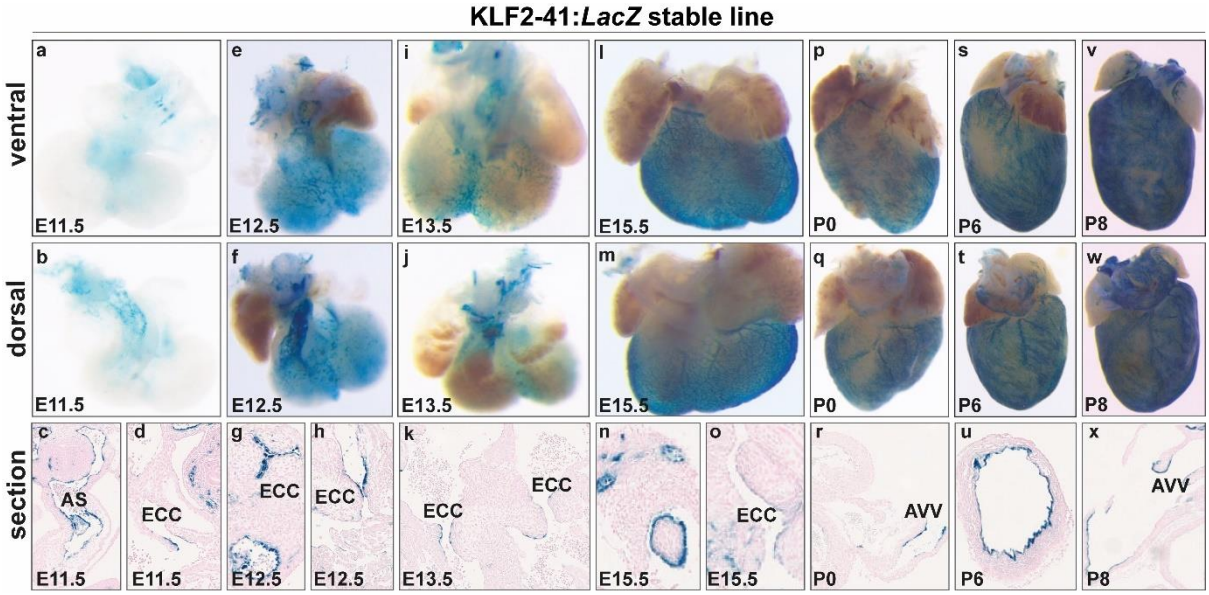


Figure 3.16: Time course analysis of KLF2-41:*LacZ* stable Line 4 to assess KLF2-41:*LacZ* transgene expression in coronary vessels during embryonic heart development. (a), (b), (c), (d) At E11.5, the KLF2-41:*LacZ* transgene was expressed within the OT, the ECCs, and the aortic sac. **(e)**, **(f)**, **(g)**, **(h)** At E12.5, The KLF2-41:*LacZ* transgene remained robustly expressed in the OT and PT valve precursors (g) as well as the AVC cushions of the heart (h), and was detectable in endocardium-derived ECs on the ventral surface. **(i)**, **(j)**, **(k)** At E13.5, the KLF2-41:*LacZ* expression persisted in the OT, the ECCs, and the endocardium-derived coronary ECs visible on the ventral side of the heart. Note that figure 3.16k is is a section of the heart from a wholemount embryo previously shown in Figure 3.15o, and not a section of an isolated wholemount heart shown in 3.16i and j. **(l)**, **(m)**, **(n)** At E15.5, the KLF2-41:*LacZ* transgene was robustly expressed in coronary vessels on the ventral and the dorsal sides of the heart. The transgene was still expressed in the ECCs. **(p)**, **(q)**, **(r)** In the P0 heart, the KLF2-41:*LacZ* transgene was detected in coronary capillaries covering the ventral and dorsal sides of the heart, with surprising expression detected within the coronary venous ECs on the dorsal side of the heart. The KLF2-41:*LacZ* transgene remained robustly expressed in the cardiac valves. **(s)**, **(t)**, **(u)** In the P6 heart, the KLF2-41:*LacZ* transgene remained robusly expressed within coronary capillaries covering the ventral and dorsal sides of the heart, with persistent expression within the coronary venous ECs on the dorsal side of the heart. **(v)**, **(w)**, **(x)** In the P8 heart, the KLF2-41:*LacZ* transgene remained robusly expressed within coronary capillaries covering the ventral and dorsal sides of the heart, with persistent expression within the coronary venous ECs on the dorsal side of the heart. Robust KLF2-41:*LacZ* transgene expression was detected within the leaflets of the atrio-ventricular valve. **AS=aortic sac, ECC=endocardial cushion, AVV=atrio-ventricular valve.**

3.6.4.2 KLF2-41:*LacZ* transgene expression analysis in post-natal organs

The analysis of the KLF2-41:*LacZ* transgene expression in the post-natal coronary vessels revealed unexpected expression in coronary venous ECs that was not detected during analysis of the KLF2-41:*LacZ* transgene in development. Vascular beds supplying blood to different organs exhibit significant structural and functional heterogeneity[313–315]. To investigate the expression of the KLF2-

41:*LacZ* transgene in different organ-specific vascular beds, different organs were dissected from a P6 mouse pup and analysed for transgene expression.

The examination of the P6 postnatal lung (**Figure 3.17a**), intercostal vasculature of the ribcage (**Figure 3.17b**), the stomach (**Figure 3.17d**), and the diaphragm (**Figure 3.17e**) showed robust KLF2-41:*LacZ* transgene expression in small diameter arteries (**Figure 3.17b**), and capillaries of the organs analysed. Interestingly, in the post-natal retinal vasculature which develops entirely post-natally, KLF2-41:*LacZ* transgene was detected both in retinal arteries and veins, as well as in the angiogenic ECs. This observation was consistent with venous transgene expression observed in the P6 coronary vasculature.

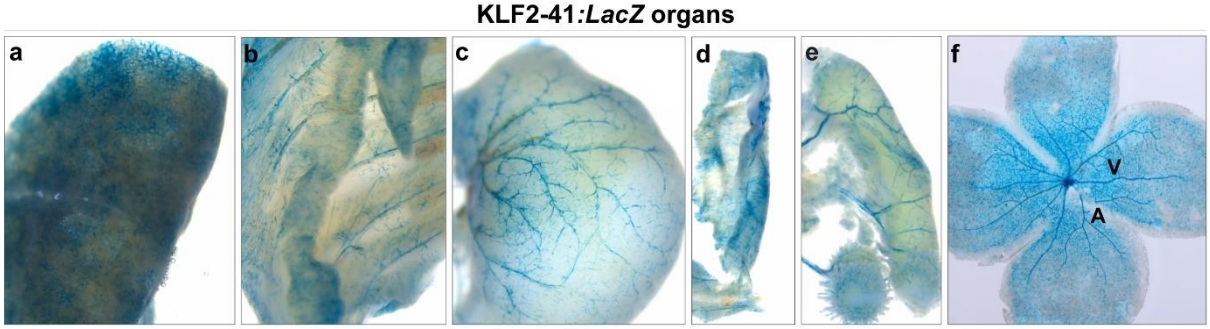


Figure 3.17: Time course analysis of KLF2-41:*LacZ* stable line 4 to assess transgene expression P6 organ-specific vasculature. The KLF2-41:*LacZ* transgene was robustly expressed in **(a)** lung capillaries, **(b)** intercostal arteries of the ribcage, **(c)** the stomach vasculature, **(d)** the blood vessels of the diaphragm, **(e)** the mesenteric vasculatre, and **f)** the arterial, venous, and angiogenic ECs of the retinal vasculature. **A=artery, V=vein.**

To conclude, the KLF2-41:*LacZ* transgene expression was restricted to arterial and angiogenic ECs, as well as arterial-derived cardiac structures and ECCs, in most aspects of embryonic and post-natal development. Notable exceptions were the pre-E10.5 embryo, the post-natal retina, where the KLF2-41:*LacZ* transgene was expressed in a pan-endothelial manner, and the coronary vasculature, where transgene expression was pre-natally confined to angiogenic ECs, and post-natally included venous ECs, but was absent from ECs of all large coronary arteries.

3.7 Analysis of KLF2-66:LacZ transgene expression in the developing mouse vasculature

3.7.1 Analysis of KLF2-66:LacZ activity during mouse development

To analyse the spatio-temporal expression pattern of KLF2-66:*LacZ* enhancer activity in mice, five stable transgenic mouse lines over-expressing the KLF2-66:*LacZ* transgene (tested transiently in Section 3.5.3.2) were generated (**Figure 3.18**). All lines were analysed for KLF2-66:*LacZ* expression, initially at E12.5-E13.5. Robust endothelial-specific transgene expression was detected in Lines 1 and 2, while no vascular transgene expression was detected in Lines 3, 4, or 5. Line 2 was found to drive more robust transgene expression than Line 1 (**Figure 3.18b**). From here onwards, Line 2 was used in of KLF2-66:*LacZ* expression analyses throughout this thesis. Variation in the intensity and pattern of transgene expression observed between the stable transgenic lines can likely be attributed to the variation in integration site and copy number of the transgene.

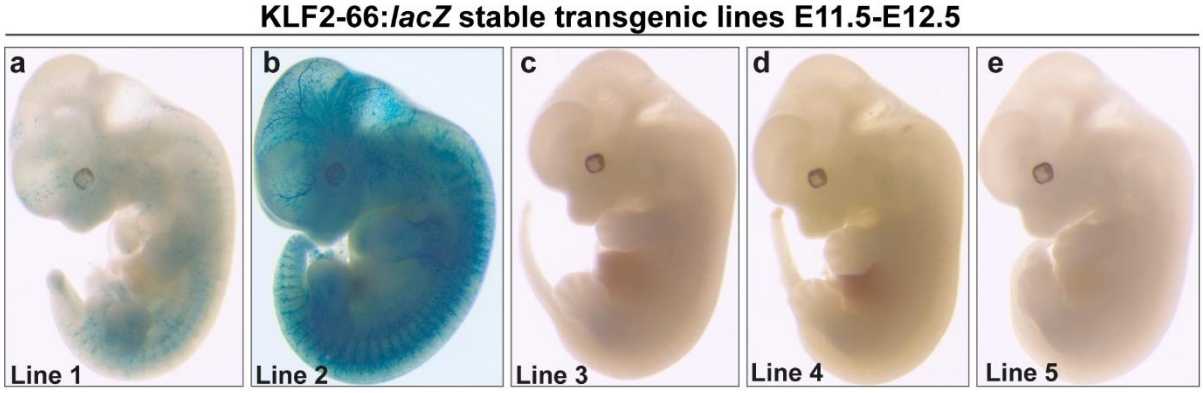


Figure 3.18: Representative whoemount X-Gal stained embryos from five independent stable KLF2-66:*LacZ* transgenic mouse lines at E11.5-E12.5. (a) KLF2-66:*LacZ* stable line 1 showed weak KLF2-66:*LacZ* transgene expression at E12.5. **(b)** KLF2-66:*LacZ* stable line 2 showed robust endothelial-specific KLF2-66:*LacZ* transgene expression at E12.5. **(c)** KLF2-66:*LacZ* stable line 3 showed no KLF2-66:*LacZ* transgene expression at E12.5. **(d)** KLF2-66:*LacZ* stable line 4 showed no KLF2-66:*LacZ* transgene expression at E12.5. **(e)** KLF2-66:*LacZ* stable line 5 showed no endothelial KLF2-66:*LacZ* transgene expression at E12.5. One representative embryo for each stable line is shown above.

3.7.2 Time-course analysis of KLF2-66:LacZ transgene expression during development

As for the KLF2-41:*LacZ* stable transgenic line, the time-course analysis of KLF2-66:*LacZ* Line 2 was also carried out between E8.5 and E15.5 (**Figure 3.19**). The KLF2-66:*LacZ* transgene was expressed specifically in ECs as early as E8.5. However, compared to the KLF2-41:*LacZ* transgene expression which was robust across all endothelial compartments at this time point, KLF2-66:*LacZ* transgene

expression was less pronounced, and was limited to individual ECs. No transgene expression was detected at an earlier E8.0 time point in the area where the paired dorsal aortae would be expected to form. No transgene expression was detected in the yolk sac vasculature.

At E9.5, transgene expression was detected throughout the paired dorsal aortae, the vitelline arteries; and first, second, and third branchial arch arteries. In the heart, transgene expression was observed in the aortic sac, the truncus arteriosus region of OT of the heart, the endothelial lining of atrio-ventricular canals, and in the remainder of the endocardium. Transgene expression was also observed in the primary head vein, vitelline veins, cardinal, and umbilical veins. The expression of the transgene in venous ECs was weaker than that of arterial ECs but was still present and less punctate than venous expression observed with the KLF2-41:*LacZ* transgene (**Figure 3.19a, b, c,** and **d**). Compared to the KLF2-41:*LacZ* stable line, overall endothelial expression at E9.5 was weaker even in arterial ECs, this was especially apparent in the lower trunk and tail regions of the embryo. Throughout the embryo, the KLF2-66:*LacZ* transgene was expressed at the angiogenic front, especially in small capillaries in the head and upper trunk regions.

At E10.5, the KLF2-66:*LacZ* transgene remained preferentially expressed in all arterial structures that were already *LacZ* positive at E9.5, but remained robustly expressed in the venous endothelium. Additional transgene expression was observed in the angiogenic branchial arches, aortic sac and the origin of the first branchial arch artery, and the endocardium of the heart (**Figure 3.19e** and **f**).

At E12.5, transgene expression remained strongest in the arteries and other structures of arterial origin, as well as the angiogenic front (hindbrain, branchial arches, developing limb buds). Robust staining was observed in middle cerebral arteries, paired dorsal aortae, brachiocephalic arteries, common carotid arteries, paired umbilical arteries, ophthalmic arteries passing distally toward foetal fissure of the optic stalk, and the aortic sac (**Figure 3.19g, h, i,** and **j**). Transgene expression was also

consistently detected in the endothelial lining of AVC tissue (future cardiac valves) and the remainder of the endocardium (**Figure 3.19k**). Despite robust transgene expression detected in the arterial endothelium, more punctate transgene expression was still observed in the endothelium of all major veins, including the primary head vein, the umbilical vein, the portal vein, and the cardinal veins. The latter stained with the least intensity, but this could be due to poorer infiltration of the stain near cardinal veins which are positioned deeper inside the embryo compared to more superficially positioned umbilical veins. Overall, KLF2-66:*LacZ* transgene expression at E12.5 fully recapitulated KLF2-66:*LacZ* transient transgene expression in strongly expressing embryos analysed in Section 3.5.3.2.

At E13.5, robust KLF2-66:*LacZ* transgene expression was observed in leaflets of now fully developed left and right atrio-ventricular (mitral and tri-cuspid, respectively) valves, consistent with previous reports of *Klf2* expression in the developing valves[192]. Transgene expression was also observed in the remainder of the endocardium, the endothelial lining of the cardiac OT, the wall of truncus arteriosus, leaflets of the pulmonary valves and the venous valve at site of entrance of inferior vena cava into the right atrium (**Figure 3.19l, m, n,** and **o**). Transgene expression in the arterial compartment of the endothelium was still robust, and was detected in cerebral arteries, internal carotid arteries, pulmonary arteries, and arch of the aorta at the entrance into the aortic duct, common carotid arteries, common iliac and umbilical arteries, and lumbar arteries. Transgene expression remained strong in the angiogenic hindbrain and other sites of active angiogenesis in the embryo. Compared to transgene expression in arterial ECs, weaker yet consistent transgene expression was observed in the venous endothelium, including in the external jugular veins, the temporal branches of the cerebral venous plexus, and the umbilical and cardinal veins. Punctate transgene expression was observed in the anterior cardinal veins and all expression was absent from the common iliac veins.

At E15.5, assessing transgene expression was challenging due to X-Gal infiltration issues that often occur when larger embryos are stained. Transgene expression could still be detected in the retinal vasculature, leaflets of left and right atrio-ventricular valves, pulmonary venous valves, and venous valve at the site of entrance of the inferior vena cava into the right atrium (data not shown). All transgene expression at the angiogenic front was entirely lost with the KLF2-66:*LacZ* transgene at E15.5, and no transgene expression was observed in superficial capillaries of the E15.5 embryo, in stark contrast KLF2-41:*LacZ* overexpressing E15.5 embryos which exhibited intense transgene expression in the ECs of small capillaries.

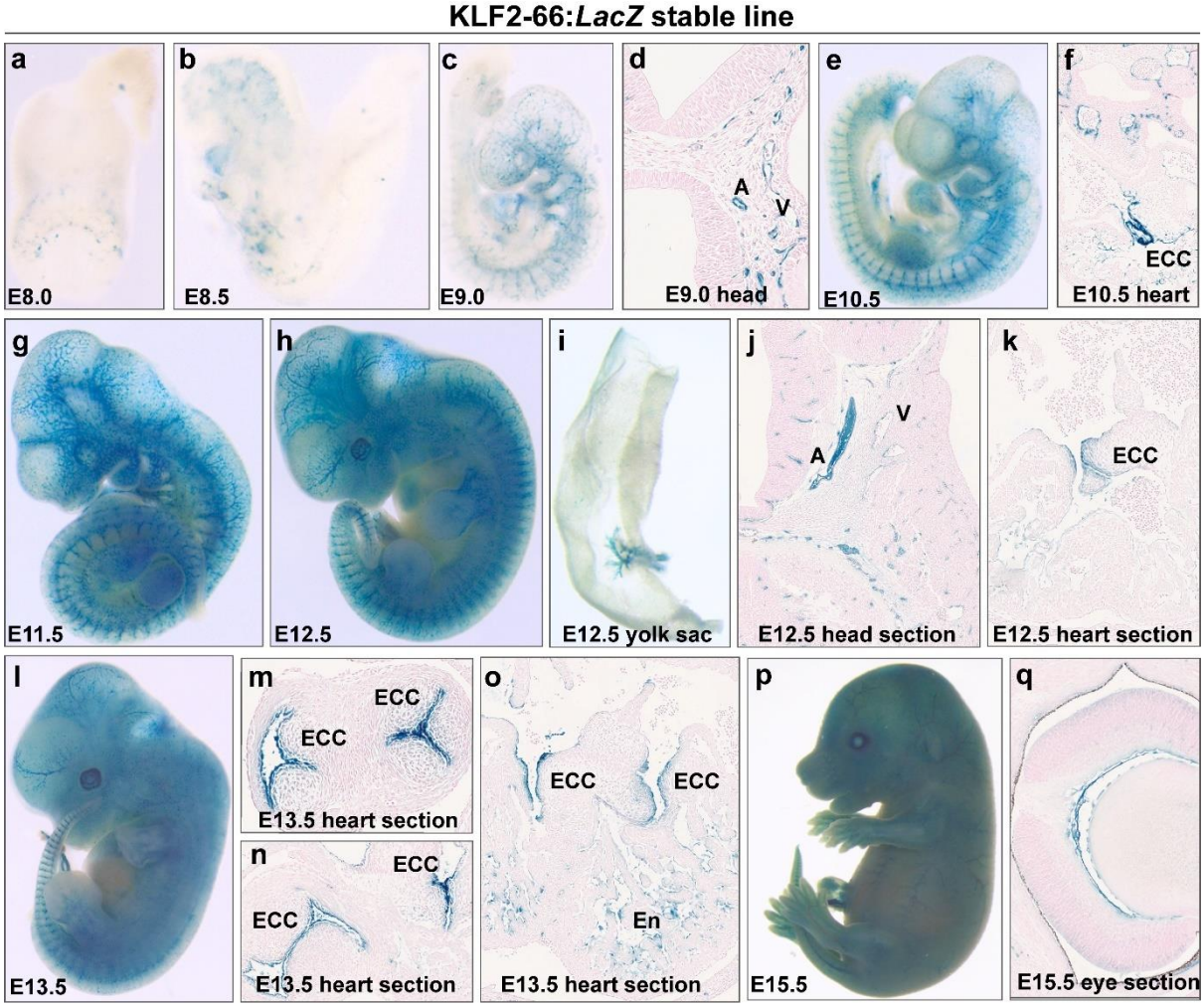


Figure 3.19: Time course analysis of KLF2-66:*LacZ* stable line 2 to assess transgene expression during embryonic development. (a) The KLF2-66:*LacZ* transgene expression was punctate and weak in a E8.0 embryo. **(b)** At E8.5, the KLF2-66:*LacZ* transgene expression was still punctate and weak in the head and trunk area. **(c), (d)** At E9.0, the KLF2-66:*LacZ* transgene expression was observed in individual cells forming the dorsal aortae, the intersegmental vessels, and the OT of the heart. A secton through the head of a E9.0 embryos showed robust

KLF2-66:*LacZ* transgene expression in arterial and venous ECs of the head. **(e), (f)** At E10.5, the KLF2-66:*LacZ* transgene was expressed in the dorsal aorta, the intersegmental vessels, and at the angiogenic front. Section analysis of the E10.5 embryo showed KLF2-66:*LacZ* transgene expression within the paired dorsal aortae; punctate, but still robust transgene expression in cardial veins; and robust transgene expression within the ECCs and the remainder of the endocardium. **(g)** At E11.5, the KLF2-66:*LacZ* transgene remains robustly expressed in arterial ECs and at the angiogenic front. **(h), (i), (j), (k)** At E12.5, the KLF2-66:*LacZ* transgene remained robustly expressed in arterial ECs and at the angiogenic front. Within the E12.5 yolk sac, the KLF2-66:*LacZ* transgene was robustly expressed in the umbilical artery and vein endothelium, but not within the yolk sac vasculature. Section analysis of the E12.5 embryo showed KLF2-66:*LacZ* transgene expression within the endothelium of the head arteries and the angiogenic hinbrain, and to a lesser extent in head veins. Robust transgene expression was also observed in the ECCs. **(l), (m), (n)** At E13.5, the KLF2-66:*LacZ* transgene remained robustly expressed in arterial ECs and at the angiogenic front. A section of the E13.5 heart showed KLF2-66:*LacZ* expression in the ECCs and the remainder of the endocardium. **(p), (q)** The KLF2-66:*LacZ* transgene remained endothelial-specific at E15.5. KLF2-66:*LacZ* expression was observed within the retinal vasculature. **A=artery, AS=aortic sac, OT=outflow tract, ECC=endocardial cushion, V=vein, PT=pulmonary trunk, En=endocardium.**

3.7.3 KLF2-66:LacZ transgene expression in the organ-specific vasculature

3.7.3.1 KLF2-66:*LacZ* transgene expression in the embryonic and post-natal heart

Using the same KLF2-66:*LacZ* stable transgenic line that was used for wholemount embryo analysis, a timecourse analysis of transgene expression in coronary vessels of the developing and post-natal heart was generated. Compared to the KLF2-41:*LacZ* stable line, KLF2-66:*LacZ* transgene expression was stronger at E11.5, and was most intense in the endocardium throughout the heart, including the OT and the AVC cushion tissue with no ECs observed on the surface of either side of the heart (**Figure 3.20a, b,** and **c**). Transgene expression at E13.5 was very similar to E11.5, but it became dramatically restricted to the atrio-ventricular valves at E15.5, best visible on the dorsal side (**Figure 3.20d, e, g,** and **h**). Robust transgene ecpression was additionally observed within the endothelium of the aorta and the pulmonary artery, which are both derived from the OT of the heart. In the early post-natal heart, robust *LacZ* expression was again observed in the coronary veins on the dorsal side, similar to what was observed with the post-natal time points of KLF2-41:*LacZ* transgene analysis in coronary vessels (**Figure 3.20j, k, m, n, p, q**). However, expression in coronary capillaries on the heart surface was of dramatically lower intensity than that observed with the KLF2-41:*LacZ* transgene. As with the KLF2-41:*LacZ* stable line, section analysis of a represetative heart at each timepoint highlighted the previously observed robust transgene expression within the OT, the ECCs, and cardiac valves throughout the timecourse (**Figure16c, d, g, h, k, n, o, r, u, x**).

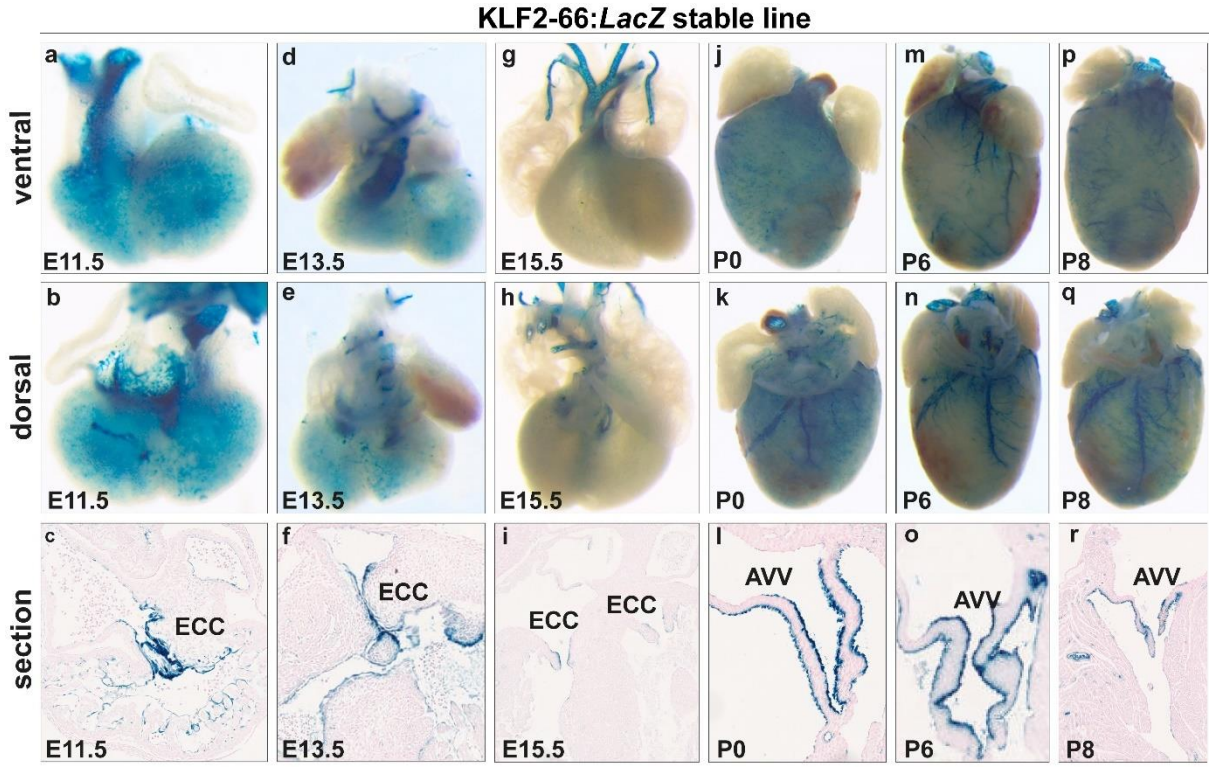


Figure 3.20: Time course analysis of KLF2-66:*LacZ* stable line 2 to assess KLF2-66:*LacZ* transgene expression in coronary vessels during embryonic heart development. (a), (b), (c) At E11.5, the KLF2-66:*LacZ* transgene was robuslt expressed within the OT and AVC cushions, and the remainder of the endocardium. **(d)**, **(e)**, **(f)** At E13.5, The KLF2-66:*LacZ* transgene remained robustly expressed in the OT and the ECCs of the heart. **(g)**, **(h)**, **(i)** At E15.5, the KLF2-66:*LacZ* expression persisted in the OT and AVC cushions. **(j)**, **(k)**, **(l)** In the P0 heart, the KLF2-66:*LacZ* transgene was detected within the coronary venous ECs on the dorsal side of the heart. The KLF2-66:*LacZ* transgene remained robustly expressed in the cardiac valves. **(m)**, **(n)**, **(o)** In the P6 heart, the KLF2-66:*LacZ* transgene remained robusly expressed within the coronary venous ECs on the dorsal side of the heart, as well as atrio-ventricular valve leaflets. **(p)**, **(q)**, **(r)** In the P8 heart, the KLF2-66:*LacZ* transgene remained robusly expressed within the coronary venous ECs on the dorsal side of the heart. Robust KLF2-66:*LacZ* transgene expression was detected within the leaflets of the atrio-ventricular valve. **ECC=endocardial cushion, AVV=atrio-ventricular valve.**

3.7.3.2 KLF2-66:*LacZ* transgene expression analysis in postnatal tissues

Examination of the P6 postnatal lung (**Figure 3.21a**), intercostal vasculature of the ribcage (**Figure 3.21b and c**), the stomach (**Figure 3.21d**), and the diaphragm (**Figure 3.21e**) showed robust transgene expression in large arteries, and low levels of expression in some veins (**Figure 3.21b**), but very limited transgene expression in small capillaries of the organs analysed. This is in line with observations of transgene expression in the post-natal heart, where small vessels on the heart surface did not express

the transgene. This is also in stark contrast to the KLF2-41:*LacZ* transgene expression which is intense in surface capillaries of the post-natal heart and other organs.

Weaker KLF2-66:*LacZ* transgene expression is observed in the P6 retina compared to KLF2-41:*LacZ* transgene expression at the same timepoint (**Figure 3.21f**). The transgene was detected both in retinal arteries and veins, but was again absent from the angiogenic portion of the developing retinal vasculature, in line with postnatal heart and other organ analyses. Post-natal expression of the KLF2-66:*LacZ* transgene was therefore both arterial and venous, but not angiogenic, in the P6 retina; largely arterial, but not angiogenic, in most other P6 organs analysed; apart from the post-natal heart, where strongest transgene expression was detected in coronary veins on the dorsal side of the P6 heart. Robust expression within the endotehlial layer of the post-natal cardiac valves (**Figure 3.20o**) was consistent with robust expression in the ECCs observed throughout development.

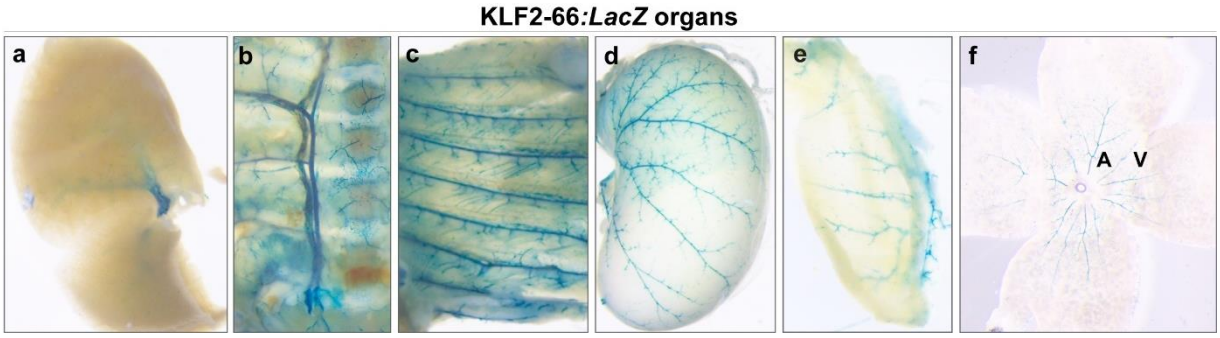


Figure 3.21: Time course analysis of KLF2-66:*LacZ* stable line 2 to assess transgene expression P6 organ-specific vasculature. The KLF2-66:*LacZ* transgene was robustly expressed in **(a)** the large diameter pulmonary artery, **(b), (c)** the intercostal arteries of the ribcage, **(d)** the stomach vasculature, **(e)** the blood vessels of the diaphragm, and **(f)** the arterial and venous, but not angiogenic, ECs of the retinal vasculature. **A=artery, V=vein.**

3.8 Putative KLF2 promoter:LacZ expression in transient transgenic mouse embryos

3.8.1 Rationale

A number of previous studies on *KLF2* transcriptional regulation have focused exclusively on the analysis of the putative KLF2 promoter sequence in cell culture systems. While the experimental systems used (cell type, reporter system, static or shear stress conditions) as well as species and length

of the putative promoter sequence analysed in these studies varied, they all reached similar conclusions on the portion of the promoter required for reporter expression. An early study looked at luciferase expression driven by a number of putative mouse Klf2 promoter pieces ranging from 1.9kb to 72bp upstream of the TSS in LA-4, a lung adenoma cell line[316]. The region between -142bp and -106bp upstream of the TSS was found to induce the highest level of luciferase expression, and further mutational analysis of this region defined the -138/-127bp and -122/-111bp regions as critical for reporter expression. A similar analysis of various mouse *Klf2* promoter pieces in the EOMA (mouse hemangioendothelioma) cell line, this time under static and shear stress (19 dyn/cm^2, 24h) conditions but also looking at luciferase reporter expression, found that the -157/-95bp region upstream of the TSS was required for luciferase upregulation following shear stress application[317]. This finding relating to the mouse putative Klf2 promoter was later validated in HUVEC, where the 30bp tripartite palindrome motif ranging from -138 to 108 bp upstream of the TSS exhibiting perfect conservation between mouse and human was found to be responsible for the reporter induction in response to shear stress[318]. Two further studies of the human putative KLF2 promoter identified a -135/-77 bp region upstream of the TSS, and the -221/-114 region upstream of the TSS, respectively, as required for driving reporter expression in both the absence and presence of shear stress[245,319]. The final study investigated KLF2 promoter inhibition by NF-κB and HDAC signalling and found that deletion between -221bp and -114bp of the human KLF2 promoter sequence resulted in loss of promoter inhibition as measured by reduction in luciferase activity in bovine aortic ECs (BAEC) in the absence of shear stress[272]. Taken together, these studies point towards the requirement for a relatively short yet perfectly conserved KLF2 promoter piece for reporter activation both in static conditions, in response to shear stress, and for inhibition of promoter-driven reporter activity in mouse and human endothelial cell lines. However, none of these studies address the relevance of the putative KLF2 promoter for reporter expression *in vivo*, which is particularly important since *KLF2* is expressed in multiple cell types in addition to ECs[273,274].

To understand whether the KLF2 promoter is able to drive reporter gene expression *in vivo*, and therefore begin to analyse the contribution of the KLF2 promoter in transcriptional regulation of *KLF2* during development, the next aim was to investigate the ability of the putative KLF2 promoter to drive *LacZ* expression in mouse embryos, using the same approach taken above to study enhancer:*LacZ* transgenes. However, in contrast to testing putative enhancers cloned upstream of the hsp68 minimal promoter which is by itself unable to drive any *LacZ* expression in ECs, the KLF2 promoter was cloned into an otherwise promoter-less pAUG-βGal vector[74]. This was an adapted version of the hsp68:*LacZ* construct lacking the hsp68 minimal promoter and with an AUG site in frame with the *LacZ*/β-gal gene. The KLF2 promoter:*LacZ* construct therefore contained no other elements but the putative promoter sequence and the *LacZ* reporter gene. This is a near-identical approach to that taken to examine the role of the Dll4 core promoter in a previously published analysis of the *Dll4* locus[58].

An important consideration was the length of the putative KLF2 promoter sequence to be tested in this transgenic assay. It is now widely accepted that core promoter sequences with embedded transcription start sites and innate capacity to recruit transcriptional machinery often span a few hundred bp[320,31]. This is consistent with a body of evidence on the putative KLF2 promoter tested in cell culture systems most of which independently identified a ~50bp evolutionarily conserved region that was able to drive reporter expression. To determine the final length of promoter to be tested, a phylogenetic foot printing approach was employed to identify the span of the KLF2 promoter immediately upstream of the TSS that was conserved from human to mouse. This analysis identified a 403 bp long putative promoter piece, which from here onwards is referred to as the 0.4 putative KLF2 promoter. This was cloned upstream of *LacZ* within the p-AUG-βGal vector and tested for transient expression in transgenic mouse embryos at E12.5 as before.

3.8.2 KLF2 promoter activity in transient transgenic mouse embryos

Out of seven KLF2-0.4:*LacZ* transiently transgenic embryos PCR positive for *LacZ*, two embryos (#85, #37) were able to drive very weak levels of *LacZ* expression in the vasculature and other compartments of the E12.5 whole-mount embryos (neural tube, eye). However, overall transgene expression in any compartment of these embryos was too weak to designate it as specific to any tissue (**Table 3.2**). In five out of seven embryos PCR positive for *LacZ*, no transgene expression was detected in any tissue (**Figure 3.22**).

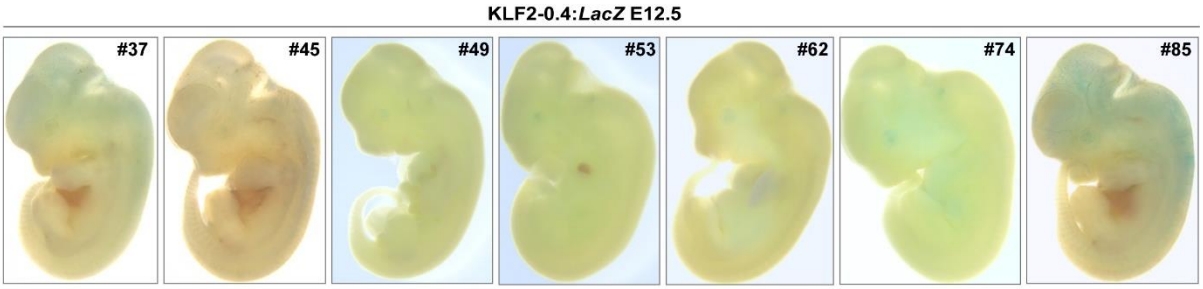


Figure 3.22: Transient transgenic KLF2-0.4:*LacZ* embryos harvested at E12.5 show no vascular transgene expression. All *LacZ* positive embryos are shown.

	KLF2-0.4:*LacZ*
Vascular	0
Ectopic	2
None	5
Total	**7**

Table 3.2: Summary of KLF2-0.4:*LacZ* transgene expression detected in E12.5 mouse embryos. Expression pattern observed was classified as one of the three categories: vascular, ectopic, or no expression.

Despite strong implications for its relevance in mediating the shear stress induction of *KLF2* in cell culture systems, the putative KLF2-0.4 promoter was unable to drive detectable levels of *LacZ* expression in E12.5 mouse embryos.

3.9 Discussion

3.9.1 KLF2 and klf2a putative enhancers

The first aim of this project was to carry out unbiased *in silico* analyses of the *KLF2* and *klf2a* gene loci in order to identify putative KLF2 and klfa2 enhancers and test them in zebrafish and mouse reporter assays *in vivo*. While none of the putative human KLF2 and zebrafish klf2a enhancers identified in the *in silico* screen were able to drive reporter gene expression in transient transgenic zebrafish before or after the onset of blood flow, two putative human *KLF2* enhancers, KLF2-66 and KLF2-41, were able to drive robust *LacZ* expression specifically in vascular ECs and ECCs in E12.5 transiently transgenic mouse embryos.

All putative *KLF2* enhancers qualified for analysis based on enhancer-characteristic histone tail modifications, residence within an OCR measured as DNaseI hypersensitivity, and evolutionary conservation of the sequence at least from human to mouse. However, evolutionary conservation does not always predict functional enhancers, and although most active enhancers are DNase I hypersensitive, DNase I hypersensitive regions are often not active enhancers. While H3K27Ac is enriched on cell type-specific enhancers, chromatin occupancy of EP300, a transcriptional co-activator that catalyses H3K27Ac deposition, has been found to more accurately predict active enhancers[86]. An adult mouse heart H3K27Ac ChIP-seq dataset (Mukesh Jain, unpublished) was analysed for the presence of peaks within the mouse *Klf2* locus and mapped onto the human *KLF2* locus. The most prominent EP300 peaks within this dataset reassuringly corresponded both to the KLF2-66 and KLF2-41 mouse orthologous enhancer regions, as well as the mouse orthologous *KLF2* promoter region. An adult mouse heart ATAC-seq dataset shared with us by the same collaborators additionally validated KLF2-66, KLF-41, and KLF2-0.4 mouse orthologous regions as open in this experimental setting.

Finally, a relevant study on bulk open chromatin profiling of straight and curved areas of the adult mouse dorsal aorta was published after my enhancer analysis was completed[321]. The ATAC-seq dataset emerging from this study revealed regions of open chromatin in the mouse *Klf2* locus which map onto the human KLF2-66 and KLF2-41 enhancers, but not any of the other putative KLF2 enhancers that did not drive *LacZ* expression in transgenic mouse embryos. This dataset therefore validated the two enhancer regions found to be active in the transgenic mouse assay. Interestingly, chromatin openness of neither the KLF2-41 nor the KLF2-66 orthologous regions changed between the "high shear stress" and "low shear stress" mouse aortic areas as defined in the study. However, a parallel RNA-seq dataset obtained in the same study found no significant difference in *Klf2* mRNA expression between the two regions either. One possible explanation is a difference in the numbers of ECs present in "high shear stress" (the whole of descending aorta and a significant proportion of the aortic arch) compared to "low shear stress" (discrete regions of the inner curvature of the aortic arch and intercostal artery branch bifurcations) areas, which combined with additional non-positional endothelial heterogeneity observed between ECs, for example within the descending aorta, might mask some differences in gene expression between the two populations.

The same dataset identified an additional region within the *KLF2* locus significantly more open in areas of high compared to areas of low shear stress. This indicated the presence of an additional, although not yet validated, KLF2 enhancer which could be responsible for aspects of *KLF2* sensitivity to shear stress. However, when this region was mapped back onto the human genome and analysed using the initial enhancer predicting criteria, it was located in a region of chromatin open both in vascular and nearly all non-vascular cell types analysed. This sequence also corresponded to peaks in E12.5 heart EP300 ChIP-seq and ATAC-seq datasets[322], but in each case peak height was lower than that observed at KLF2-66, KLF2-41, and KLF-0.4 regions.

The *KLF2* locus was also analysed using the GeneHancer database mapped onto the UCSC genome browser, a less refined enhancer-gene interaction prediction tool[323]. While tissue-specific protein-binding or open chromatin-based datasets are often more useful as they are carried out in experimental settings of interest, they are generally based on a single experimental set up and combining them with other datasets its difficult exactly because of different experimental conditions. GeneHancer on the other hand ignores the tissue or condition-specific aspect of enhancer activity but integrates candidate enhancers from many genome-wide databases, increasing the confidence in positive hits. It is also able to link candidate enhancers to cognate genes by mapping tissue co-expression of genes and enhancer RNAs (eRNAs), tissue co-expression of genes and enhancer-targeted TF genes, expression quantitative trait loci for variants within enhancers, and finally uses capture Hi-C, a promoter-specific genome conformation assay to infer interactions between different parts of the genome[323]. GeneHancer analysis of the *KLF2* locus identified one high-confidence ("elite") interaction between the KLF2-41 enhancer and the KLF2-0.4 promoter, which relates both to a high likelihood of enhancer definition and strong enhancer-gene association of the KLF2-41 enhancer. The KLF2 putative promoter region was classified as both a promoter and an enhancer. Interestingly, KLF2-66 was not classified as an enhancer using this tool, nor was it shown to interact with the KLF2 promoter.

3.9.2 Human KLF2-66 and KLF2-41 enhancers drive reporter gene expression in the endothelium throughout development and post-natally

Human KLF2-66:*LacZ* and KLF2-41:*LacZ* transgenes were both expressed specifically in the endothelium of E12.5 mouse transient transgenic embryos. In the time-course analysis of the KLF2-41 stable transgenic mouse line, the transgene expression was detected specifically in arterial and angiogenic ECs. The KLF2-66:*LacZ* transgene was also robustly expressed in arterial and angiogenic ECs throughout development, but was also consistently detected in venous ECs, although expression levels there were lower. Both enhancer:reporter transgenes were found to recapitulate most aspects of endogenous *Klf2* mRNA expression between E9.5 and E14.5. The patterning of transgene expression

was consistent between stable transgenic lines, and five to six independent transient transgenic embryos for each transgene, thereby confirming that EC-specific transgene expression observed with these transgenes reflects the true regulatory capacity of each enhancer sequence, rather than its coincidental integration into a transcriptionally permissive genomic region.

An important question was whether the developmental onset of each KLF2-66:*LacZ* and KLF2-41:*LacZ* transgene expression corresponds to the onset of shear stress in the embryo, and whether transgene expression amplifies as shear stress increases throughout development. Firstly, the precise timing of blood flow initiation in a mouse embryo remains controversial, with two main lines of evidence coming to significantly different conclusions. The first study used an ultrasound bio microscopy-Doppler imaging approach detected a heartbeat at the 5 somite (~E8.25) stage, and reliably detected aortic blood flow as early as the 8 somite (~E8.5) stage, suggesting that heartbeat initiation results in almost immediate initiation of embryonic circulation. However, the increase in infiltration of erythroblasts between 5 somite and 8 somite stages observed in this study was modest, and only detected at distal embryonic regions (head and tail)[308]. The second line of evidence comes from erythroblasts tracing in mouse embryos. This study similarly observed that some yolk sac-derived erythroblasts moved to the embryo proper as early as E8.5, which temporally corresponded with the formation of a beating heart tube and linkage of embryonic and yolk sac vasculatures. However, at E9.0, only 10% of all yolk sac-derived erythroblasts were detected in the embryo proper, suggesting that the process of establishing a fully functional circulation is far from completed at this point. Between E9 and E10, the erythroblasts in the embryo proper expanded rapidly, so that a steady state density of red blood cells within the embryo proper was finally achieved at E10.5[309]. The results of the second study fit with *in vivo* mutagenesis data for genes important for initial differentiation and expansion of the hematopoietic lineage (e.g. GATA1), the endothelial network formation (e.g. VEGFR2), and heart beat initiation (e.g. NCX1), all of which are required for the establishment of a functional circulation, yet losing neither of them is lethal before E9.5[324–326].

The KLF2-41:*LacZ* transgene was expressed in all ECs at E8.5, well before the onset of blood flow in the embryo proper. However, its expression did become fully restricted to arterial ECs and was gone from all venous ECs at E10.5 when blood flow is thought to be fully established in the embryo. On the other hand, the KLF2-66:*LacZ* transgene showed very limited expression at early time point prior to blood flow initiation. Transgene expression then increased dramatically at E10, precisely around the time point when blood flow is thought to be fully established. It could be argued that both the KLF2-41:*LacZ* and KLF2-66:*LacZ* transgene expression patterns change around the time point of shear stress onset, whether that be specification of expression to a narrower domain, or robust onset of expression. Both transgenes are still robustly expressed at E15.5, the main difference being strong angiogenic expression observed with the KLF2-41:*LacZ* transgene as opposed to strong expression in large blood vessels and no angiogenic expression observed with the KLF2-66:*LacZ* transgene. This difference carries on post-natally in most organs analysed at P6. One notable exception is the post-natal heart, where both the KLF2-41 and the KLF2-66 enhancers drive robust *LacZ* expression in coronary vein, but not in coronary artery EC, which was inconsistent with systemic transgene expression in the developing and post-natal embryo where arterial expression of the transgene was dominant with both transgenes.

The differentiation and patterning of the coronary vasculature largely occurs between E11.5 and E13.5 in the absence of blood flow, as the connection of the proximal coronary arteries to the aorta only occurs between E13.5 and E15.5[327]. Up until E15.5, KLF2-66 and KLF2-41 enhancers drove most robust *LacZ* expression in the endothelial lining of the OT, the endothelial lining, but not the invading ECs, of the developing valves, and the remainder of the endocardium, all of which are exposed to systemic blood flow. However, no KLF2-66:*LacZ* or KLF2-41:*LacZ* transgene expression was detected in the EC invading the ECCs, suggesting the transgenes are only expressed once the epithelial to mesenchymal transition crucial for ECC formation has taken place[328].

no transgene expression was detected in ECs of any large coronary vessels on the heart surface, with the KLF2-41:*LacZ* transgene detected in endocardium and SV-derived ECs at E13.5, and microvasculature derived from both sources covering the heart at E15.5. Robust transgene expression in large coronary vessels was first detected post-natally but was surprisingly limited to the endothelium of coronary veins and not coronary arteries.

Throughout development, KLF2-41:*LacZ* and KLF2-66:*LacZ* transgenes were robustly expressed at the angiogenic front. This was interesting as there is conflicting evidence about the role of shear stress, and the role of *KLF2,* in developmental angiogenesis. However, studies of the role of shear stress in angiogenesis are mostly limited to cell culture systems[329]. It has been shown *in vitro* that a certain threshold of shear stress (10dyn/cm^2) triggers endothelial sprouting. Once initiated, continuous flow is needed to sustain sprouting and prevent retraction. In local narrowings of blood vessels, shear stress increases to trigger sprouting; this is thought to normalise shear stress across the vasculature[330]. However, the E9.5 angiogenic vascular bed does not achieve a shear stress level of 10dyn/cm^2, although measuring this *in vivo* is complicated[187,331].

KLF2 has been implicated in inhibition of VEGF-A-mediated angiogenesis through inhibition of the VEGFR2 receptor and HIF1[332,333]. However, *Klf2*$^{-/-}$ mouse embryos do not show an angiogenic defect, or increased endothelial proliferation[248,334], although heterozygous *Klf2*$^{+/-}$ mice showed increased microvascular density in the brain[333] and any major angiogenic defects may have been compensated for by KLF4. In zebrafish, *KLF2* is thought to be required for angiogenesis during the development of the aortic arch[335,336]. *KLF4* expression in angiogenesis also seems to be EC sub-type and context-dependent. *KLF4* has been shown to inhibit angiogenesis and endothelial proliferation through regulation of miR-15a[337]. However, in regional microvascular ECs *KLF4* was found to promote angiogenesis via activation of the VEGF signalling pathway[338]. MEF2C, the primary direct transcriptional regulator of the *KLF2* gene studied so far, has been shown to be a key regulator of

sprouting angiogenesis in vascular development[154]. Considering this evidence, MEF2 might induce *KLF2* expression at the angiogenic front. Alternatively, *KLF2* might be differentially regulated in development and adulthood, with different transcriptional co-factors mediating different effects in angiogenesis and shear stress.

3.9.3 Human KLF2 promoter region is unable to drive LacZ expression at E12.5

Due to the previously implied role of the KLF2 promoter in transcriptional regulation of the *KLF2* gene, including its up-regulation in response to shear stress[245,246,317,339], the putative KLF2-0.4 promoter sequence was cloned upstream of *LacZ* and tested in transiently transgenic mouse embryos at E12.5. None of the embryos analysed showed any vascular, or other, *LacZ* expression at this time point. Unlike the EC culture systems, which often used adult and immortalised endothelial cell lines, the putative promoter was unable to drive detectable levels of reporter expression in an *in vivo* developmental setting. However, it should be noted that the possibility of transcriptional activation through the promoter as a feature of shear stress signalling in mature vessels, where ECs experience higher levels of shear stress compared to those experienced by ECs within embryonic blood vessels, could not be excluded based on the results shown here.

3.9.4 Developmental dynamics of the KLF2 gene locus

The identification of two KLF2 enhancers active specifically in ECs, and the finding that the promoter by itself could not drive reporter activity, raises the question of how the *KLF2* locus is regulated as a whole in development. Developmental enhancers are classically viewed as modular and additive, meaning that each enhancer normally recapitulates one aspect of endogenous gene expression pattern (modular), and the sum of all true enhancers regulating said gene together recapitulate all aspects of endogenous gene expression (additive)[87,89]. While there were some differences observed in the pattern of reporter expression driven by the KLF2-41 and KLF2-66 enhancers, there was significant overlap in their activity, and both recapitulated all aspects of the endogenous *Klf2* mRNA

expression in the vasculature during development. There are two main explanations as to why this overlap in activity was observed. Firstly, one of the two enhancers could act as a primary enhancer, while the second could act as a shadow enhancer able to compensate for the function of the first enhancer in the event the first enhancer was unable to regulate *KLF2* expression, for example due to a mutation in a TF binding motif required for its activity. Secondly, while the two enhancers might overlap in expression pattern at time points analysed, their expression might be differentially regulated in the adult setting in response to specific stimuli, for example downstream of shear stress signalling. While gene promoters generally function as docks for TFs, transcriptional co-regulators, and the transcriptional machinery that together enable transcription initiation, both KLF2 enhancers were able to drive reporter expression in the absence of the KLF2 promoter, suggesting that transcriptional information that might be encoded within the promoter sequence is not required for enhancer activity in the developmental setting. Whether the promoter by itself is able to regulate *KLF2* expression without input from the KLF2 enhancers in the adult setting remains to be determined.

3.10 Conclusion

In silico analysis of the *KLF2* locus led to the identification of two *KLF2* enhancers able to drive endothelial-specific reporter gene expression during murine vascular development in the absence of the endogenous promoter. Qualitative analysis of KLF2-41:*LacZ* and KLF2-66:*LacZ* enhancer transgenes throughout development and post-natally showed overlapping expression patterns with endogenous *Klf2* mRNA, although there were important differences in both onset and pattern of transgene expression at specific time points. Some aspects of KLF2-41:*LacZ* and KLF2-66:*LacZ* transgene expression suggest their activity could be induced by shear stress, however, this hypothesis will be directly addressed in future chapters of this thesis. Finally, the putative *KLF2* promoter sequence was unable to drive any *LacZ* expression in E12.5 embryos, suggesting that the KLF2-0.4 core promoter is not the principal regulatory region controlling *KLF2* expression during development. In

the next chapter, KLF2-41, KLF2-66, and KLF2-0.4 enhancer and promoter sequences will be analysed for TF binding motifs to identify direct transcriptional regulators of these enhancer and promoter regions during development.

4 Delineating direct transcriptional regulators of KLF2 and KLF4 enhancers

4.1 Rationale

In the first chapter, two previously undescribed KLF2 enhancers able to drive endothelial-specific reporter expression *in vivo* were identified, and the core KLF2 promoter, which was tested *in vivo* for the first time, was found to be unable to drive robust reporter gene expression during embryonic development. However, the direct transcriptional regulators of the two enhancer elements driving EC-specific reporter expression in the developing embryo remain unknown. Consequently, the aim of this chapter was to identify TFs that bind the KLF2 enhancers and modulate enhancer activity. To achieve this, the functional analysis of different species of KLF2-41 and KLF2-66 enhancers was combined with phylogenetic footprinting, identification of putative TF binding motifs by *in silico* analysis, ChIP-seq and EMSA, and validation of crucial motifs via transgenic analysis of mutant enhancer constructs.

4.2 Analysis of evolutionarily conserved KLF2-41 sequence in transgenic mouse embryos

4.2.1 KLF2-41 phylogenetic foot-printing

In order to identify putative TF binding motifs required for activity of the KLF2-41:*LacZ* transgene, species-specific sequences of the KLF2-41 enhancer were aligned under the assumption that TFBS required for KLF2-41:*LacZ* reporter expression may be more evolutionarily conserved than surrounding regions. This assumption of conservation was later validated in a transgenic mouse model (Sections 4.2.2 and 4.2.3). Upon aligning human, mouse, platypus, and zebrafish sequences using ClustalW[340–342], nine highly conserved regions (R1-9) were identified between the four species (**Figure 4.1**).

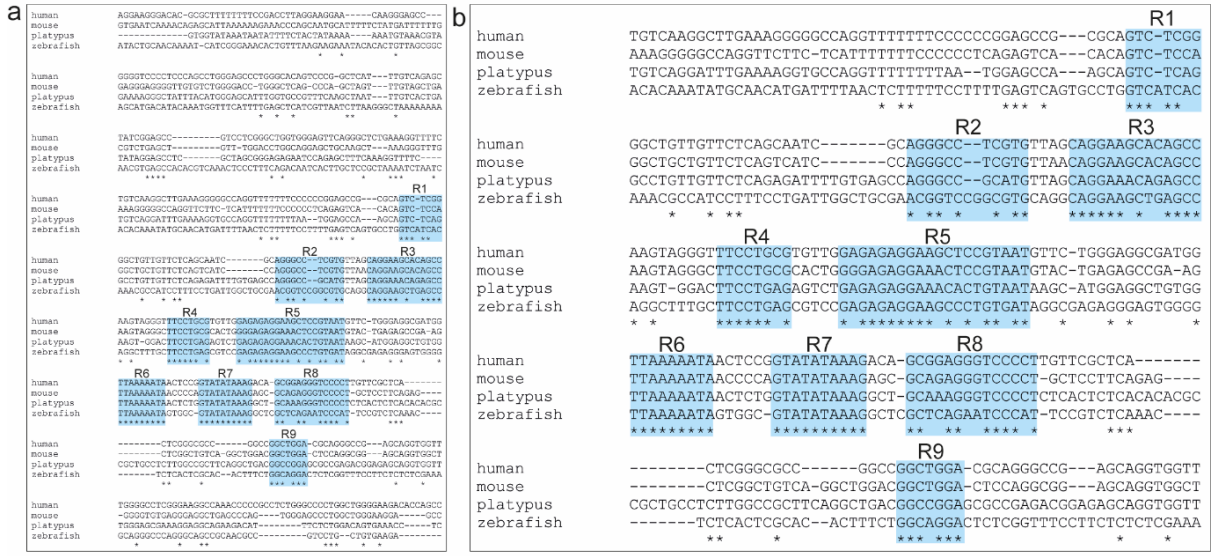


Figure 4.1: **KLF2-41 human enhancer sequence alignment with orthologous mouse, platypus, and zebrafish KLF2-41 enhancer sequences. (a)** The alignment was carried out using ClustalW software. Highly conserved regions are highlighted in blue and annotated as R1-9. The sequence shown in this alignment is a 500 bp sequence that represents the evolutionary conserved enhancer core, but does not represent the whole enhancer sequence that was cloned in the KLF2-41:*LacZ* transgenic construct. **(b)** Close-up of the KLF2-41 enhancer regions conserved between human, mouse, platypus, and zebrafish orthologous sequences. *Denotes conservation between all four species analysed using ClustalW.

Although only nine regions were highly conserved between human, mouse, platypus, and zebrafish sequences, comparing the human sequence to either the mouse, platypus, or zebrafish orthologous sequence alone yielded different results (**Figure 4.2**). For example, there were 21 highly conserved regions between the human and orthologous mouse sequence, 15 highly conserved regions between the human and orthologous platypus sequence, and nine highly conserved regions between the human and orthologous zebrafish sequence which largely, though not always, overlapped with regions conserved between all four species. While different alignment algorithms might compute alignment in slightly different ways, it is nonetheless practical to annotate highly conserved regions for experimental and validation purposes. It should be noted that there are many ways to manually group highly conserved regions based both on their length and the degree of conservation. However, the identification of these regions greatly assists in identifying putative TFBS that need to be investigated, regardless of how they are initially annotated.

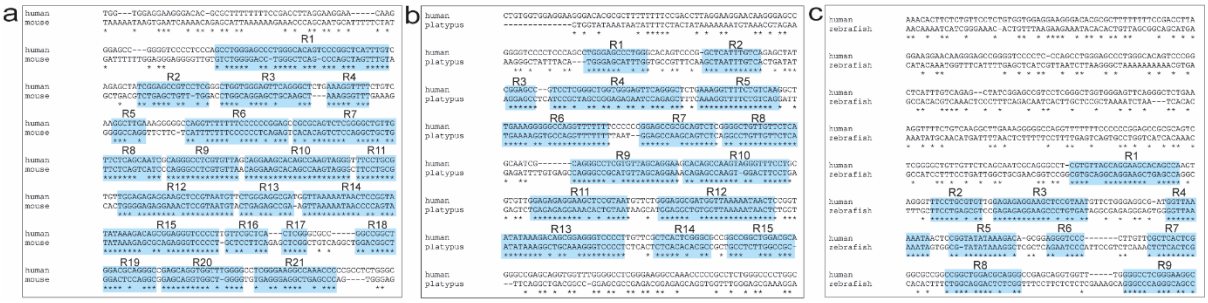


Figure 4.2: **KLF2-41 human enhancer sequence alignment with either the orthologous mouse, orthologous platypus, or orthologous zebrafish sequences demonstrates the degree of conservation of the KLF2-41 enhancer sequence decreases as species evolutionary diverge. (a)** The alignment between human and mouse orthologous KLF2-41 enhancer sequences (54.72% sequence identity) is shown. Highly conserved regions are highlighted in blue and annotated as R1-R21. **(b)** The alignment between human and platypus orthologous KLF2-41 enhancer sequences (59.46% sequence identity) is shown. Highly conserved regions are highlighted in blue and annotated as R1-R15. **(c)** The alignment between human and zebrafish orthologous KLF2-41 enhancer sequences (45.21% sequence identity) is shown. Highly conserved regions are highlighted in blue and annotated as R1-R9. *Denotes conservation between species analysed using ClustalW. The sequence shown in this alignment is a 500 bp sequence that represents the evolutionary conserved enhancer core, but does not represent the whole enhancer sequence that was cloned in the KLF2-41:*LacZ* transgenic construct.

To use phylogenetic footprinting as a technique to identify crucial TFBS, it is first essential to determine the function of the orthologous sequences. If they direct similar patterns of reporter gene expression, this suggests that evolutionarily conserved motifs are important. The analysis of reporter expression driven by orthologous sequences is equally informative if the patterns of gene expression are different, as this indicates non-conserved sequences are important. Here, in order to narrow the analysis of the KLF2-41 enhancer sequence, the assumption of functional conservation between human KLF2-41 and orthologous zebrafish sequence (termed klf2a-6 enhancer in the previous chapter) was the first to be tested as these two sequences shared the smallest number of conserved TFBS.

4.2.2 Klf2-6 enhancer:LacZ expression in transient transgenic mouse embryos

The zebrafish klf2a-6 sequence was cloned upstream of the hsp68 minimal promoter and the *LacZ* reporter and F0 transient mouse embryos transgenic for the klf2a-6:*LacZ* construct were generated, as described in the previous chapter. We received eight PCR-positive mouse embryos transgenic for the klf2a-6:*LacZ* enhancer transgene (**Table 4.1**). As explained in the previous chapter, each embryo represented an independent experiment as the site of transgene integration and the number of copies

integrated were random. Six out of eight PCR-positive embryos expressed *LacZ* as measured by X-Gal staining. Five out of six embryos were positive for *LacZ* expression in the vasculature (#10, #11, #12, #32, #66). In one out of six embryos (#48), klf2-6:*LacZ* transgene expression was detected ectopically, but appeared largely absent from all vascular and most non-vascular tissues (**Figure 4.3**). This was consistent with the numbers of *LacZ*-expressing transgenic embryos obtained with the analysis of the KLF2-41:*LacZ* transgene, where six out of seven embryos exhibited vascular *LacZ* expression (Chapter 3, Section 3.5.3.1).

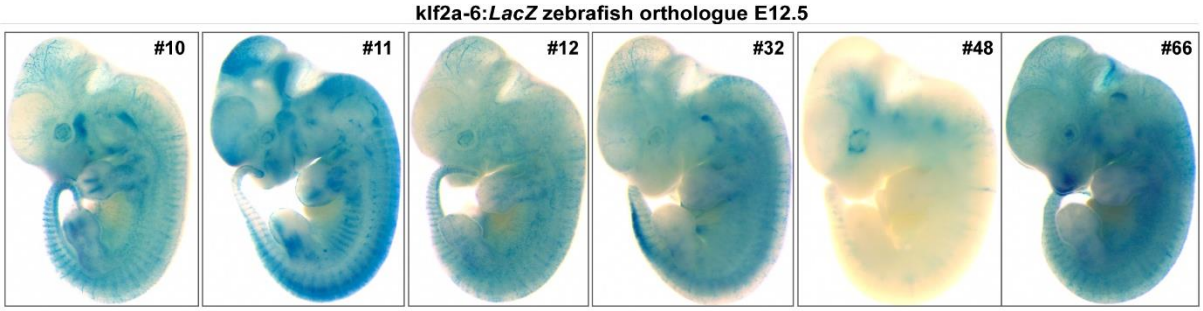


Figure 4.3: **klf2a-6, the zebrafish enhancer orthologous to KLF2-41 drives *LacZ* expression in E12.5 F0 transient transgenic embryos.** All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

While it was evident that both the human and the orthologous zebrafish versions of KLF2-41 were able to robustly drive reporter gene expression in blood vessels, it was important to determine whether the pattern of reporter expression directed by the two transgenes was the same. In order to characterise this expression pattern, four out of five embryos klf2-6:*LacZ* F0 transient embryos with vascular transgene expression (#10, #12, #32, and #66) were sectioned (as described in the previous chapter) (**Figures 4.4** and **4.5**), although detailed analysis focused on embryos #10 and #32, which had the best X-Gal infiltration. Despite differences in X-Gal infiltration, the analysis of whole-mount staining showed that transgene expression was largely consistent between the four embryos sectioned. Some embryos exhibited ectopic *LacZ* expression in the neural tube (#32), the limbs (#10), or the eye (#66), although this was not repeated across embryos suggesting ectopic expression related to the specific transgene integration sites in each embryo.

In blood vessels throughout embryos #10 and #32 (**Figures 4.4** and **4.5**), klf2-6:*LacZ* transgene expression was detected in arterial and angiogenic ECs, consistent with the expression of the KLF2-41*:LacZ* transgene analysed at the same time point. However, transgene expression of the same intensity was surprisingly also detected in venous ECs, whereas no venous transgene expression was observed in analysis of the human KLF2-41*:LacZ* transgene. This was especially apparent in the head and tail sections (**Figure 4.4c** and **g, Figure 4.5c** and **g**), presumably because blood vessels in these two parts of the embryos are easily infiltrated by X-Gal staining. In addition to transgene expression in venous ECs, transgene expression in arterial ECs was weaker overall than that observed with the human KLF2-41:*LacZ* transgene, although it is complicated to compare the intensity of X-Gal staining in F0 transgenic embryos due to differences in integration sites and variability in number of transgene copies integrated in these experiments. Weak transgene expression was also observed throughout the endocardial layer of the OT and the AVCs (**Figure 4.4d** and **e, Figure 4.5d** and **e**), which is consistent with the KLF2-41*:LacZ* transgene expression pattern.

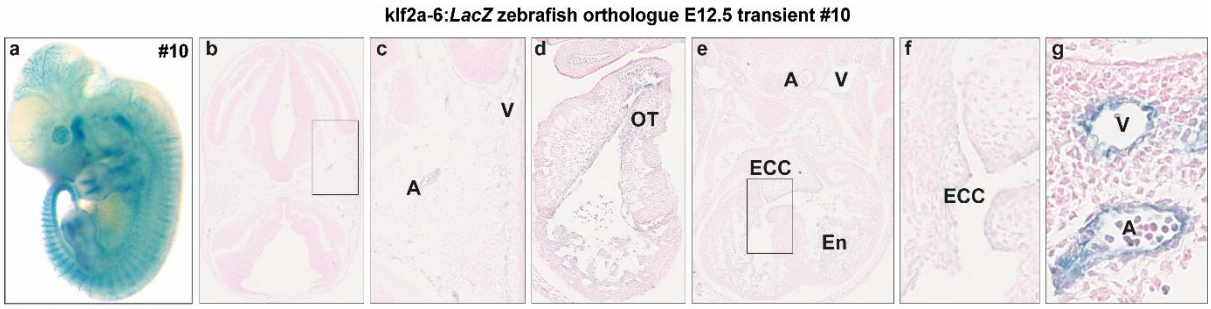


Figure 4.4: Analysis of klf2a-6:*LacZ* #10 embryo sections at E12.5 shows robust EC-specific *LacZ* expression with venous expansion. (a) The whole-mount klf2a-6:*LacZ* E12.5 F0 transient transgenic embryo #10. **(b)** Weak *LacZ* expression is present in arterial, venous, and capillary ECs in the head. **(c)** Close up of *LacZ*-expressing ECs in the head section. **(d)** Weak klf2a-6:*LacZ* transgene expression is detected throughout the endocardial layer of the OT of the heart. **(e)** Punctate klf2a-6:*LacZ* transgene expression is detected in the ECCs (here shown in AVC) and the remainder of the endocardium. **(f)** Close up of *LacZ*-expressing ECs within the ECCs. **(g)** klf2a-6:*LacZ* transgene is robustly expressed in the endothelium of the tail artery and in the endothelium of the tail vein. **A=artery, V=vein, C=capillary, OT=outflow tract, En=endocardium, ECC=endocardial cushion.**

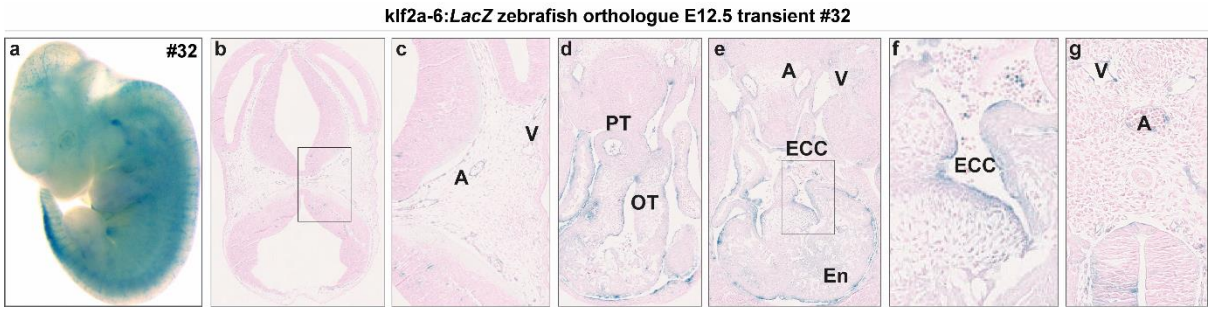


Figure 4.5: Analysis of klf2a-6:*LacZ* #32 embryo sections at E12.5 shows weak EC-specific *LacZ* expression with venous expansion. (a) A whole-mount klf2a-6:*LacZ* E12.5 F0 transient transgenic embryo #32. **(b)** Weak *LacZ* expression is present in arterial, venous, and capillary ECs in the head. **(c)** Close up of *LacZ*-expressing ECs in the head section. **(d)** Weak klf2a-6:*LacZ* transgene expression is detected in the pulmonary trunk and the endocardial layer of the OT of the heart. Robust transgene expression is detected in the epicardium of the heart. **(e)** Robust klf2a-6:*LacZ* transgene expression is detected in the ECCs (here shown in the AVC) and the remainder of the endocardium. **(f)** Close up of *LacZ*-expressing ECs within the ECCs. **(g)** Weak klf2a-6:*LacZ* transgene expression is detected in the endothelium of the tail artery and in the endothelium of the tail vein. **A=artery, V=vein, C=capillary, OT=outflow tract, En=endocardium, ECC=endocardial cushion.**

In conclusion, this analysis demonstrated that the orthologous KLF2-41 zebrafish enhancer klf2a-6 recapitulated all aspects of KLF2-41:*LacZ* activity observed at E12.5, including transgene expression in arterial ECs, capillary ECs, the OT and AVC ECCs. However, the klf2a-6 enhancer consistently drove additional reporter expression in venous ECs that was not observed with the KLF2-41:*LacZ* transgene (analysed in Chapter 3, Sections 3.5.3.1 and Section 3.6.3). This phenomenon, in which a transgene expressed in a specific domain spreads to additional domains, is referred to as expansion, and may suggest that while the TFBS conserved between the two enhancers are responsible for endothelial-specific expression activity, they do not contain the information that regulates arterial-specific activity.

4.2.3 pKlf2-41 enhancer:LacZ expression in transient transgenic mouse embryos

As the orthologous zebrafish klf2a-6 enhancer sequence did not fully recapitulate the KLF2-41:*LacZ* transgene expression pattern, it was next investigated whether a more closely conserved orthologous enhancer, the platypus pKlf2-41 enhancer, would recapitulate either the *LacZ* expression pattern driven by the human or the zebrafish versions. As demonstrated in **Figure 4.2b**, the degree of conservation between human and orthologous platypus KLF2-41 enhancer sequences (59.46%

sequence identity) was higher than that between the human and orthologous zebrafish KLF2-41 enhancer sequences (45.72% sequence identity).

The pKLF2-41 sequence was cloned upstream of the hsp68 minimal promoter and the *LacZ* reporter, as before, and analysed in F0 transgenic mouse embryos at E12.5. We received 11 PCR-positive embryos transgenic for the pKlf2-41:*LacZ* enhancer transgene (**Table 4.1**). In five out of 11 embryos, no *LacZ* expression was detected. In the remaining six embryos, robust *LacZ* expression as measured by X-Gal staining, was detected in blood vessels following whole-mount analysis (**Figure 4.6**). Additional ectopic *LacZ* expression was detected in the neural tube of some embryos (#16, #35, #39).

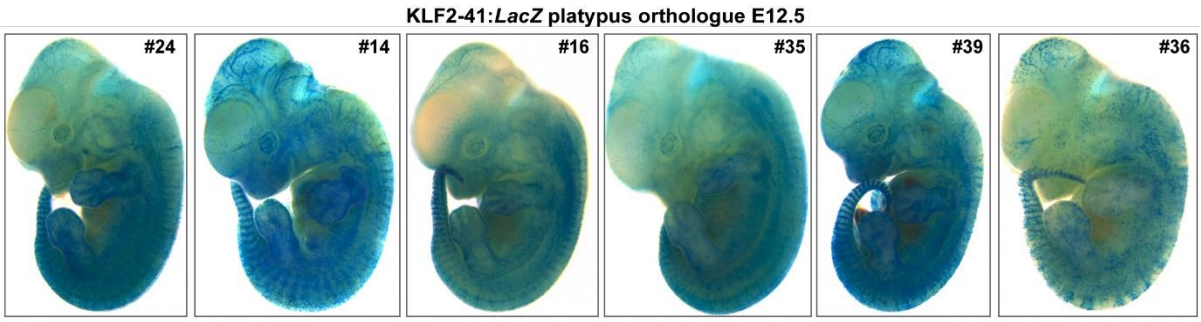


Figure 4.6: The KLF2-41 orthologous platypus sequence (pKLF2-41) drives *LacZ* expression in F0 transient transgenic mouse embryos at E12.5. All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

In order to characterise the transgene expression pattern in pKlf2-41:*LacZ* F0 transient embryos, all six embryos with vascular transgene expression (#24, #14, #16, #35, #39, and #36) were sectioned. Transgene expression pattern was largely consistent between the six embryos sectioned, although some embryos exhibited ectopic *LacZ* expression in the neural tube (#16, #35). Sections of all six embryos were analysed in the same way. Due to some variability observed between intensity of transgene expression, analysis of one representative high-expressing and one representative low-expressing embryo are focused on below.

	KLF2-41:*LacZ*	klf2a-6:*LacZ*	pKLF2-41:*LacZ*
Vascular	6	5	6
Ectopic	1	1	0
None	7	2	5
Total	13	8	11

Table 4.1: Summary of KLF2-41:*LacZ*, klf2a-6:*LacZ*, and pKLF2-41:*LacZ* transgene expression detected in E12.5 mouse embryos. The expression pattern observed was classified as one of the three categories: vascular, ectopic, or no expression.

In embryos #14 and #36, robust pKlf2-41:*LacZ* transgene expression was detected in arterial, angiogenic, and venous ECs (**Figure 4.7c** and **g; Figure 4.8c** and **g**). This pattern of reporter expression was therefore more consistent with the expression of the zebrafish rather than the human KLF2-41:*LacZ* transgene with respect to expression in the venous compartment. Robust pKlf2-41:*LacZ* transgene expression was also observed in the OT and AVC (**Figure 4.7d, e,** and **f; Figure 4.8d, e,** and **f**), consistent with both the klf2a-6*:LacZ* and KLF2-41*:LacZ* transgene expression pattern. Some pKlf2-41:*LacZ* transgene expression was detected in the remainder of the endocardium (**Figure 4.7e; Figure 4.8e**).

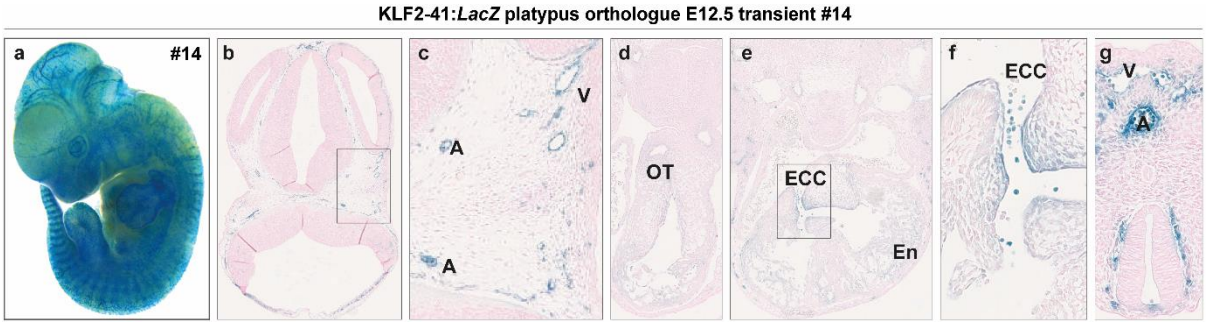


Figure 4.7: Analysis of representative strong-expressing pKLF2-41:*LacZ* embryo sections at E12.5 shows robust EC-specific *LacZ* expression with venous expansion. (a) The whole-mount pKLF2-41:*LacZ* E12.5 F0 transient transgenic embryo #14. **(b)** Robust *LacZ* expression is present in arterial, venous, and capillary ECs in the head. **(c)** Close up of *LacZ*-expressing ECs in the head section. **(d)** Weak pKLF2-41:*LacZ* transgene expression is detected in the OT of the heart. **(e)** pKLF2-41:*LacZ* transgene is robustly expressed in the OT and AVC cushions (here shown in AVC) and the remainder of the endocardium. **(f)** Close up of *LacZ*-expressing ECs within the ECCs. **(g)** pKLF2-41:*LacZ* transgene is expressed in the endothelium of the tail artery and to a lesser extent in the endothelium of the tail vein. **A=artery, V=vein, C=capillary, OT=outflow tract, En=endocardium, ECC=endocardial cushion.**

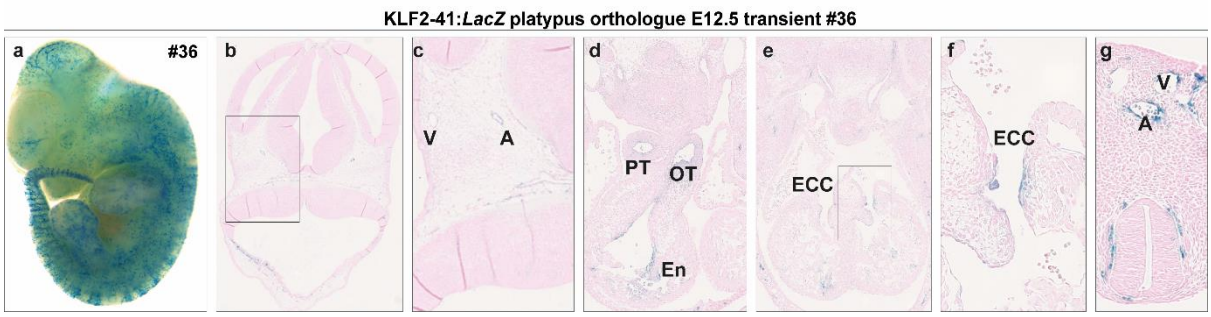


Figure 4.8: Analysis of representative weak-expressing pKLF2-41:*LacZ* embryo sections at E12.5 shows EC-specific *LacZ* expression with venous expansion. (a) The whole-mount pKLF2-41:*LacZ* E12.5 F0 transient transgenic embryo #36. **(b)** Weak *LacZ* expression is present in arterial, venous, and capillary ECs in the head. **(c)** Close up of *LacZ*$^{+}$ ECs in the head section. **(d)** Robust pKLF2-41:*LacZ* transgene expression is detected in the pulmonary trunk and the OT of the heart. **(e)** pKLF2-41:*LacZ* transgene is robustly expressed in the OT and AVC cushions (here shown in the AVC) and the remainder of the endocardium. **(f)** Close up of *LacZ*$^{+}$ ECs within the ECCs. **(g)** pKLF2-41:*LacZ* transgene is expressed in the endothelium of the tail artery and to a lesser extent in the endothelium of the tail vein. **A=artery, V=vein, C=capillary, OT=outflow tract, En=endocardium, ECC=endocardial cushion.**

In conclusion, this analysis found that the orthologous KLF2-41 platypus sequence, pKlf2-41, recapitulated all aspects of KLF2-41:*LacZ* transgene expression observed at E12.5, including expression in arterial ECs, capillary ECs, the OT and the AVC. However, pKlf2-41 consistently drove additional reporter expression in venous ECs that was not observed with the KLF2-41:*LacZ* transgene, but was observed with the zebrafish orthologous sequence klf2a-6. The persistence of pKlf2-41:*LacZ* transgene expression in venous ECs suggests both the orthologous zebrafish and platypus sequences are either missing a TF binding motif that imparts arterial specificity to reporter expression driven by the KLF2-41 enhancer, or alternatively contain a TF binding motif for a regulator that activates reporter expression in the venous compartment. This analysis led to two conclusions. Firstly, both the zebrafish and the platypus orthologous KLF2-41 enhancers contain TFBS able to direct expression specifically in ECs. Secondly, a deviation from the human enhancer sequence common both to the zebrafish and platypus orthologous sequences is responsible for regulating venous-arterial specification within the endothelium.

4.3 Identification of transcription factors interacting with the KLF2-41 enhancer

The KLF2-41 enhancer sequence was next analysed for putative TF binding motifs using a combination of manual inspection and TF binding motif prediction software, including TRANSFAC© and MEME Suite Tomtom motif comparison tool[343,344]. Following this analysis, putative TF binding motifs within the KLF2-41 enhancer sequence were annotated as shown in **Figure 4.9**. In light of the finding that some essential elements were not conserved between zebrafish/platypus and human, TFBS outside of the core conserved elements were also annotated, although previous data obtained in the lab strongly suggests that motifs responsible for endothelial-specific expression are likely to be deeply conserved.

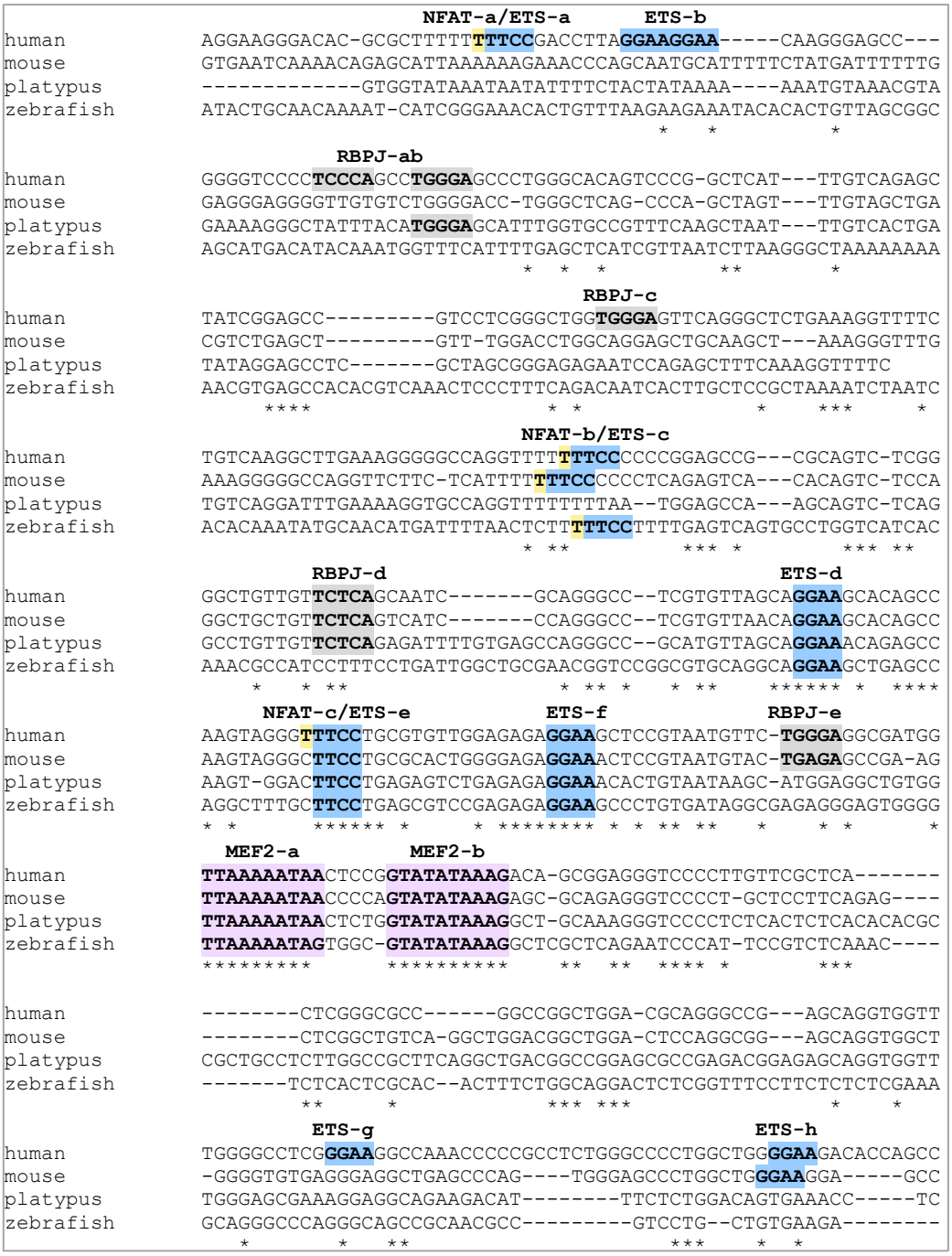


Figure 4.9: **KLF2-41 human enhancer sequence alignment with orthologous mouse, platypus, and zebrafish KLF2-41 enhancer sequences and annotated putative TF binding motifs.** The alignment was carried out using ClustalW software. Putative ETS (blue), RBPJ (grey), MEF2 (lilac), and NFAT (yellow) TFBS are annotated in alphabetical order from the 5' to 3' direction. All putative TFBS were identified through a combination of manual inspection and *in silico* analysis of the sequence. Consensus binding motifs for the TF binding motifs annotated are as follows: GGA$^{A}_{T}$ for ETS factors, $^{C}_{T}$TA$^{T}_{A}$T$^{A}_{T}$A$^{T}_{A}$TA$^{A}_{G}$ for MEF2 factors, $^{C}_{A}$TC$^{C}_{A}$T$^{T}_{C}$A for RBPJ factors, and GGAAA for NFAT factors. *Denotes conservation between all four species analysed using ClustalW. The sequence shown in this alignment is a 500 bp sequence that represents the evolutionary conserved enhancer core, but does not represent the whole enhancer sequence that was cloned in the KLF2-41:*LacZ* transgenic construct.

4.3.1 ETS transcription factors

Analysis of the KLF2-41 sequence identified eight putative ETS binding sites that matched the ETS factor consensus binding motif: GGA$^{A}_{T}$ (**Figure 4.10**). Three out of eight putative ETS binding sites within the KLF2-41 enhancer sequence were conserved between human, mouse, platypus, and zebrafish sequences, while the remaining five sites exhibited varying degrees of conservation (**Figure 4.9**). ETS factors are required for the expression of all endothelial genes and have been shown to regulate all endothelial-specific enhancers described to date[143,345,346]. In line with this notion, endothelial-specific enhancers are often densely populated with ETS binding sites. However, due to their short consensus binding motif, ETS binding sites also often occur throughout the genome by chance. It is therefore important to establish which ETS binding motifs are functional prior to further analysis of their role in regulating the enhancer:reporter transgene expression.

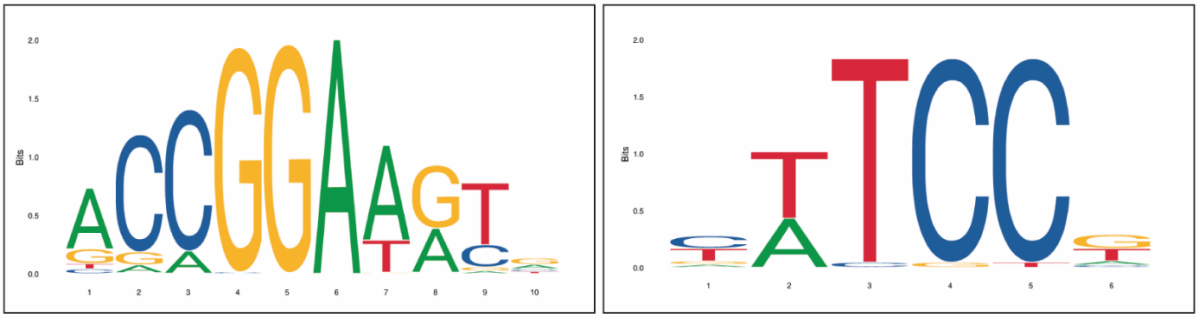


Figure 4.10: The consensus binding motifs for ETS1, a member of the ETS TF family. The consensus motifs were retrieved using the JASPAR database of TF binding profiles stored as position frequency matrices (PFMs). The ETS1 consensus binding motifs were collected to the database externally from scientific publications[347,348].

The ability of the eight putative ETS binding sites to interact with ETS factors *in vitro* was tested in an electrophoretic mobility shift assay (EMSA) using ETS1 DBD, the DNA-binding domain of ETS1. The full ETS1 DBD protein could not be used due to the presence of a repressive domain. Extensive previous experience with ETS motif analysis in the lab demonstrated that ETS1 DBD binding to putative ETS binding motifs within enhancer sequences is a good proxy for identification of functional ETS binding motifs, which can be validated using mutational analysis.

All KLF2-41 putative ETS sites were mixed with ETS1-DBD protein and a radioactively labelled control probe containing a previously validated ETS binding site already demonstrated to interact with ETS factors in EMSA[74]. The ability of each putative ETS binding motif to compete with the radioactively labelled control probe was then assessed by running the DNA-protein complexes on an electrophoretic gel. If an ETS site was functional, it would compete with the control ETS probe for ETS binding. As the unlabelled test ETS sites were added first and in great excess, this effectively prevented the binding of the ETS1DBD protein to the labelled site and resulted in the absence of the band on an EMSA gel post-exposure. In contrast, if the test ETS site was unable to compete with the control probe for ETS binding, ETS1DBD would be able to bind the labelled probe, resulting in the presence of a band on the gel, with an intensity of radioactive signal comparable to that produced by the control condition (where no competitor is added and only the radioactively labelled control probe and the protein are combined).

From all the analysed regions, motifs ETS-c, ETS-d, and ETS-e were able to robustly outcompete the radioactively labelled control probe indicating they are potentially functional ETS motifs (**Figure 4.11a**). Conversely, the inability of the other tested regions to compete strongly indicates that these motifs are not functional ETS binding sites, apart from ETS-f and ETS-h sites which were both able to partially compete for ETS DBD binding and might therefore be important for the regulation of the KLF2-41 enhancer, possibly through cooperative binding.

To assess the ability of KLF2-42 ETS-c, ETS-d, and ETS-e motifs to directly interact with ETS1 DBD, each site was radioactively labelled and then tested for its ability to interact with ETS1 DBD first in the absence of the competitor, to test its ability to directly bind ETS1 DBD. Each site was then also test for its ability to interact with ETS1 DBD in the presence of a WT or mutant unlabelled version of itself, to verify the exact bps in the sequence representing the functional ETS motif (**Figure 4.11b**). The WT and mutant sequences of all oligos used in this and other EMSA experiments can be found in Chapter 2,

Section 2.5.2. Each of the three WT ETS motifs were able to interact with ETS1 DBD without the competitor (shift observed), and to compete with a radioactively labelled version of itself (no shift observed). Conversely, the mutant versions of each motif was, as expected, unable to compete, indicating that the core ETS motifs in each sequence have been correctly identified. Interestingly, ETS-c produced comparatively stronger bands in direct EMSA, suggesting it was able to interact with ETS DBD more readily compared to the other two sites. All three functional ETS sites were conserved between human and zebrafish and were located within the highly conserved KLF2-41 enhancer core.

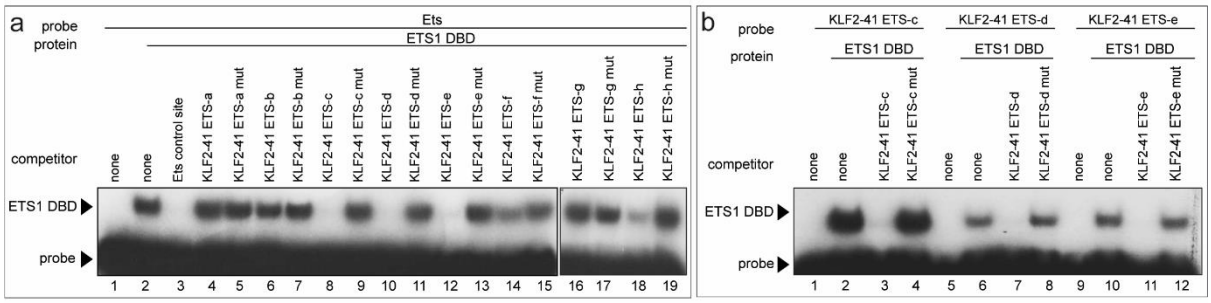


Figure 4.11: ETS factors bind the KLF2-41 enhancer in EMSA. (a) KL2-41 enhancer ETS binding sites c, d, and e compete with the radioactively labelled control probe for ETS1 DBD binding (lanes 8, 10, and 12), resulting in no shift (absence of the band). Mutant versions of the same binding motifs (lanes 7, 9, and 11) do not compete for ETS1 DBD binding, thereby allowing ETS1 DBD interaction with the radioactively labelled control site, resulting in a visible band (shift). **(b)** Radioactively labelled KLF2-41 enhancer ETS binding sites c, d, and e are able to interact with ETS1 DBD in the absence of a competitor (lanes 2, 6, and 10), resulting in a band. All three motifs also compete with radioactively labelled versions of themselves in a direct EMSA (lanes 3, 7, and 11) leading to band shifts, whereas mutant versions of the same binding motifs do not compete (lanes 4, 8, and 12), resulting in visible bands. All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

While EMSA analysis is a powerful method to identify the exact motifs that bind any protein, it is not carried out in a native genomic context. Consequently, it is helpful to support experimental results from EMSA analysis with a second method of verification. Here, a publicly available ChIP-seq dataset that investigated binding of ERG, the ETS factor most abundantly expressed in ECs after E11.5, was analysed[346,349]. This analysis, which was carried out in HUVECs[350] found a statistically significant ERG peak at the KLF2-41 enhancer. Consequently, this validates the EMSA findings that the KLF2-41 enhancer is able to robustly bind ETS factors in a different experimental system (**Figure 4.12**).

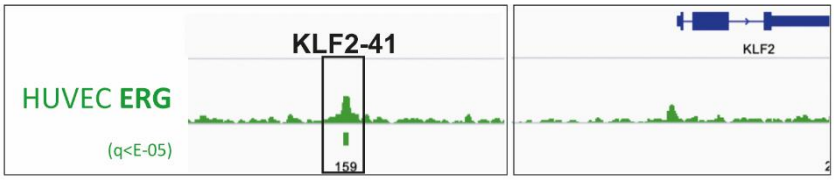


Figure 4.12: Analysis of a publicly available ChIP-seq dataset on ERG binding identifies a statistically significant peak at the KLF2-41 enhancer. Analysis of ERG binding in HUVEC[350] shows a statistically significant ERG peak overlapping with the KLF2-41 enhancer region.

Having established the ability of three conserved ETS binding motifs to bind ETS1 DBD in EMSA, the requirement of these motifs for KLF2-41 enhancer activity was tested next. In order to address this, all three ETS binding motifs that were able to interact with ETS DBD in EMSA were mutated together by exchanging the two nucleotides in the centre of the consensus sequence, e.g. GGAA was mutated to GAGA. These mutations matched the mutant oligos tested in ETS1 DBD EMSA. The KLF2-41 enhancer sequence with mutant ETS sites was then cloned upstream hsp68 and *LacZ* reporter as in previous experiments. Embryos were then analysed at E12.5 and compared to the WT KLF2-41:*LacZ* expression at the same timepoint.

	KLF2-41:*LacZ*	KLF2-41mutETScde:*LacZ*
Vascular	6	0
Ectopic	1	1
None	7	2
Total	13	3

Table 4.2: Summary of KLF2-41:*LacZ* and KLF2-41mutETS-cde:*LacZ* transgene expression detected in E12.5 mouse embryos. The expression pattern observed was classified as one of the three categories: vascular, ectopic, or no expression.

We received 3 PCR positive embryos transgenic for the KLF2-41mutETScde:*LacZ* enhancer transgene (**Table 4.2**). Vascular *LacZ* expression was not observed in any of the *LacZ*-positive mutant embryos analysed, while some ectopic expression of the transgene was observed (**Figure 4.13**). This result suggests that ETS-c, ETS-d, and ETS-e sites are required for KLF2-41:*LacZ* transgene expression, in line with previous reports of the requirement for ETS factors in regulating EC-specific enhancers[73,74,143,345,346,351].

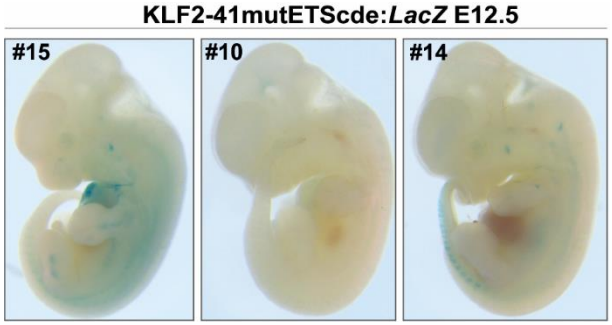


Figure 4.13: KLF2-41 enhancer conserved ETS-c, ETS-d, and ETS-e binding sites are required for KLF2-41-driven *LacZ* expression. All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

4.3.2 RBPJ transcription factor

Analysis of the KLF2-41 sequence identified four putative Recombination signal Binding Protein for immunoglobulin kappa J region (RBPJ) binding sites conforming to the RBPJ consensus binding motif: $^{C}_{A}TG^{G}_{A}T GA^{T}_{C}$ (**Figure 4.14**). While the RBPJ consensus binding motif retrieved from the JASPAR database appears rigid in the middle G nucleotide, other sources have implied a greater flexibility in the identity of this nucleotide, which could be G, A, or T[352].

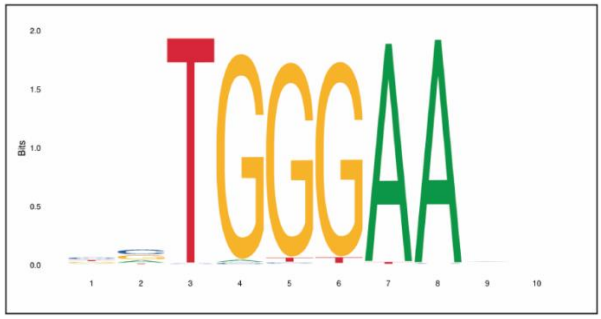


Figure 4.14: The consensus binding motifs for RBPJ. The consensus motifs were retrieved using the JASPAR database of TF binding profiles stored as position frequency matrices (PFMs). The RBPJ consensus binding motifs were collected to the database externally from scientific publication[353].

RBPJ-a, RBPJ-b and RBPJ-d binding motifs were conserved from human to platypus, while RBPJ-e motif was conserved to mouse only, and the RBPJ-c motif was only present in the human sequence (**Figure 4.9**). None of the putative RBPJ binding motifs were conserved to zebrafish. The lack of conservation of putative RBPJ binding motifs between human, platypus and zebrafish sequences represented a potential explanation for the observation that the zebrafish and platypus KLF2-41 enhancer sequences did not drive arterial-specific reporter expression. RBPJ is the transcriptional effector of all canonical

Notch signalling. Upon Notch receptor activation, the receptor intracellular domain is cleaved and imported into the nucleus where it directly interacts with RBPJ which acts as a TF[354,355]. RBPJ regulates arterio-venous differentiation: it has been shown to restrict an intronic Flk1 enhancer expression to arterial ECs through RBPJ-mediated repression of venous expression[351], and to regulate other arterial-specific enhancers, for example an enhancer for Dll4, a Notch pathway ligand[74]. It was therefore hypothesised that the RBPJ binding motifs that were conserved from human to mouse, but only partially conserved to platypus and not conserved to zebrafish, are required for arterial-specificity of KLF2-41 enhancer driven *LacZ* expression. Before this hypothesis was addressed *in vivo*, putative RBPJ motifs found within the human sequence were first tested for their ability to bind RBPJ in EMSA.

In this experiment, all KLF2-41 putative RBPJ sites were tested for their ability to compete for binding of RBPJ protein against a radioactively labelled control probe (containing a validated RBPJ site already shown to interact with RBPJ in EMSA)[74]. While the control RBPJ site was able to compete with the radioactively labelled version of itself for RBPJ binding (leading to the absence of the band), RBPJ-ab (containing both RBPJ-a and RBPJ-b), RBPJ-c, and RBPJ-d putative RBPJ motifs found within the KLF2-41 enhancer were able to only partially compete for RBPJ binding, as evidenced by less intense bands compared to the mutant versions of each site (**Figure 4.15**), and RPBJ-e (conserved only between human and mouse) did not compete at all.

Further direct EMSA experiments in which each RBPJ-d and RBPJ-e sites were each radioactively labelled and assessed for their ability to directly bind RBPJ similarly found that neither were able to interact directly with RBPJ protein (data not shown). RBPJ-d and RBPJ-e sites were both evolutionarily conserved and were therefore considered to be the most likely candidates for functional RBPJ sites, even though RBPJ-e did not appear to compete for RBPJ binding in the indirect EMSA. While both RBPJ-a+b and RBPJ-c sites were able to partially compete for RBPJ binding in the indirect experiment, neither of them was evolutionarily conserved; they were thus excluded from further analysis.

The inability of both conserved RBPJ sites to bind RBPJ in direct EMSA was surprising because both RBPJ-d and RBPJ-e sites perfectly matched the RBPJ consensus binding motif $^{C}_{A}TC^{C}_{A}{}^{T}CA^{T}_{C}$. While investigating TF binding motif-TF interactions using EMSA can be informative, the assay is limited in its recapitulation of the native genomic context, as many components important for TF-DNA interactions, including potential co-factors and co-activators, are not added to the reaction. Despite this, RBPJ has been shown to bind arterial-specific enhancers in EMSA previously without co-factors[74].

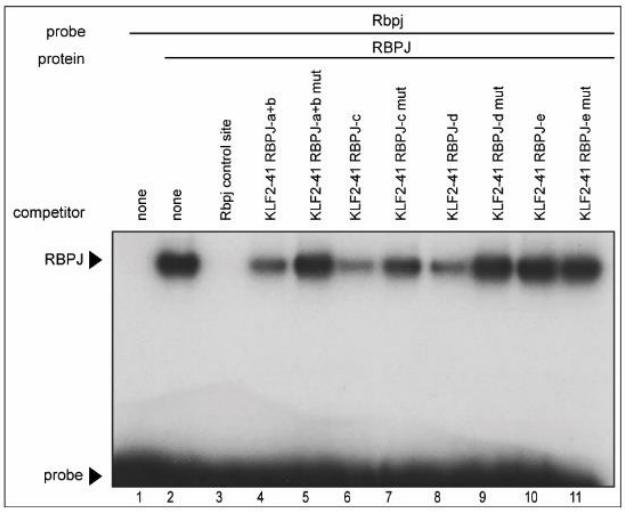


Figure 4.15: RBPJ factors weakly bind the KLF2-41 enhancer in EMSA. (a) KL2-41 enhancer RBPJ binding sites ab, c, and d compete with the radioactively labelled control probe for RBPJ binding (lanes 4, 6, and 8), resulting in the absence of the band. Mutant versions of the same binding motifs (lanes 5, 7, and 9) do not compete for RBPJ binding, thereby allowing RBPJ interaction with the radioactively labelled control site, resulting in a visible band. All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

To investigate whether RBPJ factors might interact with the KLF2-41 enhancer using a different approach, two publicly available RBPJ ChIP-seq datasets using ECs were analysed (**Figure 4.16**). In the first dataset[356], ChIP-seq was performed on HUVECs overgrown to induce cell quiescence. In this dataset, there was a statistically significant RBPJ peak at the KLF2-41 enhancer, suggesting that RBPJ might indeed bind at this region. It should be noted that while this dataset identified statistically significant RBPJ binding at other known RBPJ targets, it also identified significant peaks at genomic regions with no consensus RBPJ motifs, suggesting the number of statistically significant RBPJ targets might be over-estimated. In the second dataset, where RBPJ ChIP-seq was carried out on HUVEC stimulated with VEGFA, switching them to an activated phenotype, a statistically significant peak was not found at the KLF2-41 enhancer[357]. Again, verification of this dataset against other validated RBPJ-

bound enhancers varied: both the DLL4in3 and HEY1 enhancers contained RBPJ motifs and corresponded to statistically significant RBPJ peaks in this dataset, however the DLL4-12 enhancer, which also contains a functionally-verified RBPJ motif, did not correspond to a statistically significant RBPJ-bound peak in this assay.

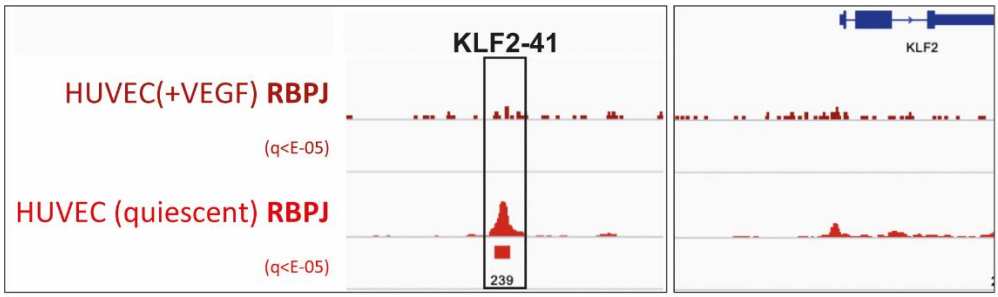


Figure 4.16: **Analysis of two publicly available ChIP-seq datasets shows conflicting results on RBPJ binding to the KLF2-41 enhancer.** Analysis of RBPJ binding in HUVECs stimulated with VEGF[357] shows no RBPJ binding at the KLF2-41 enhancer region. In contrast, analysis of RBPJ binding in quiescent HUVECs[356] shows a statistically significant RBPJ peak at the KLF2-41 enhancer region.

Although RBPJ could not interact with the putative KLF2-41 enhancer RBPJ binding motifs, and the analysis of publicly available ChIP-seq datasets led to conflicting results, the evidence from reporter expression driven by orthologous KLF2-41 enhancer sequences strongly suggested a potential role for the poorly conserved RBPJ motifs, which was sufficient to justify testing whether putative RBPJ binding motifs are required for the arterial specificity of the KLF2-41:*LacZ* transgene expression *in vivo*.

To understand whether these poorly conserved putative RBPJ sites are required for the arterial specificity of KLF2-41:*LacZ* transgene expression, the RBPJ-d and RBPJ-e sites were mutated in the human KLF2-41 enhancer sequence at the same time, while RBPJ-ab and RBPJ-c sites were left intact. These two motifs were selected due to their location within the conserved enhancer core. The KLF2-41 enhancer sequence with mutated RBPJ-d and RBPJ-e sites was then cloned upstream of hsp68 and *LacZ* as in previous experiments and analysed in F0 transient transgenic mouse embryos at E12.5.

We received nine PCR-positive embryos transgenic for the KLF2-41mutRBPJ_de:*LacZ* enhancer transgene (**Table 4.3**). In eight out of nine embryos, *LacZ* expression as measured by X-Gal staining was detected in blood vessels (**Figure 4.17**). In three out of eight embryos with vascular expression, strong *LacZ* expression was detected, whereas only weak *LacZ* expression was detected in the remaining five embryos. In order to characterise transgene expression pattern in KLF2-41mutRBPJde F0 transient embryos, section analysis of a two representative high-expressing embryos is shown below (**Figures 4.18** and **4.19**). In one (#34) out of nine embryos, strong ectopic expression was detected (**Figure 4.17**).

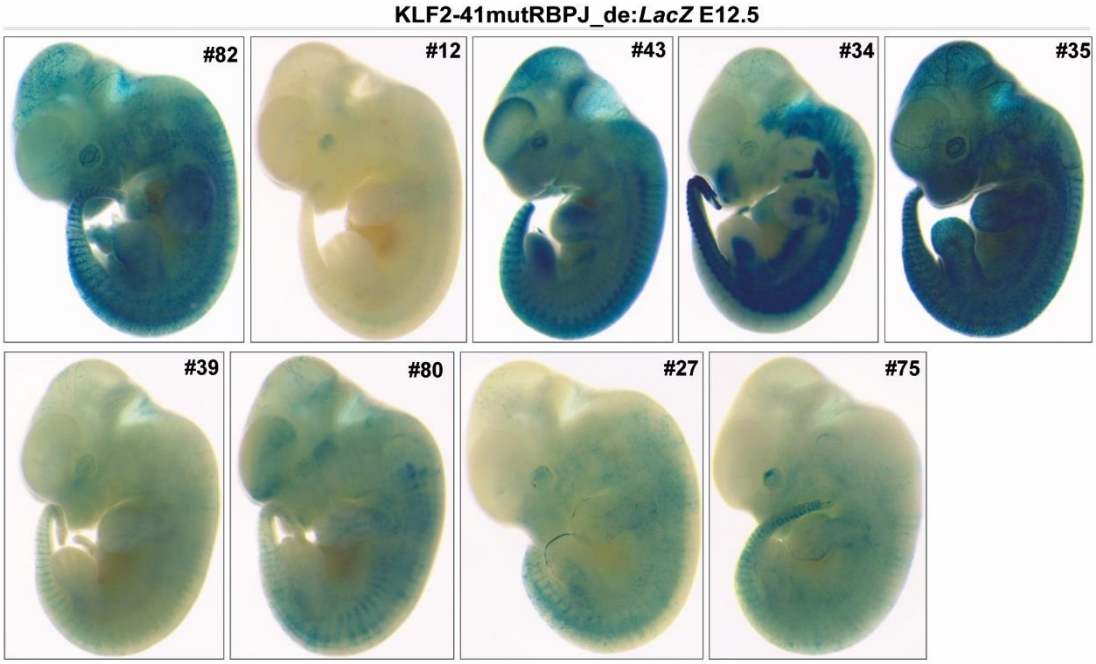


Figure 4.17: **F0 Transient transgenic KLF2-41mutRBPJ_de:*LacZ* embryos harvested at E12.5 show robust endothelial-specific *LacZ* expression with venous expansion.** All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

	Klf2-41:*LacZ*	KLF2-41mutRBPJ_de:*LacZ*
Vascular	6	3+5*
Ectopic	1	1
None	7	0
Total	13	9

Table 4.3: Summary of KLF2-41:*LacZ* and KLF2-41mutRBPJ-de:*LacZ* transgene expression detected in E12.5 mouse embryos. The expression pattern observed was classified as one of the three categories: vascular, ectopic, or no expression. *Denotes weak endothelial transgene expression.

In blood vessels throughout embryos #82 and #35, KLF2-41mutRBPJ-de:*LacZ* transgene expression was detected in arterial, angiogenic, and venous ECs (**Figure 4.18c** and **g; Figure 4.19c** and **g**). This was remarkably similar to the reporter gene expression pattern driven by both the klf2a-6:*LacZ* and pKlf2-41:*LacZ* orthologous enhancer sequences. KLF2-41mutRBPJ-de:*LacZ* transgene expression was also observed in the OT and AVCs, as well as the remainder of the endocardium (**Figure 4.18d, e,** and **f; Figure 4.19d, e,** and **f**), consistent with both the KLF2-41*:LacZ,* klf2a-6*:LacZ,* and pKlf2-41:*LacZ* transgene expression pattern.

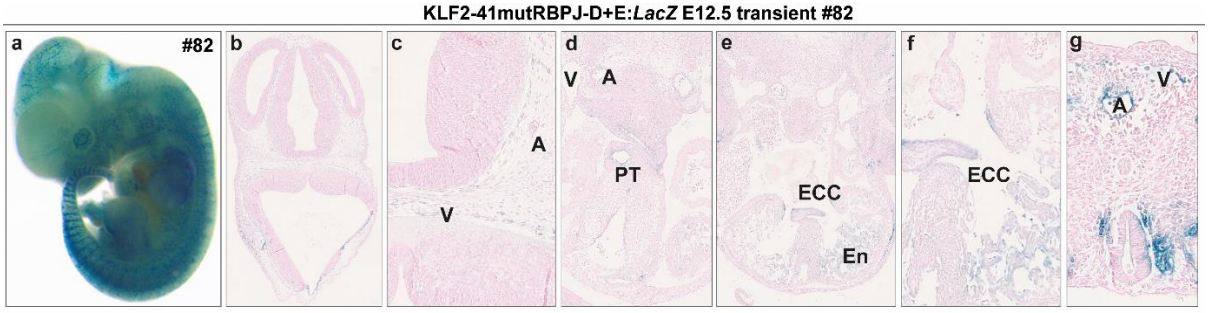


Figure 4.18: Analysis of representative KLF2-41mutRBPJ-de:*LacZ* embryo sections at E12.5 shows robust EC-specific *LacZ* expression with venous expansion. (a) The whole-mount KLF2-41mutRBPJ-d+e:*LacZ* E12.5 F0 transient transgenic embryo #82. **(b)** Weak *LacZ* expression is present in arterial, venous, and capillary ECs in the head. **(c)** Close up of *LacZ*-expressing ECs in the head section. **(d)** Weak KLF2-41mutRBPJ-d+e:*LacZ* transgene is detected in the pulmonary trunk and the OT of the heart. **(e)** Robust KLF2-41mutRBPJ-d+e:*LacZ* transgene expression is detected in the ECCs (here shown in AVC) and the remainder of the endocardium. **(f)** Close up of *LacZ*-expressing ECs within the ECCs. **(g)** KLF2-41mutRBPJ-d+e:*LacZ* transgene expression is detected in the endothelium of the tail artery and in the endothelium of the tail vein. **A=artery, V=vein, C=capillary, OT=outflow tract, En=endocardium, ECC=endocardial cushion.**

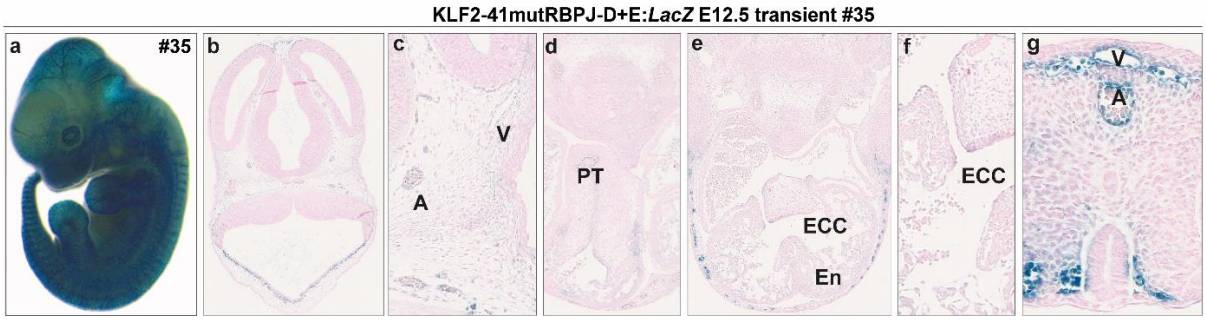


Figure 4.19: Analysis of representative KLF2-41mutRBPJ-de:*LacZ* embryo sections at E12.5 shows robust EC-specific *LacZ* expression with venous expansion. (a) The whole-mount KLF2-41mutRBPJ-d+e:*LacZ* E12.5 F0 transient transgenic embryo #35. **(b)** Weak *LacZ* expression is present in arterial, venous, and capillary ECs in the head. **(c)** Close up of *LacZ*-expressing ECs in the head section. **(d)** Weak KLF2-41mutRBPJ-d+e:*LacZ* transgene is detected in the pulmonary trunk and the OT of the heart. **(e)** Weak KLF2-41mutRBPJ-d+e:*LacZ* transgene expression is detected in the ECCs (here shown in the AVC) and the remainder of the endocardium. **(f)** Close up of *LacZ*-expressing ECs within the ECCs. **(g)** Robust KLF2-41mutRBPJ-d+e:*LacZ* transgene expression is detected

in the endothelium of the tail artery and in the endothelium of the tail vein. **A=artery, V=vein, C=capillary, OT=outflow tract, En=endocardium, ECC=endocardial cushion.**

Analysis of the KLF2-41mutRBPJde:*LacZ* transgene expression patterns recapitulated both the klf2a-6:*LacZ* and pKlf2-41:*LacZ* transgene expression patterns, suggesting that the lack of these two RBPJ sites was responsible for the lack of arterial-specificity of reporter expression driven by klf2a-6 and pKlf2-41 enhancers, and for the resultant expansion of reporter expression to the venous endothelial compartment. Further, the fact that RBPJ-d motif is conserved to platypus, whereas RBPJ-e is not, suggest that either RBPJ-e alone, or RBPJ-d and RBPJ-e sites together, are required for the specificity of reporter expression to the arterial endothelial compartment.

4.3.3 NFAT factors

Analysis of the KLF2-41 sequence identified three putative NFAT (Nuclear factors of activated T cells) binding sites that matched the NFAT factor consensus binding motif: GGAAA (**Figure 4.20**). Two out of three putative NFAT binding motifs were conserved from human to mouse, and one was further conserved to zebrafish (**Figure 4.9**). It should be noted that NFAT consensus binding motif contains an ETS binding motif, meaning that all putative NFAT binding sites are also putative ETS binding sites. Both NFAT-b and NFAT-c sites overlap with evolutionarily conserved ETS binding motifs.

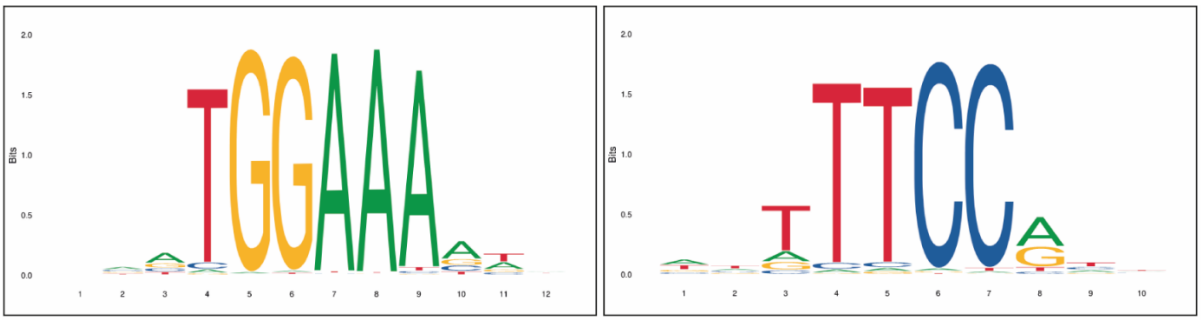


Figure 4.20: The consensus binding motifs for NFATc1. The consensus motifs were retrieved using the JASPAR database of TF binding profiles stored as position frequency matrices (PFMs). The NFATc1 consensus binding motifs were collected to the database externally from scientific publications[358,359].

NFAT factors are a TF family with important roles in the development of lymphatic endothelium[360], and additional roles in development of cardiac valves[361]. As the KLF2-41:*LacZ* transgene was robustly

expressed in ECCs throughout development (Chapter 3, Section 3.6.4.1), as well as post-natal cardiac valves (Chapter 3, Section 3.6.4.1), the regulation of the KLF2-41 enhancer sequence by NFAT factors could offer an explanation for its activity within this compartment. Additionally, NFAT factors have been implicated in regulating genes that lead to endothelial activation and are involved in EC migration, tube formation, and barrier function[362].

To test whether any of the three putative NFAT binding motifs are able to interact with NFATC1 *in vitro*, each putative motif was first tested in indirect EMSA. This analysis found that the NFAT-a motif was able to robustly compete with the radioactively-labelled control NFAT site for NFATC1 binding (**Figure 4.21**). NFAT-b motif was also able to partially compete for NFATC1 binding, whereas NFAT-c was unable to compete. KLF2-41 NFAT-a was then further tested in a direct binding assay where it was able to directly bind NFATC1 and compete with a radioactively labelled version of itself for NFATC1 binding. The inability of the mutant versions to compete again demonstrated that the correct bp had been identified. While NFATC1-b site was initially excluded from direct EMSA analysis due its ability to compete for NFATC1 binding only partially, this motif should be further investigated in the future.

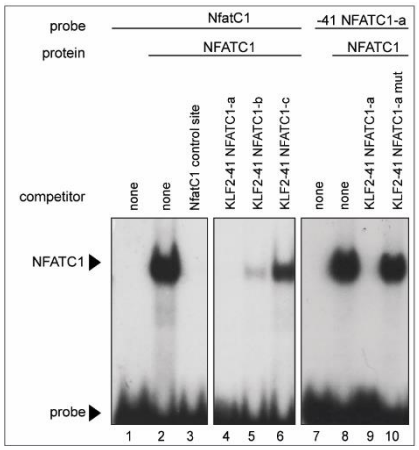


Figure 4.21: NFAT factors bind the KLF2-41 enhancer in EMSA. KLF2-41 enhancer putative NFAT binding sites a and b compete with the radioactively labelled control probe for NFATC1 binding (lanes 4 and 5), resulting in absence of the band. Putative NFAT-c site was unable to compete with the radioactively labelled control. b) Radioactively labelled KLF2-41 enhancer NFAT-a binding site was able to directly interact with NFATC1 protein (lane 8), and compete with a radioactively labelled version of itself (lane 9). The mutant version of the same site did not compete, resulting in the presence of a band (lane 10). All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

The ability of NFAT-a and NFAT-b motifs to robustly interact with NFATC1 in EMSA, especially given the conservation of the NFAT-b site from human to zebrafish, warrants further investigation. The analysis of a publicly available NFATc1 ChIP-seq dataset carried out in HUVEC both under normal

conditions and following stimulation with VEGFA[362] identified a statistically significant NFATc1 peak corresponding to the KLF2-41 enhancer region (**Figure 4.22**). This suggests that NFATc1 might regulate the KLF2-41 enhancer. Mutational analysis of NFAT binding sites NFAT-a and -b is currently in progress to establish whether these sites are required for enhancer-driven reporter expression. However, the overlap between NFAT and ETS binding motifs is expected to complicate analysis of the results.

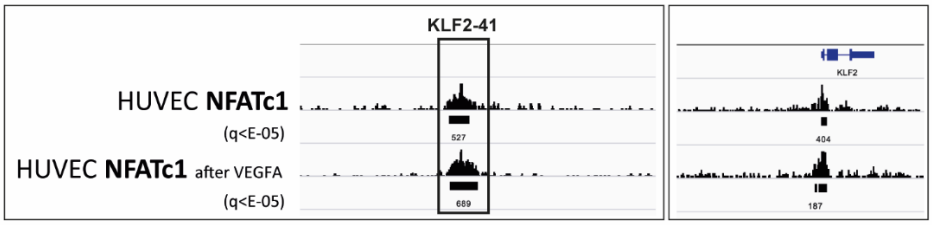


Figure 4.22: **Analysis of a publicly available ChIP-seq dataset on NFATc1 binding identifies a statistically significant peak at the KLF2-41 enhancer.** Analysis of NFATc1 binding in HUVECs under normal conditions or following activation with VEGFA[362] shows a statistically significant NFATc1 peak overlapping with the KLF2-41 enhancer region.

4.3.4 MEF2 transcription factors

Analysis of the KLF2-41 sequence identified two putative MEF2 binding sites that matched the MEF2 factor consensus binding motif: $^{C}_{T}TA^{T}_{A}T_{A}T_{A}T_{A}TA^{A}_{G}$ (**Figure 4.23**). Both putative MEF2 binding motifs were conserved from human to zebrafish, and were only five base pairs apart, a spacing which was conserved between the human and all orthologous sequences analysed (**Figure 4.9**). Cooperative MEF2 binding has been previously demonstrated at the Protein Kinase Amp-Activating Subunit Alpha 2 (*Prkaa2*) enhancer, and up to 30% of cardiac-specific enhancers have been shown to contain paired MEF2 sites[363]. The hypothesis addressing cooperative MEF2 binding to the KLF2-41 enhancer is addressed in Section 4.9.

As discussed in the Introduction, MEF2 factors have been previously implicated in regulation of *KLF2* expression in response to shear stress in cultured ECs, downstream of the ERK5-MEK5 signalling pathway[220,225,243,364]. In addition, a recent *in vivo* study demonstrated that inducible combined deletion

of *Mef2A*, *Mef2C*, and *Mef2D* in adult mice led not only to lethality but also to complete abrogation of *Klf2* and *Klf4* expression[223].

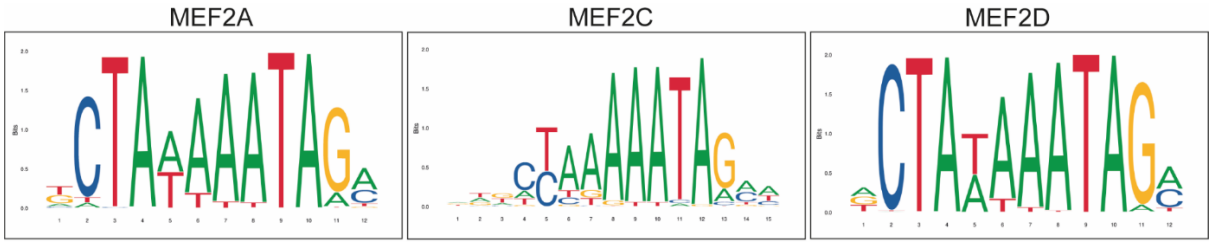


Figure 4.23: The consensus binding motifs for MEF2A, C, and D. The consensus motifs were retrieved using the JASPAR database of TF binding profiles stored as position frequency matrices (PFMs). All MEF2 consensus binding motifs were collected to the database externally from scientific publications[347].

To understand whether the putative MEF2-a and MEF2-b binding motifs interact with MEF2 factors *in vitro*, they were first tested for their ability to compete with a radioactively-labelled control MEF2 site from the Dll4-3 enhancer[74]. In this experiment, only MEF2-a, but not MEF2-b, was able to compete for MEF2C binding with the radioactively labelled control probe (**Figure 4.24a**). Further, the MEF2-a motif was also able to directly bind MEF2A, MEF2C and MEF2D (**Figure 4.24b, c,** and **d**). In contrast, MEF2-b site was unable to directly bind any of the three MEF2 factors. It was therefore concluded that only one of the putative MEF2 sites located within the KLF2-41 enhancer, MEF2-a, is a functional MEF2 binding site, whereas motif MEF2-b is not. Due to the inability of the MEF2-b motif to interact with MEF2 factors in EMSA, this motif was renamed to HBOX, due to its sequence similarity to the consensus sequence $^{G}_{C}$TATATAA$^{A}_{G}$ characteristic of the homeo-box (HBOX) TF family.

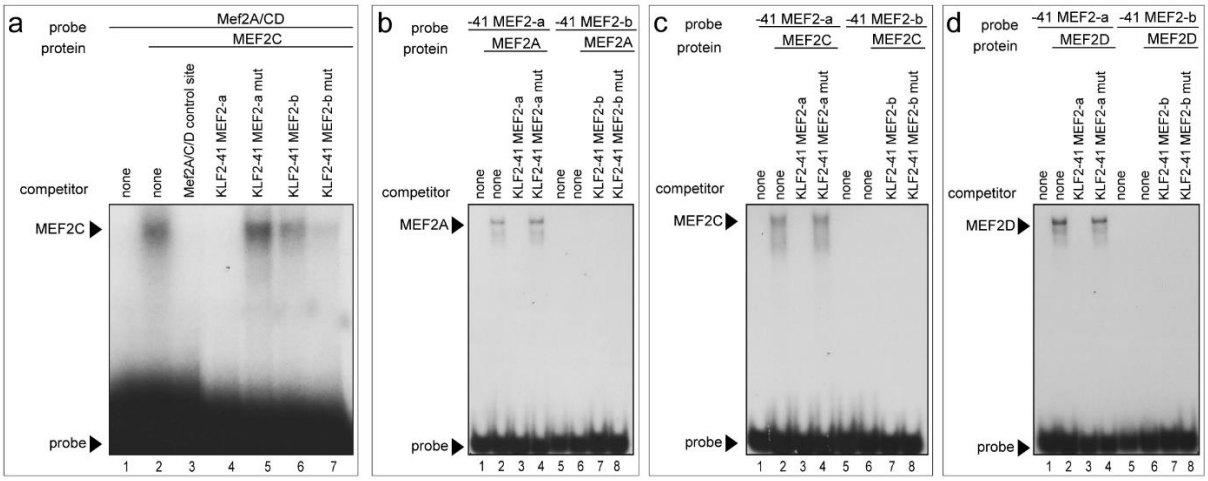


Figure 4.24: MEF2 factors bind the KLF2-41 enhancer in EMSA. (a) KL2-41 enhancer putative MEF2-a binding site competes with the radioactively labelled control probe for MEF2 binding (lanes 4), resulting in a band shift (absence of the band). The mutant version of the same binding motif (lane 5) does not compete for MEF2 binding, thereby allowing MEF2 interaction with the radioactively labelled control site, resulting in a visible band. KLF2-41 enhancer putative MEF2-b binding site is unable to compete with the radioactively labelled control MEF2 site. **(b), (c), (d)** Radioactively labelled MEF2-a site is able to interact with MEF2A, MEF2C, and MEF2D proteins in the absence of the competitor (lane 2 in each panel). MEF2-a motif is also able to with a radioactively labelled version of itself (lane 3 in each panel) leading to band shifts, whereas the mutant version of site MEF2-a is unable to compete, resulting in visible bands (lane 4 in each panel). Radioactively labelled MEF2-b motif is unable to directly interact with either MEF2A, MEF2C, or MEF2D (lane 6 in each panel), or compete with a radioactively labelled version of itself (lane 7 in each panel). All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

In line with the EMSA results demonstrating MEF2A, MEF2C, and MEF2D binding to the KLF2-41 MEF2-a motif, the analysis of MEF2A and MEF2C ChIP-seq datasets carried out in adult mouse heart and lung tissue (the data was shared with us by Prof Mukesh Jain) confirmed that both MEF2A and MEF2C are able bind the KLF2-41 enhancer in these settings (**Figure 4.25**). This confirms that the KLF2-41 enhancer is likely regulated by MEF2 factors.

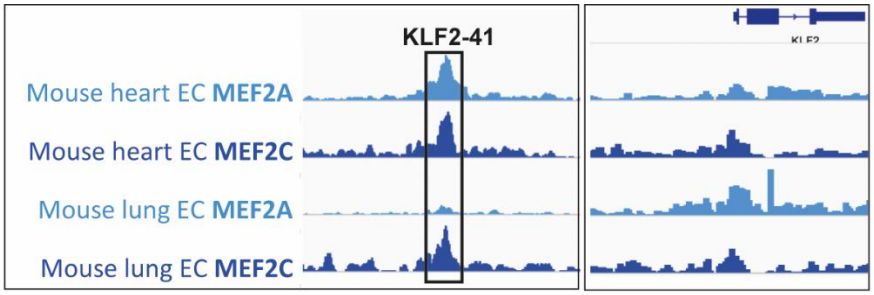


Figure 4.25: **Analysis of MEF2A and MEF2C ChIP-seq datasets carried out in adult heart and lung show MEF2 binding at KLF2-41 enhancer region.** All four unpublished MEF2A and MEF2C ChIP-seq datasets were provided by our collaborator Prof. Mukesh Jain.

Having established that the KLF2-41 enhancer MEF2-a binding motif was able to bind MEF2 factors, the role of the MEF2-a binding motif in KLF2-41 activity was next established. The KLF2-41 MEF2-a site was mutated (using the same mutation as validated in EMSA) then cloned upstream of hsp68 and *LacZ* reporter as in previous experiments. Embryos were then analysed at E12.5 and compared to the WT KLF2-41:*LacZ* expression at the same time point.

We received seven PCR-positive embryos transgenic for the KLF2-41mutMEF2-a:*LacZ* enhancer transgene (**Table 4.5**). In three out of seven embryos, ectopic *LacZ* expression was observed (**Figure 4.26**), for example in the limbs and the forebrain (#58), the neural tube (#58), and the muscle (#51). In the remaining four embryos, no *LacZ* expression was observed either in the vasculature or ectopically. None of the seven embryos analysed had any vascular *LacZ* expression, suggesting that the MEF2-a site is required for the KLF2-41:*LacZ* expression. This is in line with previous reports on the requirement for MEF2 factors in regulating *Klf2* expression and confirms that the KLF2-41 enhancer is also regulated by MEF2 factors.

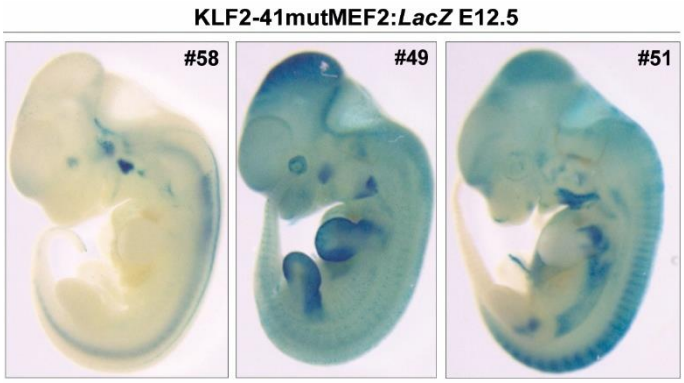


Figure 4.26: **Transient transgenic KLF2-41mutMEF2-a:*LacZ* embryos harvested at E12.5 show no vascular transgene expression.** All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

4.4 Identification of TFs interacting with the KLF2-66 enhancer

To determine the transcriptional regulators of the KLF2-66 enhancer, a similar approach to that of the KLF2-41 enhancer was taken, starting with phylogenic footprinting validated in F0 transiently transgenic mouse embryos to determine putative TF binding motifs required for activity of the KLF2-66:*LacZ* transgene.

4.4.1 Validated phylogenic footprinting of the KLF2-66 enhancer

To identify regions of conserved and non-conserved sequences, orthologous versions of the KLF2-66 enhancer from human, mouse, and Tasmanian devil genomes were aligned. Unlike the KLF2-41 enhancer, the KLF2-66 enhancer was not conserved to opossum and zebrafish, and the Tasmanian devil (*Sarcophilus harrisii*) sequence represented the least conserved species that exhibited a clear orthologous enhancer and good sequencing coverage. Upon aligning human, mouse, and Tasmanian devil sequences using ClustalW, there were 12 regions conserved between at least two of the three aligned species (**Figure 4.27**). Similarities between some of the putative TF binding motifs present in the KLF2-41 and KLF2-66 sequences were immediately obvious and were highlighted in **Figure 4.27**. Despite these similarities, the analysis of the KLF2-66 enhancer was approached in the same way to the KLF2-41 enhancer. It was first determined whether the least evolutionary conserved Tasmanian devil enhancer sequence was able to direct reporter activity in a pattern similar to the human KLF2-66 enhancer.

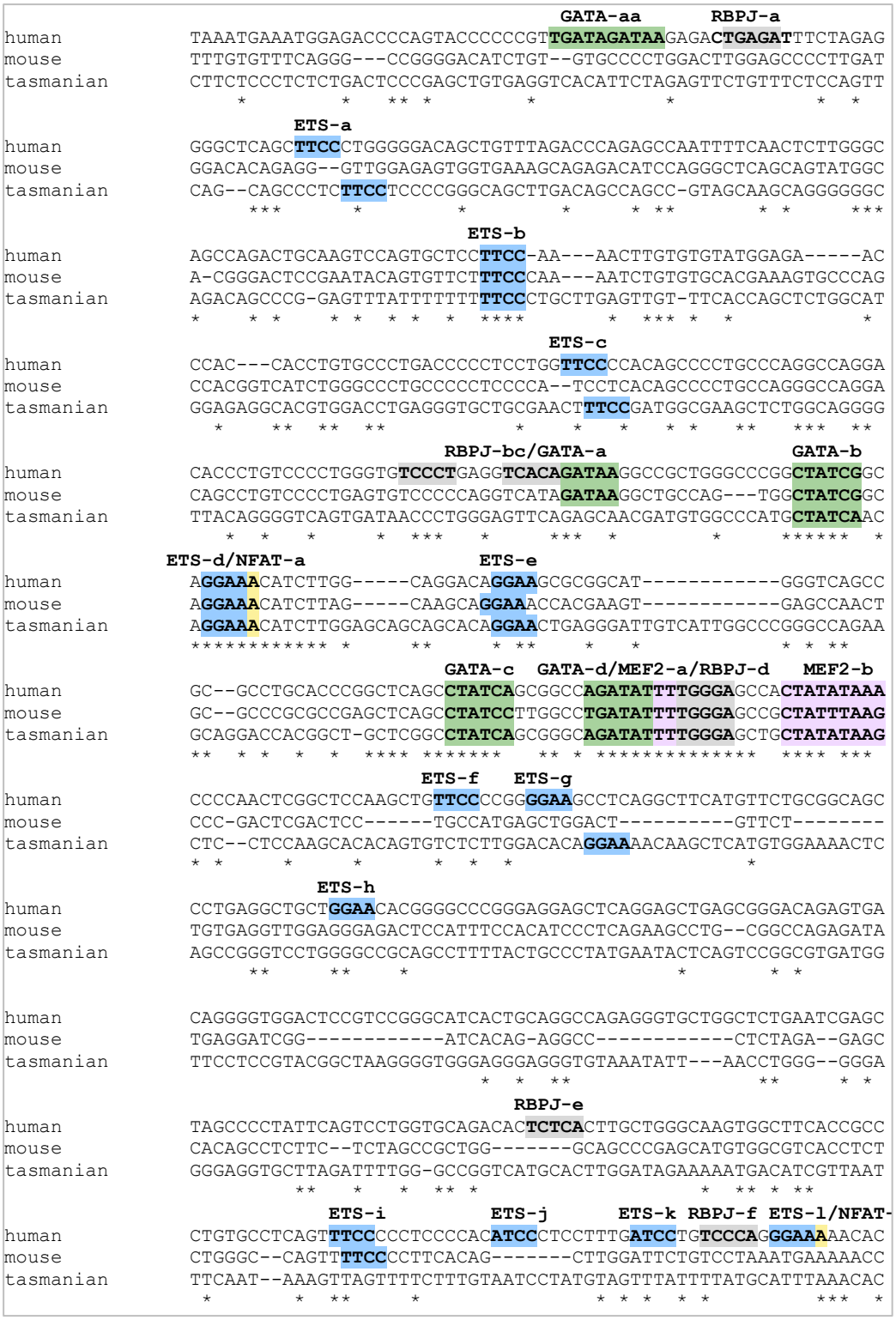


Figure 4.27: **Sequence of the KLF2-66 human enhancer aligned with the orthologous mouse and Tasmanian devil KLF2-66 enhancers alongside annotated putative TFBSs.** The alignment was carried out using ClustalW software. Putative ETS (blue), RBPJ (grey), MEF2 (lilac), NFAT (yellow), and GATA (green) TFBS are annotated in alphabetical order from the 5' to 3' direction. All putative TFBS were identified through manual inspection of the sequence by looking for consensus binding motifs: GGA$^{A}_{T}$ for ETS factors, $^{C}_{T}$TA$^{T}_{A}$T$^{A}_{T}$A$^{T}_{A}$TA$^{A}_{G}$ for MEF2 factors, $^{C}_{A}$TC$^{C}_{A}$T$^{T}_{C}$A for RBPJ factors, GGAAA for NFAT factors, and $^{A}_{T}$GATA$^{A}_{G}$ for GATA factors. *Denotes conservation between all four species analysed using ClustalW. The sequence shown in this alignment is the 500 bp sequence that represents the evolutionary conserved enhancer core, but does not represent the whole enhancer sequence that was cloned in the KLF2-66:*LacZ* transgenic construct.

As before, the Tasmanian devil orthologous enhancer, termed tdKlf2-66, was cloned upstream of hsp68 and lacZ and tested in transgenic mice. We received five PCR-positive embryos transgenic for the tdKlf2-66:*LacZ* enhancer transgene (**Table 4.4**). In three out of five embryos, no *LacZ* expression was detected. In the remaining two embryos, robust vascular transgene expression was observed (**Figure 4.28**). This was consistent with the proportion of *LacZ* expressing embryos observed in the analysis of the human KLF2-66:*LacZ* transgene, where five out of nine PCR-positive embryos expressed the transgene (Chapter 3, Section 3.5.3.2). In order to characterise the transgene expression pattern in tdKlf2-66:*LacZ* F0 transient embryos, both embryos expressing the transgene were sectioned. Transgene expression was largely consistent between the two embryos analysed, analysis of one (#21) of the two embryos is shown here (**Figure 4.29**).

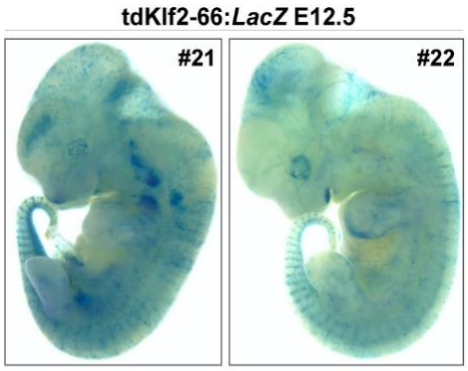


Figure 4.28: **The KLF2-66 orthologous Tasmanian Devil sequence, tdKlf2-66:*LacZ,* drives EC-specific *LacZ* expression in E12.5 embryos.** All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

Robust tdKlf2-66:*LacZ* transgene expression was detected in arterial, angiogenic, and venous ECs (**Figure 4.29c** and **h**), in the OT, AVC, and the remainder of the endocardium (**Figure 4.29d-f**). TdKlf2-66:*LacZ* transgene expression pattern was therefore consistent with the human KLF-66:*LacZ* transgene expression pattern in every aspect analysed (Chapter 3, Sections 3.5.3.2 and 3.7.2), leading to the conclusion that sequence motifs shared by both human and Tasmanian devil are likely important for all aspects of expression. Consequently, the analysis of the potential role of TF binding motifs found in both human and orthologous Tasmanian devil enhancer sequences was initiated.

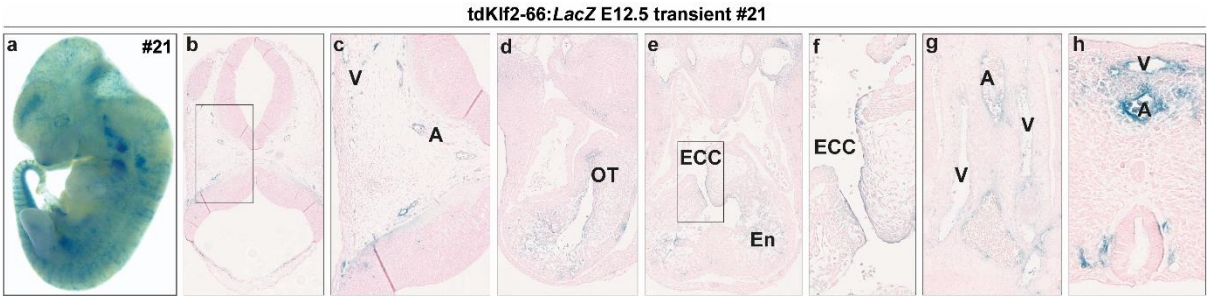


Figure 4.29: The tdKlf2-66:*LacZ* transgene expression in E12.5 sections recapitulates human KLF2-66:*LacZ* transgene expression pattern at E12.5. (a) A whole-mount tdKlf2-66:*LacZ* E12.5 F0 transient transgenic embryo #21. **(b)** Robust *LacZ* expression is present in arterial, venous, and capillary ECs in the head, with stronger expression in arterial ECs. **(c)** Close up of *LacZ*-expressing ECs in the head section. **(d)** Weak tdKlf2-66:*LacZ* transgene expression is detected in the OT of the heart. Some transgene expression is also detected in the epicardium of the heart. **(e)** Robust tdKlf2-66:*LacZ* transgene expression is detected in the ECCs (here shown in AVC) and the remainder of the endocardium. **(f)** Close up of *LacZ*-expressing ECs within the ECCs. **(g)** Robust tdKlf2-66:*LacZ* transgene expression is detected in the endothelium of the abdominal aorta, and to a lesser extent in the endothelium of the abdominal vein. **(h)** Robust tdKlf2-66:*LacZ* transgene expression is detected in the endothelium of the tail artery and to a lesser extent in the endothelium of the tail vein. **A=artery, V=vein, C=capillary, OT=outflow tract, En=endocardium, ECC=endocardial cushion.**

	KLF2-66:*LacZ*	tdKLF2-66:*LacZ*
Vascular	5	2
Ectopic	0	0
None	4	3
Total	9	5

Table 4.4: Summary of KLF2-66:*LacZ* and tdKLF2-66:*LacZ* transgene expression detected in E12.5 mouse embryos. The expression pattern observed was classified as one of the three categories: vascular, ectopic, or no expression.

4.4.2 ETS transcription factors

As with the KLF2-41 enhancer, numerous putative ETS binding sites were identified in the KLF2-66 enhancer sequence, six of which were conserved between the human and Tasmanian devil sequences (KLF2-66 ETS-a, b, c, d, e, and g). Two out of 11 putative ETS binding motifs, ETS-d and ETS-e, were able to robustly compete for ETS1 DBD binding with the control ETS probe, and to directly bind ETS1 DBD (**Figure 4.30a**). Both of these sites were conserved between human, mouse, and Tasmanian devil, and located within the highly conserved KLF2-66 enhancer core. Again, the inability of the mutated versions to compete for binding confirmed the correct bp were identified (**Figure 4.30c**)

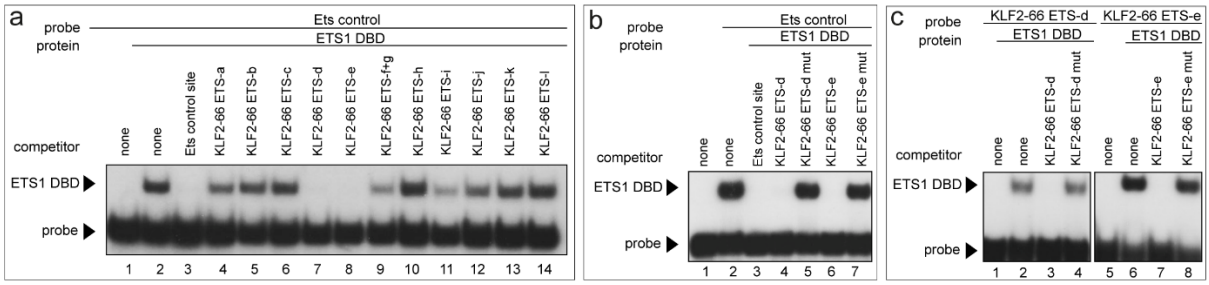


Figure 4.30: ETS factors bind the KLF2-66 enhancer in EMSA. (a) KLF2-66 enhancer ETS binding sites -d, and -e could compete with the radioactively labelled control probe for ETS1 DBD binding (lanes 7 and 8), resulting in no band shift. **(b)** Mutant versions of KLF2-66 enhancer ETS-d and -e binding sites could not compete with a radioactively labelled control probe for ETS DBD binding, although the WT versions could. **(c)** Radioactively labelled KLF2-66 ETS-d and e bound ETS1 DBD in the absence of a competitor (lanes 2 and 6), resulting in a band. All three motifs also competed with radioactively labelled versions of themselves in a direct EMSA (lanes 3 and 7), whereas mutant versions of the same binding motifs could not compete (lanes 4 and 8), resulting in a band-shift. All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

In the publicly available EC ERG ChIP-seq dataset[350], the KLF2-66 enhancer also overlapped with a statistically significant ERG peak, further validating the ability of ETS factors to bind this enhancer sequence (**Figure 4.31**).

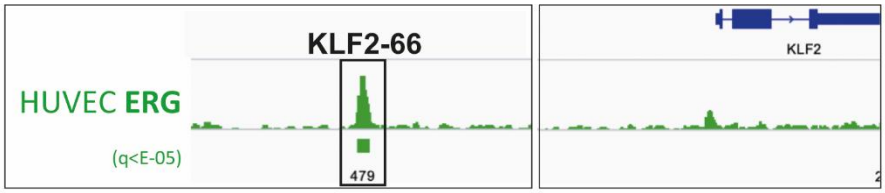


Figure 4.31: Analysis of a publicly available ChIP-seq dataset of ERG binding in HUVEC[350] identifies a statistically significant peak over the KLF2-66 enhancer.

4.4.3 RBPJ transcription factors

In the KLF2-41 enhancer, two putative RBPJ binding motifs were required for arterial specificity, although direct RBPJ binding could not be demonstrated in EMSA. Despite expression of the KLF2-66:*LacZ* transgene in venous as well as arterial ECs, several putative RBPJ binding motifs were identified in the KLF2-66 enhancer sequence, although only one of these motifs, RBPJ-c, was conserved between human and Tasmanian devil sequences. While RBPJ-c was located within a highly conserved enhancer core, it also overlapped with multiple other putative TF binding motifs. However,

none of the putative KLF2-66 enhancer RBPJ binding motifs were able to compete with a radioactively labelled RBPJ control probe in EMSA (**Figure 4.32**).

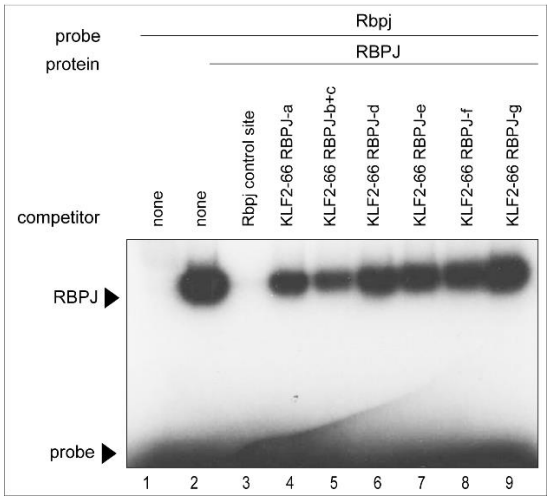


Figure 4.32: RBPJ factors did not bind the KLF2-66 enhancer in EMSA. (a) None of the putative RBPJ binding sites within the KLF2-66 enhancer sequence were able to compete with the radioactively labelled control probe for RBPJ binding. All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2. Note that KLF2-66 Rbpj-g is not shown in **Figure 4.27** due to its position outside of the conserved enhancer core.

Again, a statistically significant RBPJ binding peak corresponded to the KLF2-66 region in one of the analysed publicly available ChIP-seq datasets, although similar to KLF2-41 it was absent in the other (**Figure 4.33**)[356]. It is therefore unclear whether RBPJ motifs play a role in KLF2-66 enhancer activity. A lack of role would be consistent with the fact that KLF2-66:*LacZ* enhancer transgene was not arterial-specific, and would provide an explanation for overlapping, yet distinct *LacZ* expression patterns driven by the KLF2-66 and KLF2-41 enhancers, respectively. However, the interaction of RBPJ with the KLF2-41 enhancer was difficult to prove in EMSA, allowing for the possibility that the same limitations of the experimental approach apply to the KLF2-66 putative RBPJ binding motif analysis. A definitive approach to exclude RBPJ from the list of potential transcriptional regulators of the KLF2-66 enhancer would be to mutate the conserved RBPJ-c binding motif, and test the resulting transgene *in vivo*. However, due to the overlap between the conserved RBPJ-c and other TF binding motifs, a different site might be inadvertently mutated in the process, thereby complicating the analysis of transgenic embryos. Further, Notch regulation of arterial gene expression is already well established in the literature, and resources available for mutagenesis experiments were instead directed into the analysis of less well characterised TF binding motifs.

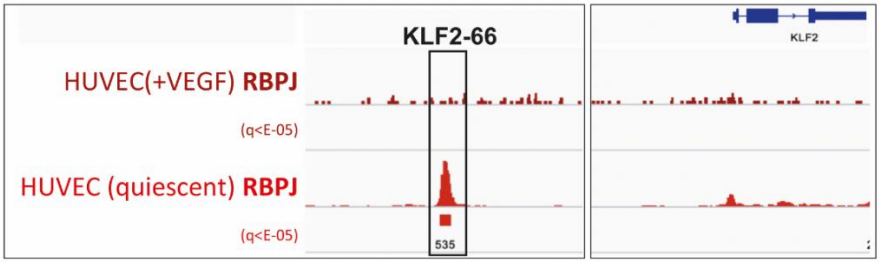


Figure 4.33: Analysis of two publicly available ChIP-seq datasets showed conflicting results on RBPJ binding to the KLF2-66 enhancer. Analysis of RBPJ binding in HUVECs stimulated with VEGF[357] showed no RBPJ binding at the KLF2-66 enhancer region. In contrast, analysis of RBPJ binding in quiescent HUVECs[356] shows a statistically significant RBPJ peak at the KLF2-66 enhancer region.

4.4.4 GATA factors bind the KLF-66 enhancer in vitro

In addition to ETS and RBPJ motifs, inspection of the KLF2-66 enhancer sequence identified multiple GATA factor binding motifs, which have a consensus binding motif of $^{A}_{T}GATA^{A}_{G}$ (**Figure 4.34**). While best known for their regulation of the hematopoietic lineage, GATA factors are known to also play a role in the regulation of EC identity[365] and valve development[366,367]. Additionally, they have been shown to be regulated by mechano-sensitive pathways, for instance during the formation of lymphatic vessels and valves[368–370].

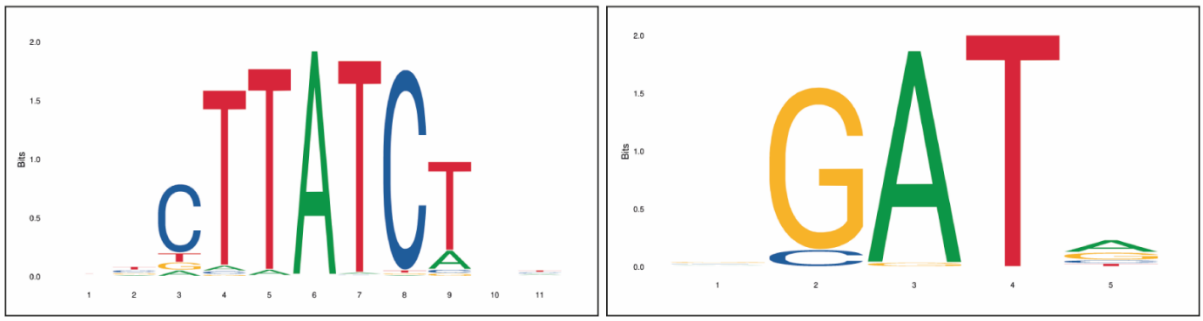


Figure 4.34: The consensus binding motifs for GATA2. The consensus motifs were retrieved using the JASPAR database of TF binding profiles stored as position frequency matrices (PFMs). The GATA2 consensus binding motifs were collected to the database externally from scientific publications[353,371].

Five putative GATA motifs were located within the KLF2-66 sequence, three of which were conserved between human and Tasmanian devil and located within the highly conserved enhancer core. In order to assess their ability to interact with GATA factors *in vitro*, these sites were tested both in indirect and direct EMSA assay. These experiments were performed entirely by Dr Sarah De Val. Putative GATA

sites -ab, -c, -e, and -f were all able to compete for GATA2 binding at least partially with the radioactively labelled control Gata2 site (**Figure 4.35a**). However, in direct EMSA experiments, only GATA-ab and GATA-c sites were able to directly bind GATA2. Their WT, but not mutant versions, were able to compete for GATA2 binding in the same direct assay (**Figure 4.35b**).

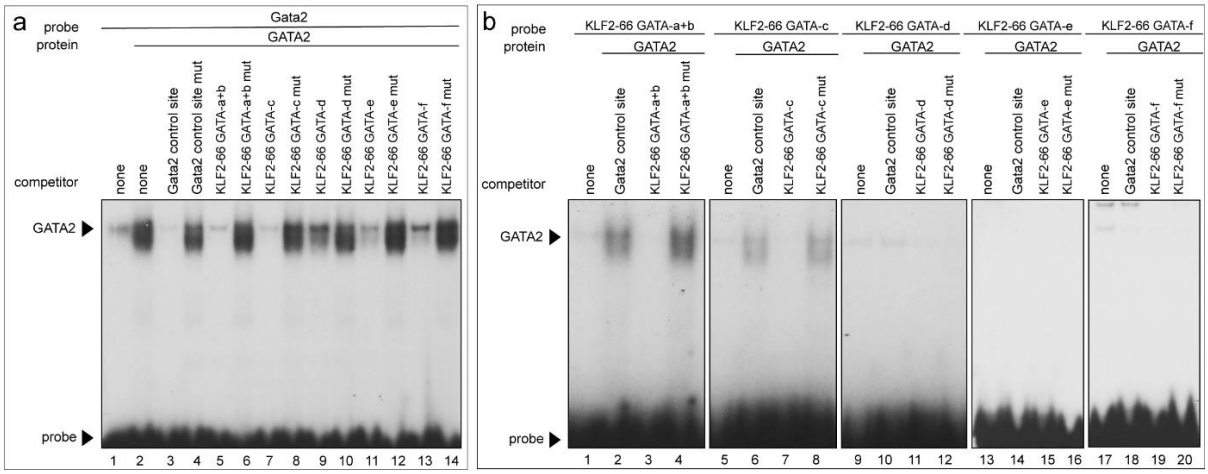


Figure 4.35: GATA factors bound the KLF2-66 enhancer in EMSA. (a) KLF2-66 enhancer putative GATAs binding sites ab and c competed with the radioactively labelled control probe for GATA2 binding in indirect (lanes 5 and 7) EMSA, resulting in absence of the band. The mutant version of the same binding motifs (lanes 6 and 8) did not compete for GATA2 binding, thereby allowing GATA2 interaction with the radioactively labelled control site, resulting in a visible band. KLF2-66 enhancer putative GATA binding sites e and f partially competed with the radioactively labelled control GATA sites (lanes 11 and 13). **(b)** Radioactively labelled Gata2 sites ab and c were able to interact with GATA2 in the absence of the competitor (lanes 2 and 6). Both motifs were also able to compete with a radioactively labelled versions of themselves (lanes 3 and 7) leading to no visible bands, whereas the mutant version of site a was unable to compete, resulting in visible bands (lanes 4 and 8). Radioactively labelled GATA-d, e, and f motifs were unable to directly interact with GATA2 (lanes 10, 14, and 18), or compete with a radioactively labelled version of themselves (lanes 11, 15, and 19). These experiments were carried out by Dr Sarah De Val. All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

In line with the EMSA result, analysis of a publicly available ChIP-seq dataset for EC GATA2 binding[372] identified a GATA2 peak over the KLF2-66 enhancer (**Figure 4.36**). However, as neither of the two functional GATA sites were evolutionarily conserved from human to Tasmanian devil, and resources were limited, these mutations were not tested in transgenic mouse embryos.

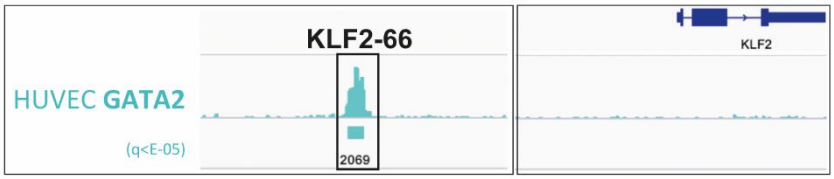


Figure 4.36: Analysis of a publicly available ChIP-seq dataset on GATA2 binding identifies a statistically significant peak at the KLF2-66 enhancer. Analysis of GATA2 binding in HUVECs[372] shows a statistically significant GATA2 peak overlapping with the KLF2-66 enhancer region.

4.4.5 MEF2 transcription factors

Analysis of the KLF2-66 sequence not only identified MEF2 motifs but identified a highly conserved double motif in the conserved core of the KLF2-66 enhancer similar to that found in KLF2-41. In a manner similar to KLF2-41, the KLF2-66 MEF2-a site was able to robustly compete with the control site, whereas the MEF2-b site was able to compete with the control probe only partially for MEF2C binding. In a direct EMSA, where MEF2-a and MEF2-b motifs were each labelled separately, only MEF2-a was able to directly interact with MEF2C, whereas MEF2-b site was unable to compete **(Figure 4.37**). Based on these results, it was concluded that only MEF2-a site was a functional MEF2 site, whereas the MEF2-b site was from here on renamed as the HBOX site, in line with the nomenclature change for the KLF2-41 enhancer. As with the KLF2-41, ChIP-seq analysis in mouse adult ECs confirmed enriched binding of MEF2A and C over the KLF2-66 enhancer.

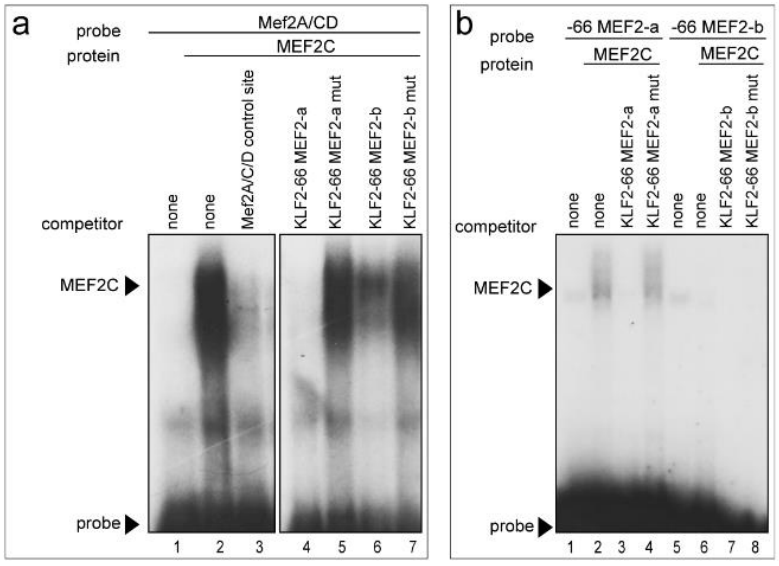


Figure 4.37: MEF2 factors bound the KLF2-66 enhancer in EMSA. (a) KLF2-66 enhancer putative MEF2-a binding site competed with the radioactively labelled control probe for MEF2 binding (lanes 4), resulting in a band shift (absence of the band). The mutant version of the same binding motifs (lane 5) did not compete for MEF2 binding, thereby allowing MEF2 interaction with the radioactively labelled control site, resulting in a visible band. KLF2-66 enhancer putative MEF2-b binding site partially competed with the radioactively labelled control MEF2 site. **(b)** Radioactively labelled MEF2-a site was able to interact with

MEF2C in the absence of the competitor (lane 2). MEF2-a motif was also able to compete with a radioactively labelled version of itself (lane 3) leading to absence of any bands, whereas the mutant version of site MEF2-a site was unable to compete, resulting in visible bands (lane 4). Radioactively labelled MEF2-b motif was unable to directly interact with MEF2C (lane 6) or compete with a radioactively labelled version of itself (lane 7). All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

In line with the EMSA results demonstrating MEF2C binding to the KLF2-66 MEF2-a motif, the analysis of MEF2A and MEF2C ChIP-seq datasets carried out in adult mouse heart and lung tissue (the data was shared with us by Prof Mukesh Jain) confirmed that both MEF2A and MEF2C were able bind the KLF2-66 enhancer in these settings (**Figure 4.38**). This confirms that the KLF2-66 enhancer is likely regulated by MEF2 factors.

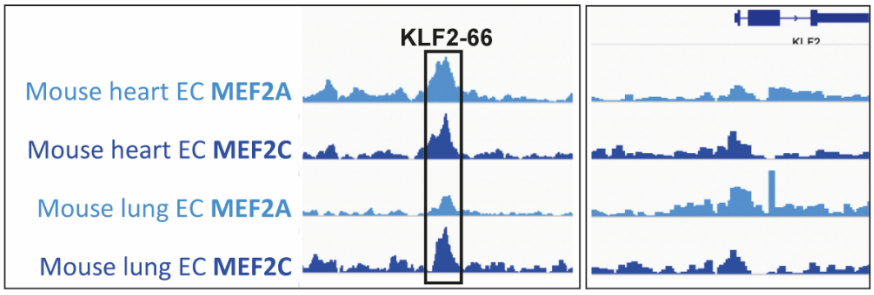


Figure 4.38: **Analysis of MEF2A and MEF2C ChIP-seq datasets carried out in adult heart and lung shows MEF2 binding at KLF2-66 enhancer region.** All four unpublished MEF2A and MEF2C ChIP-seq datasets were provided by our collaborator Prof. Mukesh Jain.

4.5 The MEF2-HBOX motif is an essential element conserved in both KLF2 enhancers

The similarity between the HBOX motifs found alongside MEF2 in both KLF2-41 and KLF2-66 enhancers was striking, and suggested a potentially important function, even though neither appeared to bind MEF2. To investigate this further, the HBOX motif was next mutated in both the KLF2-41 and KLF2-66 enhancers, and the resulting mutant constructs tested in F0 transient transgenic mouse embryos.

4.5.1 A requirement for the HBOX site in KLF2-41:LacZ transgene expression

Out of 11 embryos PCR-positive for *LacZ*, one embryo showed no *LacZ* expression, six embryos exhibited ectopic *LacZ* expression but no obvious EC expression (#13, #17, #41, #36, #11, #32) and four embryos (#39, #29, #30, #35) showed very weak vascular KLF2-41mutHBOX:*LacZ* transgene expression (**Table 4.5** and **Figure 4.39**). The proportion of KLF2-41mutHBOX:*LacZ* embryos exhibiting vascular transgene expression was considerably reduced compared to the WT KLF2-41 enhancer, which showed robust EC-specific expression in six out of seven *LacZ*-expressing embryos (**Table 4.5**). At the same time, the KLF2-41mutHBOX:*LacZ* transgene appeared to induce ectopic *LacZ* expression at a higher frequency than the WT transgene. A potential explanation for the high frequency of ectopic expression observed with the mutant transgene might be due to the endothelial-specifying function of the HBOX motif. According to this hypothesis, the high frequency of ectopic expression observed with the mutant transgene could be interpreted as an expansion of transgene expression from specifically endothelial to other cell types.

To further analyse the expression pattern of the KLF2-41mutHBOX:*LacZ* transgene, sections from all embryos with any vascular expression were analysed, with the analysis of two representative embryos, a strong and a weak expressor, shown below (**Figures 4.40** and **4.41**). Out of the four expressing embryos, one embryo showed weak expression in the same pattern as WT KLF2-41 (**Figure 4.40**), and one embryo exhibited punctate KLF2-41mutHBOX:*LacZ* transgene expression in a small number of arterial and angiogenic ECs, the ECCs and the remainder of the endocardium (**Figure 4.41c**). Minimal EC staining could be observed in the sections from the remaining two embryos.

In conclusion, although some vascular *LacZ* expression was observed in a subset of KLF2-41mutHBOX:*LacZ* embryos, the expression was considerably weaker and less penetrant than that observed with WT KLF2-41:*LacZ* transgene analysis, suggesting mutating the HBOX motif at least dramatically reduced, if not prevented, *LacZ* expression driven by the KLF2-41 enhancer.

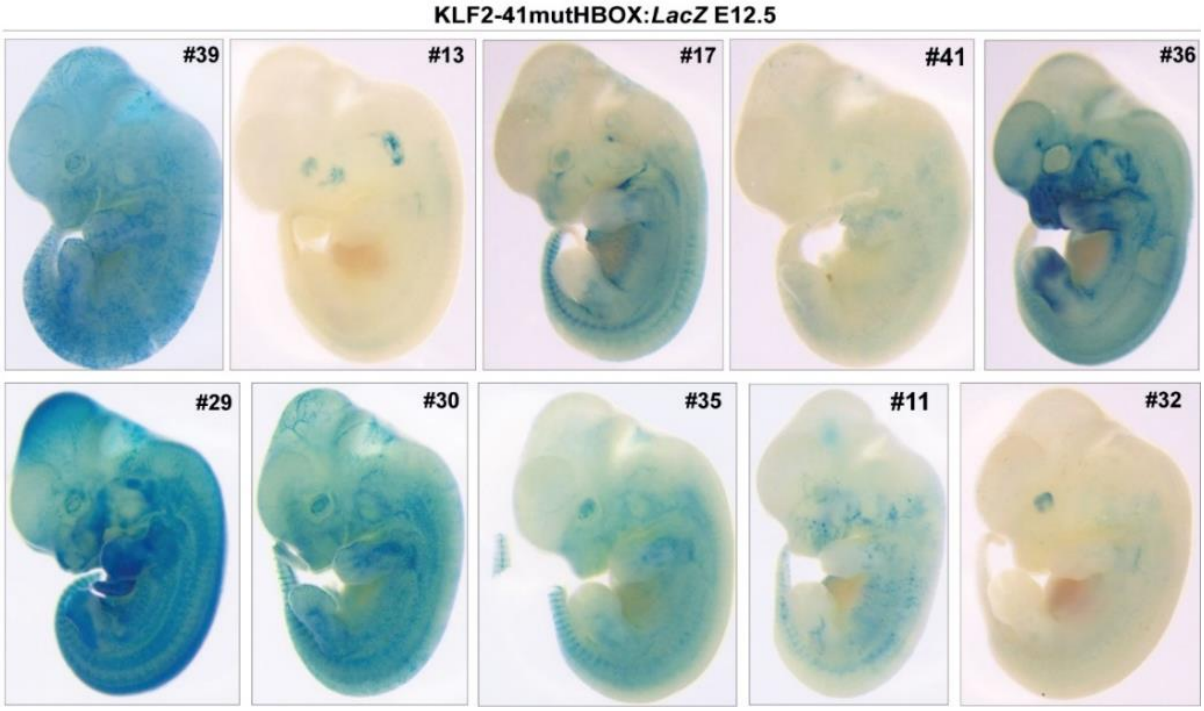


Figure 4.39: The F0 Transient transgenic KLF2-41mutHBOX:*LacZ* embryos harvested at E12.5 show limited vascular transgene expression. All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

	KLF2-41:*LacZ*	KLF2-41mutMEF2:*LacZ*	KLF2-41mutHBOX:*LacZ*
Vascular	6	0	4*
Ectopic	1	3	6
None	7	4	1
Total	13	7	11

Table 4.5: A summary of KLF2-41:*LacZ*, KLF2-41mutMEF2:*LacZ*, and KLF2-41mutHBOX:*LacZ* transgene expression detected in E12.5 mouse embryos. The expression pattern observed was classified as one of the three categories: vascular, ectopic, or no expression. *Denotes weak transgene expression.

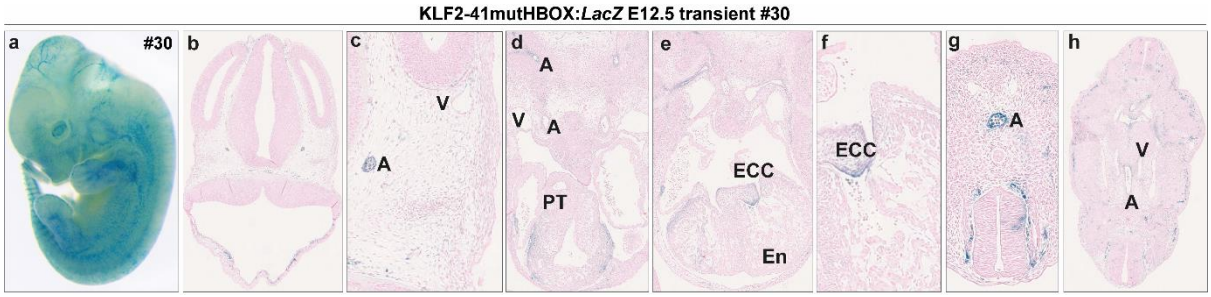


Figure 4.40: The KLF2-41mutHBOX:*LacZ* transgene is weakly expressed in a limited population of EC in E12.5 embryo sections. (a) The whole-mount KLF2-41mutHBOX:*LacZ* E12.5 F0 transient transgenic embryo #30. **(b)** Weak *LacZ* expression is present in arterial and capillary ECs, but not in venous ECs in the head. **(c)** Close up of *LacZ*-expressing ECs in the head section. **(d)** Weak KLF2-41mutHBOX:*LacZ* transgene expression is detected in the pulmonary trunk and the OT of the heart. **(e)** KLF2-41mutHBOX:*LacZ* transgene expression is detected in the

ECCs (here shown in AVC) and the remainder of the endocardium. **(f)** Close up of *LacZ*-expressing ECs within the ECCs. **(g)** KLF2-41mutHBOX:*LacZ* transgene expression is detected in the endothelium of the tail artery but not in the endothelium of the tail vein. **(h)** KLF2-41mutHBOX:*LacZ* transgene expression is detected in the endothelium of the abdominal aorta, but not in the endothelium of the abdominal vein. **A=artery, V=vein, C=capillary, OT=outflow tract, En=endocardium, ECC=endocardial cushion.**

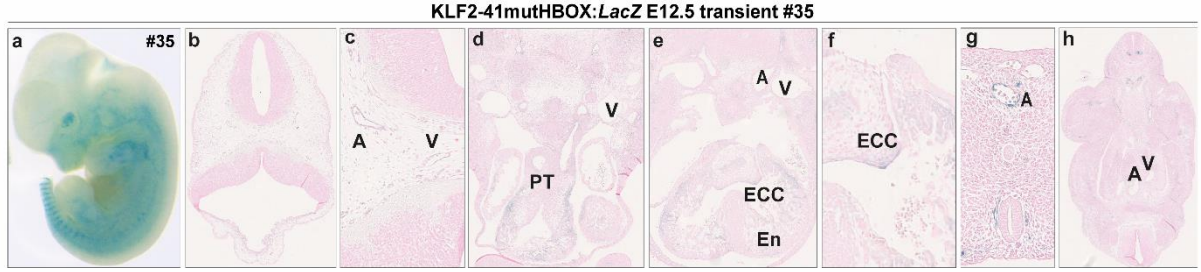


Figure 4.41: The KLF2-41mutHBOX:*LacZ* transgene is weakly expressed in a limited population of EC in E12.5 embryo sections. (a) The whole-mount KLF2-41mutHBOX:*LacZ* E12.5 F0 transient transgenic embryo #35. **(b)** Weak *LacZ* expression is present in arterial and capillary ECs, but not in venous ECs in the head. **(c)** Close up of *LacZ*-expressing ECs in the head section. **(d)** Weak KLF2-41mutHBOX:*LacZ* transgene expression is detected in the pulmonary trunk and the OT of the heart. **(e)** Weak KLF2-41mutHBOX:*LacZ* transgene expression is detected in the ECCs (here shown in the AVC) and the remainder of the endocardium of the heart. **(f)** Close up of *LacZ*-expressing ECs within the ECCs. **(g)** KLF2-41mutHBOX:*LacZ* transgene expression is detected in the endothelium of the tail artery but not in the endothelium of the tail vein. **(h)** Weak KLF2-41mutHBOX:*LacZ* transgene expression is detected in the endothelium of the abdominal aorta, but not in the endothelium of the abdominal vein. **A=artery, V=vein, C=capillary, OT=outflow tract, En=endocardium, ECC=endocardial cushion.**

4.5.2 A requirement for the HBOX site in KLF2-66:LacZ transgene expression

Out of four embryos PCR-positive for *LacZ*, one embryo showed no *LacZ* expression and the remaining three (#18, #35, #37) exhibited ectopic *LacZ* expression, but no vascular expression (**Table 4.6**). These result therefore suggests that the KLF2-66 enhancer HBOX site is required for the KLF2-66:*LacZ* transgene expression(**Figure 4.42**).

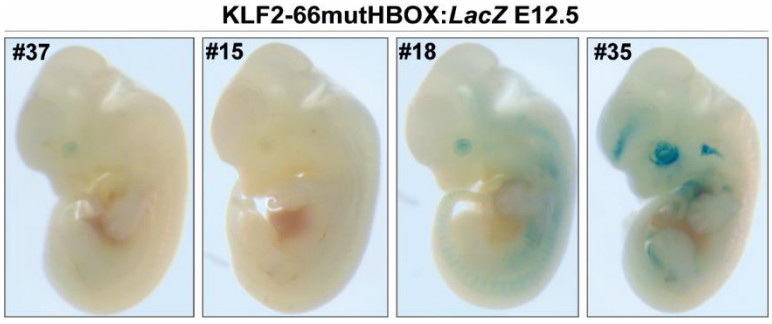


Figure 4.42: The KLF2-66 enhancer conserved HBOX binding site is required for KLF2-66-driven *LacZ* expression. All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

	KLF2-66:*LacZ*	KLF2-66mutHBOX:*LacZ*
Vascular	5	0
Ectopic	0	3
None	4	1
Total	9	4

Table 4.6: Summary of KLF2-66:*LacZ* and KLF2-66mutHBOX:*LacZ* transgene expression detected in E12.5 mouse embryos. The expression pattern observed was classified as one of the three categories: vascular, ectopic, or no expression.

4.6 The MEF2-HBOX is also found within the KLF2 promoter

Although the KLF2-0.4 core promoter did not drive significant gene expression in a mouse transgenic assay in the previous chapter, a body of literature clearly demonstrates some level of activity in cultured ECs, both under static conditions, in response to flow, or upstream activators of MEF2 signalling[245,246,317,339,373–375]. Re-analysis of this promoter region comparing the human and mouse sequences clearly demonstrated that the *KLF2* promoter sequence contains conserved ETS, MEF2, and HBOX TF binding motifs (**Figure 4.43**), although the sequence is only conserved to mouse and no further. Interestingly, the two MEF2 motifs here were configured in a similar manner to the MEF2-HBOX motifs found within the KLF2-41 and KLF2-66 enhancers.

```
human           ACGGCTCTGTGGGCCGCGGCCGGGCCCGCCCAAGCCTTATAGGCGCGGCGCGCGCGGGGG
mouse           -----TGTGTGTGCTGTGGCCCGGCCCACCCAGGCCTAGGCACAGAGG
                     * **** ** * **** ***** **** ***************  ** * * **
                                ETS-a/NFAT-a      ETS-b
human           CCGGGCCAAAGCGGCGGCGGCGGAAACGCGTCCCGGATGGGGACGAGCTCCGGGCGCAGC
mouse           GCCGGGCTAGGAGGCGTCGACGGAAACGCGTCCCGGATGGGGACGAGCTCCGGGCTCAGC
                 * ** * * * **** ** ***********************************  ****
                  MEF2-a          MEF2-b
human           CTAAATTTAGCCGCGGTATATAAGCCGGCGGGCGGCAGCGGCCGTGGCCACTGCGGCGGC
mouse           CTAAATTTAGCCGCGGTATATAAGCCTGGCGGTGGTGGCTGTGGCGGCTACTGCGGGAGC
                ************************** *  ** **  ** *  * *** *******  **

human           CTGGGCCTTGCGCTGACAGCCCGGCCCCTCGCGCGCCGCCAGGGCCGCGCCTCCCCCTCT
mouse           CCGGAC--TGCGCT------------CCTCAAGCCCCGCC--------------CCCGCG
                * ** *  ******            ****  ** *****              *** *

human           CCCGGGCCCCGCCCCCGCCCGTCTCCCCGGCGACGTTGTCAACACGCGCTCCACAACAGC
mouse           GCCGGGCCCCGCCCCCGACCGTCCCTATGGCGACGGCGTC-------------------
                 **************** ***** *   *******  ***
```

Figure 4.43: **KLF2-0.4 human core promoter sequence alignment with orthologous mouse Klf2-0.4 enhancer sequence with annotated putative TFBSs.** The alignment was carried out using ClustalW software. Putative ETS (blue), MEF2 (lilac), and NFAT (yellow) TFBS are annotated in alphabetical order from the 5' to 3' direction. All putative TFBS were identified through manual inspection of the sequence by looking for consensus binding motifs: GGA$^{A}_{T}$ for ETS factors, $^{C}_{T}$TA$^{T}_{A}$T$_{A}$T$_{A}$TA$^{A}_{G}$ for MEF2 factors, and GGAAA for NFAT factors. *Denotes conservation between all four species analysed using ClustalW. A 200bp of the human sequence that was cloned directly upstream of the TSS is represented in this alignment due to no conserved ETS, MEF2, or NFAT motifs found between -200 and -400bp upstream of the TSS. All putative TFBS are annotated in the human sequence, whereas putative TFBS in the orthologous sequence are only annotated if conserved to the corresponding human TFBS.

The ability of the two conserved putative MEF2 binding sites within this enhancer to interact with MEF factors was tested in EMSA. As with the double motif in KLF2-41 and KLF2-66, the MEF2-a site was able to robustly bind MEF2 proteins, whereas MEF2-b site was not and was renamed as a HBOX (**Figure 4.44**).

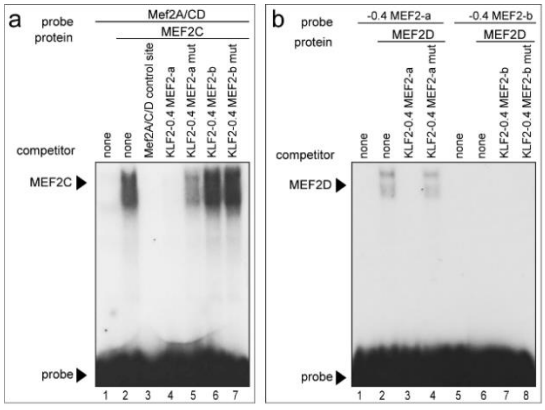


Figure 4.44: MEF2 factors bind the KLF2-0.4 core promoter in EMSA. (a) KLF2-0.4 core promoter putative MEF2 binding site a competed with the radioactively labelled control probe for MEF2 binding (lanes 4), resulting in a band shift (absence of the band). The mutant version of the same binding motifs (lane 5) did not compete for MEF2 binding, thereby allowing MEF2 interaction with the radioactively labelled control site, resulting in a visible band. KLF2-0.4 core promoter putative MEF2-b binding site did not compete with the radioactively labelled control MEF2 site. **(b)** Radioactively labelled MEF2-a site was able to interact with MEF2D in the absence of the competitor (lane 2). MEF2-a motif was also able to with a radioactively labelled version of itself (lane 3) leading to absence of the band, whereas the mutant version of MEF2-a site was unable to compete, resulting in visible bands (lane 4). Radioactively labelled MEF2-b motif was unable to directly interact with MEF2D (lane 6), or compete with a radioactively labelled version of itself (lane 7). All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

In line with the EMSA results demonstrating MEF2C binding to the KLF2-0.4 MEF2-a motif, the analysis of MEF2A and MEF2C ChIP-seq datasets carried out in adult mouse heart and lung tissue confirmed that both MEF2A and MEF2C are able bind the KLF2-0.4 enhancer in these settings (**Figure 4.45**), although the binding appeared less robust compared to KLF2-41 and KLF2-66 enhancers (the data was shared with us by Prof Mukesh Jain). This confirms that the KLF2-66 enhancer is likely regulated by MEF2 factors.

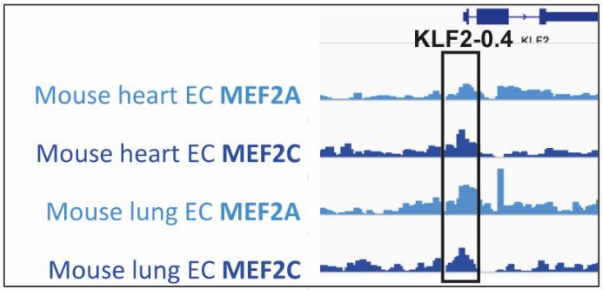


Figure 4.45: **Analysis of MEF2A and MEF2C ChIP-seq datasets carried out in adult heart and lung shows MEF2 binding at the KLF2 promoter region.** All four unpublished MEF2A and MEF2C ChIP-seq datasets were provided by our collaborator Prof. Mukesh Jain.

Next, all three MEF2-HBOX motifs were aligned using ClustalW **(Figure 4.46**). Although the MEF2 sequence varied slightly between the three elements, all loosely adhered to the consensus motif (shown in **Figure 4.23**). Further, the sequence of the HBOX motif was conserved between all three enhancers, as was the spacing between the two motifs.

```
                    MEF2                HBOX
KLF2-41         GGTTAAAAATAACTCCGGTATATAAAGA
KLF2-0.4        GCCTAAATTTAGCCGCGGTATATAAGCC
KLF2-66         AGATATTTTGGGAGCCACTATATAAACC
                  **          *   *******
```

Figure 4.46: Alignment of human KLF2-41, KLF2-66, and KLF2-0.4 MEF2-HBOX double motif sequences. Alignment was carried out using ClustalW software.

In conclusion, this analysis has so far has established that the KLF2-66 enhancer, the KLF2-41 enhancer, and the KLF2-0.4 promoter sequences all contain a MEF2-HBOX double motif. In addition to the clear requirement for both the MEF2 and HBOX motifs for the KLF2-41:*LacZ* transgene expression, as well as the requirement for the HBOX motif for KLF2-66:*LacZ* expression, the MEF2-HBOX double motif has been shown to be required for reporter expression driven by the KLF2 promoter in a cell culture system[245]. Taken together, these results demonstrate a crucial role for the MEF2-HBOX double motif in regulation of *KLF2* expression.

4.7 The MEF2-HBOX motif is not commonly found in the genome but is shared by an upstream KLF4 endothelial enhancer

In addition to *KLF2*, other MEF2-regulated endothelial-specific enhancers have been characterised in detail, although none of these enhancers were for flow-sensitive genes nor reported to be flow-sensitive enhancers[154]. To understand whether the HBOX motif is also present alongside the MEF2 motif in other MEF2-regulated endothelial enhancers, the sequences of those enhancers were analysed for potential double MEF2-HBOX motifs. However, no HBOX motif (corresponding to the $^{G}_{C}$TATATAA$^{A}_{G}$ consensus found in the *KLF2* enhancer and promoter sequences) was not found in any of these sequences.

To broaden this search to other MEF2-regulated regulatory regions in the genome, the MEF2A and MEF2C ChIP-seq datasets from mouse heart and lung ECs was interrogated for similar MEF2-HBOX motifs (analysis carried out by Dr Susann Bruche). An IUPAC version of the double motif that encompassed all three sequences was generated and then used to search all statistically significant MEF2A and MEF2C peaks in the ChIP-seq dataset for nearby HBOX motifs. However, only one statistically significant region was identified. Strikingly, this near-consensus novel MEF2-HBOX was identified far upstream of the mouse *Klf4* gene, corresponding to a region 148 kb upstream of the human *KLF4* promoter. *KLF4* is expressed in ECs in a pattern similar to *KLF2 and* is known to compensate for *KLF2* in the embryo and in the adult[241]. Further, MEF2 factors were found to be required for *Klf4* as well as *Klf2* expression in the adult mouse endothelium[223]. This putative KLF4-148 enhancer was also independently identified in a MEF2 ChIP-seq experiment investigating the role of MEF2 factors in statin regulation although its functional role as an enhancer was never investigated[376].

4.8 Identification and validation of the KLF4-148 enhancer

The presence of this highly similar MEF2-HBOX motif in the KLF4-148 enhancer indicated a potential role for KLF4-148 in regulation of *KLF4* in the endothelium. In order to investigate this, the region 148kb upstream of the *KLF4* TSS was located and assessed for enhancer marks including HUVEC-specific histone modifications, EC-specific DNaseI hypersensitivity, and evolutionary conservation. As expected, the region was marked by enriched H3K27Ac binding, a signature characteristic of active enhancers, and enriched H3K4Me1 binding, a signature characteristic of regulatory regions generally. Unexpectedly, there was also a H3K4Me3 peak characteristic of promoter regions, although this did not overlap with the entire 1kb enhancer region analysed. The putative KLF4-148 enhancer region also corresponded to high levels of DNaseI hypersensitivity (indicative of open chromatin) in endothelial cell lines, including HUVEC and HMVEC, but was also open in some non-endothelial cell lines, for example a human skeletal muscle cell line (SKMC). Finally, the KLF4-148 putative enhancer sequence was evolutionarily conserved between human, mouse, platypus, chicken, frog, and zebrafish (**Figure 4.47**). In comparison, the putative *KLF4* core promoter region was not well conserved between species, was less DNaseI-sensitive, although it was flanked by both H3K27Ac and H3K4Me3 histone marks in HUVEC, thereby carrying an epigenetic signature characteristic both of promoter and active enhancer sequences.

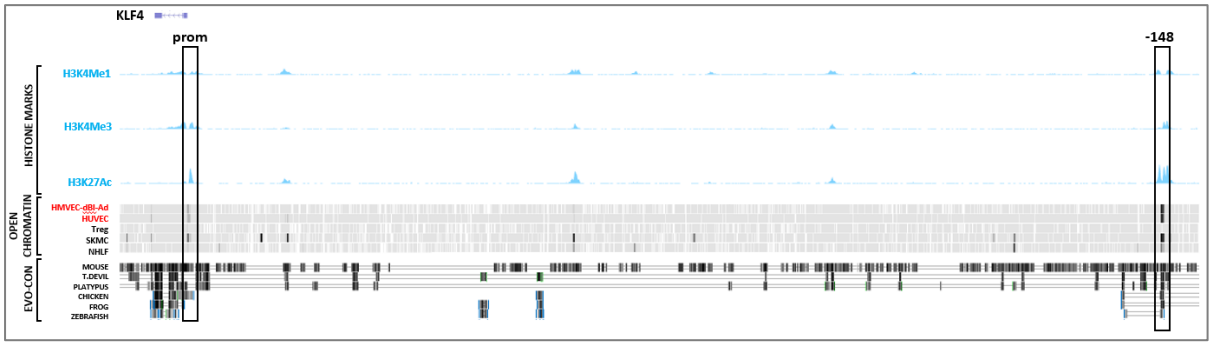


Figure 4.47: *KLF4* locus analysis and location of the KLF4-148 putative enhancer sequence. The KLF4-148 region contains robust EC-specific histone marks characteristic of enhancer regions (H3K4Me1 and H3K27Ac). In addition, the region resides in an area of open chromatin as measured by DNaseI hypersensitivity both in ECs, but also some other cell types. Finally, the KLF4-148 putative enhancer region was evolutionarily conserved from human all the way to zebrafish. Interestingly, the KLF4 putative core promoter region was not conserved to

mouse or other species and contained a histone mark signature characteristic both of enhancer and promoter regions.

4.8.1 The KLF4-148 element is an endothelial-specific enhancer

To characterise the activity of the putative KLF4-148 enhancer, this region was cloned upstream of the hsp68 promoter and *LacZ* reporter as in all previous transgenic experiments. Given the role of *Klf4* in genetic compensation for *Klf2*, and previous reports of *Klf4* activation by shear stress[377–379], these embryos were analysed at E12.5, as with the *KLF2* enhancer analysis. We received seven PCR-positive embryos (**Table 4.7**), with all seven exhibiting vascular *LacZ* expression (**Figure 4.48**). In some embryos ectopic *LacZ* expression was observed in addition to vascular expression, for example in the limbs (embryos #35 and #48) but this was not consistent between the embryos, suggesting this might be related to ectopic expression driven by the hsp68 promoter, or the transgene integration site in the case of each specific embryo.

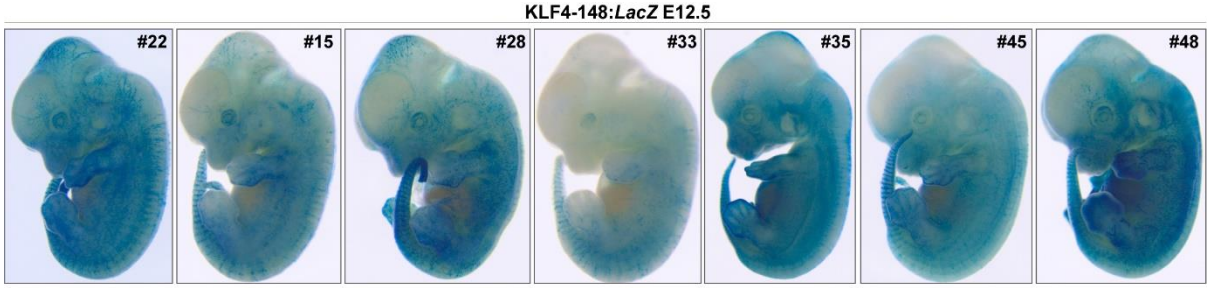


Figure 4.48: **The KLF4-148 enhancer sequence drives robust vascular-specific *LacZ* expression in E12.5 embryos.** All *LacZ* positive transgene-expressing embryos are shown. *LacZ* positive embryos that did not express the transgene are not shown.

	KLF2-41:*LacZ*	KLF2-66:*LacZ*	KLF4-148:*LacZ*
Vascular	6	5	7
Ectopic	1	0	0
None	7	4	0
Total	13	9	7

Table 4.7: Summary of KLF2-41:*LacZ*, KLF2-66:*LacZ*, and KLF4-148:*LacZ* transgene expression detected in E12.5 mouse embryos. The expression pattern observed was classified as one of the three categories: vascular, ectopic, or no expression.

In order to analyse EC subtype-specific KLF4-148:*LacZ* transgene expression, four out of seven embryos were sectioned, two representative embryos are shown here (**Figure 4.49** and **4.50).** While these embryos showed EC-specific expression similar to KLF2 enhancers, expression in the arterial ECs was generally weaker, and expression was also observed in venous ECs. Robust transgene expression was observed within the capillary and sprouting ECs (**Figure 4.49h**; **Figure 4.50h**). This analysis therefore confirmed that the MEF2-HBOX-containining KLF4-148 element was a bona-fide endothelial enhancer driving an expression pattern more similar to KLF2-66 than KLF2-41.

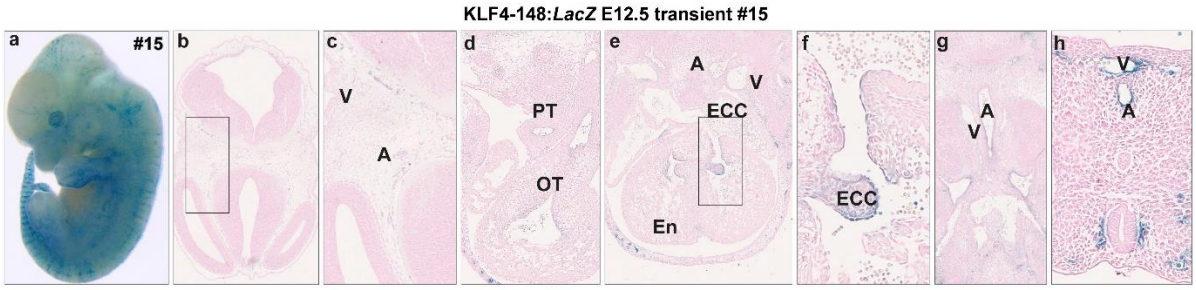


Figure 4.49: The KLF4-148:*LacZ* transgene drives robust EC and ECC-specific *LacZ* expression. (a) The whole-mount KLF4-148:*LacZ* E12.5 F0 transient transgenic embryo #15. **(b)** *LacZ* expression was detected in arterial, venous, and capillary ECs in the head. **(c)** Close up of *LacZ*-expressing ECs in the head section. **(d)** Weak KLF4-148:*LacZ* transgene expression was detected in the OT and the pulmonary trunk of the heart. **(e)** Robust KLF4-148:*LacZ* transgene expression was detected in the ECCs (here shown in AVC tissue). Some expression was present in the remainder of the endocardium. **(f)** Close up of *LacZ*-expressing ECs within the ECCs. **(g)** KLF4-148:*LacZ* transgene expression was detected in the endothelium of the abdominal aorta, but no expression was detected in the endothelium of the abdominal vein. **(h)** Robust KLF4-148:*LacZ* transgene expression was detected in the endothelium of both the tail artery and tail vein ECs. **A=artery, V=vein, C=capillary, OT=outflow tract, PT=pulmonary trunk En=endocardium, ECC=endocardial cushion.**

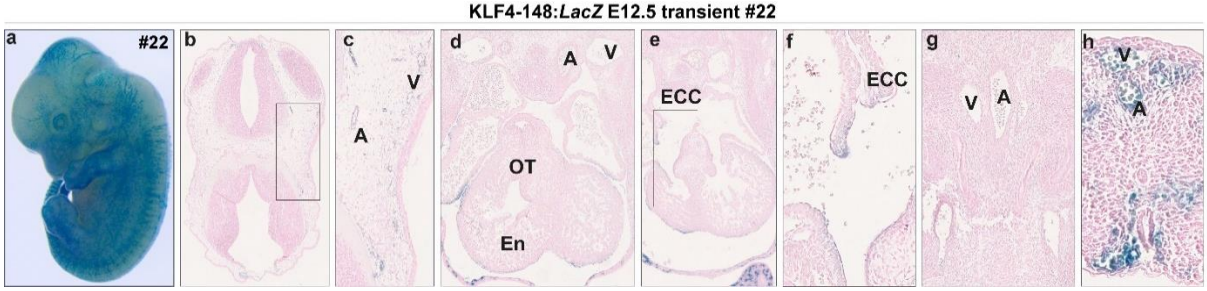


Figure 4.50: The KLF4-148:*LacZ* transgene drives robust EC and ECC-specific *LacZ* expression. (a) The whole-mount KLF4-148:*LacZ* E12.5 F0 transient transgenic embryo #22. **(b)** *LacZ* expression was detected in arterial, venous, and capillary ECs in the head. **(c)** Close up of *LacZ*$^{+}$ ECs in the head section. **(d)** No KLF4-148:*LacZ* transgene expression was detected in the OT of the heart. **(e)** Robust KLF4-148:*LacZ* transgene expression was detected in the ECCs (here shown in AVC tissue). Some expression was present in the remainder of the endocardium and the epicardium of the heart. **(f)** Close up of *LacZ*$^{+}$ ECs within the ECC. **(g)** Weak KLF4-148:*LacZ* transgene expression was detected in the endothelium of the abdominal aorta, but no expression was detected

in the endothelium of the abdominal vein. **(h)** Robust KLF4-148:*LacZ* transgene expression was detected in the endothelium of both the tail artery and tail vein ECs. **A=artery, V=vein, C=capillary, OT=outflow tract, PT=pulmonary trunk En=endocardium, ECC=endocardial cushion.**

4.8.2 Putative transcriptional regulators of the KLF4-148 enhancer

The KLF4-148 enhancer sequence was analysed for putative TF binding motifs using a conservation-based approach as described for the two *KLF2* enhancers and the promoter sequence. The human KLF4-148 enhancer sequence was compared to mouse, platypus, chicken, and zebrafish orthologous sequences, which identified multiple conserved ETS binding sites, three conserved putative NFAT binding sites, four conserved RBPJ binding sites, one conserved GATA binding site, and the MEF2-HBOX double motif (**Figure 4.51**). In the previous analysis of the *KLF2* enhancer sequences, the initial focus was on validating the conservation hypothesis by testing the least conserved orthologous sequence in F0 transient transgenic mouse embryos, then comparing the orthologous transgene expression to the human transgene expression to establish whether it is recapitulated. However, in the case of the KLF4-148 enhancer, the focus was on the similarities with the *KLF2* regulatory regions, and specifically on the conserved TF binding motifs similar to those found in the KLF2 enhancer sequences.

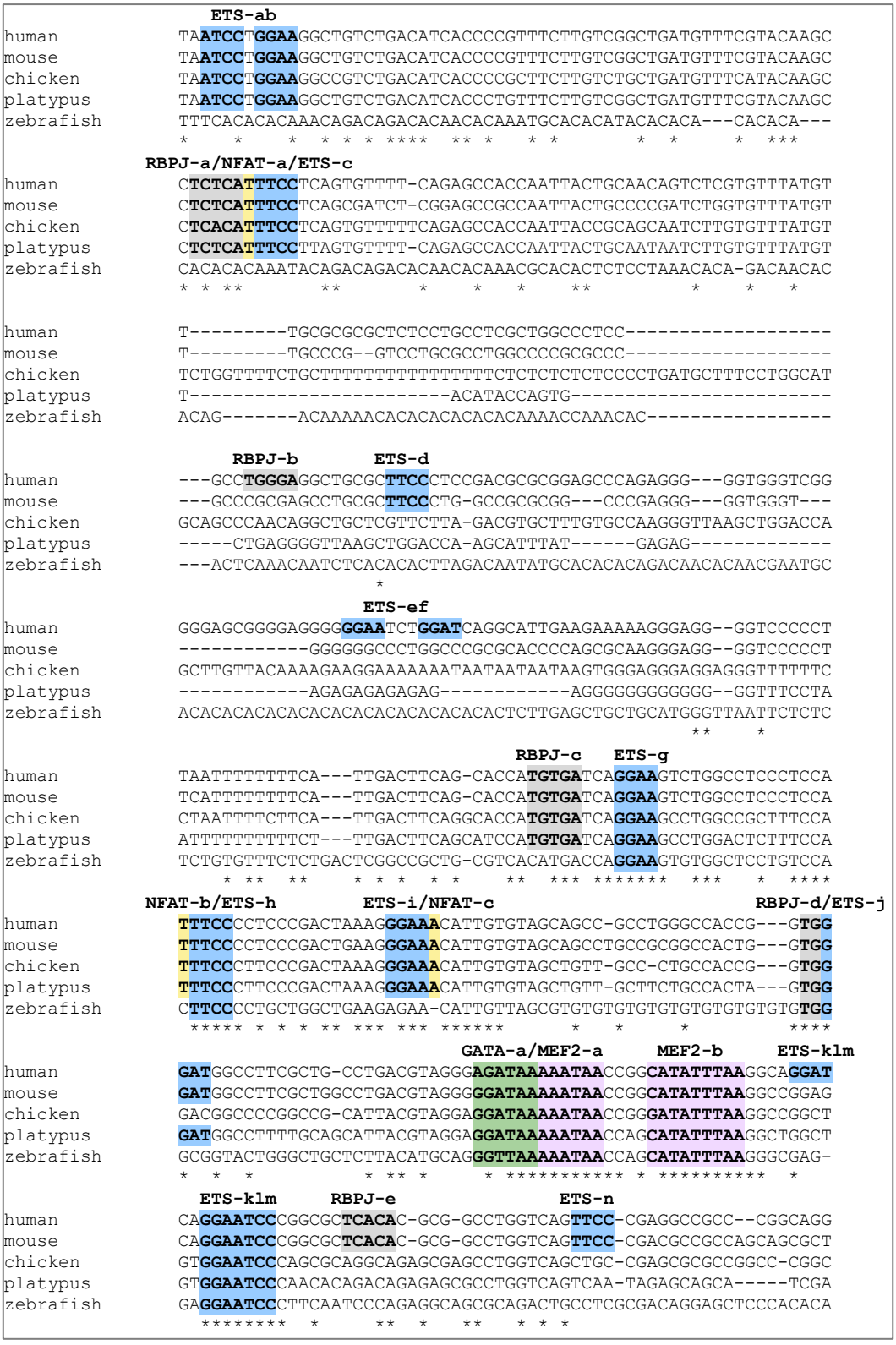


Figure 4.51: **KLF4-148 human enhancer sequence alignment with orthologous mouse, chicken, platypus, and zebrafish sequences with annotated putative TFBSs.** The alignment was carried out using ClustalW software. Putative ETS (blue), MEF2 (lilac), RBPJ (grey), GATA (green), and NFAT (yellow) TFBS are annotated in alphabetical order from the 5' to 3' direction. All putative TFBS were identified through manual inspection of the sequence by looking for consensus binding motifs: GGA$^{A}_{T}$ for ETS factors, $^{C}_{T}TA^{T}_{A}T^{A}_{T}A^{T}_{A}TA^{A}_{G}$ for MEF2 factors, $^{C}_{A}TC^{C}_{A}T^{T}CA^{T}_{C}$

for RBPJ factors, GGAAA for NFAT factors, and $^{A}_{T}GATA^{A}_{G}$ for GATA factors. *Denotes conservation between all five species analysed using ClustalW. All putative TFBS are annotated in the human sequence, whereas putative TFBS in the orthologous sequence are only annotated if conserved to the corresponding human TFBS.

4.8.3 ETS factors

In order to assess the ability of the KLF4-148 enhancer sequence to interact with ETS factors *in vitro*, the ability of each putative ETS binding motif found in the human sequence to compete for ETS1 DBD binding with a radioactively labelled control probe containing a validated ETS binding site was tested as before. Only two sites, ETS-c and ETS-g, were able to compete for ETS DBD binding with the control probe (**Figure 4.52**). Both functional ETS sites were conserved from human to platypus, with ETS-g site conserved all the way to zebrafish. Interestingly, ETS-c site, although highly conserved, was located far upstream of the conserved enhancer core.

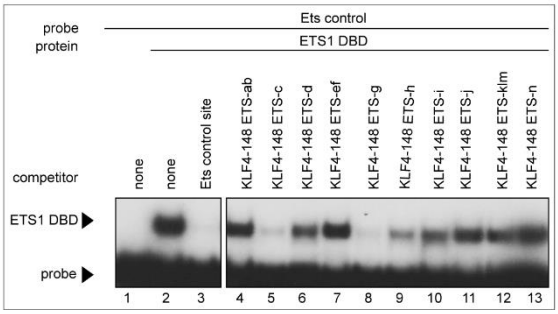


Figure 4.52: ETS factors bind the KLF4-148 enhancer in EMSA. (a) KLF4-148 enhancer ETS binding sites c and g competed with the radioactively labelled control probe for ETS1 DBD binding (lanes 5 and 8), resulting in the absence of the band. All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

This was confirmed with analysis of the publicly available ChIP-seq dataset for ERG binding in HUVECs, which found a statistically significant ERG peak over the KLF4-148 enhancer region (**Figure 4.53**), suggesting that ETS factors are indeed able to interact with the KLF4-148 enhancer.

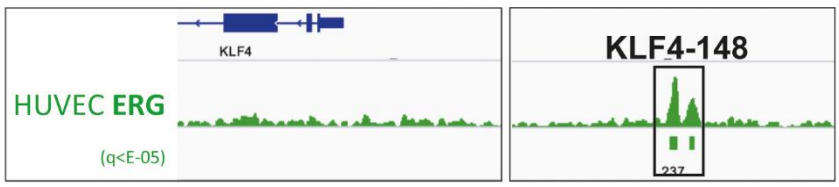


Figure 4.53: Analysis of a publicly available ChIP-seq dataset on ERG binding identified a statistically significant peak at the KLF4-148 enhancer. Analysis of ERG binding in HUVECs[350] showed a statistically significant ERG peak overlapping with the KLF4-148 enhancer region.

4.8.4 GATA factors

Upon inspection of the KLF4-148 enhancer sequence, one conserved putative GATA binding motif was identified. GATA motifs were found in the KLF2-66 but not KLF-41 enhancer. Given the similar pattern of expression between KLF2-66 and KLF4-148, this GATA motif was tested for functionality in EMSA. However, despite its robust evolutionary conservation and a perfect match to the GATA consensus binding motif, this site was unable to compete for GATA2 binding (**Figure 4.54**).

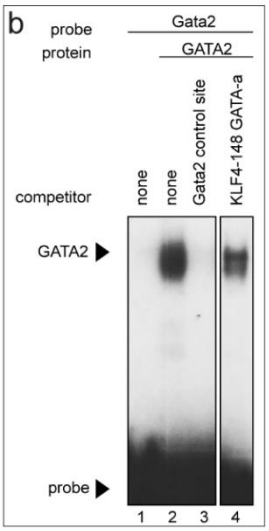


Figure 4.54: GATA2 did not bind the KLF4-148 enhancer in EMSA. The putative GATA binding site within the KLF4-148 enhancer sequence was unable to compete with the radioactively labelled control probe for GATA2 binding. This experiment was carried out by Dr Sarah De Val. All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

Based on this result, it was concluded that the KLF4-148 enhancer sequence is unlikely to interact with GATA factors. Accordingly, the KLF4-148 enhancer sequence did not overlap with a statistically significant peak for GATA2 binding in ChIP-seq (**Figure 4.55**)[372].

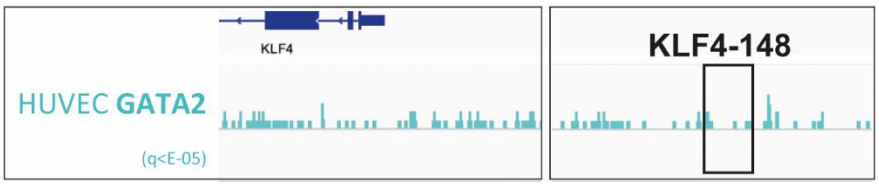


Figure 4.55: Analysis of a publicly available ChIP-seq dataset on GATA2 binding identified a statistically significant peak at the KLF4-148 enhancer. Analysis of GATA2 binding in HUVEC[372] showed no statistically significant GATA2 peak overlapping with the KLF4-148 enhancer region.

4.8.5 NFAT factors

The analysis of the KLF4-148 enhancer sequence identified three putative NFAT binding sites. Only one of the three putative NFAT binding sites, NFAT-c, was able to compete for NFATc1 binding with the radioactively labelled NFAT control probe (**Figure 4.56**). The same site was able to directly bind NFATC1 in direct EMSA and compete with a radioactively labelled version of itself. Interestingly, NFAT-c was robustly evolutionarily conserved, but did not interact with ETS DBD in EMSA, although it also contains a consensus ETS binding motif. This might suggest that NFAT-c is a true NFAT site.

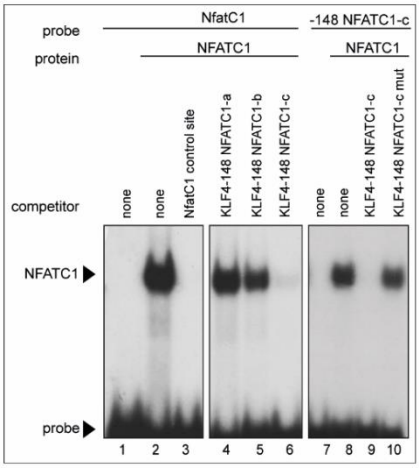


Figure 4.56: NFATC1 binds the KLF4-148 enhancer in EMSA. KLF4-148 enhancer putative NFAT binding site c competed with the radioactively labelled control probe for NFATC1 binding (lane 6), resulting in a band shift (absence of the band). Putative NFAT-a and b sites were unable to compete with the radioactively labelled control. Radioactively labelled KLF4-148 enhancer NFAT-c binding site was able to directly interact with NFATC1 protein (lane 8) and compete with a radioactively labelled version of itself (lane 9), resulting in a band shift. The mutant version of the same site did not compete, resulting in the presence of a band (lane 10). All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

The analysis of a publicly available NFATc1 ChIP-seq dataset carried out in HUVECs both under normal conditions and following stimulation with VEGFA[362] identified a statistically significant NFATc1 peak corresponding to the KLF4-148 enhancer region (**Figure 4.57**), suggesting that NFATc1 might regulate the KLF4-148 enhancer.

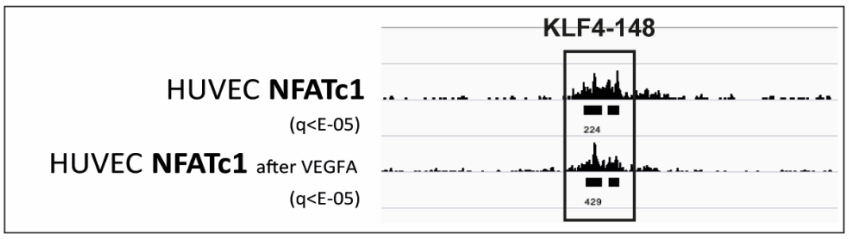


Figure 4.57: **Analysis of a publicly available ChIP-seq dataset on NFATc1 binding identifies a statistically significant peak at the KLF2-41 enhancer.** Analysis of NFATc1 binding in HUVECs under normal conditions or following activation with VEGFA[362] showed a statistically significant NFATc1 peak overlapping with the KLF2-41 enhancer region.

4.8.6 MEF2 factors

The presence of the MEF2-HBOX motif was used to identify the KLF4-148 enhancer, and this had previously been shown to be bound my MEF2 in human ChIP-seq[376]. The next question was whether the KLF4-148 MEF2-HBOX behaved in a similar manner as the KLF2 enhancer and promoter double motifs. As with previous analyses, the ability of the MEF2 and HBOX motifs within the KLF4-148 enhancer sequence to bind MEF2 factors was first assessed by EMSA (**Figure 4.58**). Although both MEF2-a and MEF2-b motifs were able to compete for binding of a control probe (**Figure 4.58a**), in direct EMSA, only the MEF2-a site was able to directly bind MEF2C and compete with a radioactively labelled version of itself for MEF2C binding (**Figure 4.58b**).

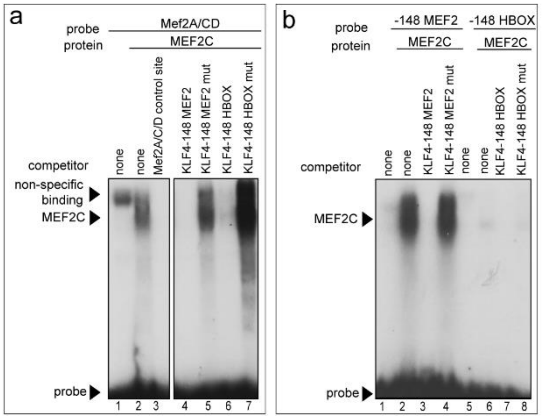


Figure 4.58: MEF2 factors bound the KLF4-148 enhancer in EMSA. (a) KLF4-148 enhancer putative MEF2 binding site a competed with the radioactively labelled control probe for MEF2 binding (lanes 4), resulting in a band shift (absence of the band). The mutant version of the same binding motifs (lane 5) did not compete for MEF2 binding, thereby allowing MEF2 interaction with the radioactively labelled control site, resulting in a visible band. KLF4-148 enhancer putative HBOX site also competed with the radioactively labelled control MEF2 site (lane 6), whereas its mutant version was unable to compete (lane 7). **(b)** The radioactively labelled MEF2 site was able to interact with MEF2C in the absence of the competitor (lane 2). The MEF2 motif was also able to compete with a radioactively labelled version of itself (lane 3) leading to a band shift, whereas the mutant version of the MEF2 site was unable to compete, resulting in visible bands (lane 4). The radioactively labelled HBOX motif was unable to directly interact with MEF2C (lane 6) or compete with a radioactively labelled version of itself (lane 7). All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

In line with the EMSA results that demonstrated MEF2 factors can interact with the KLF4-148 MEF2 motif, and previous published observations, analysis of the MEF2A and MEF2C ChIP-seq datasets from adult mouse heart and lungs ECs confirmed that these factors robustly bind the KLF4-148 enhancer

(**Figure 4.59**). Taken together, these results suggest that MEF2 factors are able to bind the KLF4-148 enhancer.

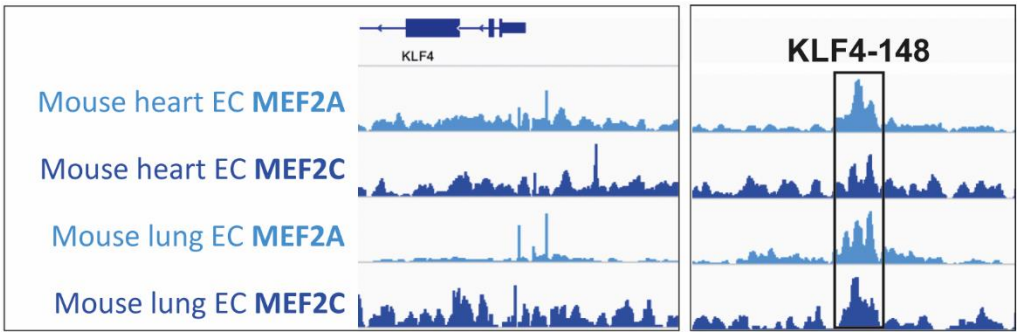


Figure 4.59: Analysis of MEF2A and MEF2C ChIP-seq datasets carried out in adult heart and lung shows MEF2 binding at KLF4-148 enhancer region. All four unpublished MEF2A and MEF2C ChIP-seq datasets were provided by our collaborator Prof. Mukesh Jain.

4.9 What binds the HBOX motif?

The initial approach to understanding the transcriptional regulation of the MEF2-HBOX binding motif was to consider the dynamics of MEF2 factors already established in the literature. Firstly, MEF2 factors have been repeatedly shown to bind as homo or heterodimers, either interacting together to enforce the robustness of the transcriptional response, or bind simultaneously with other TFs (co-factors), often with cell type-specific expression patterns, in order to regulate differentiation of a variety of different cell types[212,380–383]. Based on this knowledge, two hypotheses were formed with respect to what binds the double MEF2-HBOX motif. The first hypothesis was that the MEF2-HBOX site would be occupied by two MEF2 factors, but the second could only bind the HBOX when the first MEF2 site was occupied, in direct cooperativity between the two factors. The second hypothesis was that the HBOX motif would be occupied by an entirely different TF, which would nonetheless interact with MEF2 to integrate the upstream signalling pathways into a transcriptional response. To address the second hypothesis, putative TFs with consensus binding motifs bearing some similarity to that of the HBOX consensus, and with potential roles in shear-stress induced pathways or EC development,

were investigated. This line of inquiry led to the identification of FOX factors as potential candidates to bind with MEF2 to the MEF2-HBOX double motif.

Both these hypotheses were tested using the KLF2-41 enhancer MEF2-HBOX motif, with the expectation that any discoveries would likely be applicable to the other KLF2 regulatory regions, given the similarity in motifs between KLF2/4 regulatory regions (**Figure 4.60**), and similar behaviour of the MF2-HBOX binding motif in EMSA and mutational analyses carried out so far.



Figure 4.60: Alignment of human KLF2-41, KLF2-66, KLF2-0.4, and KLF4-148 MEF2-HBOX double motif sequences. Alignment was carried out using ClustalW software.

4.9.1 Is the MEF2-HBOX site bound by a MEF2 dimer?

The first hypothesis, namely that the MEF2-HBOX motif would be bound by a MEF2 dimer, was addressed using EMSA analysis. Instead of testing oligos containing either the MEF2 or the HBOX site alone, as was the case in previous experiments, longer oligos containing both the MEF2 and the HBOX binding motif within a single oligo were used. Four versions of the double motif were tested, the WT version where both sites were left intact, a mutated version with mutant MEF2 but WT HBOX, a mutated version with mutant HBOX but WT MEF2, and a third mutant where both sites were mutated.

In the first experiment, the WT and mutant double motifs were tested for their ability to compete for MEF2C binding with a control MEF2 probe in EMSA (**Figure 4.61a**). The WT probe and the probe with mutant HBOX site were both able to compete with the control probe for MEF2C binding, whereas the probe where the MEF2 site only was mutated, and the probe where both sited were mutated, could not compete. The results of this experiment imply that the MEF2, but not the HBOX site, is required

for the oligo to compete for MEF2C binding, as the oligo with mutant HBOX site was still able to compete.

In the second experiment, the same set-up was used to test whether a MEF2C/D heterodimer, rather than a MEF2 homodimer, binds the MEF2-HBOX double motif. In this experiment, the WT KLF2-41 MEF2-HBOX double motif was radioactively labelled, then tested to determine which oligos can interact with MEF2C alone, MEF2D alone, or a combination of MEF2C and D together. The WT version of the motif and two mutant versions, where either the MEF2 site or HBOX site alone were mutated, were then tested for their ability to compete with WT double motif for MEF2C, MEF2D, or MEF2C/D binding. The WT version of the double motif was able to directly bind MEF2C (data not shown), MEF2D and a mixture of MEF2C and D, and was also able to compete with a radioactively labelled version of itself for MEF2C, MEF2D, and MEF2C/D binding (**Figure 4.61b**). In a similar manner, the KLF2-41 double motif in which the HBOX site alone was mutated, was also able to compete for MEF2C, MEF2D, and MEF2C/D binding with the radioactively labelled WT double motif. In contrast, the double motif in which the MEF2 site alone was mutated was unable to compete with the WT probe for MEF2C, MEF2D, and MEF2C/D binding. While the results of both these experiments confirmed that the presence of MEF2 site, irrespective of the HBOX site, was required for the MEF2-HBOX oligo to compete for MEF2C binding, both these experiments included competition, which when successful results in a complete absence of a band-shift. This was a problem, as it is difficult to interpret the difference between two bands that are not there, for example between the KLF2-41 MEF2-HBOX WT and KLF2-41 MEF2-mutHBOX in competition with either the MEF2 control or the KLF2-41 MEF2-HBOX WT radioactively labelled probe. Therefore, to test the ability of each version of the double motif to interact with MEF2 factors in a direct way, a third experiment was designed.

In the third experiment, each version of the double motif probe was separately radioactively labelled, then tested for its ability to interact with MEF2C in the absence of any competitor, with care taken to

use similar amounts of each probe and similar levels of the radioactive label. Interaction between the radioactively labelled probe and the MEF2C protein in this experiment would be measured as the presence, not the absence of the band-shift, and therefore allow for the difference in band intensity to be the readout of interaction between the double-motif and MEF2C. In this experiment, the WT version of the double motif and the version in which HBOX site alone was mutated were both able to interact with MEF2C, as evidenced by the presence of a strong band (**Figure 4.61c**). In contrast, the double motif in which the MEF2 site alone, or both the MEF2 and the HBOX sites were mutated, was not able to interact with MEF2C at all, as evidenced by the absence of the bands. Further, there was no difference in band intensity between the MEF2-HBOX WT double motif and the double motif where the HBOX motif was mutated, suggesting that the presence of a WT HBOX motif does not increase the binding capacity of MEF2C to the MEF2-HBOX double motif.

Based on the KLF2-41 enhancer double motif data shown in this section, it seems unlikely that a MEF2 homo- or heterodimer would interact with the MEF2-HBOX double motif. It should however be noted that the experiments described in this section are limited in their capacity to answer the question of whether the MEF2-HBOX double motif is bound by a MEF2 dimer, particularly if the interaction of MEF2 with the HBOX motif requires initial occupancy of the MEF2 motif. Further, if the double motif is bound by a heterodimer, pre-incubation with MEF2A, C, or D, followed by incubation with different combinations of the same factors might be necessary to address this question. In order to refine the experimental approach, an increasing amount of protein could be used to visualize the transition of the gel shift from single to a double band.

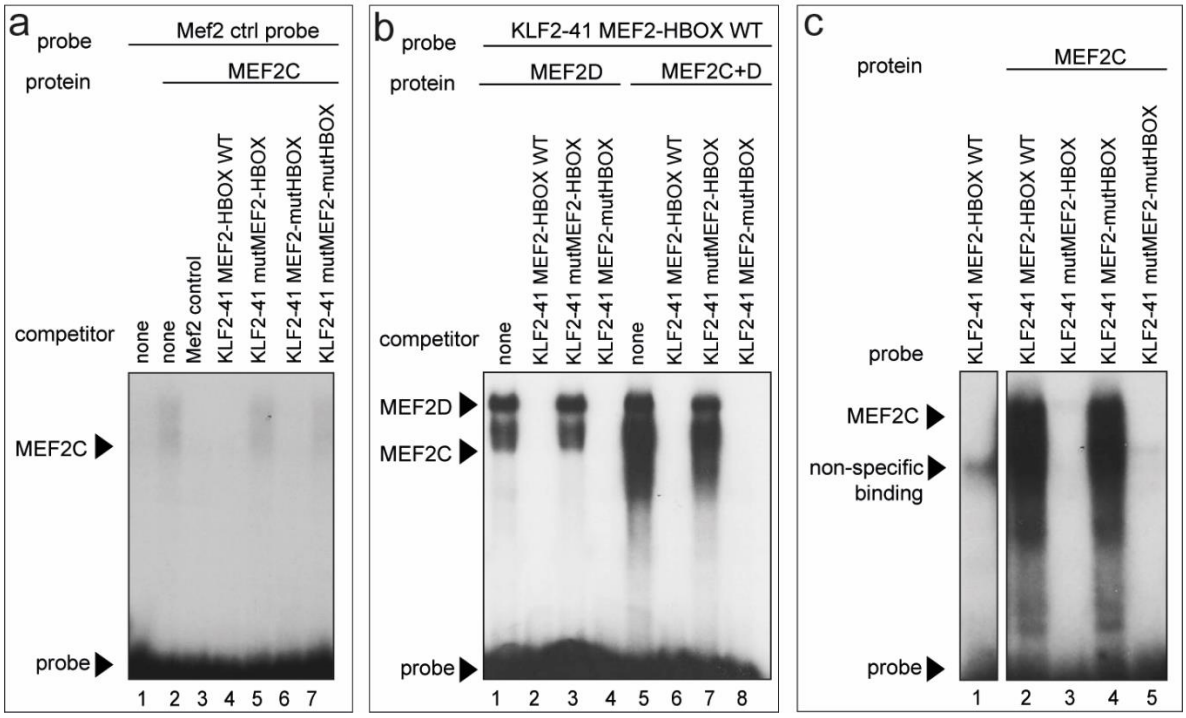


Figure 4.61: Analysis of the KLF2-41 MEF2-HBOX double motif and its interaction with MEF2 factors in EMSA. **(a)** The KLF2-41 WT MEF2-HBOX double motif was able to compete with the radioactively labelled Mef2 control probe for MEF2C binding (Lane 4). The MEF2-HBOX where the HBOX motif only has been mutated was also able to compete (Lane 6), whereas the mutant version with mutated MEF2 motif only, or where MEF2 and HBOX motifs had been mutated at the same time, were unable to compete with the control Mef2 probe for MEF2C binding (Lanes 5 and 7, respectively). **(b)** The KLF2-41 WT MEF2-HBOX double motif was able to compete with the radioactively labelled version of itself for MEF2D binding, and a combination of MEF2C and MEF2D proteins (Lanes 2 and 6). The mutated version of the same site where the HBOX site alone had been mutated was also able to compete (Lanes 4 and 8), whereas the site with mutated MEF2 alone did not compete for either MEF2D alone or a mixture of MEF2C and MEF2D proteins (Lanes 3 and 7). **(c)** THE KLF2-41 WT MEF2-HBOX double motif directly bound MEF2C when no competitor was present (Lane 2). The double motif version in which the HBOX site alone was mutated could also directly interact with MEF2C (Lane 4). Mutant versions of the double motif where either the MEF2 motif only was mutated, or both the MEF2 and HBOX motifs were mutated together, could not directly bind MEF2C (Lanes 3 and 5). All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

4.9.2 FOX factors interact with the MEF2, but not the HBOX motif

The next aim was to focus on other potential TFs that bind consensus motifs with similarities to the HBOX motif. Extensive experience in analysing of putative TF binding motifs in our lab, in combination with TRANSFAC and MEME suite Tomtom tool, led to realisation that the HBOX motif was fairly similar to the FOXC factor consensus binding motif: $^{A}_{G}{}^{C}_{T}{}^{A}_{C}AA^{C}_{T}A$ (**Figure 4.62**). FOXC transcription factors have been shown to collaborate with ETS factors to regulate many EC-specific enhancers during

development[143], and have been further implicated in regulation of shear stress signalling in lymphatic valves[384,385].

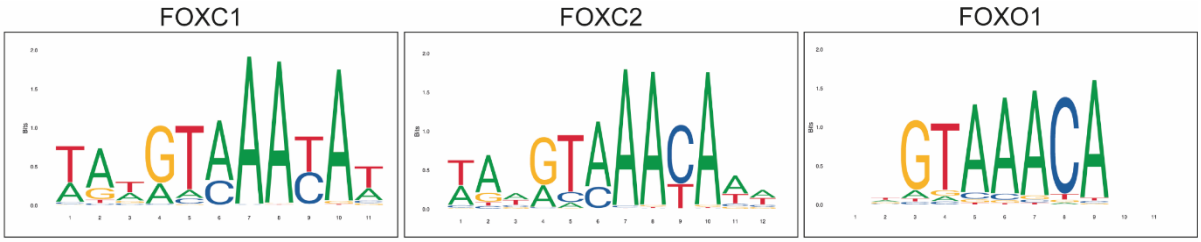


Figure 4.62: The consensus binding motifs for FOXC1, FOXC2, and FOXO1. The consensus motifs were retrieved using the JASPAR database of TF binding profiles stored as position frequency matrices (PFMs). These consensus binding motifs were collected to the database externally from scientific publications[347].

The hypothesis that FOX factors might interact with the HBOX motif was tested in EMSA. First, the KLF2-41 MEF2 and HBOX motifs were separately tested for their ability to compete with a radioactively labelled FOXC1/2 control motif probe, which robustly bound both FOXC1 and FOXC2 proteins. Unexpectedly, the WT version of the KLF2-41 MEF2 site was able to compete both for FOXC1 and FOXC2 binding, whereas the HBOX site could not (**Figure 4.63a**).

To further explore this result, four versions of the KLF2-41 MEF2-HBOX double motif were tested for their ability to compete with a radioactively labelled FOXC1/2 control probe for FOXC2 binding. Similar to the MEF2 dimer analysis described in the previous section, the WT version of the double motif and the version in which the HBOX site was mutated were able to compete for FOXC2 binding, whereas the version with a mutated MEF2 site, and a double mutant version, could not (**Figure 4.63b**). Finally, the ability for each version of the double motif to interact with FOXC2 directly was tested. As with the MEF2 dimer experiments, the two motifs where the MEF2 site was intact (WT double motif and mutant HBOX double motif), were able to interact with FOXC2 directly. On the other hand, the versions where MEF2 site was mutated could not bind FOXC2 (**Figure 4.63c**).

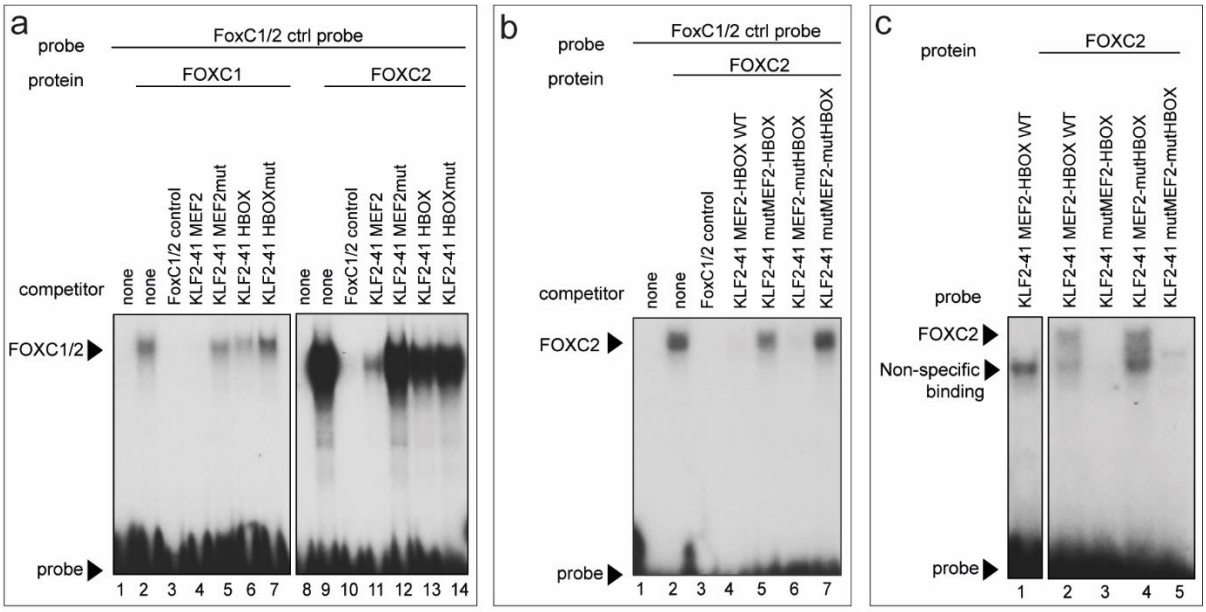


Figure 4.63: FOXC factors bind the KLF2-41 enhancer at the MEF2, but not the HBOX site. (a) The KLF2-41 MEF2 motif competed with the radioactively labelled FoxC1/2 control probe for FOXC1 and FOXC2 binding (Lanes 4 and 11), whereas the HBOX motif did not compete (Lanes 6 and 13). Mutant versions of either the MEF2 or HBOX motif also did not compete for FOXC1 or FOXC2 binding (Lanes 5, 7, 12, and 14). **(b)** The KLF2-41 WT double MEF2-HBOX motif competed with the radioactively labelled FoxC1/2 control probe for FOXC2 binding (Lane 4). The mutant version of the KLF2-41 MEF2-HBOX motif in which HBOX motif alone was mutated also competed for FOXC2 binding with the control probe (Lane 7). In contrast, mutant versions of the double motif in which either the MEF2 site alone, or both the HBOX and MEF2 sites, were mutated, was unable to compete with the control probe for FOXC2 binding (Lanes 6 and 8). **(c)** The KLF2-41 WT MEF2-HBOX motif directly bound FOXC2 (Lane 2). The double motif in which the HBOX site alone was mutated could also bind FOXC2 (Lane 3), whereas mutant version of the double motif in which either the MEF2 site, or both the MEF2 and the HBOX sites, were mutated, were unable to directly bind FOXC2 in the absence of a competitor (Lanes 2 and 5). All WT and mutant oligo sequences used in these experiments are listed in Chapter 2, Section 2.5.2.

Although no applicable publicly available FOXC ChIP-seq dataset were available, multiple FOXO1 ChIP-seq datasets were identified (**Figure 4.63**). Like FOXC1/2, FOXO proteins are expressed in ECs and play a number of crucial roles in EC development, although they have not been implicated in shear stress-induced signalling pathways or *KLF* gene regulation. In HUVECs, FOXO1 peaks overlapped with KLF2-66, KLF2-41, and KLF4-148 enhancer regions, and also with the *KLF2* and *KLF4* promoters[386]. In another dataset from adult mouse hearts (which included all cells, not just coronary and endocardial ECs), all KLF2/4 regulatory regions were also marked by FOXO1 peaks[387]. This confirmed that FOX factors might indeed be able to interact with the *KLF2/4* regulatory regions and implicates them in the regulation of these enhancers, although the exact role of the HBOX in this binding is unclear.

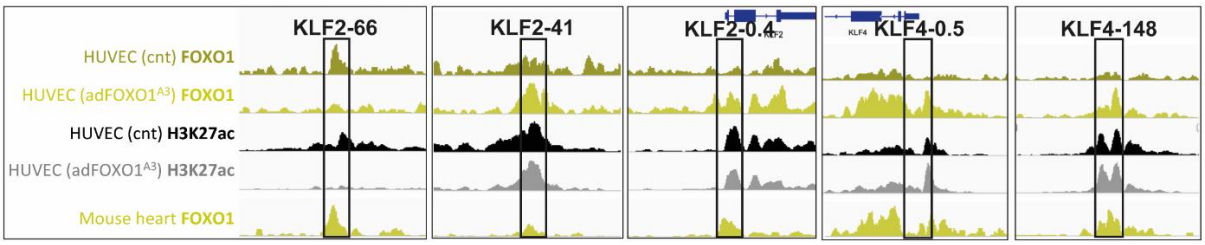


Figure 4.63: **Analysis of publicly available FOXO1 ChIP-seq datasets carried out in HUVECs and the adult mouse heart at the KLF2/4 enhancer and promoter regions.** A FOXO1 peak overlapped with all KLF2 regulatory regions in the control HUVEC dataset[386], but not with the KLF4 promoter or the KLF4-148 enhancer. In the corresponding dataset where HUVECs were treated with adFOXO1, FOXO1 peaks overlapped with KLF4 promoter and KLF4-148 enhancer regions, remained present at the KLF2-0.4 promoter and KLF2-41 enhancer, and was interestingly absent from the KLF2-66 enhancer region. Corresponding to the same datasets described above was ChIP-seq dataset of H3K27ac histone mark characteristic of enhancer regions. In both the control and adFOXO1 datasets, KLF2-41 and KLF4-148 enhancer regions overlapped with a H3K27ac peak. Interestingly, H3K27ac peaks also corresponded to the KLF2 and KLF4 promoter regions. As with the FOXO1 ChIP-seq dataset, the H3K27ac peak was absent from the adFOXO1 condition. Finally, FOXO1 binding in adult mouse hearts[387] was found at all KLF2/4 regulatory regions.

4.10 Discussion

The overall aim of this chapter was to analyse both *KLF2* enhancer sequences, KLF2-66 and KLF2-41, to understand how they are transcriptionally regulated. In order to identify direct transcriptional regulators of each enhancer, a combination of phylogenetic foot printing, mutational analysis, literature-based and *in silico* binding motif identity prediction, and validation using both EMSA and publicly available ChIP-seq datasets was employed. The key findings from these experiments are discussed below.

4.10.1 KLF2/4 regulatory regions are transcriptionally regulated by ETS and MEF2 factors

4.10.1.1 ETS factors

All *KLF2/4* regulatory sequences analysed contained multiple putative ETS binding motifs. This was expected due to the known role of ETS factors in transcriptional regulation of all endothelial-specific genes[143,345,346]. The ETS family of TFs contains at least 30 different proteins in mammals, each with distinct expression patterns and physiological roles. For example, ETV2, an ETS factor required for differentiation of endothelial and hematopoietic lineages, is expressed as early as E7.0 in a mouse embryo, though its expression is entirely lost by E11.5[388]. In contrast, endothelial expression of the equally important ETS factor ERG starts at E8.5 but remains high in the endothelium throughout development and adult life[349].

In EMSA analysis of the KLF2/4 enhancer and promoter regions, the highly conserved ETS1 DNA binding domain (ETS1DBD) was used to test the ability of putative ETS sites to interact generally with ETS factors. Despite many consensus ETS motifs occurring in each enhancer, there were only a few functional ETS motifs within each, although it should be noted that motifs with the ability to partially compete for ETS1 DBD binding in indirect experiments were not further investigated. The motifs that were able to directly bind ETS1 DBD were always evolutionarily conserved. One possible explanation

as to why some putative ETS binding motifs are functional and others are not, despite the same consensus sequence, is that the surrounding sequence varies between different putative sites. It is also possible that the large number of ETS sites provide robustness: given that ETS factors are essential for vascular development and physiology, the presence of many putative ETS sites within important gene regulatory regions might protect against mutations in functional ETS sites. Alternatively, some ETS sites may only become functional where specific co-factors are present or absent from the genomic context. This would potentially confound EMSA analysis, where no co-factors or chromatin modifiers are added.

Mutating the three ETS sites within the KLF2-41 enhancer sequence able to interact with ETS1DBD in EMSA resulted in loss of *LacZ* expression in the endothelium. This demonstrates that ETS factors are required for KLF2-41:*LacZ* expression, although it is impossible to firmly conclude whether one, two, or all three of these ETS sites are required. It should also be noted that ETS-c site located within the KLF2-41 enhancer overlapped with a conserved NFAT site, allowing for the possibility that the loss of transgene expression observed when conserved ETS sites were mutated could be partially or wholly attributed to the loss of NFAT binding. This possibility should be further investigated in the future.

So far, the requirement for ETS binding motifs in driving KLF2 activity has only been investigated in the context of KLF2-41:*LacZ*, although further mutational analysis of functional ETS binding sites within the KLF2-66 enhancer sequence is in progress. It is anticipated that the results of that analysis will mimic the outcome observed with the KLF2-41:*LacZ* transgene.

4.10.1.2 MEF2 factors

Mutational analysis of the *KLF2* promoter in cell culture systems by other groups already demonstrated that the conserved MEF2 binding motif within the *KLF2* promoter sequence is required for promoter activity, both under static and shear stress conditions[246,374]. Further, alterations to

pathways upstream of MEF2 factors significantly affected *Klf2* expression *in vivo* and *in vitro*[245,389], and recently induced EC-specific compound deletion of *Mef2a*/c/d in the adult mouse resulted in loss of *Klf2/4* expression[223]. In line with this, the mutation of the highly conserved MEF2 site within the KLF2-41 enhancer sequence resulted in complete abrogation of KLF2-41:*LacZ* transgene expression, confirming a requirement for the MEF2 binding motif for enhancer activity during development. While further mutational analysis to validate this finding in other *KLF2/4* enhancer regions is in progress, the *in vitro* analysis of conserved MEF2 binding motifs in all four sequences has clearly demonstrated that MEF2 factors are able to directly bind all *KLF2/4* regulatory elements investigated. I found that MEF2A, MEF2C, and MEF2D factors were equally able to bind the KLF2-41 enhancer. In addition, the MEF2A and MEF2C ChIP-seq datasets from mouse heart and lung ECs (provided by Prof. Mukesh Jain) validates these finding, suggesting that the MEF2 factors may genetically compensate for each other in a scenario where one of them was not expressed. This is consistent with evidence of MEF2A/C/D genetic compensation in other settings[212], and would explain the contrast in endothelial phenotypes between individual *Mef2* null mice and the induced *Mef2a/c/d* deletion[223].

Detailed analysis of MEF2 motifs within each *KLF2/4* regulatory region investigated revealed some sequence differences between the separate motifs. While the MEF2 motifs within the KLF2-41 enhancer, the KLF2-0.4 core promoter, and the KLF4-148 sequence corresponded perfectly to the MEF2 consensus binding motif $^{C}_{T}TA^{T}_{A}T^{A}_{T}A^{T}_{A}TA^{A}_{G}$ (**Figures 4.23 and 4.60**), the MEF2 motif within the KLF2-66 enhancer deviated from the consensus and was instead: GATATTTTGGGA. This sequence could therefore be considered composed of three putative binding motifs: a GATA site, a degenerate MEF2 site, and an RBPJ site. However, analysis of this motif for MEF2, GATA, and RBPJ binding in EMSA resulted found only robust MEF2 binding, with no not GATA or RBPJ binding detected. It is however possible that different TFs interact with this motif in response to specific cues, something that is complicated to test in EMSA, but could be addressed using advanced proteomic analyses.

4.10.2 RBPJ, GATA, and NFAT factors and regulation of EC subtype-specific expression

Mutational analysis of putative ETS and MEF2 binding motifs, combined with analysis of EMSA binding and the analysis of publicly available ChIP-seq datasets together demonstrated that ETS and MEF2 factors are both required for the regulation of KLF2-41:*LacZ* transgene expression. Further comparative analysis of other KLF2/4 regulatory regions suggests that, at least based on EMSA and ChIP-seq analysis, these two findings are applicable to all *KLF2/4* regions investigated. While functional ETS and MEF2 motifs were common to all regions, there were also notable differences in conserved TF binding motifs between different *KLF2/4* sequences which might be responsible for specific aspects of enhancer:reporter expressions patterns observed.

4.10.2.1 RBPJ

RBPJ, the transcriptional effector of the Notch signalling pathway, has long been thought to play a crucial role in the regulation of arterial identity. Although a direct gene activation role for RBPJ has recently been disputed, with data instead suggesting it functions through inhibition of cell cycle which in turn promotes arterial identity[176], many experimental models strongly link RBPJ-Notch to arterio-venous differentiation[390]. Further, the Notch signalling pathway can be induced by high shear stress in EC culture, which results in upregulation of *Cx37*, an arterial-expressed gene which is thought to inhibit cell cycle and thereby impart arterial specificity[198]. However, inhibition of different components of this signalling pathways in this *Cx37* study did not affect *KLF2* expression patterns [198].

Phylogenetic foot-printing of the KLF2-41 enhancer sequence combined with experimental validation of multiple orthologous enhancer versions led to the identification of two putative RBPJ binding motifs with the KLF2-41 enhancer sequence. When these two motifs were mutated in the human enhancer, the arterial specificity of KLF2-41:*LacZ* enhancer transgene expression was lost, and expansion to venous ECs was observed. This indicates a possible role for these putative RBPJ binding sites in maintaining the arterial specificity of the KLF2-41:*LacZ* transgene expression through suppression of

venous expression, as has been previously described in the RBPJ-regulated arterial-specific *Flk1* enhancer[275]. However, while these putative RBPJ motifs both matched the RBPJ consensus motif, the attempt to validate the two putative RBPJ sites in EMSA was not successful. Further, analysis of publicly available ChIP-seq datasets for RBPJ in ECs yielded contradicting results, as each dataset seemed to either overestimate or underestimate RBPJ binding at known RBPJ targets, making it complicated to draw conclusions relating to the *KLF2/4* regulatory regions.

In both the KLF2-66 and the KLF4-148 enhancers, a number of putative RBPJ motifs were also present. The KLF2-66 sequence contained only one conserved RBPJ motif that overlapped with the functional MEF2 binding site, whereas the KLF4-148 enhancer contained three conserved RBPJ motifs conserved all the way to zebrafish. While EMSA analysis did not demonstrate any of these motifs as functional RBPJ sites, this was also true at the KLF2-41 enhancer, perhaps suggesting that the RBPJ EMSA set up required optimisation. However, neither the KLF2-66 or the KLF4-148 enhancer drove arterial-specific transgene expression, and instead drove additional expression in the venous endothelium. This could therefore either suggest the RBPJ motifs were not functional, or that these enhancers also contain TF binding motifs that actively direct expression in the venous compartment. Mutational analyses of RBPJ binding motifs within these two enhancers are currently ongoing, and the result from these experiments is expected to further clarify the role of RBPJ in *KLF2/4* regulation.

4.10.2.2 KLF2-66 enhancer contains a functional GATA binding motif

GATA factors have been implied in regulation of EC identity[365,369] and valve development[366,391], in addition to their well-defined roles in the regulation of hemtopoiesis[126]. Further, exposing human lymphatic endothelial cells (LEC) to oscillatory shear stress in cell culture led to upregulation of *GATA2* mRNA expression and increased GATA2 protein levels in cell nuclei[392], suggesting *GATA2* expression in LEC is regulated by shear stress.

Analysis of the KLF2-66 enhancer sequence revealed four putative conserved GATA binding motifs, in addition to one conserved GATA motif identified within the KLF4-148 enhancer sequence. However, EMSA analysis found that only a single conserved GATA motif from the KLF2-66 enhancer appeared functional. Surprisingly, the functional GATA motif within the KLF2-66 enhancer sequence was located at the very edge of the conserved enhancer core and was itself poorly conserved, in contrast to the ETS and MEF2 motifs.

While the non-conserved GATA motifs located outside the conserved enhancer core are unlikely to be required for KLF2/4 enhancer activity, this analysis cannot rule out the possibility that other GATA motifs within the KLF2-66 enhancer interact with GATA factors other than GATA2. For example, GATA6 has been shown to be activated by laminar shear stress in cultured human coronary artery ECs and inhibit the expression of the urokinase-type plasminogen activator, a gene activated in response to oscillatory shear stress whose expression is upregulated within atherosclerotic lesions[393]. In contrast, GATA1 and GATA4 expression levels corresponded with oscillatory-flow activated transient receptor potential (TRP) channel 6 mRNA expression in HUVECs[394].

Alternatively, GATA2 binding might require co-factors not included in EMSA setup. For example, one GATA motif within the KLF2-66 enhancer sequence was adjacent to an ETS binding site, and in both the KLF2-66 and the KLF4-148 enhancers, a GATA site overlapped with the functional MEF2 site. The fact that both these enhancers drove pan-endothelial reporter expression suggests that these GATA sites may potentially play a role in venous transgene expression, although the existing data cannot confirm or rule this out. In addition, GATA factors have been implicated in valve development, although it is unlikely they would be responsible for transgene expression in ECCs which is driven by all *KLF2/4* enhancers analysed, as the KLF2-41 enhancer contains no putative GATA binding sites.

Further mutational analysis of the KLF2-66 enhancer GATA sites should help elucidate whether these sites are required for KLF2-66:*LacZ* and KLF4-148:*LacZ* transgene expression. Evaluating the ability of

different GATA factors to interact with these regions in EMSA would provide further clarity on the role of GATA factors in *KLF2/4* regulation.

4.10.2.3 KLF2-41 and KLF4-148 enhancer contain functional NFAT binding motifs

NFAT factors regulate valve[395–397], blood vessel[362], and lymphatic vessel development[360], and have been shown to lie downstream of mechanotransduction signalling in non-endothelial cell types, including T cells[398] and osteoblasts[399]. Although NFAT motifs (GGAAA consensus) are highly similar to ETS (GGA^{A}_{T}), the sequence analysis revealed the presence of putative NFAT binding sites, with EMSA analysis confirming functional NFAT motifs within the KLF2-41 and the KLF4-148 enhancer sequences. While by definition each NFAT motif is also a putative ETS motif, neither of these motifs bound ETS1DBD in EMSA, suggesting they were not functional ETS binding sites. Mutational analysis of the KLF2-41 enhancer NFAT sites is in progress.

4.10.3 A previously unidentified HBOX motif found in KLF2/4 regulatory regions is required for KLF2/4 enhancer:reporter transgene expression

The HBOX motif found adjacent to the functional MEF2 motif in all investigated *KLF2/4* regulatory sequences was initially thought to also bind MEF2 due to its similarity to the MEF2 consensus motif. However, no HBOX motif bound MEF2 factors in EMSA, although mutational analysis of both the KLF2-41 and KLF2-66 enhancer regions confirmed that this motif is required for reporter expression.

Interestingly, the HBOX motif within the *KLF2* promoter sequence was previously identified as a required motif for ERK5-induced *KLF2* expression under static conditions[245], although the simultaneous discovery of the MEF2 motif meant this observation was never followed up at the time. However, these experiments clearly demonstrated that mutating either the MEF2 or the HBOX motif alone results in abrogation of reporter activity driven by the *KLF2* promoter in response to ERK5 activation. Remarkably, the HBOX motif sequence was almost perfectly conserved between all four

KLF2/4 regulatory regions tested, suggesting the same factor is interacting with all four sequences. Precise spacing between the MEF2 and HBOX binding motifs suggests rigid enhancer grammar between these two sites. Further, the small spacing of only five nucleotides between the two sites suggests the two factors might bind cooperatively. However, despite its requirement for the activity of all *KLF2/4* regulatory elements, efforts to identify the factor interacting with the HBOX motif have so far been unsuccessful. The initial *in vitro* study had suggested that the HBOX motif might be bound by TEAD factors, although the consensus binding motif for TEAD factors is a poor match to the HBOX motif[245] and no aspect of TEAD biology supports an obvious role in KLF regulation.

4.10.4 Identifying the transcriptional regulators of the MEF2-HBOX double motif

Analysis of MEF2-bound regions found no other occurrences for this double MEF2-HBOX motif aside from *KLF2/4* regulatory regions studied in this project. This aligns with the observation that loss of MEF2A/C/D in adult tissue very specifically affects KLF2/4 driven pathways, suggesting that this specificity may occur via the MEF2-HBOX motif and demonstrating the importance of understanding which TFs regulate the HBOX site.

Previous studies of MEF2 factor binding dynamics demonstrated that MEF2 factors often bind DNA as homo- or heterodimers[212,382,383]. While the preliminary EMSA analysis of the MEF2-HBOX binding motif within the KLF2-41 enhancer found no evidence in support of the MEF2 dimer hypothesis, these experiments did not adequately address the possibility of a requirement for the occupancy of the first MEF2 site prior to the binding of the second MEF2 factor. Additional analysis using the same system is currently in progress and includes testing all *KLF2/4* double motifs for their ability to bind MEF2A, C, and D proteins alone and in combination in direct EMSA with no competitors present.

The potential of FOX factors to interact with the HBOX motif was also tested in EMSA, since the FOX factor consensus binding motifs was similar to that of HBOX. Surprisingly, FOXC factors were unable

to interact with the HBOX binding motif but robustly bound the MEF2 motif in the KLF2-41 enhancer. This finding suggest that MEF2 and FOX factors might occupy the MEF2 site under different conditions. EMSA analysis where different combinations of MEF2 and FOX factors were added to a radioactively labelled KLF2-41 MEF2-HBOX double motif showed that the presence of both factors increased MEF2 binding, but no specific FOX band was produced (data not shown). As with the MEF2 dimer hypothesis, these experiments were limited in their capacity to fully address the question as the requirement for the occupancy of one motif to allow binding at the second motif was not tested. Further studies testing a combination of FOX and MEF2 factors with MEF2/HBOX double motifs are in progress.

4.10.5 A model of transcriptional regulation at the KLF2 gene locus

An important remaining question is why three distinct regulatory elements that control the expression of a single gene might all be regulated by the same array of transcriptional regulators: ETS, MEF2, and HBOX. One possible explanation is that the KLF2 promoter, while not by itself able to drive reporter expression during development, serves as a dock for these factors brought over by the KLF2-66 and KLF2-41 enhancers. Another potential explanation is that the enhancers, true to their function, act as amplifiers of low baseline *KLF2* transcription mediated by the KLF2 promoter through its minimal assortment of functional binding motifs: one ETS, one MEF2, and one HBOX site. The mechanism of signal amplification in this case might be simply the presence of more TF molecules, which might then be able to accumulate at the promoter in greater numbers, and potentially interact with each other though dimer formation as has been previously suggested in other MEF2-regulated genes. Further, this could be stimulus- and context-dependent, for example induced by especially high levels of shear stress in a specific vascular bed. In order to investigate this possibility, it would be interesting to analyse the whole *KLF2* locus containing both enhancers and the promoter in a cell culture experiment. If the whole locus was fused with a minimal promoter and a reporter gene, mutational analysis of different MEF2 and HBOX sites within different elements and the resulting reporter expression changes could be used to understand the dynamic of MEF2 and HBOX binding at this locus,

and how MEF2 binding behaves under different conditions, for example after application of different shear stress patterns and magnitudes.

4.11 Conclusion

The results of this chapter strongly support a requirement for MEF2 factors in regulation of *KLF2* expression. However, in a developmental setting, MEF2 factors primarily regulate the activity of two previously uncharacterised enhancers located far upstream the *KLF2* TSS, and not the *KLF2* promoter. Detailed analysis of KLF2/4 regulatory regions using a combination of mutational analysis, *in vitro* validation, and analysis of publicly available ChIP-seq datasets identified additional roles for ETS, RBPJ, NFAT, and FOX factors in regulation of KLF2/4 enhancers. Crucially, this analysis identified an HBOX motif common to all *KLF2/4* regulatory regions, that was required for the expression of KLF2-41:*LacZ* and KLF-66:*LacZ* transgenes. Initial EMSA analysis for the identity of the putative HBOX motif was inconclusive, suggesting a different approach is required to enable the identification of the TF regulating this site.

5 KLF2 enhancer activity in response to shear stress

5.1 Rationale

In the previous chapters, the KLF2-41 and KLF2-66 enhancers were identified and characterised during development, the key TF binding motifs were elucidated and the cognate TFs able to interact with these motifs *in vitro* were found. However, an important remaining question in this analysis of the KLF2-41 and KLF2-66 enhancers is whether they play a role in shear stress-induced gene expression. This chapter therefore aims address this question using the KLF2-41:*LacZ* and KLF2-66:*LacZ* stable transgenic mouse lines characterised in Chapter 3 of this thesis.

The notion of *KLF2* regulation by shear stress through interaction of MEF2 with the KLF2 promoter has been wholly based on *in vitro* experiments in which shear stress application to cultured ECs overexpressing a KLF2 promoter-reporter gene construct resulted in induction of reporter expression[245,246,374]. Application of shear stress to cultured ECs has various advantages over measuring shear stress responses *in vivo*. For example, parallel plate flow systems used in these settings allow for precise calculations of shear stress applied[400–403]. Cell lines are easily transfected with different constructs which allows for parallel testing of WT and mutant sequences. Further, the hallmark shear stress responses in specific endothelial cell lines, for example HUVECs, have been well characterised in the literature[198,404,405]. These include upregulation of specific genes at specific timepoints, and phenotypic adaptations induced by shear stress such as cell alignment with the direction of flow[210].

However, there are also several limitations to studying shear stress responses in cell culture. ECs within different vascular beds, blood vessels, and even parts of a single blood vessel, all differ significantly in both the shear stress magnitude and shear stress pattern that they experience[406,407]. In the context of *in vitro* shear stress studies, these differences have often been disregarded for practical reasons. For instance, HUVECs, which are of venous origin, have been widely used to apply arterial shear stress magnitudes to study haemodynamic regulation of arterial-expressed genes, due to their

phenotypic properties which allow for good adhesion when exposed to shear stress[198,408,409]. While HUVECs are widely available and applicable to human biology, endothelial cell lines from different species, including human, mouse, and bovine ECs, are often used interchangeably in the same study due to their varying phenotypic properties that allow for investigation of different aspects of the shear stress response[410]. These properties include the expression of membrane mechanosensors, adhesion molecules, and primary cilia, which have rarely been systematically defined and compared between different cell types, but are known to influence the shear stress response, including gene expression and cell alignment. Further, due to the differences in cell types and shear stress conditions used in these studies, minimal overlap is observed both between datasets generated *in vitro*, as well as between datasets obtained in cell culture and *in vivo*[210,321].

In addition to disregard for cell origin and phenotypic properties of different cell lines, adult immortalised endothelial cell lines are routinely applied to analysis of developmental gene expression. The differences in shear stress dynamics between embryonic and adult blood vessels are widely acknowledged in the literature, although the precise relationship between embryonic and adult cardiac output, vessel diameter, and blood viscosity, all of which affect shear stress levels, have not been systematically investigated[183,187,411,412]. The difficulty in translating these variables, well characterised in certain adult blood vessels such as the dorsal aorta, into the developmental setting lies in the fact that measuring shear stress in live embryos is experimentally complicated due to accessibility difficulties and inability to culture mouse embryos beyond early stages of vascular development (E9.5)[331,413].

Despite the many shortcomings of studying the endothelial shear stress response in cell culture, manipulating shear stress *in vivo* using genetic, pharmacological, or surgical approaches is not straightforward either. Systemic genetic and pharmacological approaches that cease cardiac contractions in embryos ablate shear stress as a secondary effect and come with additional secondary

effects including ablation of nutrient and oxygen supply, thereby inducing nutrient starvation and hypoxia, both of which influence EC behaviour[414]. In adult animals, surgical manipulations, such as carotid artery ligation, are more targeted, but suffer from other disadvantages: only certain blood vessels can be efficiently manipulated in this way, defining appropriate controls can be complicated, and the reduction in shear stress levels is difficult to measure[415,416].

The dorsal aorta is a stereotypical example of a blood vessel where differential geometry of the vessel, namely the straight descending aorta and the curved aortic arch, dictates differential magnitudes and patterns of shear stress that have been well characterised in the literature[182,188,417–419]. Within the straight descending aorta, ECs are thought to experience high laminar shear, which predisposes them to a quiescent phenotype and is negatively correlated with atherosclerotic plaque formation. In contrast, ECs within the inner curvature of the aortic arch, and vascular bifurcations of carotid arteries at the outer curvature, are thought to experience oscillatory shear stress linked to an inflammatory EC phenotype and development of atherosclerotic plaques in this area. Analysis of *KLF2* mRNA expression in the human dorsal aorta samples has demonstrated that *KLF2* expression is high within the endothelium of the descending aorta but decreases in the inner curvature of the aortic arch and at vascular bifurcations[420]. Further, a mouse study using the aortic cuff model to increase shear stress level within the descending aorta by reducing its diameter found that *Klf2* mRNA expression increased in the area where shear stress magnitude was higher[249].

In addition to the well characterised difference in shear stress magnitudes within the dorsal aorta, shear stress magnitudes experienced by different vascular beds are defined in the literature. Small diameter resistance arteries and capillaries experience shear stress of the highest magnitude, in the range of 40-60 dyn/cm^2. The dorsal aorta and small diameter veins are though to experience shear stress between 10-15 dyn/cm^2. Finally, large diameter veins are thought to experience shear stress of

around 5 dyn/cm^2 [186]. In line with this, *Klf2* mRNA expression has been robustly detected in arterial and, to a lesser extent, in venous ECs[248].

5.1.1 Aim

This chapter aims to use the KLF2-41:*LacZ* and KLF2-66:*LacZ* stable transgenic lines to characterise transgene expression in the post-natal (P8) and adult dorsal aorta, and investigate whether transgene expression patterns correspond to differential shear stress magnitudes and differential *KLF2* mRNA expression in this setting. Additionally, a pharmacological approach was taken to reduce embryonic shear stress and characterise transgene expression in drug-treated embryos compared to vehicle-treated controls.

5.2 KLF2 enhancer:LacZ transgene expression in the dorsal aorta and adult organs

In order to analyse KLF2-66:*LacZ* and KLF2-41:*LacZ* transgene expression, the dorsal aorta was dissected from either a mouse pup or an adult mouse from each transgenic mouse line defined in Chapter 3. Following staining and fixation, the aorta was further dissected to allow for the visualisation of the endothelial layer within the whole vessel, referred to as *en face* analysis. The initial hypothesis was that transgene expression would be high in the descending aorta where shear stress is high, and lower in the inner curvature of the aortic arch, consistent with previously published *Klf2* mRNA expression data. Reduction of transgene expression was also hypothesised to occur at vascular bifurcations at which the brachiocephalic artery, left common carotid artery, and left subclavian artery, emerge from the aortic arch, and at inter-costal artery branch points that emerge from the descending aorta.

5.2.1 KLF2-41:LacZ expression in the post-natal dorsal aorta

The KLF2-41:*LacZ* transgene expression in P8 mouse was detected alongside the dorsal aorta in whole-mount analysis (**Figure 5.1a**). *En face* analysis of the descending part of the dorsal aorta confirmed

that the KLF2-41:*LacZ* transgene is robustly expressed in the endothelium of the P8 dorsal aorta. Upon closer inspection, especially strong X-Gal staining was observed within inter-costal arteries, which are high-resistance small diameter arteries that branch off the descending aorta and supply blood to the ribcage. While it initially seemed like it is the branch points that exhibit more staining (**Figure 5.1b** and **c**), analysis of the whole-mount aorta in fact demonstrated that high intensity staining is detected throughout the intercostal arteries, and not just the branch points (**Figure 5.1a**). Within the aortic arch, the KLF2-41:*LacZ* transgene was expressed throughout the outer curvature, including the brachiocephalic artery, left common carotid artery, and left subclavian arteries (**Figure 5.1d**). However, a reduction of KLF2-41:*LacZ* transgene expression was detected in the inner curvature of the dorsal aorta (**Figure 5.1d** and **e**), where ECs are thought to experience a lower magnitude of shear stress with an oscillatory pattern, and *KLF2* expression is reduced.

The most important determinant of shear stress magnitude in different blood vessels within the same organism with constant heartbeat is vessel diameter, which is inversely proportional to the shear stress magnitude. Small diameter vessels thus experience a higher shear stress magnitude than large diameter vessels. The observation of higher X-Gal staining intensity within intercostal arteries compared to the dorsal aorta within the same mouse pup might therefore be a reflection of higher shear stress magnitude experiences by ECs in the small diameter, high resistance intercostal arteries compared to the larger diameter dorsal aorta. Further, robust KLF2-41:*LacZ* expression in the outer curvature of the aortic arch, with a reduction of expression in the inner curvature, is also consistent with both the shear stress magnitudes experienced by ECs in these areas and *KLF2* expression pattern described previously.

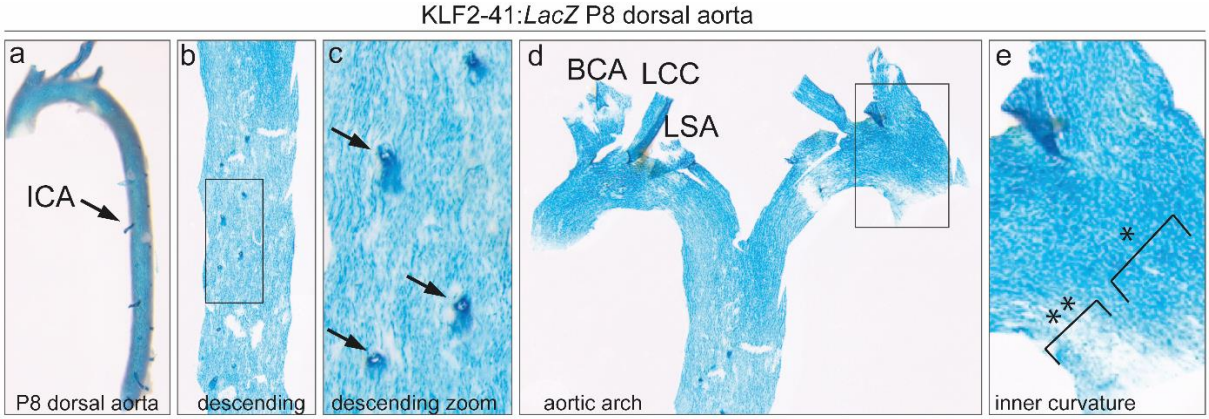


Figure 5.1: Analysis of KLF2-41:*LacZ* transgene expression in the post-natal P8 dorsal aorta. (a) Whole-mount dorsal aorta dissected from a KLF2-41:*LacZ* stable line P8 mouse pup. **(b)** *En face* view of the KLF2-41:*LacZ* transgene expression in the P8 descending aorta. **(c)** Close-up *en face* view of the descending aorta with inter-costal artery branch points exhibiting especially high KLF2-41:*LacZ* expression. **(d)** *En face* view of the KLF2-41:*LacZ* transgene expression in the P8 aortic arch. **(e)** Close-up *en face* view of the aortic arch, with different degrees of KLF2-41:*LacZ* expression between the inner (**) and outer (*) curvatures of the aortic arch. **ICA=intercostal artery, BCA=brachiocephalic artery, LCC=left common carotid artery, LSA=left subclavian artery.**

5.2.2 KLF2-66:LacZ expression in the post-natal dorsal aorta

Similar to the results of KLF2-41:*LacZ* expression in the P8 dorsal aorta, the KLF2-66:*LacZ* transgene was expressed throughout the P8 dorsal aorta, including the descending aorta and the aortic arch (**Figure 5.2a**). In the *en face* analysis, the KLF2-66:*LacZ* transgene was expressed throughout the endothelium of the descending aorta, although the expression was less continuous compared to KLF2-41:*LacZ* transgene, with visible patches of tissue with low levels of *LacZ* expression (**Figure 5.2b**). In line with KLF2-41:*LacZ* expression, KLF2-66:*LacZ* expression was also stronger in the inter-costal arteries observed along the descending aorta (**Figure 5.2a, b**), including branch points and the rest of the vessel emerging from the descending aorta (**Figure 5.2c**). In the aortic arch, the KLF2-66:*LacZ* transgene was expressed in the outer curvature, including the brachiocephalic artery, left common carotid artery, and left subclavian artery, all of which extend from the outer curvature of the aortic arch (**Figure 5.2d**). In the inner curvature of the aortic arch, reduced KLF2-66:*LacZ* transgene expression was observed (**Figure 5.2d** and **e**), which is consistent both with the KLF2-41:*LacZ* transgene expression, as well as lower shear stress magnitudes and lower levels of *KLF2* expression expected to occur in this area. However, as observed within the descending aorta, the KLF2-66:*LacZ*

transgene expression was more patchy compared to KLF2-41:*LacZ* expression, which was also observed at the bottom of the aortic arch as it transitions into descending aorta.

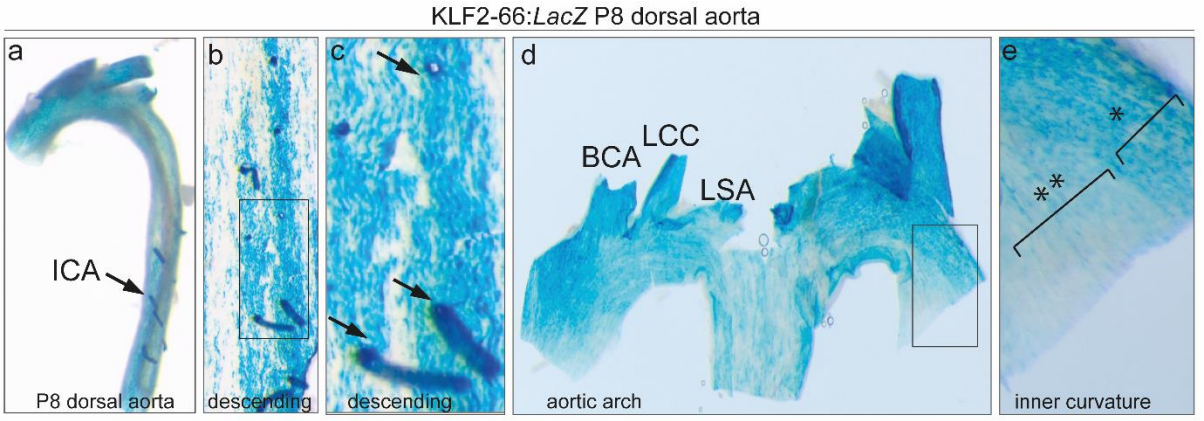


Figure 5.2: Analysis of KLF2-66:*LacZ* transgene expression in the post-natal P8 dorsal aorta. (a) Whole-mount dorsal aorta dissected from a KLF2-66:*LacZ* stable line P8 mouse pup. **(b)** *En face* view of the KLF2-66:*LacZ* transgene expression in the P8 descending aorta. **(c)** Close-up *en face* view of the descending aorta with inter-costal artery branch points exhibiting especially high KLF2-66:*LacZ* expression. **(d)** *En face* view of the KLF2-66:*LacZ* transgene expression in the P8 aortic arch. **(e)** Close-up *en face* view of the aortic arch, with different degrees of KLF2-66:*LacZ* expression between the inner (**) and outer (*) curvatures of the aortic arch. **ICA=intercostal artery, BCA=brachiocephalic artery, LCC=left common carotid artery, LSA=left subclavian artery.**

Taken together, in the analysis of KLF2-66:*LacZ* and KLF2-41:*LacZ* transgene expression in the post-natal dorsal aorta, both enhancers directed *LacZ* expression in a pattern consistent both with shear stress magnitude and pattern, as well as *KLF2* expression. To further explore the correlation between the ability of both enhancers to drive reporter expression consistent with shear stress dynamics, their expression within the adult dorsal aorta was investigated next.

5.2.3 KLF2-41:LacZ transgene expression in the adult dorsal aorta

Having observed that KLF2-41:*LacZ* and KLF2-66:*LacZ* transgene expression patterns in the post-natal mouse dorsal aorta corresponded both to shear stress magnitude and *KLF2* expression levels expected in different regions of the dorsal aorta, the next aim was to investigate *LacZ* expression driven by these enhancers in the adult dorsal aorta. Shear stress magnitudes experienced by ECs in the adult dorsal aorta are estimated to be between 10 and 15 dyn/cm^2, whereas embryonic shear stress magnitude is

not thought to exceed 5 dyn/cm[187]. There are no shear stress magnitude data currently available on the post-natal aorta, although the value would probably lie between the embryonic and adult magnitudes. ECs within adult small diameter and high resistance arterioles, such as intercostal arteries, are thought to experience significantly higher shear stress magnitudes, ranging from 30-50 dyn/cm[186]. Based on this information, it was hypothesised that enhancer:*LacZ* expression within the endothelium of the adult dorsal aorta would be detected at high levels, with differential expression between the descending aorta and the inner curvature of the aortic arch. High transgene expression levels were also expected within the endothelium of intercostal arteries.

Analysis of the KLF2-41:*LacZ* transgene within the adult dorsal aorta did not confirm this hypothesis. Although some sparse and punctate *LacZ* expression was observed both within the aortic arch and the descending aorta, no consistent *LacZ* expression was observed within the endothelium of the descending aorta or the aortic arch, and no difference was observed between the inner and the outer curvatures (**Figure 5.3a, b,** and **c**). Upon closer inspection of the inter-costal artery branch points, where KLF2-41:*LacZ* expression was especially pronounced post-natally, a few *LacZ* positive cells were identified in some of the inter-costal artery branch points. However, this expression was inconsistent, and it was difficult to say whether expression overlapped with ECs or was instead just a consequence of random speckled staining observed across the entire adult dorsal aorta (**Figure 5.3d** and **e).**

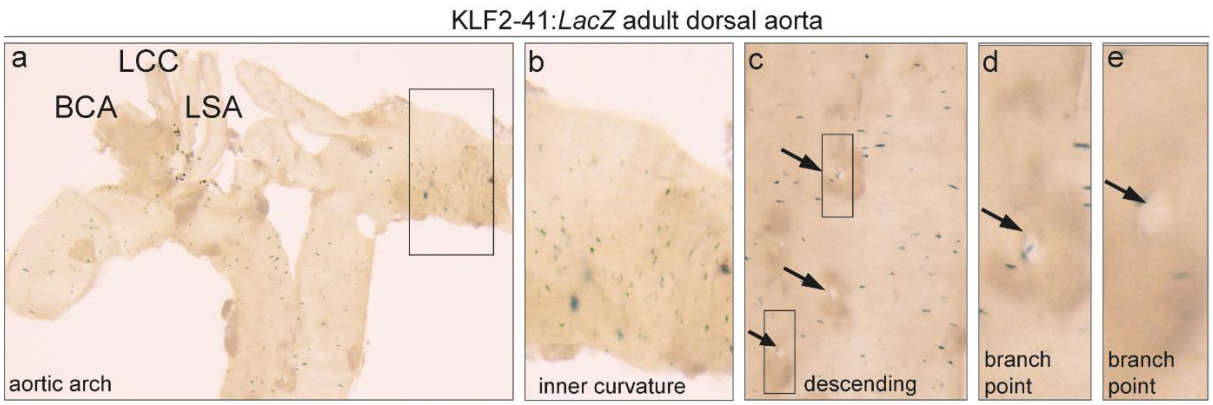


Figure 5.3: Analysis of KLF2-41:*LacZ* transgene expression in the adult dorsal aorta. (a) *En face* view of the KLF2-41:*LacZ* transgene expression in the adult aortic arch. **(b)** Close-up *en face* view of the aortic arch. **(c)** *En face*

view of the KLF2-41:*LacZ* transgene expression in the adult descending aorta. **(d), (e)** Close-up *en face* view of the descending aorta with inter-costal artery branch points exhibiting especially high KLF2-41:*LacZ* expression. **BCA=brachiocephalic artery, LCC=left common carotid artery, LSA=left subclavian artery.**

5.2.4 KLF2-66:LacZ transgene expression in the adult dorsal aorta

Similar to the KLF2-41:*LacZ* transgene, the KLF-66:*LacZ* transgene was only sporadically expressed in the adult dorsal aorta. *LacZ* positive speckles were observed within the aortic arch (**Figure 5.4a** and **b**), but seemed consistent between the outer and inner curvatures. Further, it was difficult to tell whether *LacZ* positive cells overlap with ECs. Within the descending aorta, there was less *LacZ* staining overall, but there was consistent staining within multiple cells of each inter-costal artery branch point (**Figure 5.4c, d** and **e**). This result demonstrates that the KLF2-66 enhancer is able to drive *LacZ* expression in the area that experiences the highest shear stress magnitude within the dorsal aorta.

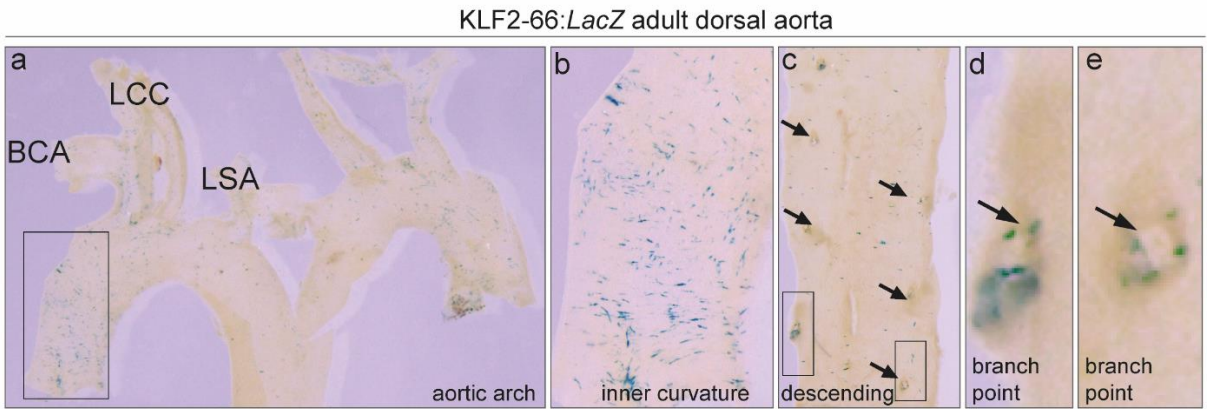


Figure 5.4: Analysis of KLF2-66:*LacZ* transgene expression in the adult dorsal aorta. (a) *En face* view of the KLF2-66:*LacZ* transgene expression in the adult aortic arch. **(b)** Close-up *en face* view of the aortic arch. **(c)** *En face* view of the KLF2-66:*LacZ* transgene expression in the adult descending aorta. **(d), (e)** Close-up *en face* view of the descending aorta with inter-costal artery branch points exhibiting especially high KLF2-66:*LacZ* expression. **BCA=brachiocephalic artery, LCC=left common carotid artery, LSA=left subclavian artery.**

5.2.5 KLF2-41:LacZ expression in adult organs

Although minimal KLF2-41:*LacZ* expression was detected in the adult dorsal aorta, some *LacZ* expression in inter-costal artery branch points suggested that *LacZ* expression might be detected in smaller diameter arteries that experience higher shear stress magnitudes. To investigate this hypothesis, adult mouse organs known to be highly vascularised, including the brain, the lung, the heart, the diaphragm, and the ribcage, were dissected and stained with X-Gal. While some *LacZ*

expression was observed in some of the tissues analysed, including punctate expression in the brain, the ribcage, and the diaphragm (**Figure 5.5a, b,** and **c**), no vascular *LacZ* expression was observed in any of the organs analysed (**Figure 5.5**).

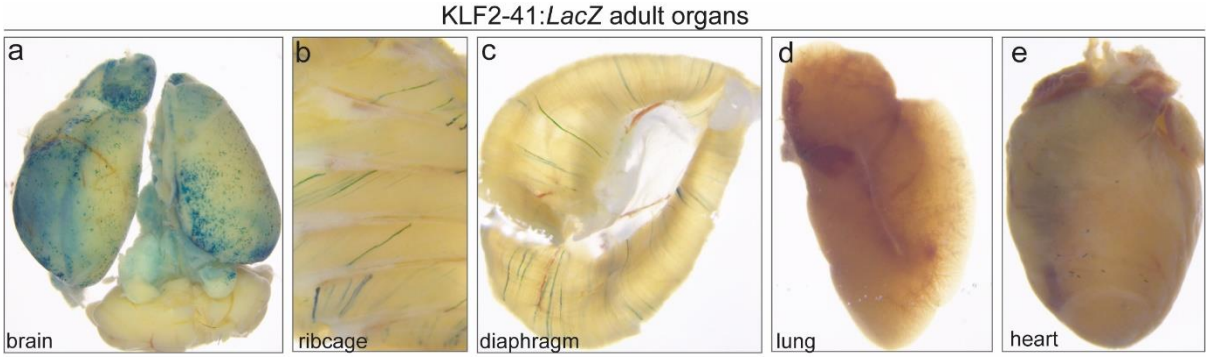


Figure 5.5: Analysis of KLF2-41:*LacZ* transgene expression in adult organs. (a) Ectopic KLF2-41:*LacZ* expression in the adult mouse brain, while no expression is detected in the vasculature. **(b)** No KLF2-41:*LacZ* transgene expression is detected in intercostal blood vessels within the adult ribcage. **(c)** No KLF2-41:*LacZ* transgene expression is detected in blood vessels of the adult diaphragm. **(d)** No KLF2-41:*LacZ* transgene expression is detected within the adult lung vasculature. **(e)** No KLF2-41:*LacZ* transgene expression is detected within the coronary vessels of the adult heart.

In addition to expression in ECs, *Klf2* mRNA is also expressed in ECCs and later cardiac valves[192,248]. Both KLF2-66:*LacZ* and KLF2-41:*LacZ* transgenes were robustly expressed in the endothelial lining of embryonic ECC and of the post-natal heart valve leaflets (Chapter 3, Sections 3.5.4 and 3.6.3). Upon dissection of the adult mouse heart, some KLF2-41 expression was detected in the cardiac valve endothelial lining, although this was dispersed and punctate (**Figure 5.6a** and **b**). It was therefore difficult to conclude that this expression corresponds to valve ECs. Further *Klf2* mRNA expression analysis would be needed to understand whether the positive *LacZ* signal overlaps with *Klf2* mRNA expression in adult cardiac valves.

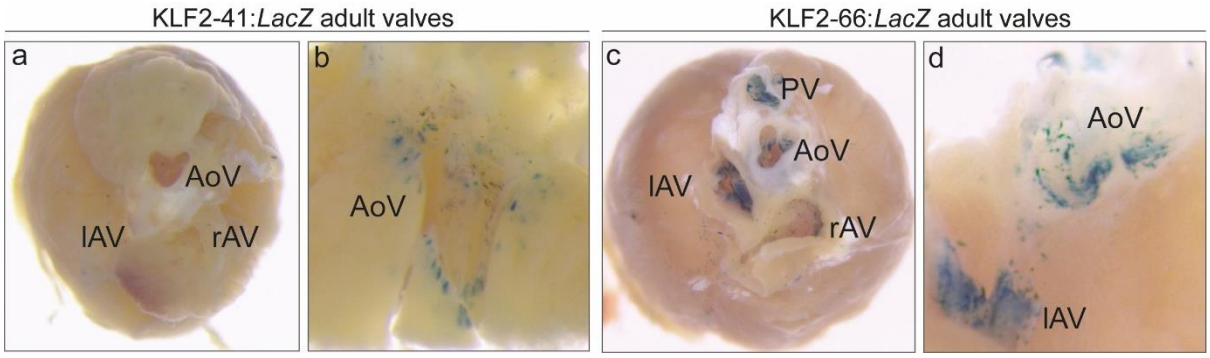


Figure 5.6: KLF2-41:*LacZ* and KLF2-66:*LacZ* transgene expression analysis in the adult mouse cardiac and aortic valves. (a) Top view of the KLF2-41:*LacZ* stable line whole-mount adult mouse heart with no transgene expression detected within the cardiac or aortic valves. **(b)** Cross-section of the KLF2-41:*LacZ* adult atrio-ventricular valve shows punctate transgene expression around the edges of valve tissue. **(c)** Top view of the KLF2-66:*LacZ* stable line whole-mount adult mouse heart with robust transgene expression detected within the cardiac and aortic valves. **(d)** Cross-section of the KLF2-66:*LacZ* adult atrio-ventricular valve shows robust transgene expression along the surface of both the aortic and the atrio-ventricular valves. **lAV=left atrioventricular valve, rAV=right atrioventricular valve, PV=pulmonary valve, AoV=aortic valve.**

5.2.6 KLF2-66:LacZ expression in adult organs

Analysis of the KLF2-66:*LacZ* expression in adult organs yielded similar results to the analysis of KLF2-41:*LacZ* expression. No vascular LacZ expression was detected in any of the organs analysed, including the brain, the ribcage, the diaphragm, the lung, and the heart (**Figure 5.7**). However, examination of KLF2-66:*LacZ* transgene expression in cardiac valves revealed robust transgene expression lining the valve aortic, pulmonary, right, and left atrio-ventricular valves (**Figure 5.6c** and **d**). Shear stress dynamics within the adult cardiac valves are complex. Valves probably experience the highest shear stress of all components of the cardiovascular system, but shear stress pattern in this area is thought to be oscillatory[412]. As with KLF2-41:*LacZ* transgene, further expression analysis is required to definitively conclude that *LacZ* positive cells correspond to *Klf2* mRNA expressing cells.

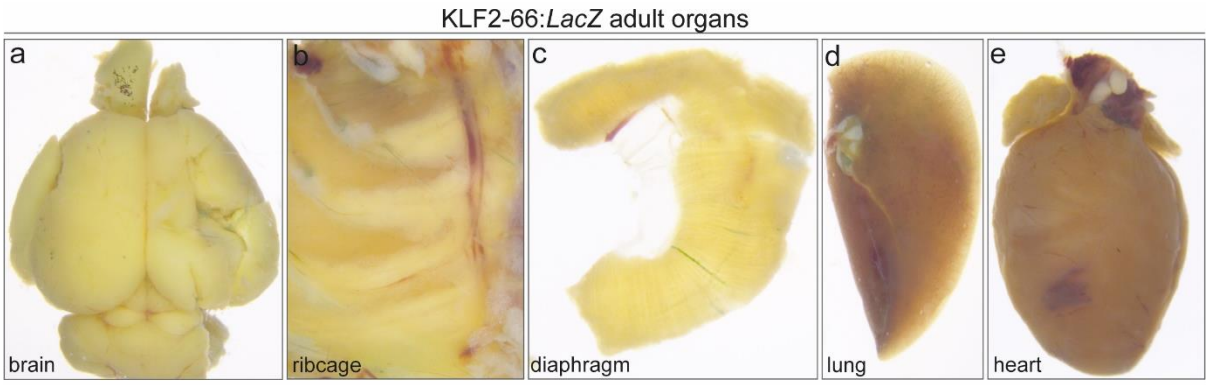


Figure 5.7: Analysis of KLF2-66:*LacZ* transgene expression in adult organs. (a) No KLF2-66:*LacZ* expression is detected in the adult mouse brain vasculature. **(b)** No KLF2-66:*LacZ* transgene expression is detected in intercostal blood vessels within the adult ribcage. **(c)** No KLF2-66:*LacZ* transgene expression is detected in blood vessels of the adult diaphragm. **(d)** No KLF2-66:*LacZ* transgene expression is detected within the adult lung vasculature. **(e)** No KLF2-66:*LacZ* transgene expression is detected within the coronary vessels of the adult heart.

5.3 KLF2-41:LacZ and KLF2-66:LacZ expression following pharmacological manipulation of blood flow

Due to the lack of KLF2-41:*LacZ* and KLF2-66:*LacZ* transgene expression in adult vasculature, including the dorsal aorta and organ-specific vessels, it was concluded that the adult vasculature is not an appropriate model to further investigate the shear stress-mediated induction of *LacZ* expression driven by *KLF2* enhancers. The analysis of *LacZ* expression driven by both enhancers in the embryonic and early post-natal vasculature seemed a more appropriate model to address this question.

5.3.1 Rationale

In order to test whether KLF2-41:*LacZ* and KLF2-66:*LacZ* transgene expression changes when shear stress is manipulated, a pharmacological approach was used. Pharmacological inhibition of shear stress has certain limitations. Manipulating cardiac contractions has systemic effects on the whole embryo and does not allow for selective manipulation of specific blood vessels, for example the dorsal aorta. In addition, while a reduction in blood flow, and therefore shear stress magnitude, is one of the secondary effects of inhibiting cardiac contractions, additional secondary effects, such as reduction in oxygen and nutrient supply also accompany this intervention.

Despite these limitations, the pharmacological approach was chosen given the limited capacity for surgical manipulation of embryonic blood vessels. An alternative approach could be to use a genetically modified model in which genes encoding proteins required for cardiac contractions, such as *Ncx1* or *Mlc2a*, are deleted. However, due to complete cessation of cardiac contractions in these genetic models, the limitations were the same to that of a pharmacological approach. Because they completely lack a heartbeat, *Ncx1*$^{-/-}$ and *Mlc2a*$^{-/-}$ embryos are not viable past E9.5, which would complicate enhancer:*LacZ* transgene analysis past this timepoint. Finally, the high cost and long timeline for generation of embryos transgenic for each enhancer:*LacZ* construct on the *Ncx1*$^{-/-}$ or *Mlc2a*$^{-/-}$ genetic background was considered an additional setback. A pharmacological approach was therefore the most feasible given the limited time and resources available.

Dofetilide is a class III anti-arrhythmic drug that has been demonstrated to induce bradycardia and secondary hypoxia in rodent embryos through inhibition of the rapid component of the delayed rectifying potassium current (I_{Kr}). Unlike the human adult heart, the rodent adult hearts are not dependent on I_{Kr} for their function. Dofetilide administration to pregnant mice thus only affects the embryos', but not the mother's cardiac function[421]. In rat embryos, administration of 2.5mg/kg or 4mg/kg of Dofetilide on E11 caused a 60% reduction in heart rate two hours post-administration, as measured by Doppler ultrasound. At 24 hours post-administration, the embryonic heart rate was still reduced by 26% following the administration of 2.5mg/kg dose, or by 51% following the administration of the 4mg/kg dose. The 4mg/kg dose was lethal for 19% of rat embryos within each litter, whereas the low dose was lethal only for 2.7% of embryos within each litter[422].

A Dofetilide fetotoxicity study carried out in mice divided 20 pregnant females into three treatment groups, each dosed with either 0.5mg/kg, 2mg/kg, or 5mg/kg Dofetilide once a day between E6 to E13. The control group was administered with vehicle on the same schedule. Both the drug and the vehicle were administered by oral gavage, and embryos were dissected at E18. In the 5mg/kg dose

group, all litters were reabsorbed by pregnant females, while a 2mg/kg dose was lethal for 30.1% of all embryos analysed. Both the 0.5mg/kg dose and the vehicle were lethal for 10% of all embryos within the litters analysed. There was no maternal toxicity associated with any of the doses tested. The 2mg/kg dose was associated with increased vertebral and sternebral abnormalities compared to the low dose and control groups at E18[423].

Due to the lack of a comprehensive study that would measure heart rate in response to Dofetilide treatment in mouse embryos, it was difficult to estimate exactly what dose would result in a significant reduction in heart rate and therefore blood flow. A single 2mg/kg dose of Dofetilide administered at E9.5 was shown to induce heart defects in 10% of treated embryos at E17.5, while the baseline rate of heart defects observed in untreated embryos of the same strain was approximately 5%[424]. Taken together with the results of the fetotoxicity study described above, 2mg/kg was taken as a starting point that compared to the 2.5mg/kg dose that induced a 26% decrease in heart rate of rat embryos 24h post-administration. A higher dose might have been preferable in terms of long-term heart rate reduction that does not decrease significantly after 24h (only from 60% to 51% in rats dosed with 4mg/kg). However, increasing the dose in mice would likely lead to more litter reabsorption, depending on the additive effects of the drug. With 2mg/kg Dofetilide dosing regimen for five consecutive days (E6.5-E10.5), embryonic lethality up to 30% was expected based on previous studies, however, no maternal toxicity was expected to occur.

To avoid some variation in transgene expression even within a single transgenic line, where transgene expression was compared between different litters, the same stud male carrying the transgene was used for timed mating. Additionally, to account for two other major sources of variability in transgene expression observed, all litters within a comparison group were pre-fixed and stained with X-Gal for the same number of hours. Where embryos were developmentally delayed, they were pre-fixed for the amount of time equivalent to that of an untreated embryo of the same developmental stage.

5.3.2 KLF2-41:LacZ expression in Dofetilide-treated embryos

For the pilot experiment, a dose of 2mg/kg was administered by oral gavage to pregnant females once daily from E6.5 to E10.5. The first administration at E6.5 was chosen because this time point precedes the first cardiac contractions in mouse development[425,426]. The end point of E10.5 was chosen as this is a time point during cardiovascular development where the vasculature is fully perfused[427]. In this initial experiment, one pregnant female was treated with a vehicle control on the same schedule while a second pregnant female which was treated with Dofetilide.

In the KLF2-41:*LacZ* litter from the vehicle-treated female, three out of four embryos exhibited robust KLF2-41:*LacZ* expression (**Figure 5.8**) consistent with the previously characterised expression pattern (Chapter 3, Section 3.6.3). However, one embryo from this litter showed markedly reduced transgene expression (**Figure 5.8b**). Such internal variation had been observed during the characterisation of this stable transgenic line and could not be circumvented by testing different stud males (data not shown).

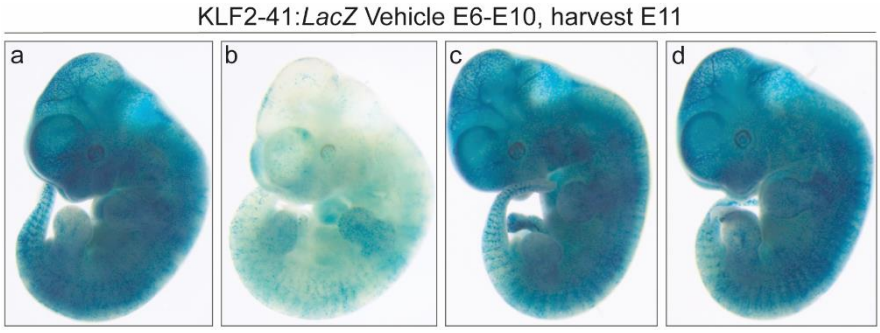


Figure 5.8: Analysis of KLF2-41:*LacZ* transgene expression in embryos treated with vehicle once a day between E6 and E10, and harvested at E11. All *LacZ* expressing embryos from the litter are shown. Embryos a, c, and d exhibited comparable transgene expression, whereas embryo b showed more limited, though still EC-specific, transgene expression pattern and intensity.

Although the KLF2-41:*LacZ* transgene expression in the vehicle-treated control embryos was expected to recapitulate KLF2-41:*LacZ* expression at the same timepoint in embryos that have received no vehicle or drug treatment (Chapter 3, Section 3.6.3), a representative vehicle-treated control embryo was still sectioned to characterise KLF2-41:*LacZ* transgene expression for the purpose of comparison

in *LacZ* expression between vehicle-treated and Dofetilide-treated embryos. In the control embryo (**Figure 5.9**), KLF2-41:*LacZ* expression was detected in the endothelium of head arteries, capillaries, and the angiogenic hindbrain, but not in the endothelium of head veins (**Figure 5.9b** and **c**). Robust KLF2-41:*LacZ* expression was also detected in the OT and AVC, and the remainder of the endocardium (**Figure 5.9d** and **e**). In line with previous analyses, enhancer activity was present within the endothelium of the dorsal aorta, but not within the cardinal vein (**Figure 5.9e**). Finally, expression was observed within the abdominal aorta and the angiogenic vessels surrounding the neural tube (**Figure 5.9g**). In conclusion, KLF2-41:*LacZ* transgene expression in the E11.5 embryo treated with vehicle once daily between E6 and E10 recapitulated KLF2-41:*LacZ* transgene expression observed in the WT E11.5 embryo characterised in Chapter 3, Section 3.6.3.

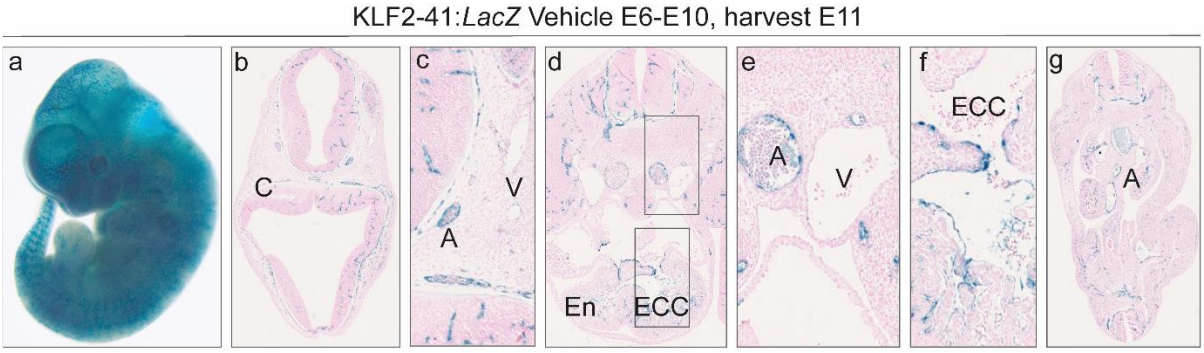


Figure 5.9: Section analysis of KLF2-41:*LacZ* transgene expression in vehicle-treated control embryo. (a) Whole-mount vehicle-treated control embryo. **(b)** *LacZ* expression in detected in the head vessels, including the endothelium of small diameter arteries and capillaries, but not in veins. **(c)** Close-up of *LacZ* expression in head arteries. **(d)** KLF2-41:*LacZ* transgene is robustly expressed within the ECCs of the vehicle-treated embryo. It is also expressed within the endothelium of the dorsal aorta, but not within the endothelium of the cardinal vein. **(e)** Close-up of *LacZ* expression in the endothelium of the dorsal aorta, and absence of expression in the endothelium of the cardinal vein. **(f)** Close-up of *LacZ* expression within the ECCs and the remainder of the endocardium. **(g)** *LacZ* expression is detected in large abdominal vessels and capillaries. **A=artery, V=vein, ECC=endocardial cushion, En=endocardium.**

In the KLF2-41:*LacZ* litter dissected at E11.5 from a female administered with 2mg/kg Dofetilide once daily between E6.5 and E10.5, a variety of Dofetilide-induced phenotypes were observed (**Figure 5.10**). Embryo e appeared phenotypically normal and expressed the KLF2-41:*LacZ* transgene levels comparable to that of the vehicle-treated controls upon whole-mount analysis (**Figure 5.10e**). Embryo

d appeared developmentally delayed compared to e, and exhibited weaker KLF2-41:*LacZ* transgene expression (**Figure 5.10d**), as well as some ectopic transgene expression not observed with vehicle controls. The remaining three *LacZ* positive embryos (**Figure 5.10 a, b, and c**) were severely developmentally delayed, with dramatically lower levels of KLF2-41:*LacZ* expression compared to both embryos d and e from the same litter, as well as compared to vehicle-treated control embryos. While some of this may have been explained by developmental delays, these embryos exhibited less transgene expression than seen in the untreated E9.5 KLF2-41:*LacZ* embryos analysed in Chapter 3 (Section 3.6.3), which may better match their developmental stage/size.

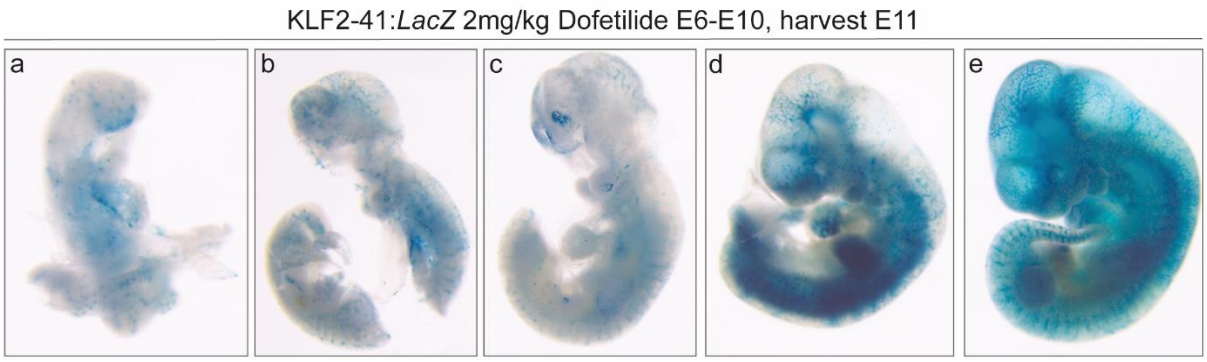


Figure 5.10: Analysis of KLF2-41:*LacZ* transgene expression in embryos treated with 2mg/kg Dofetilide once a day between E6 and E10, and harvested at E11. All *LacZ* expressing embryos from the litter are shown. Different embryos from the litter exhibited a variety of Dofetilide treatment-induced phenotypes, ranging from seemingly unaffected (embryo e) to developmentally delayed (embryo d) and severely developmentally delayed (embryos a, b, c).

To further analyse KLF2-41:*LacZ* expression in each of the phenotypes observed, one embryo representing each phenotype was sectioned, and the sections were analysed as described in previous chapters. Analysis of embryo e, which exhibited no obvious developmental delay, showed transgene expression largely consistent with vehicle-treated E11.5 KLF2-41:*LacZ* embryos (**Figure 5.9**). Robust KLF2-41:*LacZ* expression was observed in head capillary and arterial ECs, although punctate expression was also observed in venous ECs, which was not present in vehicle-treated control embryos (**Figure 5.11b and c**). Transgene expression was also detected within the OT and AVC and the remainder of the endocardium, consistent with transgene expression in the vehicle control (**Figure 5.11 d and f**).

Transgene expression was observed in the endothelium of the dorsal aorta, but not the endothelium of the cardinal vein (**Figure 5.11e**). Transgene expression was also detected in abdominal arteries and the angiogenic limb bud endothelium (**Figure 5.11g**).

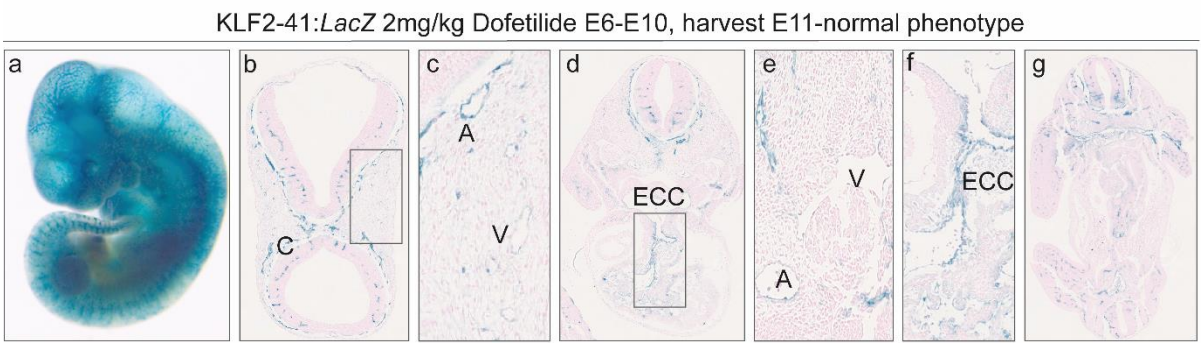


Figure 5.11: Section analysis of KLF2-41:*LacZ* transgene expression in Dofetilide-treated embryo exhibiting no developmental phenotype. (a) Whole-mount Dofetilide-treated embryo. **(b)** *LacZ* expression in detected in the head vessels, including the endothelium of large vessels and capillaries. **(c)** Close-up of *LacZ* expression in head vessels showing robust *LacZ* expression within the endothelium of arteries, and to a lesser extent in endothelium of head veins. **(d)** KLF2-41:*LacZ* transgene is robustly expressed within the ECCs of the Dofetilide-treated embryo, and in large blood vessels surrounding the heart. **(e)** Close-up of *LacZ* expression in the endothelium of the dorsal aorta, and punctate expression in the endothelium of the cardinal vein. **(f)** Close-up of *LacZ* expression within the ECCs. **(g)** *LacZ* expression is detected in large abdominal vessels and capillaries. **A=artery, V=vein, ECC=endocardial cushion.**

The analysis of embryo d, which exhibited some developmental delay and reduced KLF2-41:*LacZ* expression, showed that some aspects of KLF2-41:*LacZ* transgene expression were consistent with the control E11.5 KLF2-41:*LacZ* embryo, but there were notable deviations. Most importantly, the KLF2-41:*LacZ* transgene expression was overall less continuous within the endothelial layer, and was not present in all structures of the same type, for example not all arteries within the head were *LacZ* positive. Transgene expression was also reduced in head capillary ECs, and almost no expression was observed in the angiogenic hindbrain, in stark contrast to robust transgene expression detected in the angiogenic hindbrain of vehicle-treated embryos. Transgene expression was detected in the endothelium of head arteries, although this was continuous in some, punctate in others, and completely absent from certain arteries. No transgene expression was detected in the endothelium of head veins. Ectopic *LacZ* expression that did not seem to correspond to ECs was also observed in the

head sections, for example within the hindbrain (**Figure 5.12b** and **c**). Punctate transgene expression was observed in the endothelium of the dorsal aorta, while no expression was observed in the endothelium of the cardinal vein (**Figure 5.12e**). Transgene expression was also detected in individual ECs within the abdominal arteries, and the angiogenic limb bud endothelium, where transgene expression was particularly of particularly strong intensity (**Figure 5.11g**).

Within the ECCs, transgene expression was present, but was more punctate compared to the vehicle control, and was consistently observed within the ECC tissue in addition to the endothelium lining the cushion leaflets (**Figure 5.12d** and **f**). The same expansion of transgene expression from the leaflet lining to the interior of the cushion was never observed with WT or vehicle-treated embryos. The punctate expression of the transgene within the cushion tissue might correspond to the invading EC which undergo EMT to line the AVC leaflets[428]. This result suggests that Dofetilide treatment, and by extension a reduction in shear stress levels, might lead to the loss of restriction of KLF2-41:*LacZ* transgene expression from the endothelial lining of the ECC. This interpretation of results is consistent with the evidence demonstrating that the change in hemodynamic load alters EMT during ECC formation [429].

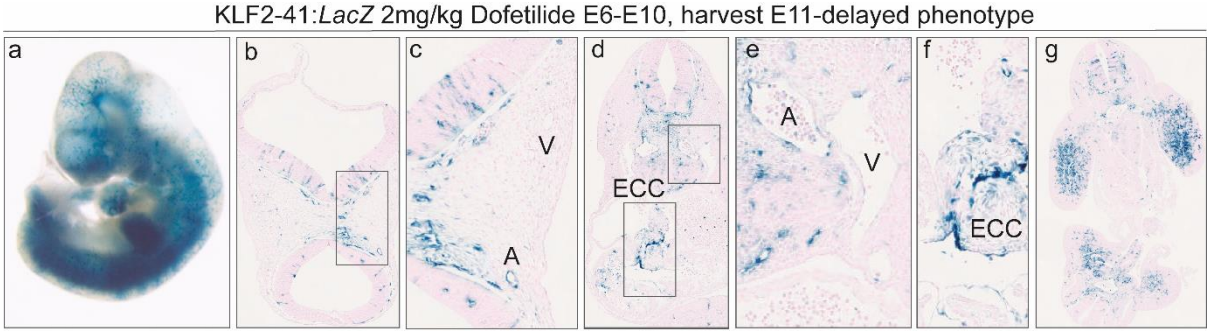


Figure 5.12: Section analysis of KLF2-41:*LacZ* transgene expression in Dofetilide-treated embryo exhibiting some developmental delay. (a) Whole-mount Dofetilide-treated embryo. **(b)** Some *LacZ* expression in detected within the endothelium of some, though not all, head arteries. *LacZ* expression was also detected in the endothelium of some capillaries but was more punctate and less consistent compared to the vehicle control embryo. No *LacZ* expression was observed in the venous endothelium. **(c)** Close-up of *LacZ* expression in head vessels showing robust *LacZ* expression within the endothelium of some arteries, and punctate expression in capillaries. Ectopic *LacZ* expression in the surrounding tissue was also observed. **(d)** Robust and punctate KLF2-41:*LacZ* transgene is observed within the ECCs of the Dofetilide-treated embryo. Punctate *LacZ* expression is also observed within the endothelium of the dorsal aorta. No *LacZ* expression is detected within the cardinal vein endothelium. **(e)** Close-up of punctate *LacZ* expression in the endothelium of the dorsal aorta. **(f)** Close-up

of robust and punctate *LacZ* expression within the ECCs. **(g)** Punctate ectopic *LacZ* expression is detected within limb bud capillaries. **A=artery, V=vein, ECC=endocardial cushion.**

Within the sections of embryo c, which exhibited severe developmental delay that seemed to correspond to E9.5 stage of development, KLF2-41:*LacZ* transgene expression was difficult to assess due to severe Dofetilide-induced malformations, including cranio-facial, tail, and limb developmental defects, and potentially early stages of tissue decay (**Figure 5.13**). Punctate KLF2-41:*LacZ* transgene expression was observed within different parts of the embryo, but due to lack of proper development of most vascular structures it was impossible to say whether staining corresponded to the endothelium. While the heart was formed, the ECCs did not exhibit their characteristic morphology, and no transgene expression was detected in the area where they were expected to form (**Figure 5.13e**).

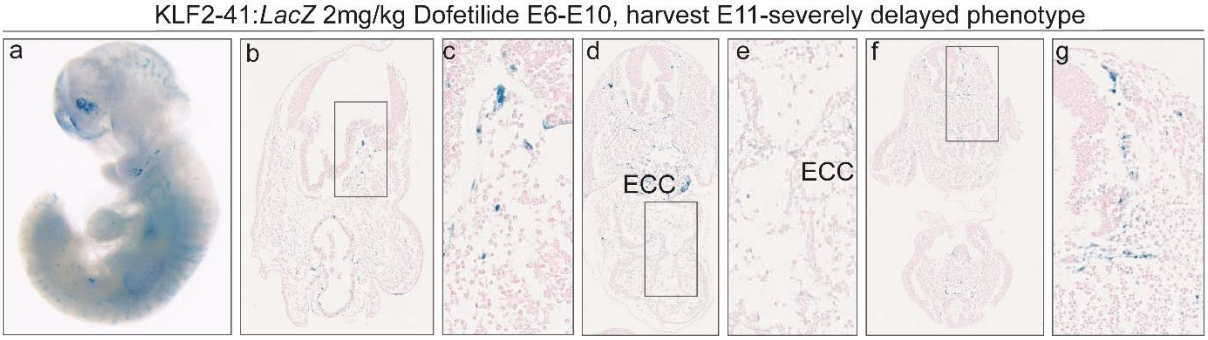


Figure 5.13: Section analysis of KLF2-41:*LacZ* transgene expression in Dofetilide-treated embryo exhibiting severe developmental delay. (a) Whole-mount Dofetilide-treated embryo. **(b)** Some *LacZ* expression in detected within the head of the embryo, though it is impossible to relate them to blood vessels due to severe defects in blood vessel formation observed in this embryo. **(c)** Close-up of *LacZ* expression in the head. **(d)** Weak *LacZ* expression is detected in the ECC area. **(e)** Close-up of ECC area with weak *LacZ* expression. **(f)** Some *LacZ* expression is detected in the area surrounding the neural tube normally rich in capillaries. **(g)** Close up of *LacZ* positive areas surrounding the neural tube. **ECC=endocardial cushion.**

In conclusion, administration of Dofetilide to mouse embryos between E6 and E10 led to a reduction of KLF2-41:*LacZ* transgene expression that was proportional to the severity of the developmental phenotype induced by Dofetilide. In the embryo exhibiting some developmental delay, the KLF2-41:*LacZ* transgene expression was specifically reduced and punctate within capillary and artery ECs, with expansion of *LacZ* expression from the surface of ECCs to surrounding non-endothelial tissue.

These results support the hypothesis stating that *LacZ* expression driven by the KLF2-41 enhancer is regulated by shear stress.

5.3.3 KLF2-66:LacZ expression in Dofetilide-treated embryos

The same pilot experiment was repeated with the KLF2-66:*LacZ* stable line. Pregnant females were administered either vehicle, or 2mg/kg Dofetilide, once daily between E6 and E10, and were harvested at E11. In the KLF2-66:*LacZ* litter dissected form the vehicle-treated female, all three *LacZ* positive embryos exhibited robust transgene expression that was consistent between the embryos (**Figure 5.14**) and consistent with the expression of this transgene characterised in Results Chapter 3 (Section 3.7.2).

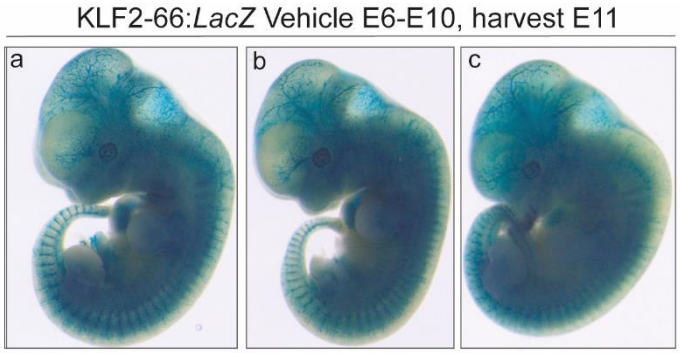


Figure 5.14: Analysis of KLF2-66:*LacZ* transgene expression in embryos treated with vehicle once a day between E6 and E10, and harvested at E11. All *LacZ* expressing embryos from the litter are shown. All three embryos exhibited comparable transgene expression levels.

As with KLF2-41:*LacZ*, KLF2-66:*LacZ* expression in the vehicle-treated control embryos was expected to recapitulate KLF2-66:*LacZ* expression at the same timepoint in untreated embryos. Despite this, a representative vehicle-treated control embryo was still sectioned to characterise KLF2-66:*LacZ* transgene expression for the purpose of comparison in *LacZ* expression between vehicle-treated and Dofetilide-treated embryos. In the representative embryo (**Figure 5.15**), KLF2-66:*LacZ* expression was detected in the endothelium of head arteries, and to a lesser extent head veins. Robust expression was detected in head capillaries, and within the angiogenic hindbrain (**Figure 5.15b** and **c**). Robust KLF2-66:*LacZ* expression was also detected in the OT and AVC, and the remainder of the endocardium

of the heart (**Figure 5.15d** and **e**). In line with previous analyses, KLF2-66:*LacZ* transgene expression was present within the endothelium of the dorsal aorta, and to a lesser extent within the cardinal vein (**Figure 5.15e**). In conclusion, KLF2-66:*LacZ* transgene expression in the E11.5 embryo treated with vehicle once daily between E6 and E10 recapitulated KLF2-66:*LacZ* transgene expression observed in the WT E11.5 embryo characterised in Chapter 3, Section 3.7.2.

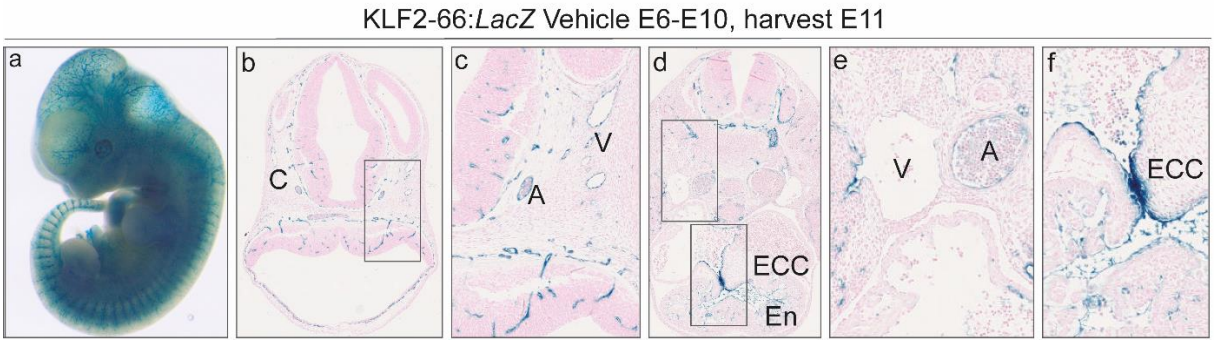


Figure 5.15: Section analysis of KLF2-66:*LacZ* transgene expression in vehicle-treated control embryo. (a) Whole-mount vehicle-treated control embryo. **(b)** *LacZ* expression in detected in the head vessels, including the endothelium of small diameter arteries and capillaries, and to a lesser extent in the endothelium of head veins. **(c)** Close-up of *LacZ* expression in head arteries, veins, and capillaries. **(d)** KLF2-66:*LacZ* transgene is robustly expressed within the ECCs of the vehicle-treated embryo. It is also expressed within the endothelium of the dorsal aorta, and to a lesser extent within the endothelium of the cardinal vein. **(e)** Close-up of *LacZ* expression in the endothelium of the dorsal aorta, and punctate expression in the endothelium of the cardinal vein. **(f)** Close-up of *LacZ* expression within the ECCs and the remainder of the endocardium of the heart. **A=artery, V=vein, ECC=endocardial cushion, En=endocardium.**

In the KLF2-66:*LacZ* litter dissected at E11.5 from a female administered with 2mg/kg Dofetilide once daily between E6 and E10, a variety of Dofetilide-induced phenotypes were observed within the litter, just as with KLF2-41:*LacZ*. Embryo c appeared phenotypically normal and expressed the KLF2-66:*LacZ* transgene levels comparable to that of the vehicle-treated controls upon whole-mount analysis (**Figure 5.16c**). Embryo b appeared slightly developmentally delayed compared to embryo c, and exhibited weaker KLF2-66:*LacZ* transgene expression (**Figure 5.16b**). Embryo a was severely developmentally delayed, with dramatically lower levels of KLF2-66:*LacZ* expression compared to both embryos b and c from the same litter, as well as compared to vehicle-treated control embryos. This embryo also exhibited less KLF2-66:*LacZ* transgene expression compared to E9.5 embryos transgenic for the KLF2-66:*LacZ* construct analysed in Chapter 3 (Section 3.7.2), which matches the stage of their

developmental delay. In whole-mount analysis, ectopic *LacZ* expression was observed, for example at the tip of the tail.

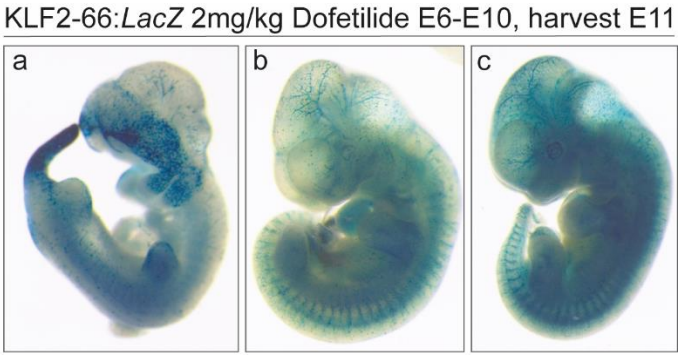


Figure 5.16: Analysis of KLF2-66:*LacZ* transgene expression in embryos treated with 2mg/kg Dofetilide once a day between E6 and E10, and harvested at E11. All *LacZ* expressing embryos from the litter are shown. Different embryos from the litter exhibited a variety of Dofetilide treatment-induced phenotypes, ranging from seemingly unaffected (embryo c) to developmentally delayed (embryo b) and severely developmentally delayed (embryo a).

Each of the embryos shown in **Figure 5.16** was sectioned and analysed as described in the previous chapters. Embryo c, which exhibited no obvious developmental delay, showed KLF2-66:*LacZ* transgene expression that was largely consistent with vehicle E11.5 KLF2-66:*LacZ* embryos: robust expression was observed in head capillary, arterial, and vein ECs, the OT, AVC, and the remainder of the endocardium (**Figure 5.17b** and **c**). Transgene expression was also observed in the endothelium of the dorsal aorta, but not the endothelium of the cardinal vein (**Figure 5.17e**).

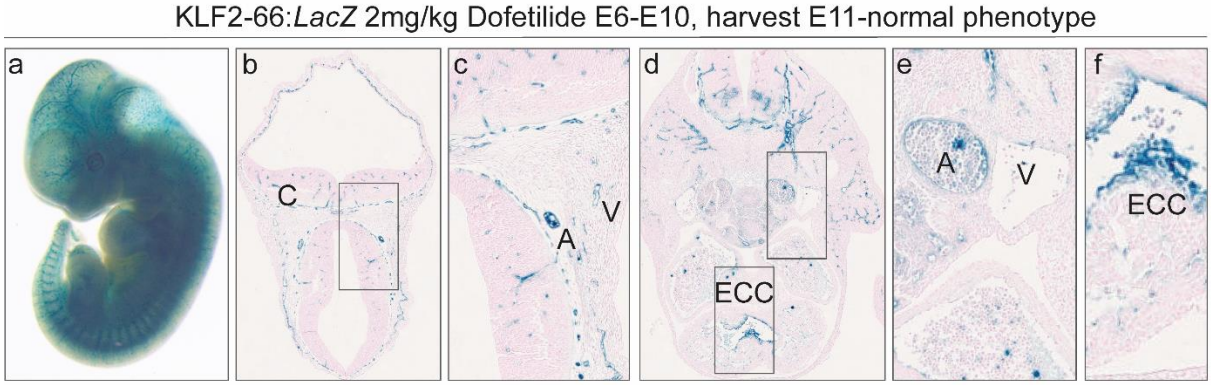


Figure 5.17: Section analysis of KLF2-66:*LacZ* transgene expression in Dofetilide-treated embryo exhibiting no developmental phenotype. (a) Whole-mount Dofetilide-treated embryo. **(b)** *LacZ* expression in detected in the head vessels, including the endothelium of large vessels and capillaries. **(c)** Close-up of *LacZ* expression in head vessels showing robust *LacZ* expression within the endothelium of arteries, and to a lesser extent in endothelium

of head veins. **(d)** KLF2-66:*LacZ* transgene is robustly expressed within the ECCs of the Dofetilide-treated embryo, and in large blood vessels surrounding the heart. **(e)** Close-up of *LacZ* expression in the endothelium of the dorsal aorta, and punctate expression in the endothelium of the cardinal vein. **(f)** Close-up of *LacZ* expression within the ECCs. **A=artery, V=vein, ECC=endocardial cushion.**

The analysis of embryo b, which exhibited some developmental delay and reduced KLF2-66:*LacZ* expression, shared some aspects of KLF2-66:*LacZ* transgene expression consistent with control and WT E11.5 KLF2-66:*LacZ* embryos, but its expression was less continuous within the endothelial layer, and overall weaker, particularly in small diameter vessels, including capillaries (**Figure 5.18a-c**). There was only limited expression observed in the angiogenic hindbrain, although transgene expression in the endothelium of head veins and in the dorsal aorta was consistent with the control. Within the ECCs, robust transgene expression was present, but appeared more punctate and less specific to the ECC surface (**Figure 5.18d** and **f**). Transgene expression was however detected in the remainder of the endocardium of the heart in a pattern similar to vehicle control (**Figure 5.18f**).

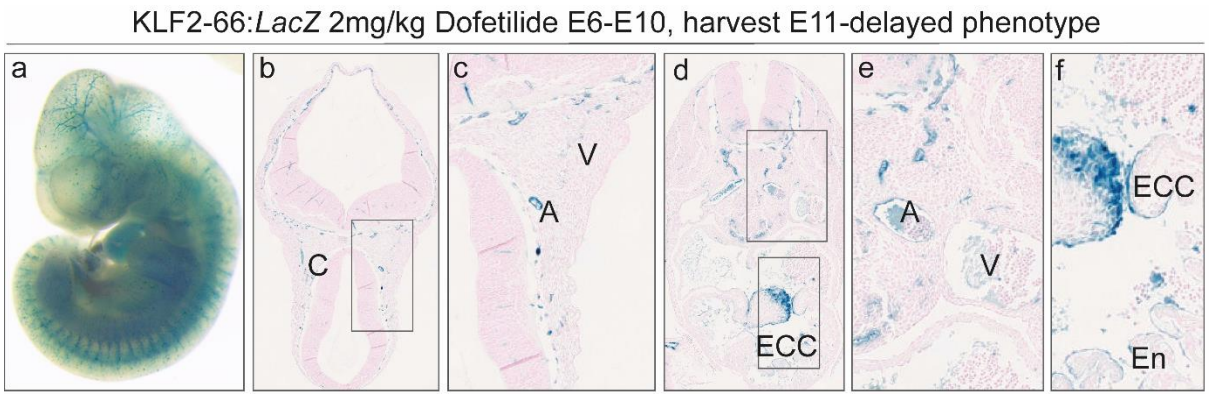


Figure 5.18: Section analysis of KLF2-66:*LacZ* transgene expression in Dofetilide-treated embryo exhibiting some developmental delay. (a) Whole-mount Dofetilide-treated embryo. **(b)** Robust *LacZ* expression in detected within the endothelium of most head arteries. *LacZ* expression was also detected in the endothelium of some capillaries, but less consistent compared to the vehicle control embryo. Punctate *LacZ* expression was observed in the venous endothelium of the head. **(c)** Close-up of *LacZ* expression in head vessels showing robust *LacZ* expression within the endothelium of some arteries, and punctate expression in capillaries. Some ectopic *LacZ* expression in the surrounding tissue was observed. **(d)** Robust, though punctate KLF2-66:*LacZ* transgene is observed within the ECCs of the Dofetilide-treated embryo. Punctate *LacZ* expression is also observed within the endothelium of the dorsal aorta. Low level *LacZ* expression is detected within the cardinal vein endothelium. **(e)** Close-up of punctate *LacZ* expression in the endothelium of the dorsal aorta and the cardinal vein. **(f)** Close-up of robust and punctate *LacZ* expression within the ECCs and the remainder of the endocardium of the heart. **A=artery, V=vein, ECC=endocardial cushion.**

Within the sections of embryo a, which exhibited severe developmental delay that seemed to correspond to E9.5 stage of development, KLF2-66:*LacZ* transgene expression was difficult to assess due to severe Dofetilide-induced malformations observed in almost all embryonic structures, as with the KLF2-41:*LacZ* transgene (**Figure 5.19**). Punctate KLF2-66:*LacZ* transgene expression was notably ectopically within the embryonic wall, but due to lack of proper development or tissue decay of most vascular structures it was impossible to say whether staining corresponded to the endothelium. Some transgene expression was detected within the area where the ECCs are expected to develop, although this was difficult to say due to the lack of their characteristic shape and issue with heart development (**Figure 5.19e**).

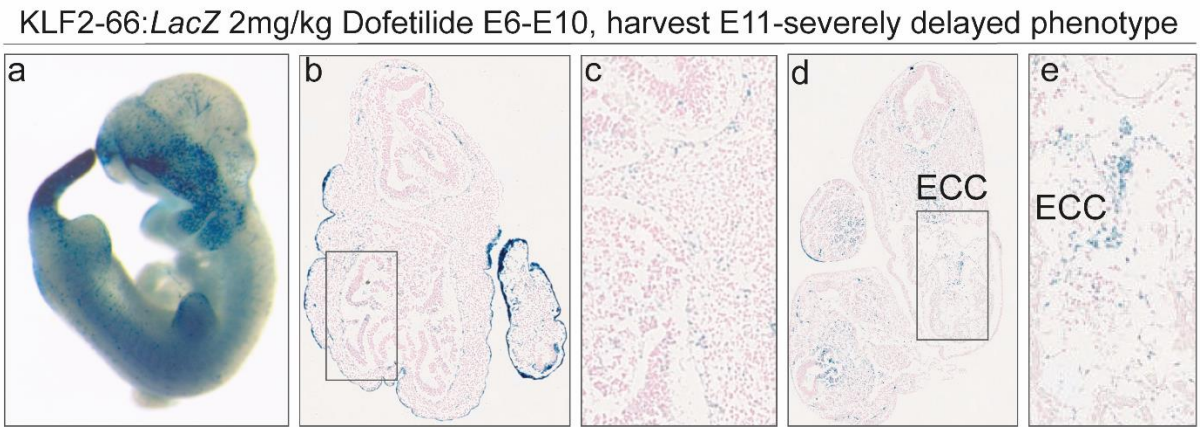


Figure 5.19: Section analysis of KLF2-66:*LacZ* transgene expression in Dofetilide-treated embryo exhibiting severe developmental delay. (a) Whole-mount Dofetilide-treated embryo. **(b)** Some *LacZ* expression in detected within the head of the embryo, though it is impossible to relate them to blood vessels due to severe defects in blood vessel formation observed in this embryo. **(c)** Close-up of *LacZ* expression in the head. **(d)** Weak *LacZ* expression is detected in the ECC area. **(e)** Close-up of ECC area with weak *LacZ* expression. **ECC=endocardial cushion.**

In conclusion, administration of Dofetilide to mouse embryos between E6 and E10 led to a reduction of KLF2-66:*LacZ* transgene expression that was proportional to the severity of the developmental phenotype induced by Dofetilide, similar to the result observed with KLF2-41:*LacZ* embryos that received the same treatment. In the embryo exhibiting some developmental delay, the KLF2-66:*LacZ* transgene expression was specifically reduced and punctate within capillary and artery ECs, with expansion of *LacZ* expression from the surface of ECCs to surrounding non-endothelial tissue. These

results support the hypothesis stating that *LacZ* expression driven by the KLF2-66 enhancer is regulated by shear stress.

Due to the variability of KLF2-41:*LacZ* expression within embryos of the same litter, and the resultant difficulty in interpreting whether the reduction of transgene expression is due to Dofetilide treatment or baseline expression variability, further dose adjustment experiments were carried out solely with the KLF2-66:*LacZ* transgenic line.

In order to avoid obtaining a severe developmental delay phenotype affecting the morphogenesis of the whole embryo due to teratogenic effects of Dofetilide, the first Dofetilide administration was moved from E6 to E7. Four 2mg/kg doses of Dofetilide were administered between E7 and E10, twice a day. Vehicle treated control litter followed the same schedule with vehicle administration. The control litter was administered with two doses of vehicle control/day between E7 and E10 and appeared phenotypically normal, with KLF2-66:*LacZ* expression patterns consistent with previous analysis (Chapter 3, Section 3.7.2) (**Figure 5.20**). In order to allow for detection of subtle differences in KLF2-66:*LacZ* expression, this litter was stained for 8h at 37°C, which was 4h less than the control litter in the first pilot experiment. As with the KLF2-41:*LacZ* transgenic line, all litters within a comparison group were pre-fixed and stained with X-Gal for the same number of hours. Where embryos were developmentally delayed, they were pre-fixed for the amount of time equivalent to that of an untreated embryo of the same developmental stage.

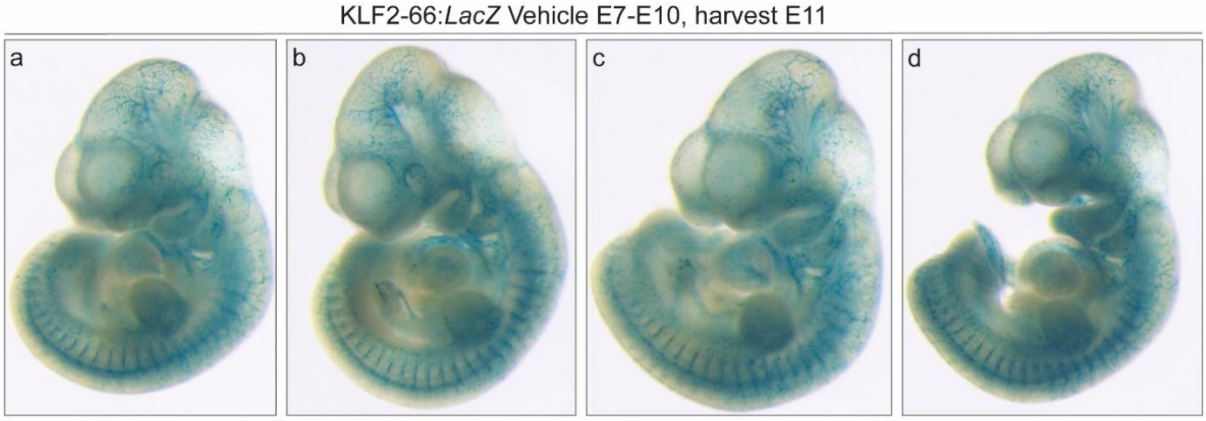


Figure 5.20: Analysis of KLF2-66:*LacZ* transgene expression in embryos treated with vehicle twice a day between E7 and E10, and harvested at E11. All *LacZ* expressing embryos from the litter are shown. All four embryos exhibited comparable transgene expression levels.

Due to a change in vehicle and Dofetilide administration schedule, and a change in the staining time, a representative vehicle-treated control KLF2-66:*LacZ* embryo was sectioned and its transgene expression assessed, although it was expected to recapitulate the KLF2-66:*LacZ* expression at the same timepoint in other vehicle treated embryos (Section 5.3.3). In the representative embryo, KLF2-66:*LacZ* transgene expression was detected in the endothelium of head arteries, and to a lesser extent head veins, in head capillaries and the angiogenic hindbrain. Robust KLF2-66:*LacZ* expression was also detected in the OT and AVC ECCs, the remainder of the endocardium, and in the endothelium of the dorsal aorta, and to a lesser extent within the cardinal vein (**Figure 5.21**). In conclusion, KLF2-66:*LacZ* transgene expression in the E11.5 embryo treated with vehicle twice daily between E7 and E10 recapitulated KLF2-66:*LacZ* transgene expression observed in the WT E11.5 embryo characterised in Chapter 3, Section 3.7.2, and vehicle treated embryo characterised in Section 5.3.3.

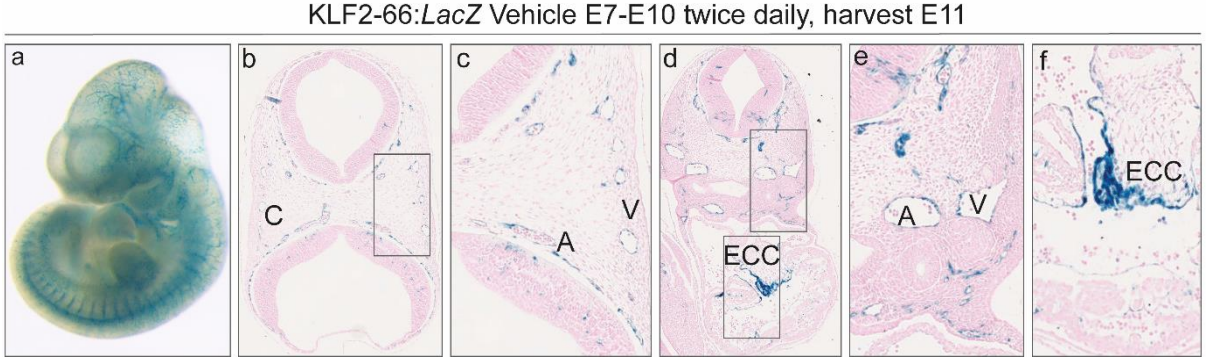


Figure 5.21: Section analysis of KLF2-66:*LacZ* transgene expression in vehicle-treated control embryo. (a) Whole-mount vehicle-treated control embryo. **(b)** *LacZ* expression in detected in the head vessels, including the

endothelium of small diameter arteries and capillaries, and to a lesser extent in the endothelium of head veins. **(c)** Close-up of *LacZ* expression in head arteries, veins, and capillaries. **(d)** KLF2-66:*LacZ* transgene is robustly expressed within the ECCs of the vehicle-treated embryo. It is also expressed within the endothelium of the dorsal aorta, and to a lesser extent within the endothelium of the cardinal vein. **(e)** Close-up of *LacZ* expression in the endothelium of the dorsal aorta, and punctate expression in the endothelium of the cardinal vein. **(f)** Close-up of *LacZ* expression within the ECCs and the remainder of the endocardium of the heart. **A=artery, V=vein, ECC=endocardial cushion, En=endocardium.**

Within the Dofetilide-treated group, three litters were analysed (**Figure 5.22**). In all of the transgenic expressing embryos within each litter, severe developmental delay was observed, and the embryonic stages of these embryos corresponded to anywhere between E9 to E10. In addition, many of these embryos appeared dead, suggesting 2 mg/kg administration of Dofetilide twice a day between E7 and E10 was lethal for these embryos. Embryonic malformations observed in these litters ranged from cranio-facial malformations (**Figure 5.22d, e, h, i, o**), complete lack of limbs (**Figure 5.22e, j, l, m, n, p**), and tail development defects which were observed in every embryo. However, the heart and blood vessels still developed at least to some extent, and *LacZ* expression was detectable in these embryos. Whole-mount and section analysis of KLF2-66:*LacZ* transgene expression was therefore carried out.

In embryos where blood vessels were clearly visible (**Figure 5.22b, c, f, g, k, m, o**), transgene expression appeared either reduced or entirely absent from some blood vessels, most notably capillaries, large vessels of the head and intersegmental vessels throughout the embryo, compared to vehicle control embryos (**Figure 5.19b, c, f, k, o**). In some embryos, *LacZ* positive the dorsal aortae and the cardinal veins were visible (**Figure 5.22c, g, m**), and in some embryos, OT and AV ECC staining could be observed even in whole-mount analysis (**Figure 5.22b, c, g, j, l, o**). In addition, ectopic *LacZ* expression was observed in non-vascular tissues of many of these embryos, notably at the tip of the tail and within the limb buds (**Figure 5.22d, f, g, l, k, l**).

To investigate KLF2-66:*LacZ* expression in these embryos in more detail, three embryos with visible vascular *LacZ* expression were sectioned (**Figure 5.22c, m, and o**), and compared to *LacZ* expression within the vehicle-treated control embryo.

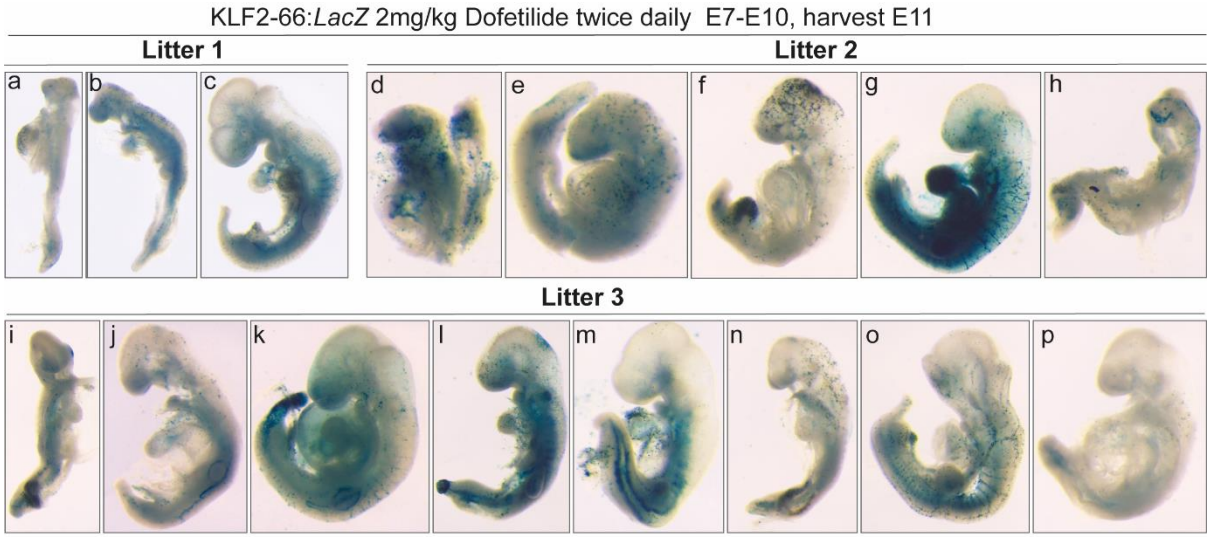


Figure 5.22: Analysis of KLF2-66:*LacZ* transgene expression in embryos treated with 2mg/kg Dofetilide twice a day between E7 and E10, and harvested at E11. All *LacZ* expressing embryos from three different litters using the same stud male are shown. Different embryos from different litters exhibited a range of Dofetilide treatment-induced phenotypes, though all were severely developmentally delayed, and most appeared dead.

Within the head of embryo c, *LacZ* expression was detected within arterial and vein endothelium, although the expression was more punctate compared to vehicle control. *LacZ* expression in the angiogenic hindbrain was barely present. Within the endothelium of the dorsal aorta and the cardinal vein, very limited *LacZ* expression was detected in single ECs (**Figure 5.23d** and **f).** In the angiogenic tissue surrounding the neural tube, some *LacZ* expression was detected, although this was less pronounced than that observed within the vehicle control. More *LacZ* expression was detected in the area where ECCs have formed (**Figure 5.23d** and **e**), although *LacZ* positive cells were dispersed throughout the cushion tissue and did not form a continuous layer on the surface as with the control embryo. Within the abdominal aorta, dispersed *LacZ* expression was observed in what appeared to be blood cells collecting in this area of the embryo, potentially suggesting that KLF2-66:*LacZ* transgene expression has expanded to blood cells following Dofetilide treatment.

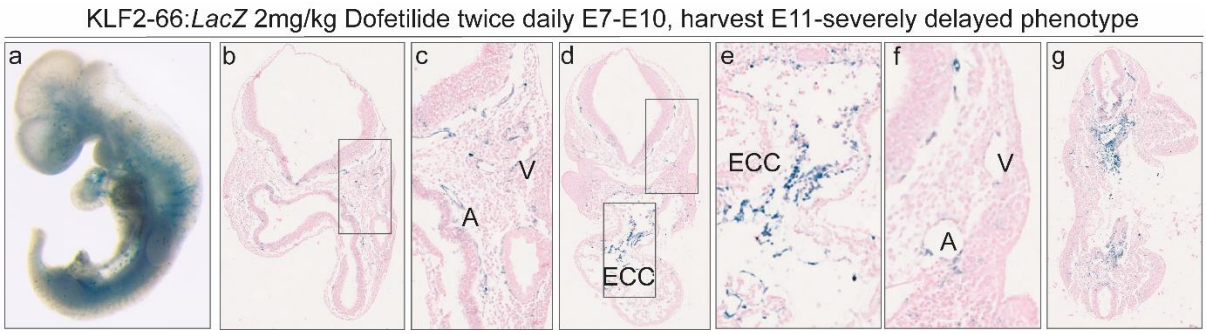


Figure 5.23: Section analysis of KLF2-66:*LacZ* transgene expression in Dofetilide-treated embryo exhibiting a severe developmental phenotype shows reduced transgene expression in the vascular endothelium and ECCs. **(a)** Whole-mount Dofetilide-treated embryo. **(b)** Punctate *LacZ* expression in detected within the endothelium of most head arteries. *LacZ* expression was also detected in the endothelium of some capillaries, but less consistent compared to the vehicle control embryo. Punctate *LacZ* expression was observed in the venous endothelium of the head. **(c)** Close-up of *LacZ* expression in head vessels showing robust *LacZ* expression within the endothelium of some arteries, and punctate expression in capillaries. Some ectopic *LacZ* expression in the surrounding tissue was observed. **(d)** Robust, though punctate KLF2-66:*LacZ* transgene expression is observed within the ECCs of the Dofetilide-treated embryo. Low level punctate *LacZ* expression is also observed within the endothelium of the dorsal aorta. Low level *LacZ* expression is detected within the cardinal vein endothelium. **(e)** Close-up of robust and punctate *LacZ* expression within the ECCs and the remainder of the endocardium of the heart. **(f)** Close-up of punctate *LacZ* expression in the endothelium of the dorsal aorta and the cardinal vein. **A=artery, V=vein, ECC=endocardial cushion.**

Next, the KLF2-66:*LacZ* expression was analysed in embryos m and o. In both embryos, *LacZ* expression was only detected within occasional ECs in head vessels, although it was difficult to define the vessels based on anatomical locations due to developmental delay (**Figures 5.24b, c** and **5.25b, c**). No *LacZ* expression was observed within the angiogenic vessels of the hindbrain. Within the endothelium of the dorsal aorta and the cardinal vein, very limited *LacZ* expression, if any, was detected in single ECs (**Figure 5.24** and **5.25d, f**), whereas more *LacZ* expression was detected within the aortic sac and the both the OT and AVC ECCs (**Figure 5.24** and **5.25d, e**), although *LacZ* positive cells were dispersed throughout the cushion tissue and did not form a continuous layer on the surface as with the vehicle control embryo. Some *LacZ* expression was observed within the endothelium tail vessels, although these vessels were difficult to identify, alongside dispersed *LacZ* expression in blood cells, again suggesting that KLF2-66:*LacZ* transgene expression has expanded to blood cells following Dofetilide treatment.

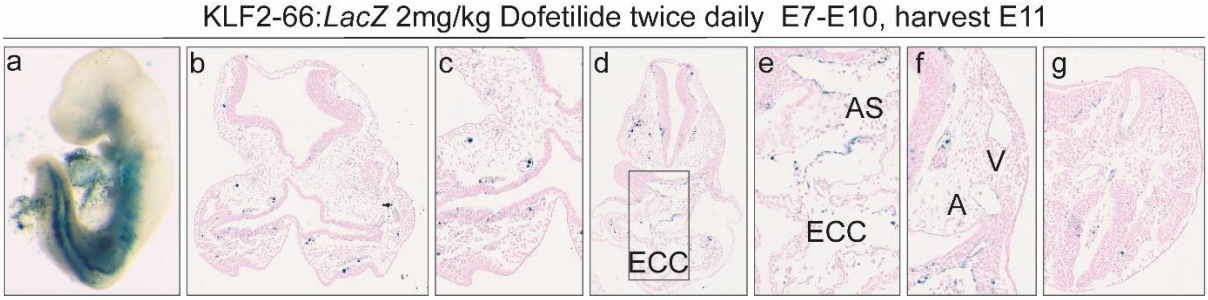


Figure 5.24: Section analysis of KLF2-66:*LacZ* transgene expression in Dofetilide-treated embryo exhibiting a severe developmental phenotype shows reduced transgene expression in the vascular endothelium and ECCs. **(a)** Whole-mount Dofetilide-treated embryo. **(b)** *LacZ* expression in detected in individual ECs within a minority of head vessels. *LacZ* expression was lacking from the endothelium of capillaries. **(c)** Close-up of *LacZ* expression in head vessels. Some ectopic *LacZ* expression in the blood cells was observed. **(d)** Low intensity KLF2-66:*LacZ* expression is observed within the aortic sac and the area where ECCs are expected to form. Barely any *LacZ* expression is also observed within the endothelium of the dorsal aorta and the cardinal vein. **(e)** Close-up of robust and punctate *LacZ* expression within the aortic sac, the ECCs, and the remainder of the endocardium of the heart. **(f)** Close-up of punctate *LacZ* expression in the endothelium of the dorsal aorta and the cardinal vein. **(g)** Some *LacZ* expression is detected within the endothelium of the tail vasculature. **A=artery, V=vein, ECC=endocardial cushion, AS=aortic sac.**

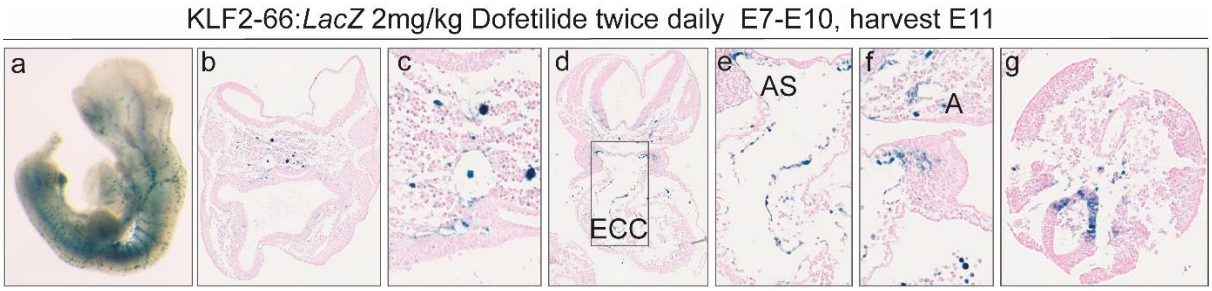


Figure 5.25: Section analysis of KLF2-66:*LacZ* transgene expression in Dofetilide-treated embryo exhibiting a severe developmental phenotype shows reduced transgene expression in the vascular endothelium and ECCs. **(a)** Whole-mount Dofetilide-treated embryo. **(b)** *LacZ* expression in detected in individual ECs within a minority of head vessels. *LacZ* expression was lacking from the endothelium of capillaries. **(c)** Close-up of *LacZ* expression in head vessels. Some ectopic *LacZ* expression in the blood cells was observed. **(d)** Low intensity KLF2-66:*LacZ* expression is observed within the aortic sac and the area where ECCs are expected to form. Barely any *LacZ* expression is also observed within the endothelium of the dorsal aorta and the cardinal vein. **(e)** Close-up of robust and punctate *LacZ* expression within the aortic sac, the ECCs, and the remainder of the endocardium of the heart. **(f)** Close-up of punctate *LacZ* expression in the endothelium of the dorsal aorta and the cardinal vein. **(g)** Some *LacZ* expression is detected within the endothelium of the tail vasculature. **A=artery, V=vein, ECC=endocardial cushion, AS=aortic sac.**

Taken together, administration of 2mg/kg Dofetilide twice daily between E7 and E10 reduced *LacZ* expression driven by the KLF2-66 enhancer. This dosing regimen led to severe developmental defects in all embryos within the three litters analysed, which complicated the analysis of KLF2-66:*LacZ* transgene expression due to the absence or malformation of many anatomical structures, including many blood vessels. This appeared lethal for most embryos analysed. Despite this, Dofetilide

treatment resulted in a clear reduction in KLF2-66:*LacZ* transgene expression within the endothelium of blood vessels and vascular structures that did form in Dofetilide-treated embryos. The expression of the transgene was also reduced within the ECCs of the embryos analysed. Future optimisation of the Dofetilide dosing regimen could potentially allow for the analysis of KLF2-66:*LacZ* transgene expression within a larger array of vascular structures.

5.4 Discussion

In this chapter, the analysis of KLF2 enhancer:*LacZ* transgene expression in response to changes in shear stress was first investigated in the post-natal and adult dorsal aorta, which is a stereotypical example of a blood vessel within which vessel geometry dictates distinct shear stress magnitudes and patterns experienced by ECs depending on their location. Following this analysis, pharmacological inhibition of cardiac contractions using a teratogenic class III anti-arrhythmic drug, Dofetilide, was used to establish the change in KLF2-66:*LacZ* and KLF2-41:*LacZ* transgene expression pattern in response to Dofetilide treatment.

5.4.1 KLF2-66:LacZ and KLF2-41:LacZ transgene expression corresponds to shear stress magnitudes in the post-natal dorsal aorta

Within the P8 post-natal dorsal aorta, the activity of both the KLF2-66 and KLF2-41 enhancers corresponded to shear stress magnitudes within the dorsal aorta. Both transgenes were robustly expressed within the descending aorta, where shear stress magnitude and *Klf2* mRNA expression are thought to be at their highest, with reporter gene intensity especially strong within the inter-costal arteries, which have a small diameter and therefore experience shear stress magnitude 3 or 4-fold higher compared to the descending aorta. Both the KLF2-41:*LacZ* and the KLF2-66:*LacZ* transgenes were also robustly expressed within the aortic arch, especially along the outer curvature of the aortic arch, and carotid arteries that emerge from the aortic arch. However, some reduction in the expression of both transgenes was observed within the inner curvature of the dorsal aorta, where ECs are thought to experience a lower shear stress magnitude compared to the descending aorta, and an oscillatory shear stress pattern, compared to laminar shear stress pattern experienced by descending aorta. It should however be noted that the difference in transgene expression intensity between the outer and inner curvature of the aortic arch was more pronounced with the KLF2-41:*LacZ* than the KLF2-66:*LacZ* transgene. Taken together with the analysis of KLF2-41:*LacZ* and the KLF2-66:*LacZ* transgene expression within post-natal cardiac valves, where *Klf2* mRNA is also robustly expressed, it

was concluded that *LacZ* expression driven by the KLF2-41 and KLF2-66 enhancers may correspond both to shear stress magnitude and pattern, as well as *Klf2* mRNA expression at P8. These results support the hypothesis that the KLF2-41 and KLF2-66 enhancers could regulate *KLF2* expression in response to shear stress.

5.4.2 KLF2-66:LacZ and KLF2-41:LacZ transgene expression is absent from the adult dorsal aorta

The analysis of KLF2-41:*LacZ* and the KLF2-66:*LacZ* transgene expression in the adult dorsal aorta, where shear stress levels are expected to be higher compared to the embryonic or the post-natal dorsal aorta, resulted in an unexpected absence of *LacZ* expression driven by both KLF2 enhancers. While some *LacZ* expression was dispersed throughout the dorsal aortae of each stable line, these *LacZ* positive cells did not seem to correspond to ECs. Further, this punctate expression was consistent between the descending aorta and the inner curvature of the aortic arch, suggesting this might be attributable to ectopic expression driven by the hsp68 promoter. Despite the lack of reporter expression within the descending aorta and the aortic arch, *LacZ* expression was observed within the endothelium of the inter-costal arteries that branch off the descending aorta. This was especially pronounced with the KLF2-66:*LacZ* transgene. Due to the inter-costal arteries experiencing a higher magnitude of shear stress compared to the dorsal aorta, this result might suggest that in the adult vasculature, the KLF2-66 enhancer drives *LacZ* expression only in response to a high shear stress threshold. However, further analysis of KLF2-66 and KLF2-41 driven *LacZ* expression in other adult vascular beds, including the lung, the brain, and the heart vasculature, found no *LacZ* expression in these vascular beds, many of which are predominantly composed of capillaries and small diameter arterioles thought to experience shear stress magnitudes comparable to that of intercostal arteries. Further, analysis of *LacZ* expression driven by each enhancer within inter-costal arteries of the ribcage found no *LacZ* expression within the intercostal artery endothelium. This could suggest that *LacZ*

expression in the adult intercostal arteries is limited to their branch points, which are thought to experience an oscillatory shear stress pattern.

The analysis of KLF2-66- and KLF2-41-driven *LacZ* within cardiac valves revealed that, unlike most other tissues, both enhancers drive some *LacZ* expression within this compartment even in the adult. The expression driven by the KLF2-41 enhancer was punctate and appeared dispersed throughout the valve tissue, whilst the KLF2-66 enhancer drove robust *LacZ* expression along the endothelium of both the aortic and the atrio-ventricular valves. The shear stress magnitude experienced by the valve endothelium is thought to be among the highest experienced in any organ system. However, due to the pulsatility of valve activity within small cardiac chambers, the pattern of shear stress experienced by the valve endothelium is thought to be highly oscillatory. The ability of the KLF2-66 enhancer to drive *LacZ* expression in adult inter-costal artery branch points and cardiac valves, both areas of oscillatory shear stress, might suggest that in the adult setting, the KLF2-66 enhancer responds to a specific shear stress pattern, namely oscillatory shear stress. However, oscillatory shear stress pattern also occurs within the inner curvature of the aortic arch, where no *LacZ* expression was detected.

While the lack of KLF2 enhancer-driven *LacZ* expression within the adult vasculature might have an explanation that is related to shear stress levels, there are other explanations for this observation that are equally possible. Enhancers are well known to regulate genes expressed in different cell types in response to lineage-specifying or environmental cues during development. However, the role of enhancers in regulating the expression of the same genes in the adult setting has been less well characterised. It is conceivable that KLF2 enhancers, while active during developmental and early post-natal stages, become inactivated in the adult, where EC phenotypes are generally less plastic. Expression of KLF2 in the adult setting might proceed through the KLF2 promoter, which, while entirely unable to drive *LacZ* during development, still contains key transcriptional information, namely and ETS, a MEF2, and an HBOX motif, that could allow it to recruit this transcription factors required for

LacZ expression driven by KLF2 enhancer during development, to regulate *KLF2* expression in the adult vasculature. This hypothesis would also explain the ability of the KLF2 promoter to drive reporter expression in response to shear stress in cell culture systems, which generally utilise endothelial cell lines derived from mature blood vessels.

The suppression of *LacZ* expression driven by the KLF2 enhancers in the adult setting might occur on the level of transcription, for example through inhibition of MEF2 by HDAC4/5 which has been widely described in the literature[212,247,430,431]. An alternative explanation for enhancer repression would be through a chromatin remodelling mechanism, where nucleosomes would re-populate the otherwise open chromatin within the KLF2 enhancer sequences, thereby rendering them transcriptionally inactive. However, one of the datasets utilised for KLF2 enhancer prediction was a mouse ATAC-seq dataset carried out in the endothelium of the adult mouse dorsal aorta[321]. In this dataset, both the KLF2-66 and KLF2-41 enhancer regions were open. This suggests that KLF2 enhancers are unlikely to be inactive due to being inaccessible in the adult setting. An interesting future direction would be to attempt to re-activate KLF2-66:*LacZ* and KLF2-41:*LacZ* transgenes in the adult, for example using a pharmacological approach. Statins are HMG-CoA reductase inhibitors whose therapeutic lipid-lowering action has been partially attributed to induction of KLF2 expression through activation of the ERK5-MEF2 signalling pathway[373,432–436]. Administration of statins to adult mice expressing the KLF2-66:*LacZ* and KLF2-41:*LacZ* transgenes would address the question whether the activity of these enhancers can be induced in the adult setting. If so, this would suggest that inactivity of KLF2-66:*LacZ* and KLF2-41:*LacZ* transgenes is due to the inhibition of MEF2 factors, and HDAC inhibitors could be administered to mice to validate this hypothesis.

5.4.3 KLF2-66:LacZ and KLF2-41:LacZ transgene expression is reduced following Dofetilide treatment

The absence of KLF2-66:*LacZ* and KLF2-41:*LacZ* transgene expression within the adult dorsal aorta and small diameter arteries meant that a surgical approach, such as carotid artery ligation, to test the response of each transgene to shear stress manipulation was not applicable. Instead, a pharmacological approach using Dofetilide, a class III anti-arrhythmic drug, was used to assess KLF2-66:*LacZ* and KLF2-41:*LacZ* transgene expression in the embryonic setting, where both transgenes were robustly expressed.

The administration of 2mg/kg Dofetilide between E6 and E10 resulted in in a variety of developmental phenotypes observed within the KLF2-41:*LacZ* litter harvested, including phenotypically normal, developmentally delayed, and severely developmentally delayed embryos. The expression of each transgene corresponded to the severity of phenotype observed. Annotating the location of each embryo within the uterus during dissection revealed no correlation between location and severity of developmental delay.

In developmentally delayed embryos, transgene expression was most notably reduced in small diameter arteries, capillary ECs, and the angiogenic endothelium of the hindbrain. Transgene expression within the endothelium of the dorsal aorta was only present within some ECs, compared to a more continuous expression in the vehicle-treated control. Transgene expression in the ECCs appeared less specific to the ECC surface and was instead more dispersed throughout the surrounding tissue. Ectopic transgene expression was observed in blood cells, which was especially evident in the abdominal vessels of the embryo. Similar results were observed with the KLF2-66:*LacZ* transgenic line, where 2mg/kg Dofetilide administration between E6 and E10 resulted in reduction of transgene expression within ECs of small diameter arteries, veins, and capillaries, accompanied by reduction of transgene expression at the angiogenic front and within the ECCs.

Administration of 2mg/kg Dofetilide dose to pregnant females carrying the from E7-E10, twice daily, was implemented to test KLF2-66:*LacZ* transgene expression in response to an adapted dosing regimen. Delaying the time of first administration was expected to circumvent the early developmental defects likely attributable to the teratogenic nature of Dofetilide action. Administration twice daily was expected to result in more stable levels of the drug within the circulation, thereby increasing the inhibition of cardiac contractions and blood flow. In the three litters treated according to this regimen, all embryos exhibited severe developmental delay. Despite this, *LacZ* expression in blood vessels of the majority of these embryos could still be observed during whole-mount analysis. Section analysis of representative embryos showed reduced levels of transgene expression in vessels that formed. In some embryos analysed, no KLF2-66:*LacZ* transgene expression was observed within blood vessels with a distinct endothelial layer, including the dorsal aortae and the cardinal veins. While transgene expression was still present within cells surrounding the ECCs, transgene expression appeared less confined to the ECC surface, although ECCs did not appear entirely correctly formed. Robust ectopic expression in blood was observed in many embryos during whole mount analysis.

5.5 Conclusion

The expression of KLF2-66:*LacZ* and KLF-41:*LacZ* expression in the post-natal dorsal aorta appeared to correspond to shear stress magnitudes experienced by different parts of the blood vessel, although the lack of transgene expression in the adult dorsal aorta warrants further investigation. Dofetilide treatment reduced KLF2-66:*LacZ* and KLF2-41:*LacZ* transgene expression within the endothelium, led to expansion of transgene expression within the ECCs, and induced ectopic transgene expression in blood cells. The results of these experiments together suggest that *LacZ* expression driven by the KLF2-41 and KLF2-66 enhancers might indeed be regulated by shear stress-induced signalling pathways.

6 Conclusion

6.1 Overall conclusions

KLF2 is a transcriptional master regulator of the endothelial shear stress response, with crucial roles in the maintenance of endothelial health[225,228,234,243,255,437]. The dysregulation of *KLF2* expression in response to pathological shear stress dynamics has been linked to the development of multiple vascular diseases[194,223,250,254,256,373,438–445]. Until recently, our knowledge of how *KLF2* expression is regulated during vascular development was limited. The aim of this thesis was to investigate how endothelial *KLF2* expression is regulated in development through the identification and characterisation of endothelial-specific KLF2 enhancers.

In Chapter 3, I identified two novel KLF2 enhancers located far upstream of the *KLF2* TSS. Both the KLF2-41 and KLF2-66 enhancers were able to drive reporter expression in transgenic mouse embryos specifically in ECs and ECCs independently of the KLF2 promoter. Both enhancers recapitulated aspects of endogenous *Klf2* expression patterns, although they differed in onset and pattern of enhancer activity in development. The KLF2 promoter, which has been previously implicated in *KLF2* regulation by shear stress, was by itself unable to drive reporter expression in development.

In Chapter 4, I have shown that MEF2 factors are required for the activity of both KLF2 enhancers. Detailed analysis of both enhancer sequences identified additional roles for ETS, RBPJ, GATA, NFAT, and FOX factors in regulation of enhancer activity. The analysis of KLF2 enhancer and promoter sequences led to the identification of a MEF2-HBOX motif which was present in all KLF2 regulatory regions and was required for KLF2-41:*LacZ* and KLF2-66:*LacZ* transgene expression. The search for other regulatory regions containing this double motif identified a novel distal KLF4-148 enhancer, which also drove endothelial-specific reporter expression. Further investigation is required to identify which TF interacts with the HBOX motif.

In Chapter 5, I investigated the KLF2-41:*LacZ* and KLF2-66:*LacZ* transgene expression in the dorsal aorta, and in response to pharmacological manipulation of cardiac contractions. Both enhancers appeared to drive reporter expression that corresponded to shear stress magnitudes and *Klf2* expression patterns within the post-natal dorsal aorta. Further, reporter expression driven by both enhancers was reduced following pharmacological inhibition of cardiac contractions in development, suggesting that both enhancers might be regulated by shear stress-induced TFs.

In conclusion, these results demonstrate the complexity of developmental gene regulation, and increase our understanding of transcriptional regulation of *KLF2*. Further characterisation of KLF2 enhancers provides an opportunity to identify novel transcriptional regulators of shear stress signalling in the endothelium.

6.2 Limitations

6.2.1 Statistical analysis

Statistical analysis should have been applied particularly in the analysis of mutant enhancer:*LacZ* transgene expression in comparison with WT enhancer:*LacZ* expression in mouse embryos (Chapter 4). As the transgene expression pattern was determined to fall under distinct categories (“vascular”, “ectopic”, and “none”) with no natural order or correlation between them, the type of data featuring in this analysis can be described as nominal categorical data, rather than ordinal categorical data which can be put into a ranked order[446]. In order to determine whether there is a statistically significant difference between the WT and each mutant enhancer:*LacZ* transgene expression pattern, the appropriate statistical test to use would be the Pearson’s chi-square test. Due to low embryo n numbers in some of these experiments, additional tests such as Yates’ correction or Fischer’s exact test might need to be applied.

6.2.2 Blinding

Blinding is often applied to limit the unintentional bias involved in conducting experiments and interpreting the results. Depending on the nature of the experiment, blinding can include, but is not limited to 1) concealing the treatment allocation, 2) designating another person with no investment in the experimental outcome to assess the results, especially when there is a subjective element to result assessment, and 3) blinding during the analysis of data[447,448].

There are two sections of this thesis where blinding might have been especially beneficial to ensure a high standard of scientific integrity in data acquisition and interpretation. The first is the zebrafish enhancer screen described in Section 3.3, where putative enhancer:GFP transgene expression was assessed and annotated in each embryo as “vascular”, “ectopic”, or “none”. Due to a large number of embryos analysed in each experiment (between 23 and 132), and only a subset of representative embryos imaged (five to ten per experiment), there is a potential element of subjectivity in assigning the level of reporter expression as well as selecting the representative embryos to image. It would have been beneficial for researchers unrelated to this project to count and assign each fish to the respective GFP expression level categories.

The second section of this thesis where blinding might have been beneficial is Dofetilide studies described in Section 5.3. In these experiments, allocation concealment could have been applied to the vehicle and Dofetilide-treated groups to minimize selection bias which might lead to preconceptions about the differences between animals assigned to the two distinct groups. As gavage administration was carried out by the animal technicians, the strategies to limit selection bias might include using opaque tape and coloured syringes to conceal the difference in the appearance of vehicle and drug solutions due to drug precipitation and assigning treatment and vehicle to each mouse numerically or alphabetically rather than using the words “drug” and “vehicle”. To reduce the potential for selection bias in vehicle and drug solution preparation, a different researcher could have been assigned to

prepare and aliquot the solutions. To reduce the potential for selection bias in analysis of enhancer:*LacZ* transgene expression in vehicle and Dofetilide-treated embryos, a different researcher could have been assigned to monitor the timing of fixation and staining, while maintaining a blinded assignment of drug and vehicle-treated groups until image acquisition and data analysis have been completed.

6.2.3 Zebrafish transgenic studies

The zebrafish enhancer screen described in Section 3.3 did not lead to the identification of any human or zebrafish enhancers able to drive EC-specific GFP expression. However, additional controls should have been included in these experiments to conclusively rule out the ability of the putative human and zebrafish enhancer regions to drive EC-specific reporter expression, especially given that many of the sequences analysed drove either low levels or no GFP expression in any tissue at different time points analysed. The control conditions that would have been useful for the interpretation of results in these experiments include: 1) co-injecting a control construct previously established to drive robust EC- specific reporter expression at the time points analysed to serve as a positive control for pro-nuclear injection and Tol2 transposase-mediated integration of the transgene and 2) injecting the putative enhancer:GFP constructs in a stable transgenic zebrafish line expressing a non-GFP EC-specific enhancer:reporter construct to control for fluorescence detection during imaging. Finally, the use of an altogether different imaging technology, such as light-sheet microscopy, could be considered in the future to discount the possibility of negative results due to high GFP detection threshold in imaging set-ups utilised.

6.2.4 Mouse transgenic studies

An inclusion of a negative control, for instance a scrambled WT enhancer sequence, could have been used in mouse transgenic studies. However, scrambling a sequence can lead to generation of artificial functional binding motifs which might complicate the interpretation of results.

6.2.5 Image annotation

Scale bars are often desirable on micrographs. This could have been usefully applied to zebrafish images, histological analysis of embryo sections, and whole-mount mouse embryo images.

6.3 Future directions

To investigate whether the KLF2-66 and KLF2-41 enhancers are required for *KLF2* expression, CRISPR-Cas9 gene editing will be used to generate stable endothelial cell lines in which KLF2-41 enhancer alone, KLF2-66 enhancer alone, or both enhancers together, will be deleted. The resulting KO cell lines will be characterised in terms of *KLF2* expression under static and shear stress conditions using the parallel plate flow chamber. These experiments will be complemented by analysis of WT and mutant KLF2 enhancer sequences cloned upstream different types of promoters and a *LacZ* reporter gene, which will be transfected into human aortic endothelial cell line (telo-HAEC) and tested for their ability to drive reporter expression both under static and shear stress conditions. This analysis is expected to result in a model of *KLF2* locus dynamics under static and shear stress conditions and provide further information on the function of different TF binding motifs for KLF2 enhancer activity in the context of shear stress.

In order to identify which TF binds the HBOX motif, mass spectrometry analysis will be employed as previously described[449–452]. Briefly, the 50 bp conserved core of the KLF2-41 enhancer sequence will be used as a bait to pull down TFs that interact with this sequence, using the telo-HAEC lysate as the source of nuclear protein. This experiment will be carried out using both a WT and a mutant probe in which the HBOX site has been mutated. The dissociated TFs will then be subjected to mass spectrometry analysis, and the resultant lists of TFs pulled down following analysis will be compared to identify differentially pulled down TFs, which will then be validated using *in vitro* techniques such as EMSA and ChIP-qPCR. This experiment will be expanded to include nuclear protein from cell lysates

exposed to static or shear stress conditions to allow for the possibility that shear stress induces TF binding to the putative HBOX motif.

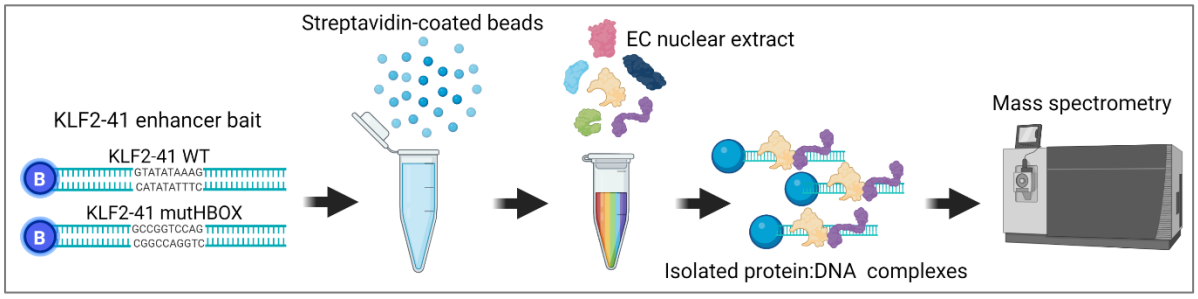


Figure 6.1: A mass spectrometry approach for identification of TFs that bind the HBOX motif. The WT and mutant versions of the KLF2-41 enhancer oligos will be immobilised onto streptavidin-coated beads, then used as baits for EC nuclear extract obtained from the telo-HAEC line exposed to static or shear stress conditions. The isolated TF:DNA complexes will be subjected to mass spectrometry analysis as previously described[451]. The analysis of differences between TFs pulled down by the WT vs. the mutant enhancer between static and shear stress conditions will determine which TF interacts with the HBOX motif, and whether its binding to the enhancer sequence is shear-stress dependent. The figure is adapted from Vecellio *et al*. (2021)[451]. This figure was created with BioRender.com

7 References

1. Kornberg, R. D. *Chromatin structure: a repeating unit of histones and DNA*. https://www.science.org (1974) doi:10.1126/science.184.4139.868.

2. Luger, K., Mä Der, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. *Crystal structure of the nucleosome core particle at 2.8 A ˚ resolution*. *Nature © Macmillan Publishers Ltd* vol. 389 (1997).

3. Ang, J. C., Barkley, M. D. & Bourgeois, S. *Subunit structure of chromatin*. *Nature new Biol* vol. 212 (1966).

4. Thomas, J. O. & Kornberg, R. D. An octamer of histones in chromatin and free in solution. *Proc Natl Acad Sci U S A* **72**, 2626–2630 (1975).

5. Struck, M. M., Klug, A. & Richmond, T. J. Comparison of X-ray structures of the nucleosome core particle in two different hydration states. *J Mol Biol* **224**, 253–264 (1992).

6. Uberbacher, E. C. & Bunick, G. J. X-ray structure of the nucleosome core particle. *J Biomol Struct Dyn* **2**, 1033–1055 (1985).

7. Ramakrishnan, V., Fincht, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. Crystal structure of globular domain of histone HS and its implications for nucleosome binding. *Nature* (1993).

8. Linker Histones Graphical, by *et al.* Structural Mechanisms of Nucleosome Recognition Article Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol Cell* **59**, 628–638 (2015).

9. Fyodorov, D. v., Zhou, B. R., Skoultchi, A. I. & Bai, Y. Emerging roles of linker histones in regulating chromatin structure and function. *Nature Reviews Molecular Cell Biology 2017 19:3* **19**, 192–206 (2017).

10. Dimitri, P., Bozzetti, M. P., Eissenberg, J., Furlan-Magaril, M. & Penagos-Puig, A. Heterochromatin as an Important Driver of Genome Organization. *Front Cell Dev Biol* (2020) doi:10.3389/fcell.2020.579137.

11. Grewal, S. I. S. & Jia, S. Heterochromatin revisited. *Nat Rev Genet* **8**, 35 (2007).

12. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* (2011) doi:10.1101/gad.176826.111.

13. Adam, R. C. *et al.* Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* **521**, 366–370 (2015).

14. Huebert, D. J., Kuan, P.-F., Keleş, S. & Gasch, A. P. Dynamic Changes in Nucleosome Occupancy Are Not Predictive of Gene Expression Dynamics but Are Linked to Transcription and Chromatin Regulators. *Mol Cell Biol* **32**, 1645–1653 (2012).

15. Oruba, A., Saccani, S. & van Essen, D. Role of cell-type specific nucleosome positioning in inducible activation of mammalian promoters. *Nat Commun* (2020) doi:10.1038/s41467-020-14950-5.

16. Millamp, G., Burton, A., Bannister, A. J. & Schneider, R. Histone post-translational modifications — cause and consequence of genome function. *Nat Rev Genet* (2022) doi:10.1038/s41576-022-00468-7.

17. ENCODE Project Consortium, T. *et al.* An integrated encyclopedia of DNA elements in the human genome ENCODE Encyclopedia of DNA Elements. *Nature* (2012) doi:10.1038/nature11247.

18. Moore, J. E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 25 (2020).

19. Cremer, T. *et al. Rabl's Model of the Interphase Chromosome Arrangement Tested in Chinese Hamster Cells by Premature Chromosome Condensation and Laser-UV-Microbeam Experiments*. *Hum Genet* vol. 60 (1982).

20. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **292**, (2001).

21. Tanabe, H. *et al.* Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A* **99**, 4424–4429 (2002).

22. Parada, L. A., McQueen, P. G., Munson, P. J. & Misteli, T. Conservation of relative chromosome positioning in normal and cancer cells. *Current Biology* **12**, 1692–1697 (2002).

23. Cremer, T. & Cremer, M. Chromosome Territories. *Cold Spring Harb Perspect Biol* (2010) doi:10.1101/cshperspect.a003889.

24. Finlan, L. E., Sproul, D., Thomson, I., Boyle, S. & Kerr, E. Recruitment to the Nuclear Periphery Can Alter Expression of Genes in Human Cells. *PLoS Genet* **4**, 1000039 (2008).

25. Fritz, A. J., Sehgal, N., Pliss, | Artem, Xu, J. & Berezney, R. Chromosome territories and the global regulation of the genome. *Genes Chromosomes Cancer* (2019) doi:10.1002/gcc.22732.

26. Nikolova, V. *et al.* Defects in nuclear structure and function promote dilated cardiomyopathy in lamin A/C-deficient mice. *J Clin Invest* **113**, 357–369 (2004).

27. Mattick, J. S. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep* **2**, 986–991 (2001).

28. Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. doi:10.1038/s41580-020-00315-9.

29. Roeder, R. G. The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem Sci* **16**, 402–408 (1991).

30. Nikolov, D. B. & Burley, S. K. RNA polymerase II transcription initiation: A structural view. *Proc Natl Acad Sci U S A* **94**, 15–22 (1997).

31. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology* vol. 19 621–637 Preprint at https://doi.org/10.1038/s41580-018-0028-8 (2018).

32. Cormack, B. P. & Struhl, K. The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell* **69**, 685–696 (1992).

33. Jacobi, U. G. *et al.* TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation. *EMBO J* **26**, 3900–3909 (2007).

34. Nikolov, D. B. *et al. Crystal structure of a human TATA box-binding protein/TATA element complex*. *Cell Biology* vol. 93 https://www.pnas.org (1996).

35. Patikoglou, G. A. *et al.* TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev* (1999).

36. Yang, M. Q. & Elnitski, L. L. Diversity of core promoter elements comprising human bidirectional promoters. *BMC Genomics* **9**, 1–8 (2008).

37. Gupta, H. *et al.* Highly diversified core promoters in the human genome and their effects on gene expression and disease predisposition. *BMC Genomics* **21**, 1–12 (2020).

38. Kutach, A. K. & Kadonaga, J. T. The Downstream Promoter Element DPE Appears To Be as Widely Used as the TATA Box in Drosophila Core Promoters. *Mol Cell Biol* **20**, 4754 (2000).

39. Vo ngoc, L., Jack Cassidy, C., Yunjing Huang, C., Duttke, S. H. & Kadonaga, J. T. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev* (2017) doi:10.1101/gad.293837.116.

40. Burke, T. W. & Kadonaga, J. T. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAF II 60 of Drosophila. *Genes Dev* (1996).

41. Verrijzer, C. P., Chen, J. L., Yokomori, K. & Tjian, R. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81**, 1115–1125 (1995).

42. Hantsche, M. & Cramer, P. Conserved RNA polymerase II initiation complex structure. *Curr Opin Struct Biol* **47**, 17–22 (2017).

43. Dollinger, R. & Gilmour, D. S. Regulation of Promoter Proximal Pausing of RNA Polymerase II in Metazoans. *J Mol Biol* **433**, (2021).

44. Tang, H., Liu, Y., Madabusi, L. & Gilmour, D. S. Promoter-proximal pausing on the hsp70 promoter in Drosophila melanogaster depends on the upstream regulator. *Mol Cell Biol* **20**, 2569–2580 (2000).

45. Gilchrist, D. A. *et al.* Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. *Genes Dev* (2012) doi:10.1101/gad.187781.112.

46. Hendrix, D. A., Hong, J.-W., Zeitlinger, J., Rokhsar, D. S. & Levine, M. S. Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. *PNAS* (2008).

47. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* (2018) doi:10.1038/s41586-018-0442-2.

48. Gilchrist, D. A. *et al.* NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev* **22**, 1921–1933 (2008).

49. Chopra, V. S., Cande, J., Hong, J. W. & Levine, M. Stalled Hox promoters as chromosomal boundaries. *Genes Dev* **23**, 1505 (2009).

50. Rahl, P. B. *et al.* C-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).

51. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**, (2015).

52. Danko, C. G. *et al.* Signaling Pathways Differentially Affect RNA Polymerase II Initiation, Pausing, and Elongation Rate in Cells. *Mol Cell* **50**, 212–222 (2013).

53. Venkatesh, S. & Workman, J. L. RNA polymerase II (Pol II) faithfully transcribes DNA templates into RNA with the Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol* (2015) doi:10.1038/nrm3941.

54. Veloso, A. *et al.* Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res* **24**, 896–905 (2014).

55. Gromak, N., West, S. & Proudfoot, N. J. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol* **26**, 3986–3996 (2006).

56. Roy, A. L. & Singer, D. S. Core promoters in transcription: Old problem, new insights. *Trends Biochem Sci* **40**, 165–171 (2015).

57. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).

58. Wythe, J. D. *et al.* ETS factors regulate Vegf-dependent arterial specification. *Dev Cell* **26**, 45–58 (2013).

59. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**, (2014).

60. Shi, W. & Zhou, W. Frequency distribution of TATA Box and extension sequences on human promoters. *BMC Bioinformatics* **7**, 1–12 (2006).

61. Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. & Martinez, E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389**, 52–65 (2007).

62. Yan Lim, C. *et al.* The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev* (1999) doi:10.1101/gad.1193404.

63. Yella, V. R. & Bansal, M. DNA structural features of eukaryotic TATA-containing and TATA-less promoters. *FEBS* (2017) doi:10.1002/2211-5463.12166.

64. van Heeringen, S. J. *et al.* Nucleotide composition-linked divergence of vertebrate core promoter architecture. *Genome Res* **21**, 410 (2011).

65. Vavouri, T. & Lehner, B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biol* **13**, 1–12 (2012).

66. Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *PNAS* **90**, 11995–11999 (1993).

67. Antequera, F. Review Structure, function and evolution of CpG island promoters. *CMLS, Cell. Mol. Life Sci* **60**, 1647–1658 (2003).

68. Juven-Gershon, T., Hsu, J. Y., Theisen, J. W. & Kadonaga, J. T. The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol* **20**, 253–259 (2008).

69. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* (2011) doi:10.1101/gad.2037511.

70. Zabidi, M. A. *et al.* Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2015).

71. Schaffner, W. Enhancers, enhancers – from their discovery to today's universe of transcription enhancers. *Biol Chem* **396**, 311–327 (2015).

72. Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797–803 (2005).

73. Neal, A. *et al.* Venous identity requires BMP signalling through ALK3. *Nat Commun* **10**, 453 (2019).

74. Sacilotto, N. *et al.* Analysis of Dll4 regulation reveals a combinatorial role for Sox and Notch in arterial development. *Proceedings of the National Academy of Sciences* **110**, 11893–11898 (2013).

75. Tuvikene, J. *et al.* Intronic enhancer region governs transcript-specific bdnf expression in rodent neurons. *Elife* **10**, 1–28 (2021).

76. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).

77. Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet* (2020) doi:10.1038/s41576-019-0209-0.

78. Drouin, J. Minireview: Pioneer Transcription Factors in Cell Fate Specification. *Molecular Endocrinology* **28**, 989 (2014).

79. Spicuglia, S. & Vanhille, L. Chromatin signatures of active enhancers salvatore spicuglia & Laurent Vanhille Chromatin signatures of active enhancers. *Nucleus* (2012) doi:10.4161/nucl.19232.

80. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *PNAS* (2010) doi:10.1073/pnas.1016071107.

81. Karlić, R., Chung, H. R., Lasserre, J., Vlahoviček, K. & Vingron, M. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A* **107**, 2926–2931 (2010).

82. Dong, X. & Weng, Z. The correlation between histone modifications and gene expression. *Epigenomics* **5**, 113 (2013).

83. Paauw, N. D. *et al.* H3K27 acetylation and gene expression analysis reveals differences in placental chromatin activity in fetal growth restriction. *Clin Epigenetics* **10**, 1–11 (2018).

84. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev* **43**, 73–81 (2017).

85. Gerritsen, M. E. *et al.* CREB-binding protein/p300 are transcriptional coactivators of p65. *Proc Natl Acad Sci U S A* **94**, 2927–2932 (1997).

86. Zhou, P. *et al.* Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. *Elife* **6**, (2017).

87. Long, H. K., Prescott, S. L. & Wysocka, J. Leading Edge Review Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170–1187 (2016).

88. Honnell, V. *et al.* Identification of a modular super-enhancer in murine retinal development. *Nat Commun* (2022) doi:10.1038/s41467-021-27924-y.

89. Hay, D. *et al.* Genetic dissection of the α-globin super-enhancer in vivo. *Nat Genet* **48**, 895 (2016).

90. Menke, D. B., Guenther, C. & Kingsley, D. M. Dual hindlimb control elements in the Tbx4 gene and region-specific control of bone size in vertebrate limbs. *Development* **135**, 2543–2553 (2008).

91. Cannavò, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol* **26**, 38–51 (2016).

92. Hörnblad, A. *et al.* Dissection of the Fgf8 regulatory landscape by in vivo CRISPR-editing reveals extensive intra-and inter-enhancer redundancy. *Nat Commun* (2021) doi:10.1038/s41467-020-20714-y.

93. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* (2018) doi:10.1038/nature25461.

94. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat Rev Genet* **19**, 789 (2018).

95. van Arensbergen, J., van Steensel, B. & Bussemaker, H. J. In search of the determinants of enhancer-promoter interaction specificity. *Trends in Cell Biology* vol. 24 695–702 Preprint at https://doi.org/10.1016/j.tcb.2014.07.004 (2014).

96. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

97. Beagan, J. A. & Phillips-Cremins, J. E. On the existence and functionality of topologically associating domains. *Nat Genet* (2020) doi:10.1038/s41588-019-0561-1.

98. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature 2012 485:7398* **485**, 376–380 (2012).

99. McArthur, E. & Capra, J. A. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am J Hum Genet* **108**, 269–283 (2021).

100. Morey, C., da Silva, N. R., Perry, P. & Bickmore, W. A. Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development* **134**, 909–919 (2007).

101. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).

102. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the-globin locus. (2006) doi:10.1101/gad.399506.

103. Hyle, J. *et al.* Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Res* **47**, 6699–6713 (2019).

104. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).

105. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).

106. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**, 2038–2049 (2016).

107. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**, E6456–E6465 (2015).

108. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra-and interchromosomal interactions. (2006) doi:10.1038/ng1891.

109. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**, E6456–E6465 (2015).

110. Banigan, E. J., van den Berg, A. A., Brandão, H. B., Marko, J. F. & Mirny, L. A. Chromosome organization by one-sided and two-sided loop extrusion. *Elife* **9**, (2020).

111. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252–263 (2009).

112. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e16 (2018).

113. Wilkinson, A. C., Nakauchi, H. & Gö Ttgens, B. Cell Systems Perspective Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. *Cell Syst* **5**, 319–331 (2017).

114. Buecker, C. & Wysocka, J. Enhancers as information integration hubs in development: Lessons from genomics. *Trends in Genetics* **28**, 276–284 (2012).

115. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how and why? *Mol Cell* (2013) doi:10.1016/j.molcel.2013.01.038.

116. Farley, E. K., Olson, K. M., Zhang, W., Rokhsar, D. S. & Levine, M. S. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc Natl Acad Sci U S A* **113**, 6508–6513 (2016).

117. Panne, D., Maniatis, T. & Harrison, S. C. An atomic model of the interferon-beta enhanceosome. *Cell* **129**, 1111–1123 (2007).

118. Liu, F. & Posakony, J. W. Role of Architecture in the Function and Specificity of Two Notch-Regulated Transcriptional Enhancer Modules. *PLoS Genet* **8**, e1002796 (2012).

119. Goldwater, M. B., Markman, A. B. & Stilwell, C. H. Structural Rules and Complex Regulatory Circuitry Constrain Expression of a Notch- and EGFR-Regulated Eye Enhancer. *Dev Cell* **18**, 359–370 (2010).

120. Slattery, M. *et al.* Absence of a simple code: How transcription factors read the genome. *Trends Biochem Sci* **39**, 381–399 (2014).

121. Vockley, C. M., Mcdowell, I. C., D'ippolito, A. M. & Reddy, T. E. A long-range flexible billboard model of gene activation. *Transcription* (2017) doi:10.1080/21541264.2017.1317694.

122. Smith, R. P. *et al.* Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nature Genetics 2013 45:9* **45**, 1021–1028 (2013).

123. Kulkarni, M. M. & Arnosti, D. N. Information display by transcriptional enhancers. *Development* **130**, 6569–6575 (2003).

124. Brown, C. D., Johnson, D. S. & Sidow, A. Functional architecture and evolution of transcriptional elements that drive gene coexpression. *Science* **317**, 1557–1560 (2007).

125. Rastegar, S. *et al.* The words of the regulatory code are arranged in a variable manner in highly conserved enhancers. *Dev Biol* **318**, 366–377 (2008).

126. Tijssen, M. R. *et al.* Genome-wide Analysis of Simultaneous GATA1/2, RUNX1, FLI1, and SCL Binding in Megakaryocytes Identifies Hematopoietic Regulators. *Dev Cell* **20**, 597–609 (2011).

127. Junion, G. *et al.* A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148**, 473–486 (2012).

128. de Almeida, B. P., Reiter, F., Pagani, M. & Stark, A. DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of enhancers. *bioRxiv* (2022) doi:10.1101/2021.10.05.463203.

129. Yadav, V. K. *et al.* Promoter-proximal transcription factor binding is transcriptionally active when coupled with nucleosome repositioning in immediate vicinity. *Nucleic Acids Res* **42**, 9602 (2014).

130. Miller, J. A. & Widom, J. Collaborative competition mechanism for gene activation in vivo. *Mol Cell Biol* **23**, 1623–1632 (2003).

131. Mirny, L. A. & Onuchic, J. N. Nucleosome-mediated cooperativity between transcription factors. *PNAS* (2010) doi:10.1073/pnas.0913805107/-/DCSupplemental.

132. Udan, R. S., Culver, J. C. & Dickinson, M. E. Understanding vascular development: WIRE Developmental Biology (2012). *Wiley Interdiscip Rev Dev Biol* **2**, 327 (2013).

133. Chappell, J. C. & Bautch, V. L. Chapter Two - Vascular Development—Genetic Mechanisms and Links to Vascular Disease. *Current Topics in Developmental Biology - Volume 90* **90**, 43–72 (2010).

134. Bautch, V. L. & Caron, K. M. Blood and lymphatic vessel formation. *Cold Spring Harb Perspect Biol* **7**, a008268 (2015).

135. Culver, J. C. & Dickinson, M. E. The effects of hemodynamic force on embryonic development. *Microcirculation* **17**, 164 (2010).

136. de Val, S. Key transcriptional regulators of early vascular development. *Arterioscler Thromb Vasc Biol* **31**, 1469–1475 (2011).

137. Inman, K. E. & Downs, K. M. The Murine Allantois: Emerging Paradigms in Development of the Mammalian Umbilical Cord and Its Relation to the Fetus. *Genesis* (2007) doi:10.1002/dvg.20281.

138. Harris, I. S. & Black, B. L. Development of the Endocardium. *Pediatr Cardiol* (2010) doi:10.1007/s00246-010-9642-8.

139. Schmidt, A., Brixius, K. & Bloch, W. Endothelial precursor cell migration during vasculogenesis. *Circ Res* **101**, 125–136 (2007).

140. Xiong, J.-W. Molecular and Developmental Biology of the Hemangioblast. *Developmental Dynamics* **237**, 1218–1231 (2008).

141. Gritz, E. & Hirschi, K. K. Specification and function of hemogenic endothelium during embryogenesis. *Cellular and Molecular Life Sciences* (2016) doi:10.1007/s00018-016-2134-0.

142. Ferguson, J. E., Kelley, R. W. & Patterson, C. Mechanisms of endothelial differentiation in embryonic vasculogenesis. *Arterioscler Thromb Vasc Biol* **25**, 2246–2254 (2005).

143. de Val, S. *et al.* Combinatorial Regulation of Endothelial Gene Expression by Ets and Forkhead Transcription Factors. *Cell* **135**, 1053–1064 (2008).

144. Potente, M. & Mäkinen, T. Vascular heterogeneity and specialization in development and disease. *Nature Reviews Molecular Cell Biology 2017 18:8* **18**, 477–494 (2017).

145. Lee, H. S., Han, J., Bai, H. J. & Kim, K. W. Brain angiogenesis in developmental and pathological processes: regulation, molecular and cellular communication at the neurovascular interface. *FEBS J* **276**, 4622–4635 (2009).

146. Wu, B. *et al.* Endocardial cells form the coronary arteries by angiogenesis through myocardial-endocardial VEGF signaling. *Cell* **151**, 1083–1096 (2012).

147. Milde, F., Lauw, S., Koumoutsakos, P. & Iruela-Arispe, M. L. The mouse retina in 3D: quantification of vascular growth and remodeling. *Integrative Biology* **5**, 1426 (2013).

148. Potente, M., Gerhardt, H. & Carmeliet, P. Basic and therapeutic aspects of angiogenesis. *Cell* **146**, 873–887 (2011).

149. Eelen, G., Treps, L., Li, X. & Carmeliet, P. Basic and Therapeutic Aspects of Angiogenesis Updated. *Circ Res* **127**, 310–329 (2020).

150. Lobov, I. B., Brooks, P. C. & Lang, R. A. Angiopoietin-2 displays VEGF-dependent modulation of capillary structure and endothelial cell survival in vivo. *Proc Natl Acad Sci U S A* **99**, 11205–11210 (2002).

151. Lobov, I. B. *et al.* Delta-like ligand 4 (DII4) is induced by VEGF as a negative regulator of angiogenic sprouting. *Proc Natl Acad Sci U S A* **104**, 3219–3224 (2007).

152. Jakobsson, L. *et al.* Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. *Nat Cell Biol* (2010) doi:10.1038/ncb2103.

153. Roukens, M. G. *et al.* Control of endothelial sprouting by a Tel-CtBP complex. *Nat Cell Biol* (2010) doi:10.1038/ncb2096.

154. Sacilotto, N. *et al.* MEF2 transcription factors are key regulators of sprouting angiogenesis. *Genes Dev* **30**, 2297–2309 (2016).

155. de Spiegelaere, W. *et al.* Intussusceptive Angiogenesis: A Biologically Relevant Form of Angiogenesis. *J Vasc Res* **49**, 390–404 (2012).

156. Jones, E. A. V., le Noble, F. & Eichmann, A. What determines blood vessel structure? Genetic prespecification vs. hemodynamics. *Physiology* **21**, 388–395 (2006).

157. Davis, G. E. & Bayless, K. J. An Integrin and Rho GTPase-Dependent Pinocytic Vacuole Mechanism Controls Capillary Lumen Formation in Collagen and Fibrin Matrices. *Microcirculation* (2003).

158. Strilić, B. *et al.* The Molecular Basis of Vascular Lumen Formation in the Developing Mouse Aorta. *Dev Cell* **17**, 505–515 (2009).

159. Xu, K. *et al.* Blood Vessel Tubulogenesis Requires Rasip1 Regulation of GTPase Signaling. *Dev Cell* **20**, 526–539 (2011).

160. Gaengel, K., Genové, G., Armulik, A. & Betsholtz, C. Endothelial-mural cell signaling in vascular development and angiogenesis. *Arterioscler Thromb Vasc Biol* **29**, 630–638 (2009).

161. Pardali, E., Goumans, M. J. & ten Dijke, P. Signaling by members of the TGF-β family in vascular morphogenesis and disease. *Trends Cell Biol* **20**, 556–567 (2010).

162. Lindblom, P. *et al.* Endothelial PDGF-B retention is required for proper investment of pericytes in the microvessel wall. *Genes Dev* **17**, 1835–1840 (2003).

163. Fish, J. E. & Wythe, J. D. The molecular regulation of arteriovenous specification and maintenance. *Developmental Dynamics* **244**, 391–409 (2015).

164. Fang, J. S., Hirschi, K. K. & Cleaver, O. Open Peer Review Molecular regulation of arteriovenous endothelial cell specification. *F1000 Reserach* (2019) doi:10.12688/f1000research.16701.1.

165. Adams, R. H. *et al.* Roles of ephrinB ligands and EphB receptors in cardiovascular development: demarcation of arterial/venous domains, vascular morphogenesis, and sprouting angiogenesis. *Genes Dev* (1999).

166. Gerety, S. S., Wang, H. U., Chen, Z. F. & Anderson, D. J. Symmetrical mutant phenotypes of the receptor EphB4 and its specific transmembrane ligand ephrin-B2 in cardiovascular development. *Mol Cell* **4**, 403–414 (1999).

167. Hogan, B. M. & Schulte-Merker, S. How to Plumb a Pisces: Understanding Vascular Development and Disease Using Zebrafish Embryos. *Dev Cell* **42**, 567–583 (2017).

168. Ungos, J. & Weinstein, B. M. Vascular Development in the Zebrafish. *Advances in Developmental Biology* **18**, 301–332 (2007).

169. Kohli, V., Schumacher, J. A., Desai, S. P., Rehn, K. & Sumanas, S. Arterial and Venous Progenitors of the Major Axial Vessels Originate at Distinct Locations. *Dev Cell* **25**, 196–206 (2013).

170. Udan, R. S., Culver, J. C. & Dickinson, M. E. Understanding vascular development. *Wiley Interdiscip Rev Dev Biol* **2**, 327–346 (2013).

171. Chong, D. C., Koo, Y., Xu, K., Fu, S. & Cleaver, O. Stepwise Arteriovenous Fate Acquisition During Mammalian Vasculogenesis. *Developmental Dynamics* (2011) doi:10.1002/dvdy.22706.

172. Krebs, L. T. *et al.* Notch signaling is essential for vascular morphogenesis in mice. *Genes and Developement* (2000).

173. Lawson, N. D. *et al.* Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development* **128**, 3675–3683 (2001).

174. Holderfield, M. T. & Hughes, C. C. W. Crosstalk Between Vascular Endothelial Growth Factor, Notch, and Transforming Growth Factor-β in Vascular Morphogenesis. *Circ Res* **102**, 637–652 (2008).

175. Fischer, A., Schumacher, N., Maier, M., Sendtner, M. & Gessler, M. The Notch target genes Hey1 and Hey2 are required for embryonic vascular development. *Genes Dev* **18**, 901–911 (2004).

176. Luo, W. *et al.* Arterialization requires the timely suppression of cell growth. *Nature* **589**, 437–441 (2021).

177. Pendeville, H. *et al.* Zebrafish Sox7 and Sox18 function together to control arterial-venous identity. *Dev Biol* **317**, 405–416 (2008).

178. Corada, M. *et al.* Sox17 is indispensable for acquisition and maintenance of arterial identity. *Nat Commun* (2013) doi:10.1038/ncomms3609.

179. Hayashi, H. & Kume, T. Foxc Transcription Factors Directly Regulate Dll4 and Hey2 Expression by Interacting with the VEGF-Notch Signaling Pathways in Endothelial Cells. *PLoS One* **3**, e2401 (2008).

180. Seo, S. *et al.* The forkhead transcription factors, Foxc1 and Foxc2, are required for arterial specification and lymphatic sprouting during vascular development. *Dev Biol* **294**, 458–470 (2006).

181. You, L. R. *et al.* Suppression of Notch signalling by the COUP-TFII transcription factor regulates vein identity. *Nature 2005 435:7038* **435**, 98–104 (2005).

182. Hahn, C. & Schwartz, M. A. Mechanotransduction in vascular physiology and atherogenesis. *Nat Rev Mol Cell Biol* **10**, 53–62 (2009).

183. Jones, E. A. v. The initiation of blood flow and flow induced events in early vascular development. *Semin Cell Dev Biol* **22**, 1028–1035 (2011).

184. Friedman, M. H. Variability of arterial wall shear stress, its dependence on vessel diameter and implications for Murray's Law. *Atherosclerosis* **203**, 47–48 (2009).

185. Pries, A. R., Secomb, T. W. & Gaehtgens, P. Design Principles of Vascular Beds. *Circ Res* **77**, 1017–1023 (1995).

186. Givens, C. & Tzima, E. Endothelial Mechanosignaling: Does One Sensor Fit All? *Antioxid Redox Signal* **25**, 373–388 (2016).

187. Jones, E. A. v, Baron, M. H., Fraser, S. E. & Dickinson, M. E. Measuring hemodynamic changes during mammalian development. *Am J Physiol Heart Circ Physiol* **287**, 1561–1569 (2004).

188. Humphrey, J. D., Schwartz, M. A., Tellides, G. & Milewicz, D. M. Role of mechanotransduction in vascular biology: Focus on thoracic aortic aneurysms and dissections. *Circ Res* **116**, 1448–1461 (2015).

189. Baeyens, N., Bandyopadhyay, C., Coon, B. G., Yun, S. & Schwartz, M. A. Endothelial fluid shear stress sensing in vascular health and disease. *Journal of Clinical Investigation* **126**, 821–828 (2016).

190. Chapman, W. B. The effect of the heart-beat upon the development of the vascular system in the chick. *American Journal of Anatomy* **23**, 175–203 (1918).

191. Lucitti, J. L. *et al.* Vascular remodeling of the mouse yolk sac requires hemodynamic force. *Development* **134**, 3317–3326 (2007).

192. Goddard, L. M. *et al.* Hemodynamic Forces Sculpt Developing Heart Valves through a KLF2-WNT9B Paracrine Signaling Axis. *Dev Cell* **43**, 274-289.e5 (2017).

193. Hove, J. R. *et al.* Intracardiac fluid forces are an essential epigenetic factor for embryonic cardiogenesis. *Nature 2003 421:6919* **421**, 172–177 (2003).

194. Souilhol, C. *et al.* Endothelial responses to shear stress in atherosclerosis: a novel role for developmental genes. *Nat Rev Cardiol* 1–12 (2019) doi:10.1038/s41569-019-0239-5.

195. Hwa, J. J. *et al.* Abnormal arterial-venous fusions and fate specification in mouse embryos lacking blood flow. *Sci Rep* **7**, 11965 (2017).

196. le Noble, F. *et al.* Flow regulates arterial-venous differentiation in the chick embryo yolk sac. *Development* **131**, 361–375 (2004).

197. Buschmann, I. *et al.* Pulsatile shear and Gja5 modulate arterial identity and remodeling events during flow-driven arteriogenesis. *Development* **137**, 2187–2196 (2010).

198. Fang, J. S. *et al.* Shear-induced Notch-Cx37-p27 axis arrests endothelial cell cycle to enable arterial specification. *Nat Commun* **8**, (2017).

199. Bussmann, J., Wolfe, S. A. & Siekmann, A. F. Arterial-venous network formation during brain vascularization involves hemodynamic regulation of chemokine signaling. *Development* **138**, 1717–1726 (2011).

200. Chouinard-Pelletier, G., Jahnsen, E. D. & Jones, E. A. v. Increased shear stress inhibits angiogenesis in veins and not arteries during vascular development. *Angiogenesis* (2013) doi:10.1007/s10456-012-9300-2.

201. Song, J. W., Munn, L. L., Steele, E. L. & Chien, S. Fluid forces control endothelial sprouting. *PNAS* (2011) doi:10.1073/pnas.1105316108.

202. Kochhan, E., Lenard, A., Ellertsdottir, E., Herwig, L. & Affolter, M. Blood Flow Changes Coincide with Cellular Rearrangements during Blood Vessel Pruning in Zebrafish Embryos. *PLoS One* **8**, 75060 (2013).

203. Elworthy, S., Savage, A. M., Wilkinson, R. N., Malicki, J. J. & Chico, T. J. A. The role of endothelial cilia in postembryonic vascular development. *Dev Dyn* **248**, 410–425 (2019).

204. Egorova, A. D. *et al.* Lack of primary cilia primes shear-induced endothelial-to-mesenchymal transition. *Circ Res* **108**, 1093–1101 (2011).

205. Chen, X., Gays, D., Milia, C. & Santoro, M. M. Cilia Control Vascular Mural Cell Recruitment in Vertebrates. *Cell Rep* **18**, 1033–1047 (2017).

206. Tzima, E. *et al.* A mechanosensory complex that mediates the endothelial cell response to fluid shear stress. *Nature* **437**, 426–431 (2005).

207. Zhou, J., Li, Y. S. & Chien, S. Shear stress-initiated signaling and its regulation of endothelial function. *Arterioscler Thromb Vasc Biol* **34**, 2191–2198 (2014).

208. Topper, J. N. & Gimbrone, M. A. Blood flow and vascular gene expression: Fluid shear stress as a modulator of endothelial phenotype. *Mol Med Today* **5**, 40–46 (1999).

209. Braddock, M. *et al.* Fluid shear stress modulation of gene expression in endothelial cells. *News in Physiological Sciences* **13**, 241–246 (1998).

210. Li, J. Y.-S. *et al.* Systems biology analysis of longitudinal functional response of endothelial cells to shear stress. *Proceedings of the National Academy of Sciences* **114**, 10990–10995 (2017).

211. Mccormick, S. M. *et al.* DNA microarray reveals changes in gene expression of shear stressed human umbilical vein endothelial cells. *PNAS* (2001).

212. Potthoff, M. J. & Olson, E. N. MEF2: A central regulator of diverse developmental programs. *Development* vol. 134 4131–4140 Preprint at https://doi.org/10.1242/dev.008367 (2007).

213. de Val, S. *et al.* Mef2c is activated directly by Ets transcription factors through an evolutionarily conserved endothelial cell-specific enhancer. *Dev Biol* **275**, 424–434 (2004).

214. Bi, W., Drake, C. J. & Schwarz, J. J. The transcription factor MEF2C-null mouse exhibits complex vascular malformations and reduced cardiac expression of angiopoietin 1 and VEGF. *Dev Biol* **211**, 255–267 (1999).

215. Lin, Q., Schwarz, J., Bucana, C. & Olson, E. N. Control of Mouse Cardiac Morphogenesis and Myogenesis by Transcription Factor MEF2C. *Science (1979)* **276**, 1404–1407 (1997).

216. Materna, S. C., Sinha, T., Barnes, R. M., Lammerts van Bueren, K. & Black, B. L. Cardiovascular development and survival require Mef2c function in the myocardial but not the endothelial lineage. *Dev Biol* **445**, 170–177 (2019).

217. Naya, F. J. *et al.* Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nature Medicine 2002 8:11* **8**, 1303–1309 (2002).

218. Kim, Y. *et al.* The MEF2D transcription factor mediates stress-dependent cardiac remodeling in mice. *J Clin Invest* **118**, (2008).

219. Liu, N. *et al.* Requirement of MEF2A, C, and D for skeletal muscle regeneration. *PNAS* (2014) doi:10.1073/pnas.1401732111.

220. Regan, C. P. *et al.* Erk5 null mice display multiple extraembryonic vascular and embryonic cardiovascular defects. *PNAS* (2002).

221. Mottet, D. *et al.* Histone Deacetylase 7 Silencing Alters Endothelial Cell Migration, a Key Step in Angiogenesis. *Molecular Medicine* (2007) doi:10.1161/CIRCRESAHA.107.149377.

222. Chang, S. *et al.* Histone Deacetylase 7 Maintains Vascular Integrity by Repressing Matrix Metalloproteinase 10. *Cell* **126**, 321–334 (2006).

223. Lu, Y. W. *et al.* MEF2 (Myocyte Enhancer Factor 2) Is Essential for Endothelial Homeostasis and the Atheroprotective Gene Expression Program. *Arterioscler Thromb Vasc Biol* **41**, 1105–1123 (2021).

224. Preiss, A., Rosenberg, U. B., Kienlin, A., Seifert, E. & Jackie, H. *Molecular genetics of Kriippel, a gene required for segmentation of the Drosophila embryo*. *Antimicrob. Agents Chemother* vol. 38 (1984).

225. Atkins, G. B. & Jain, M. K. Role of Krüppel-like transcription factors in endothelial biology. *Circ Res* **100**, 1686–1695 (2007).

226. Bieker, J. J. Krüppel-like factors: three fingers in many pies. *J Biol Chem* **276**, 34355–8 (2001).

227. Novodvorsky, P. & Chico, T. J. A. The role of the transcription factor KLF2 in vascular development and disease. *Prog Mol Biol Transl Sci* **124**, 155–188 (2014).

228. Feinberg, M. W., Lin, Z., Fisch, S. & Jain, M. K. An emerging role for Krüppel-like factors in vascular biology. *Trends in Cardiovascular Medicine* Preprint at https://doi.org/10.1016/j.tcm.2004.06.005 (2004).

229. Lee, J. S. *et al.* Klf2 Is an Essential Regulator of Vascular Hemodynamic Forces In Vivo. *Dev Cell* **11**, 845–857 (2006).

230. Novodvorsky, P. *et al.* Klf2ash317mutant zebrafish do not recapitulate morpholino-induced vascular and haematopoietic phenotypes. *PLoS One* **10**, e0141611 (2015).

231. Kuo, C. T. *et al.* The LKLF transcription factor is required for normal tunica media formation and blood vessel stabilization during murine embryogenesis. *Genes Dev* **11**, 2996–3006 (1997).

232. Chiplunkar, A. R., Lung, T. K., Alhashem, Y., Koppenhaver, B. A. & Salloum, F. N. Krü ppel-Like Factor 2 Is Required for Normal Mouse Cardiac Development. *PLoS One* **8**, 54891 (2013).

233. Chiplunkar, A. R. *et al.* The Krüppel-like factor 2 and Krüppel-like factor 4 genes interact to maintain endothelial integrity in mouse embryonic vasculogenesis. *BMC Dev Biol* **13**, 1–9 (2013).

234. Sangwung, P. *et al.* KLF2 and KLF4 control endothelial identity and vascular integrity. *JCI Insight* **2**, e91700 (2017).

235. Segre, J. A., Bauer, C. & Fuchs, E. Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nature Genetics 1999 22:4* **22**, 356–360 (1999).

236. Xu, Z. *et al.* Transcription factor MEF2C suppresses endothelial cell inflammation via regulation of NF-kappaB and KLF2. *J Cell Physiol* **230**, 1310–1320 (2015).

237. Jha, P. & Das, H. KLF2 in Regulation of NF-kappaB-Mediated Immune Cell Function and Inflammation. *Int J Mol Sci* **18**, (2017).

238. Xu, Z. *et al.* Transcription Factor MEF2C Suppresses Endothelial Cell Inflammation via Regulation of NFkB and KLF2. *J Cell Physiol* **230**, 1310–1320 (2015).

239. Pathak, R. *et al.* IKKbeta regulates endothelial thrombomodulin in a Klf2-dependent manner. *J Thromb Haemost* **12**, 1533–1544 (2014).

240. Lin, Z. *et al.* KLF2 regulates endothelial barrier function. *Arterioscler Thromb Vasc Biol* **30**, 1952–1959 (2010).

241. Sangwung, P. *et al.* KLF2 and KLF4 control endothelial identity and vascular integrity. *JCI Insight* **2**, (2017).

242. SenBanerjee, S. *et al.* KLF2 Is a novel transcriptional regulator of endothelial proinflammatory activation. *J Exp Med* **199**, 1305–1315 (2004).

243. Nayak, L., Lin, Z. & Jain, M. K. *'Go with the flow': how Krüppel-like factor 2 regulates the vasoprotective effects of shear stress. Antioxidants & redox signaling* vol. 15 (2011).

244. Jain, M. K., Sangwung, P. & Hamik, A. Regulation of an inflammatory disease: Krüppel-like factors and atherosclerosis. *Arterioscler Thromb Vasc Biol* **34**, 499–508 (2014).

245. Sohn, S. J., Li, D., Lee, L. K. & Winoto, A. Transcriptional Regulation of Tissue-Specific Genes by the ERK5 Mitogen-Activated Protein Kinase. *Mol Cell Biol* **25**, 8553–8566 (2005).

246. Parmar, K. M. *et al.* Integration of flow-dependent endothelial phenotypes by Kruppel-like factor 2. *Journal of Clinical Investigation* **116**, 49–58 (2006).

247. Wang, W. *et al.* Fluid shear stress stimulates phosphorylation-dependent nuclear export of HDAC5 and mediates expression of KLF2 and eNOS. *Blood* **115**, 2971–2979 (2010).

248. Lee, J. S. *et al.* Klf2 is an essential regulator of vascular hemodynamic forces in vivo. *Dev Cell* **11**, 845–857 (2006).

249. Dekker, R. J. *et al.* Endothelial KLF2 links local arterial shear stress levels to the expression of vascular tone-regulating genes. *Am J Pathol* **167**, 609–618 (2005).

250. Dai, G. *et al.* Distinct endothelial phenotypes evoked by arterial waveforms derived from atherosclerosis-susceptible and -resistant regions of human vasculature. *PNAS* (2004).

251. Nigro, P., Abe, J.-I. & Berk, B. C. Flow shear stress and atherosclerosis: a matter of site specificity. *Antioxid Redox Signal* **15**, 1405–1414 (2011).

252. Schwartz, M. A., Vestweber, D. & Simons, M. A unifying concept in vascular health and disease. *Science (1979)* **360**, 270–271 (2018).

253. Atkins, G. B. *et al.* Hemizygous deficiency of krüppel-like factor 2 augments experimental atherosclerosis. *Circ Res* **103**, 690–693 (2008).

254. Zhou, G. *et al.* Endothelial Kruppel-like factor 4 protects against atherothrombosis in mice. *J Clin Invest* **122**, 4727–4731 (2012).

255. Shi, H. *et al.* Kruppel-like factor 2 protects against ischemic stroke by regulating endothelial blood brain barrier function. *Am J Physiol Heart Circ Physiol* **304**, (2013).

256. Bondareva, O. *et al.* Identification of atheroprone shear stress responsive regulatory elements in endothelial cells. *Cardiovasc Res* **115**, 1487–1499 (2019).

257. Ajami, N. E. *et al.* Systems biology analysis of longitudinal functional response of endothelial cells to shear stress. *Proceedings of the National Academy of Sciences* **114**, 201707517 (2017).

258. Kuhlencordt, P. J. *et al.* Role of endothelial nitric oxide synthase in endothelial activation: Insights from eNOS knockout endothelial cells. *Am J Physiol Cell Physiol* **286**, (2004).

259. McCormick, S. M. *et al.* DNA microarray reveals changes in gene expression of shear stressed human umbilical vein endothelial cells. *Proc Natl Acad Sci U S A* **98**, 8955–8960 (2001).

260. Sorescu, G. P. *et al.* Bone morphogenic protein 4 produced in endothelial cells by oscillatory shear stress stimulates an inflammatory response. *Journal of Biological Chemistry* **278**, 31128–31135 (2003).

261. Wang, X.-L., Fu, A., Raghavakaimal, S. & Lee, H.-C. Proteomic analysis of vascular endothelial cells in response to laminar shear stress. *Proteomics* **7**, 588–596 (2007).

262. Maimari, N., Pedrigi, R. M., Russo, A., Broda, K. & Krams, R. Integration of flow studies for robust selection of mechanoresponsive genes. *Thromb Haemost* **115**, 474–483 (2016).

263. Li, Y. S. J., Haga, J. H. & Chien, S. Molecular basis of the effects of shear stress on vascular endothelial cells. *Journal of Biomechanics* vol. 38 1949–1971 Preprint at https://doi.org/10.1016/j.jbiomech.2004.09.030 (2005).

264. Kusinski, L. C. *et al.* eNOS knockout mouse as a model of fetal growth restriction with an impaired uterine artery function and placental transport phenotype. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **303**, R86–R93 (2012).

265. Yanagisawa, H., Clouthier, D. E. & Yanagisawa, M. Role of Endothelin-1/Endothelin-A receptor-mediated signaling pathway in the aortic arch patterning in mice. *J Clin Invest* **102**, 22–33 (1998).

266. Chiplunkar, A. R. *et al.* The Krüppel-like factor 2 and Krüppel-like factor 4 genes interact to maintain endothelial integrity in mouse embryonic vasculogenesis. *BMC Dev Biol* **13**, 40 (2013).

267. Atochin, D. N. & Huang, P. L. Endothelial nitric oxide synthase transgenic models of endothelial dysfunction. *Pflugers Archiv European Journal of Physiology* vol. 460 965–974 Preprint at https://doi.org/10.1007/s00424-010-0867-4 (2010).

268. Kisanuki, Y. Y. *et al.* Low blood pressure in endothelial cell-specific endothelin 1 knockout mice. *Hypertension* **56**, 121–128 (2010).

269. Bagnall, A. J. *et al.* Deletion of Endothelial Cell Endothelin B Receptors Does Not Affect Blood Pressure or Sensitivity to Salt. **48**, 286–293 (2006).

270. Nakajima, H. & Mochizuki, N. Flow pattern-dependent endothelial cell responses through transcriptional regulation. *Cell Cycle* **16**, 1893–1901 (2017).

271. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional Regulatory Elements in the Human Genome. *Annual Reviews* **7**, 29–59 (2006).

272. Kumar, A., Lin, Z., Senbanerjee, S. & Jain, M. K. Tumor Necrosis Factor Alpha-Mediated Reduction of KLF2 Is Due to Inhibition of MEF2 by NF-B and Histone Deacetylases. *Mol Cell Biol* **25**, 5893–5903 (2005).

273. Sebzda, E., Zou, Z., Lee, J. S., Wang, T. & Kahn, M. L. Transcription factor KLF2 regulates the migration of naive T cells by restricting chemokine receptor expression patterns. *Nat Immunol* (2008) doi:10.1038/ni1565.

274. Yin, K. J., Hamblin, M., Fan, Y., Zhang, J. & Chen, Y. E. Krüpple-like factors in the central nervous system: novel mediators in Stroke. *Metab Brain Dis* **30**, 401–410 (2015).

275. Becker, P. W. *et al.* An intronic Flk1 enhancer directs arterial-specific expression via RBPJ-mediated venous repression. *Arterioscler Thromb Vasc Biol* **36**, 1209–1219 (2016).

276. Becker, P. W. *et al.* An Intronic *Flk1* Enhancer Directs Arterial-Specific Expression via RBPJ-Mediated Venous RepressionHighlights. *Arterioscler Thromb Vasc Biol* **36**, 1209–1219 (2016).

277. Payne, S. *et al.* Regulatory pathways governing murine coronary vessel formation are dysregulated in the injured adult heart. *Nat Commun* **10**, 1–19 (2019).

278. Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. Enhancers: five essential questions. *Nature Perspectives* (2013) doi:10.1038/nrg3458.

279. Yáñez-Cuna, J. O., Kvon, E. Z. & Stark, A. Deciphering the transcriptional cis-regulatory code. *Trends in Genetics* vol. 29 11–22 Preprint at https://doi.org/10.1016/j.tig.2012.09.007 (2013).

280. Calo, E. & Wysocka, J. Modification of Enhancer Chromatin: What, How, and Why? *Mol Cell* **49**, 825–837 (2013).

281. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters ENCODE Encyclopedia of DNA Elements. *Nature* **489**, (2012).

282. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics* vol. 20 207–220 Preprint at https://doi.org/10.1038/s41576-018-0089-8 (2019).

283. Wang, R. *et al.* FBW7 regulates endothelial functions by targeting KLF2 for ubiquitination and degradation. *Cell Res* **23**, 803–819 (2013).

284. Pennacchio, L. A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).

285. Lieschke, G. J. & Currie, P. D. Animal models of human disease: Zebrafish swim into view. *Nature Reviews Genetics* vol. 8 353–367 Preprint at https://doi.org/10.1038/nrg2091 (2007).

286. Quaife, N. M., Watson, O. & Chico, T. J. A. Zebrafish: an emerging model of vascular development and remodelling. *Curr Opin Pharmacol* **12**, 608–614 (2012).

287. Ishibashi, M., Mechaly, A. S., Becker, T. S. & Rinkwitz, S. Using zebrafish transgenesis to test human genomic sequences for specific enhancer activity. *Methods* **62**, 216–225 (2013).

288. Kvon, E. Z. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics* vol. 106 185–192 Preprint at https://doi.org/10.1016/j.ygeno.2015.06.007 (2015).

289. Jovana Serbanovic-Canic, Amalia de Luca, Christina Warboys, Pedro F. Ferreira, L. A. L., Sarah Hsiao, Ismael Gauci, Marwa Mahmoud, Shuang Feng, Celine Souilhol, N. B., John-Paul Ashton, Henning Walczak, David Firmin, Rob Krams, J. C. M. & Dorian O. Haskard, Spencer Sherwin, Victoria Ridger, Timothy J.A. Chico, P. C. E. Zebrafish Model For Functional Screening of Flow-Responsive Genes. *Arterioscler Thromb Vasc Biol* (2016) doi:10.1161/ATVBAHA.116.308502/-/DC1.

290. Isogai, S., Horiguchi, M. & Weinstein, B. M. The Vascular Anatomy of the Developing Zebrafish: An Atlas of Embryonic and Early Larval Development. *Dev Biol* (2001) doi:10.1006/dbio.2000.9995.

291. Begeman, I. J., Emery, B., Kurth, A. & Kang, J. Regeneration and developmental enhancers are differentially compatible with minimal promoters. *Dev Biol* (2022) doi:10.1016/J.YDBIO.2022.09.007.

292. Clark, K. J., Urban, M. D., Skuster, K. J. & Ekker, S. C. Transgenic Zebrafish Using Transposable Elements. *Methods Cell Biology* (2011) doi:10.1016/B978-0-12-374814-0.00008-2.

293. Nakajima, H. & Mochizuki, N. Flow pattern-dependent endothelial cell responses through transcriptional regulation. *Cell Cycle* **16**, 1893–1901 (2017).

294. Wang, L. L. *et al.* Ablood flow-dependent klf2a-NO signaling cascade is required for stabilization of hematopoietic stem cell programming in zebrafish embryos. *Blood* **118**, 4102–4110 (2011).

295. Nicoli, S. *et al.* MicroRNA-mediated integration of haemodynamics and Vegf signalling during angiogenesis. *Nature* **464**, 1196–1200 (2010).

296. Antonellis, A. *et al.* Identification of Neural Crest and Glial Enhancers at the Mouse Sox10 Locus through Transgenesis in Zebrafish. *PLoS Genet* **4**, e1000174 (2008).

297. Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L. & McCallion, A. S. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science (1979)* **312**, 276–279 (2006).

298. Postlethwait, J. H. *et al.* Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* **10**, 1890–1902 (2000).

299. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).

300. Xue, Y., Gao, S. & Liu, F. Genome-wide analysis of the zebrafish Klf family identifies two genes important for erythroid maturation. *Dev Biol* **403**, 115–127 (2015).

301. Li, L. & Wunderlich, Z. An enhancer's length and composition are shaped by its regulatory task. *Front Genet* **8**, (2017).

302. Chatterjee, S., Bourque, G. & Lufkin, T. Conserved and non-conserved enhancers direct tissue specific transcription in ancient germ layer specific developmental control genes. *BMC Dev Biol* **11**, 63 (2011).

303. Culshaw, L. H., Savery, D., Greene, N. D. E. & Copp, A. J. Mouse whole embryo culture: Evaluating the requirement for rat serum as culture medium. *Birth Defects Res* **111**, 1165–1177 (2019).

304. Brachtendorf, G. *et al.* Early expression of endomucin on endothelium of the mouse embryo and on putative hematopoietic clusters in the dorsal aorta. *Developmental Dynamics* **222**, 410–419 (2001).

305. Kothary, R. *et al.* Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105**, 707–714 (1989).

306. Kang, J. *et al.* Modulation of tissue repair by regeneration enhancer elements. (2016) doi:10.1038/nature17644.

307. Song, D. L., Chalepakis, G., Gruss, P. & Joyner, A. L. Two Pax-binding sites are required for early embryonic brain expression of an Engrailed-2 transgene. *Development* **122**, 627–635 (1996).

308. Ji, R. P. *et al.* Onset of Cardiac Function During Early Mouse Embryogenesis Coincides With Entry of Primitive Erythroblasts Into the Embryo Proper Results UBM-Doppler Imaging of Early-Stage Embryos Initiation of Early Cardiovascular Function and Correlation With Erythroblast Entry. *Circ Res* **92**, 133–135 (2003).

309. McGrath, K. E. *et al.* Circulation is established in a stepwise pattern in the mammalian embryo. *Blood* **101**, 1669–76 (2003).

310. Serebriiskii, I. G. & Golemis, E. A. Uses of lacZ to Study Gene Function: Evaluation of β-Galactosidase Assays Employed in the Yeast Two-Hybrid System. *Anal Biochem* **285**, 1–15 (2000).

311. Kaufman, M. H. *The atlas of mouse development*. (Academic Press, 1992).

312. EMAP Home. http://www.emouseatlas.org/emap/home.html.

313. Augustin, H. G. & Koh, Y. Organotypic vasculature: From descriptive heterogeneity to functional pathophysiology. *Science (1979)* (2017) doi:10.1126/science.aal2379.

314. Marcu, R. *et al.* Human Organ-Specific Endothelial Cell Heterogeneity. *iScience* **4**, 20–35 (2018).

315. Jambusaria, A. *et al.* Endothelial heterogeneity across distinct vascular beds during homeostasis and inflammation. *Elife* **9**, (2020).

316. Schrick, J. J., Hughes, M. J., Anderson, K. P., Croyle, M. L. & Lingrel, J. B. Characterization of the lung Kruppel-like transcription factor gene and upstream regulatory elements. *Gene* **236**, 185–195 (1999).

317. Huddleson, J. P., Srinivasan, S., Ahmad, N. & Lingrel, J. B. Fluid shear stress induces endothelial KLF2 gene expression through a defined promoter region. *Biol Chem* **385**, 723–729 (2004).

318. Huddleson, J. P., Ahmad, N., Srinivasan, S. & Lingrel, J. B. Induction of KLF2 by fluid shear stress requires a novel promoter element activated by a phosphatidylinositol 3-kinase-dependent chromatin-remodeling pathway. *Journal of Biological Chemistry* **280**, 23371–23379 (2005).

319. Lin, Z. *et al.* Kruppel-like factor 2 (KLF2) regulates endothelial thrombotic function. *Circ Res* **96**, e48-57 (2005).

320. Levine, M. Transcriptional enhancers in animal development and evolution. *Current Biology* vol. 20 R754 Preprint at https://doi.org/10.1016/j.cub.2010.06.070 (2010).

321. Engelbrecht, E. *et al.* Sphingosine 1-phosphate-regulated transcriptomes in heterogenous arterial and lymphatic endothelium of the aorta. *Elife* (2020) doi:10.7554/eLife.52690.

322. Zhou, P. *et al.* Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. *Elife* **6**, (2017).

323. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, 1–17 (2017).

324. Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A* **93**, 12355 (1996).

325. Wakimoto, K. *et al.* Targeted disruption of Na+/Ca2+ exchanger gene leads to cardiomyocyte apoptosis and defects in heartbeat. *J Biol Chem* **275**, 36991–36998 (2000).

326. Shalaby, F. *et al.* Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* **376**, 62–66 (1995).

327. Ivins, S. *et al.* The CXCL12/CXCR4 Axis Plays a Critical Role in Coronary Artery Development. *Dev Cell* **33**, 455–468 (2015).

328. Macgrogan, D. *et al.* How to Make a Heart Valve: From Embryonic Development to Bioengineering of Living Valve Substitutes. doi:10.1101/cshperspect.a013912.

329. Garcia, M. D. & Larina, I. v. Vascular development and hemodynamic force in the mouse yolk sac. *Front Physiol* **5 AUG**, 308 (2014).

330. Galie, P. A. *et al.* Fluid shear stress threshold regulates angiogenic sprouting. **111**, (2014).

331. Jones, E. A. v, Baron, M. H., Fraser, S. E. & Dickinson, M. E. Imaging of Mammalian Hematovascular Development Dynamic In Vivo Imaging of Mammalian Hematovascular Development Using Whole Embryo Culture. in *Methods in Molecular Medicine* vol. 105 (2010).

332. Bhattacharya, R. *et al.* Inhibition of vascular permeability factor/vascular endothelial growth factor-mediated angiogenesis by the Kruppel-like factor KLF2. *J Biol Chem* **280**, 28848–51 (2005).

333. Kawanami, D. *et al.* Kruppel-like factor 2 inhibits hypoxia-inducible factor 1alpha expression and function in the endothelium. *J Biol Chem* **284**, 20522–20530 (2009).

334. Wu, J., Bohanan, C. S., Neumann, J. C. & Lingrel, J. B. KLF2 transcription factor modulates blood vessel maturation through smooth muscle cell migration. *J Biol Chem* **283**, 3942–3950 (2008).

335. Meadows, S. M., Salanga, M. C. & Krieg, P. A. Kruppel-like factor 2 cooperates with the ETS family protein ERG to activate Flk1 expression during vascular development. *Development* **136**, 1115–1125 (2009).

336. Renz, M. *et al.* Regulation of β1 Integrin-Klf2-Mediated Angiogenesis by CCM Proteins. *Dev Cell* **32**, 181–190 (2015).

337. Zheng, X. *et al.* Key role of microRNA-15a in the KLF4 suppressions of proliferation and angiogenesis in endothelial and vascular smooth muscle cells. *Biochem Biophys Res Commun* **437**, 625–631 (2013).

338. Wang, Y. *et al.* KLF4 promotes angiogenesis by activating VEGF signaling in human retinal microvascular endothelial cells. *PLoS One* **10**, (2015).

339. Huddleson, J. P., Ahmad, N., Srinivasan, S. & Lingrel, J. B. Induction of KLF2 by fluid shear stress requires a novel promoter element activated by a phosphatidylinositol 3-kinase-dependent chromatin-remodeling pathway. *J Biol Chem* **280**, 23371–23379 (2005).

340. Chenna, R. *et al.* Multiple sequence alignment with the Clustal series of programs. *Nucleic Acid Research* (2003) doi:10.1093/nar/gkg500.

341. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. **23**, 2947–2948 (2007).

342. Thompson, J. D., Higgins, D. G. & Gibson, T. J. *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. *Nucleic Acids Research* vol. 22 https://academic.oup.com/nar/article/22/22/4673/2400290 (1994).

343. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res* **43**, W39–W49 (2015).

344. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol* **8**, 1–9 (2007).

345. Neal, A. *et al.* ETS factors are required but not sufficient for specific patterns of enhancer activity in different endothelial subtypes. *Dev Biol* **473**, 1–14 (2021).

346. Meadows, S. M., Myers, C. T. & Krieg, P. A. Regulation of endothelial cell development by ETS transcription factors. *Semin Cell Dev Biol* **22**, 976–984 (2011).

347. Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).

348. Woods, D. B., Ghysdael, J. & Owen, M. J. Identification of nucleotide preferences in DNA sequences recognised specifically by c-Ets-1 protein. *Nucleic Acids Res* **20**, 699–704 (1992).

349. Shah, A. v., Birdsey, G. M. & Randi, A. M. Regulation of endothelial homeostasis, vascular development and angiogenesis by the transcription factor ERG. *Vascul Pharmacol* **86**, 3–13 (2016).

350. Sissaoui, S. *et al.* Genomic characterization of endothelial enhancers reveals a multifunctional role for NR2F2 in regulation of arteriovenous gene expression. *Circ Res* 875–888 (2020) doi:10.1161/CIRCRESAHA.119.316075.

351. Philipp W. Becker,* Natalia Sacilotto,* Svanhild Nornes, Alice Neal, Max O. Thomas, Ke Liu, Chris Preece, Indrika Ratnayaka, Benjamin Davies, George Bou-Gharios, S. D. V. An Intronic Flk1 Enhancer Directs Arterial-Specific Expression via RBPJ-Mediated Venous Repression. *Arterioscler Thromb Vasc Biol* **36**, 1209–1219 (2016).

352. Gonzalez-Lopez, O. *et al.* A herpesvirus transactivator and cellular POU proteins extensively regulate DNA binding of the host Notch signaling protein RBP-J to the virus genome. *Journal of Biological Chemistry* **294**, 13073–13092 (2019).

353. Chèneby, J. *et al.* ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res* **46**, 267–275 (2017).

354. Mack, J. J. & Luisa Iruela-Arispe, M. NOTCH regulation of the endothelial cell phenotype. *Curr Opin Hematol* (2018) doi:10.1097/MOH.0000000000000425.

355. Nielsen, C. M. *et al.* Deletion of Rbpj from postnatal endothelium leads to abnormal arteriovenous shunting in mice. *Development* (2014) doi:10.1242/dev.108951.

356. Poulsen, L. L. C. *et al.* Inhibition of Endothelial NOTCH1 Signaling Attenuates Inflammation by Reducing Cytokine-Mediated Histone Acetylation at Inflammatory Enhancers. *Arterioscler Thromb Vasc Biol* **38**, 854–869 (2018).

357. Wang, S. *et al.* A dynamic and integrated epigenetic program at distal regions orchestrates transcriptional responses to VEGFA. *Genome Res* (2019) doi:10.1101/gr.239053.118.

358. Wada, H. *et al.* Calcineurin-GATA-6 pathway is involved in smooth muscle-specific transcription. *J Cell Biol* **156**, 983–991 (2002).

359. Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).

360. Kulkarni, R. M., Greenberg, J. M. & Akeson, A. L. NFATc1 regulates lymphatic endothelial development. *Mech Dev* **126**, 350–365 (2009).

361. Wu, B., Baldwin, H. S. & Zhou, B. Nfatc1 directs the endocardial progenitor cells to make heart valve primordium. *Trends Cardiovasc Med* **23**, 294–300 (2013).

362. Suehiro, J. I. *et al.* Genome-wide approaches reveal functional vascular endothelial growth factor (VEGF)-inducible nuclear factor of activated T cells (NFAT) c1 binding to angiogenesis-related genes in the endothelium. *Journal of Biological Chemistry* **289**, 29044–29059 (2014).

363. Anderson, C. M. *et al.* Cooperative activation of cardiac transcription through myocardin bridging of paired MEF2 sites. (2017) doi:10.1242/dev.138487.

364. Sohn, S. J., Li, D., Lee, L. K. & Winoto, A. Transcriptional Regulation of Tissue-Specific Genes by the ERK5 Mitogen-Activated Protein Kinase. *Mol Cell Biol* **25**, 8553–8566 (2005).

365. Kappel, A. *et al.* Role of SCL/Tal-1, GATA, and Ets transcription factor binding sites for the regulation of Flk-1 expression during murine vascular development. *Blood* **96**, 3078–3085 (2000).

366. Whitcomb, J., Gharibeh, L. & Nemer, M. From embryogenesis to adulthood: Critical role for GATA factors in heart development and function. *Life* (2019) doi:10.1002/iub.2163.

367. Rivera-Feliciano, J. *et al.* Development of heart valves requires Gata4 expression in endothelial-derived cells. *Development* **133**, 3607–3618 (2006).

368. The Journal of Clinical Investigation. doi:10.1172/JCI78888.

369. Linnemann, A. K., O'geen, H., Keles, S., Farnham, P. J. & Bresnick, E. H. Genetic framework for GATA factor function in vascular biology. *PNAS* (2011) doi:10.1073/pnas.1108440108.

370. Mahamud, M. R. *et al.* GATA2 controls lymphatic endothelial cell junctional integrity and lymphovenous valve morphogenesis through miR-126. *Development* (2019) doi:10.1242/dev.184218.

371. Merika, M. & Orkin, S. H. DNA-binding specificity of GATA family transcription factors. *Mol Cell Biol* **13**, 3999–4010 (1993).

372. Linnemann, A. K., O'Geen, H., Keles, S., Farnham, P. J. & Bresnick, E. H. Genetic framework for GATA factor function in vascular biology. *Proceedings of the National Academy of Sciences* **108**, 13641–13646 (2011).

373. Parmar, K. M. *et al.* Statins Exert Endothelial Atheroprotective Effects via the KLF2 Transcription Factor. *Journal of Biological Chemistry* **280**, 26714–26719 (2005).

374. Kumar, A., Lin, Z., SenBanerjee, S. & Jain, M. K. Tumor necrosis factor alpha-mediated reduction of KLF2 is due to inhibition of MEF2 by NF-kappaB and histone deacetylases. *Mol Cell Biol* **25**, 5893–5903 (2005).

375. Huddleson, J. P., Ahmad, N. & Lingrel, J. B. Up-regulation of the KLF2 transcription factor by fluid shear stress requires nucleolin. *J Biol Chem* **281**, 15121–15128 (2006).

376. Maejima, T. *et al.* Direct Evidence for Pitavastatin Induced Chromatin Structure Change in the KLF4 Gene in Endothelial Cells. *PlosONE* (2014) doi:10.1371/journal.pone.0096005.

377. Denis, J. F. *et al.* KLF4-induced connexin40 expression contributes to arterial endothelial quiescence. *Front Physiol* **10**, 80 (2019).

378. Villarreal Jr, G. *et al.* Defining the Regulation of Klf4 Expression and Its Downstream Transcriptional Targets in Vascular Endothelial Cells. *Biochem Biophys Res Commun* **391**, 984–989 (2010).

379. Clark, P. R. *et al.* MEK5 is Activated by Shear Stress, Activates ERK5 and Induces KLF4 to Modulate TNF Responses in Human Dermal Microvascular Endothelial Cells. *Microcirculation* **18**, 102–117 (2011).

380. Lu, J., Mckinsey, T. A., Nicol, R. L., Olson, E. N. & Mcknight, S. L. Signal-dependent activation of the MEF2 transcription factor by dissociation from histone deacetylases. *PNAS* (2000).

381. Wu, W., de Folter, S., Shen, X., Zhang, W. & Tao, S. Vertebrate Paralogous MEF2 Genes: Origin, Conservation, and Evolution. *PlosONE* (2011) doi:10.1371/journal.pone.0017334.

382. Molkentin, J. D., Li, L. & Olson, E. N. Phosphorylation of the MADS-Box transcription factor MEF2C enhances its DNA binding activity. *J Biol Chem* **271**, 17199–17204 (1996).

383. Molkentin, J. D., Black, B. L., Martin, J. F. & Olson, E. N. Mutational Analysis of the DNA Binding, Dimerization, and Transcriptional Activation Domains of MEF2C. *Mol Cell Biol* **16**, 2627–2636 (1996).

384. Papanicolaou, K. N., Izumiya, Y. & Walsh, K. Forkhead Transcription Factors and Cardiovascular Biology. *Circ Res* (2008) doi:10.1161/CIRCRESAHA.107.164186.

385. Sabine, A. *et al.* FOXC2 and fluid shear stress stabilize postnatal lymphatic vasculature. *J Clin Invest* **125**, 3861–3877 (2015).

386. Andrade, J. *et al.* Control of endothelial quiescence by FOXO-regulated metabolites. *Nat Cell Biol* **23**, (2021).

387. Pfleger, J. *et al.* Genomic binding patterns of forkhead box protein O1 reveal its unique role in cardiac hypertrophy. *Circulation* 882–898 (2020) doi:10.1161/CIRCULATIONAHA.120.046356.

388. Oh, S. Y., Kim, J. Y. & Park, C. The ETS factor, ETV2: A master regulator for vascular endothelial cell development. *Mol Cells* **38**, 1029–1036 (2015).

389. Kato, Y. *et al.* BMK1/ERK5 regulates serum-induced early gene expression through transcription factor MEF2C. *EMBO J* **16**, 7054–7066 (1997).

390. Ren, B. *et al.* Developmental Perspectives on Arterial Fate Specification. *Front Cell Dev Biol* (2021) doi:10.3389/fcell.2021.691335.

391. (No Title). doi:10.1242/dev.02519.

392. Kazenwadel, J. *et al.* GATA2 is required for lymphatic vessel valve development and maintenance. *J Clin Invest* **125**, 2979 (2015).

393. Sokabe, T. *et al.* Differential regulation of urokinase-type plasminogen activator expression by fluid shear stress in human coronary artery endothelial cells. *Am J Physiol Heart Circ Physiol* **287**, (2004).

394. Thilo, F. *et al.* Pulsatile atheroprone shear stress affects the expression of transient receptor potential channels in human endothelial cells. *Hypertension* **59**, 1232–1240 (2012).

395. Zhou, B. *et al.* Characterization of Nfatc1 regulation identifies an enhancer required for gene expression that is specific to pro-valve endocardial cells in the developing heart. *Development* **132**, 1137–1146 (2005).

396. Wu, B., Baldwin, H. S. & Zhou, B. Nfatc1 directs the endocardial progenitor cells to make heart valve primordium. *Trends Cardiovasc Med* **23**, 294–300 (2013).

397. Gunawan, F., Gentile, A., Gauvrit, S., Stainier, D. Y. R. & Bensimon-Brito, A. Nfatc1 promotes interstitial cell formation during cardiac valve development in Zebrafish. *Circ Res* 968–984 (2020) doi:10.1161/CIRCRESAHA.119.315992.

398. Celil Aydemir, A. B. *et al.* Nuclear factor of activated T cells mediates fluid shear stress- and tensile strain-induced Cox2 in human and murine bone cells. *Bone* **46**, 167–175 (2010).

399. Ding, N. *et al.* Fluid shear stress promotes osteoblast proliferation through the NFATc1-ERK5 pathway. *Connect Tissue Res* (2018) doi:10.1080/03008207.2018.1459588.

400. Wang, Y. X. *et al.* A multi-component parallel-plate flow chamber system for studying the effect of exercise-induced wall shear stress on endothelial cells. *Biomed Eng Online* **15**, 659–672 (2016).

401. Brown, D. C. & Larson, R. S. Improvements to parallel plate flow chambers to reduce reagent and cellular requirements. *BMC Immunol* **2**, 1–7 (2001).

402. Wong, A. K., LLanos, P., Boroda, N., Rosenberg, S. R. & Rabbany, S. Y. A Parallel-Plate Flow Chamber for Mechanical Characterization of Endothelial Cells Exposed to Laminar Shear Stress. *Cell Mol Bioeng* **9**, 127–138 (2016).

403. Fallon, M. E., Mathews, R. & Hinds, M. T. In Vitro Flow Chamber Design for the Study of Endothelial Cell (Patho)Physiology. *J Biomech Eng* (2022) doi:10.1115/1.4051765.

404. DeStefano, J. G. *et al.* Real-time quantification of endothelial response to shear stress and vascular modulators. *Integr Biol (Camb)* **9**, 362 (2017).

405. Thilo, F. *et al.* Pulsatile atheroprone shear stress affects the expression of transient receptor potential channels in human endothelial cells. *Hypertension* **59**, 1232–1240 (2012).

406. Liu, Z. *et al.* Single-Cell RNA Sequencing Reveals Endothelial Cell Transcriptome Heterogeneity under Homeostatic Laminar Flow. *Arterioscler Thromb Vasc Biol* **41**, 2575–2584 (2021).

407. Hong, D., Jaron, D., Buerk, D. G. & Barbee, K. A. Heterogeneous response of microvascular endothelial cells to shear stress. *Am J Physiol Heart Circ Physiol* **290**, 2498–2508 (2006).

408. Deng, H. *et al.* Activation of Smad2/3 signaling by low fluid shear stress mediates artery inward remodeling. *PNAS* (2021) doi:10.1073/pnas.2105339118/-/DCSupplemental.

409. Sheikh, S., Rainger, G. E., Gale, Z., Rahman, M. & Nash, G. B. Exposure to fluid shear stress modulates the ability of endothelial cells to recruit neutrophils in response to tumor necrosis factor-α: a basis for local variations in vascular sensitivity to inflammation. *Blood* **102**, 2828–2834 (2003).

410. Mehta, V. *et al.* The guidance receptor plexin D1 is a mechanosensor in endothelial cells. *290 | Nature |* **578**, (1979).

411. Papaioannou, T. G. & Stefanadis, C. Vascular wall shear stress: basic principles and methods. *Hellenic J Cardiol* **46**, 9–15 (2005).

412. Yap, C. H., Saikrishnan, N., Tamilselvan, G. & Yoganathan, A. P. Experimental measurement of dynamic fluid shear stress on the aortic surface of the aortic valve leaflet. *Biomech Model Mechanobiol* **11**, 171–182 (2012).

413. Culshaw, L. H., Savery, D., Greene, N. D. E. & Copp, A. J. Mouse whole embryo culture: Evaluating the requirement for rat serum as culture medium. *Birth Defects Res* (2019) doi:10.1002/bdr2.1538.

414. Garcia, M. D. & Larina, I. v. Vascular development and hemodynamic force in the mouse yolk sac. *Front Physiol* **5 AUG**, 308 (2014).

415. Nam, D. *et al.* Innovative Methodology Partial carotid ligation is a model of acutely induced disturbed flow, leading to rapid endothelial dysfunction and atherosclerosis. *Am J Physiol Heart Circ Physiol* **297**, 1535–1543 (2009).

416. Kuhlmann, M. T. *et al.* Implantation of a carotid cuff for triggering shear-stress induced atherosclerosis in mice. *J Vis Exp* 1–6 (2012) doi:10.3791/3308.

417. Feintuch, A. *et al.* Hemodynamics in the mouse aortic arch as assessed by MRI, ultrasound, and numerical modeling. *Am J Physiol Heart Circ Physiol* **292**, 884–892 (2007).

418. Trachet, B. *et al.* Wall shear stress in the mouse aortic arch : Does size matter? *IFMBE Proc* **22**, 1994–1998 (2008).

419. Suo, J. *et al.* Hemodynamic shear stresses in mouse aortas: Implications for atherogenesis. *Arterioscler Thromb Vasc Biol* **27**, 346–351 (2007).

420. Dekker, R. J. *et al. Endothelial KLF2 Links Local Arterial Shear Stress Levels to the Expression of Vascular Tone-Regulating Genes*. *American Journal of Pathology* vol. 167 http://frodo.wi.mit.edu/cgi-bin/primer3/primer3www.cgi (2005).

421. Abrahamsson, C. *et al.* Induction of rhythm abnormalities in the fetal rat heart. A tentative mechanism for the embryotoxic effect of the class I11 antiarrhythmic agent almokalant. *Cardiovasc Res* **28**, 337–344 (1994).

422. H, R. *et al.* The Effect of Dofetilide on the Heart Rate of GD11 and GD13 Rat Embryos, in vivo, Using Ultrasound. *Birth Defects Res B Dev Reprod Toxicol* **104**, 196–203 (2015).

423. P. Gill-Kumar. Dofetilide fetotoxicity study. *https://www.accessdata.fda.gov/drugsatfda_docs/nda/99/20-931_Tikosyn_pharmr_P1.pdf* (1999).

424. Chapman, G. *et al.* Functional genomics and gene-environment interaction highlight the complexity of congenital heart disease caused by Notch pathway variants. *Hum Mol Genet* **29**, 566–579 (2020).

425. Tyser, R. C. V. *et al.* Calcium handling precedes cardiac differentiation to initiate the first heartbeat. *Elife* **5**, (2016).

426. Ji, R. P. *et al.* Onset of cardiac function during early mouse embryogenesis coincides with entry of primitive erythroblasts into the embryo proper. *Circ Res* **92**, 133–135 (2003).

427. Walls, J. R., Coultas, L., Rossant, J. & Henkelman, R. M. Three-Dimensional Analysis of Vascular Development in the Mouse Embryo. *PLoS One* **3**, 2853 (2008).

428. von Gise, A. & Pu, W. T. Review Endocardial and Epicardial Epithelial to Mesenchymal Transitions in Heart Development and Disease. (2012) doi:10.1161/CIRCRESAHA.111.259960.

429. Barrett, C. J. *et al.* Increased Hemodynamic Load in Early Embryonic Stages Alters Endocardial to Mesenchymal Transition. *Frontiers in Physiology | www.frontiersin.org* **8**, 56 (2017).

430. McKinsey, T. A. Therapeutic Potential for HDAC Inhibitors in the Heart. *Annu Rev Pharmacol Toxicol* **52**, 303–319 (2012).

431. Tian, R. *et al.* Metformin ameliorates endotoxemia-induced endothelial pro-inflammatory responses via AMPK-dependent mediation of HDAC5 and KLF2. *Biochim Biophys Acta Mol Basis Dis* **1865**, 1701–1712 (2019).

432. Bu, D. *et al.* Statin-induced Kruppel-like factor 2 expression in human and mouse T cells reduces inflammatory and pathogenic responses. *J Clin Invest* **120**, 1961–1970 (2010).

433. Ali, F. *et al.* Statin-mediated cytoprotection of human vascular endothelial cells: a role for Kruppel-like factor 2-dependent induction of heme oxygenase-1. *J Thromb Haemost* **5**, 2537–2546 (2007).

434. Owens, A. P. 3rd & Mackman, N. The antithrombotic effects of statins. *Annu Rev Med* **65**, 433–445 (2014).

435. Trebicka, J. & Schierwagen, R. Statins, Rho GTPases and KLF2: New mechanistic insight into liver fibrosis and portal hypertension. *Gut* **64**, 1349–1350 (2015).

436. Sen-Banerjee, S. *et al.* Kruppel-like factor 2 as a novel mediator of statin effects in endothelial cells. *Circulation* **112**, 720–726 (2005).

437. Lin, Z. *et al.* Kruppel-like factor 2 regulates endothelial barrier function. *Arterioscler Thromb Vasc Biol* **30**, 1952–1959 (2010).

438. Thum, T. & Bauersachs, J. Sports or statins for atheroprotection? New insight from Kruppel-like factor 2. *Cardiovascular research* vol. 72 193–195 Preprint at https://doi.org/10.1016/j.cardiores.2006.08.016 (2006).

439. Niu, N., Xu, S., Xu, Y., Little, P. J. & Jin, Z.-G. Targeting Mechanosensitive Transcription Factors in Atherosclerosis. *Trends Pharmacol Sci* **40**, 253–266 (2019).

440. Fang, Y. & Davies, P. F. Site-specific microRNA-92a regulation of Kruppel-like factors 4 and 2 in atherosusceptible endothelium. *Arterioscler Thromb Vasc Biol* **32**, 979–987 (2012).

441. Shaked, I. & Ley, K. Protective role for myeloid specific KLF2 in atherosclerosis. *Circulation research* vol. 110 1266 Preprint at https://doi.org/10.1161/CIRCRESAHA.112.270991 (2012).

442. Alaiti, M. A., Orasanu, G., Tugal, D., Lu, Y. & Jain, M. K. Kruppel-like factors and vascular inflammation: Implications for atherosclerosis. *Curr Atheroscler Rep* **14**, 438–449 (2012).

443. Atkins, G. B. *et al.* Hemizygous deficiency of Kruppel-like factor 2 augments experimental atherosclerosis. *Circ Res* **103**, 690–693 (2008).

444. Deng, Y. *et al.* ERK5/KLF2 activation is involved in the reducing effects of puerarin on monocyte adhesion to endothelial cells and atherosclerotic lesion in apolipoprotein E-deficient mice. *Biochim Biophys Acta Mol Basis Dis* **1864**, 2590–2599 (2018).

445. Lingrel, J. B. *et al.* Myeloid-specific Kruppel-like factor 2 inactivation increases macrophage and neutrophil adhesion and promotes atherosclerosis. *Circ Res* **110**, 1294–1302 (2012).

446. Xu, B., Feng, X. & Burdine, R. D. Categorical data analysis in experimental biology. *Dev Biol* **348**, 3–11 (2010).

447. Doig, G. S. & Simpson, F. Randomization and allocation concealment: a practical guide for researchers. *J Crit Care* **20**, 187–191 (2005).

448. Macleod, M. R. *et al.* Good Laboratory Practice. *Stroke* **40**, (2009).

449. Butter, F. *et al.* Proteome-Wide Analysis of Disease-Associated SNPs That Show Allele-Specific Transcription Factor Binding. *PlosONE* (2012) doi:10.1371/journal.pgen.1002982.

450. Tacheny, A. *et al.* Unbiased proteomic analysis of proteins interacting with the HIV-1 5 0 LTR sequence: role of the transcription factor Meis. *Nucleic Acid Research* (2012) doi:10.1093/nar/gks733.

451. Vecellio, M. *et al.* Functional Genomic Analysis of a RUNX3 Polymorphism Associated With Ankylosing Spondylitis. *Arthritis & Rheumatology* **73**, 980–990 (2021).

452. Mittler, G., Butter, F. & Mann, M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res* (2009) doi:10.1101/gr.081711.108.