Investigation of the cis-regulation of vaso-intestinal peptide receptor 2 in zebrafish epicardial development

Chong Hui Lua

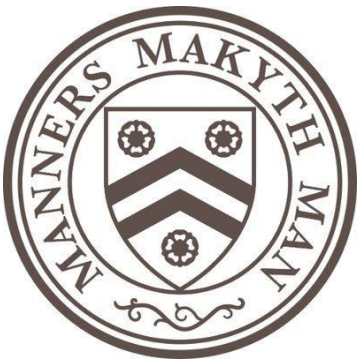
New College

A Thesis submitted to the Division of Medical Sciences, University of Oxford, in partial fulfilment of the requirements for the degree of Master of Science by research

Michaelmas 2019

Department of Physiology, Anatomy and Genetics University of Oxford





Abstract

The human heart is known to have limited regeneration capability, and myocardial infarction leads to irreparable cardiac damage and pathological remodelling. The epicardial cell layer has been identified as a viable resident tractable source of multipotent progenitors in the heart which may be harnessed for cardiac regeneration. During development, the epicardium is known to be an important source of cardiomyogenic paracrine factors as well as supportive cardiac progenitors and thereafter becoming quiescent in adulthood. Interestingly, embryonic epicardial genes are re-expressed following cardiac injury and have been shown to be an important part of the zebrafish cardiac regeneration process. However, little is known about the upstream elements which underlie both epicardial development and regeneration processes, which may be utilized to improve current treatment regimes. In this report, my primary focus is centred on elucidating the cis-regulatory and trans-regulatory elements acting on an epicardial gene locus of interest during development. Previously generated RNA sequencing results show that the G-protein coupled receptor, Vaso-intestinal peptide receptor 2 (*vipr2*) was expressed in epicardial cells at 5 days post fertilization (5dpf). *Vipr2* is also differentially expressed during cardiac regeneration at 3 days post-injury (dpi) of cryoinjured zebrafish hearts. Here, I show with in situ techniques that *vipr2* is expressed in both a rare subpopulation of epicardial cells as well as in most cardiomyocytes in the myocardium. Chromatin accessibility assays and in vivo validation revealed a functional enhancer within *vipr2* intronic regions, enhancer 14, which was found to be active from the 21-somite stage and exhibiting heart-specific activity from 3dpf to 5dpf during development. Enhancer 14 was further shown to be active in *vipr2*-expressing epicardial cells, as well as the cardiomyocytes of the primordial layer of the heart. Through combined transcription factor

binding motif identification, site-directed mutagenesis and transient transgenesis, two transcription factors, myeloid ecotropic viral integration site 1 homolog (MEIS1) and helicase-like transcription factor (HLTF), were found to be critical for enhancer 14 activity. In the future, *vipr2* and enhancer 14 knockout zebrafish lines could be generated to elucidate gene and enhancer function during epicardial and myocardial development and regeneration.

I dedicate this thesis to:

My dearest mum,

without whom I will not be who I am today.

I love and miss you dearly, and you will always be in my heart.

My dearest dad,

for your constant guidance and unwavering support.

My pillar of strength.

My dearest sister,

with whom I share the fondest memories.

Acknowledgements

First of all, my work would not have been possible without the guidance and generosity of Professor Paul Riley. Thank you very much for accepting me into your group and providing guidance for my work amongst your extremely busy schedule.

I am also very thankful to Michael, Filipa and Caroline, who have taught me all aspects of zebrafish work, as well as your patience and guidance in helping me make the right experimental decisions.

Also, a big thank you to Dr Tatjana Sauka-Spengler and her the TSS lab members, Ruth, Mara, Vanessa, Irving, Ivan and Martyna. Thank you for hosting me in the lab throughout my project, and I thank you all for your individual guidance and insights into many aspects of my work. I have had an unforgettable experience in the TSS lab with all of you.

Thank you to all Riley lab members, especially Katie, Dee and Tertia, for providing much-needed support, welfare and administration. I would also especially like to thank Christophe and Sophia, for all your unwavering support, guidance and the great friendship we have and will continue to share.

A huge thank you to the aquarium family, Alison, Andy, Greg, Adam and Kelly. Thank you for taking care of my fish with great care and dedication. The aquarium was never a dull place because of all of you.

Thanks go to Arif and Aldo, for all the discussions that we had on life and the future.

Thanks also to WIMM facility managers, Chris and Jana for your expertise on microscopy, as well as Dominic and Ulrike for guidance on image analysis.

Acknowledgements are also in order to New College for funding my studies, as well as for providing me with a home away from home in Oxford.

I want to also thank my wonderful girlfriend, for your unwavering patience and belief in me, every step of the way as I went through all the ups and downs of this exciting journey.

Table of Contents

Abstract **2**

Dedication **4**

Acknowledgements **5**

Table of Contents **6**

List of figures **9**

List of abbreviations **10**

Chapter 1: General introduction **15**

1.1 Cardiovascular disease background 15
1.2 The zebrafish as a model of heart regeneration 16
1.2.1 Early morphogenetic processes in the formation of the early zebrafish heart 18
1.2.2 Overview of the post-looping cardiac morphogenesis 21
1.2.3 Overview of the zebrafish cardiac conduction system 23
1.2.4 The formation and role of the epicardium during cardiac development 24
1.3 Role of the epicardium in regeneration 25
1.4 Overview of the eukaryotic transcriptional process 26
1.5.1 Proximal promoter and core promoter elements 28
1.5.2 Silencer elements 30
1.5.3 Insulator elements 30
1.5.4 Role of enhancers in gene regulation 31
1.6 Enhancer regulation of epicardial genes during development and regeneration 33
1.7 Identification of enhancers 35
1.7.1 Histone modifications, variants and co-activators 35
1.7.1.1 Epigenetic histone modifications 35
1.7.1.2 H3.3 histone variant 36
1.7.1.3 Co-activator: P300 37
1.7.2 Chromatin accessibility assays 38
1.7.2.1 DNase-seq 38
1.7.3 Enhancer RNA 40
1.7.4 Chromatin conformation capture method and variants 41
1.8 Phylogenetic foot-printing to identify putative cis-regulatory regions 43
1.9 Testing enhancer activity through transient transgenesis 44
1.10 Thesis aims 46
1.10.1 Working hypothesis 46
1.10.2 Aims 46

Chapter 2: Materials and methods 48

2.1 Zebrafish husbandry 48
2.2 Zebrafish lines 48
2.3 Enhancer amplification and cloning 48
2.4 Transcription of Ac RNA for micro-injection procedure 51
2.5 Transgenesis micro-injection procedure 51
2.6 Immunofluorescence and Imaging 52
2.7 Hybridization chain reaction v3.0 53
2.9 *In silico* binding site analysis and binding site perturbation 55

Chapter 3: Identification of the *vipr2* gene as a putative regulator of epicardial development 57

3.1 Introduction 57
3.1.1 *In vivo* bio-tagging toolkit for the isolation of cells for RNA-seq 57
3.1.2 VIP and VIPR2: evolutionarily conserved neurotransmitter system 57
3.2.3 VIP and VIPR2: cardiovascular effects 59
3.2 Results 63
3.2.1 Differential expression analysis to identify genes of interest 63
3.2.2 Developmental expression of *vipr2* in *tcf21+* epicardial cells and *myl7+* cardiomyocytes 65
3.3 Discussion 67

Chapter 4: Genome-wide accessibility assays, verification of putative enhancers and elucidation of TF-enhancer interactions in *vipr2* expression. 69

4.1 Introduction 69
4.1.1 *In silico* analysis of putative TFBS on enhancer sequences 69
4.1.2 Background of putatively important TFBS and corresponding TFs 70
4.1.3 *In silico* analysis of TFBS within an enhancer 72
4.2 Results 72
4.2.1 Characterizing spatio-temporal enhancer-reporter activity and generation of the enhancer 14 stable transgenic line 72
4.2.2 Immunofluorescence investigation of enhancer 14 expression 75
4.2.3 Determining activity of enhancer 14 against *vipr2* expression domain 76
4.2.4 PWM-based TFBS mutation strategy to elucidate important TF-enhancer interactions 79
4.2.5 Determination of TFBS-mutation on enhancer 14 activity 81
4.2.6 High-resolution spinning disc imaging and Imaris analysis of *hltf, stat1/6* and *hic1/2* TFBS- mutation on enhancer 14 activity 83
4.3 Discussion 86

Chapter 5: General discussion and future directions 95

5.1 Transcriptomic and genome-wide accessibility to elucidate cis-regulatory elements involved in epicardial gene expression 95
5.2 Characterising *vipr2* expression and activity patterns of putative enhancers 96
5.3 Elucidating important TF-enhancer interactions underlying enhancer 14 activity 101
5.4 *Meis1* and *hltf* TFBS mutations abrogates enhancer 14 activity 103
5.5 Summary 107
References 109

List of figures

Figure 1.1: Development of the three myocardial muscle layers in the zebrafish ventricle …… 22
Figure 2.1: Primers for cloning putative enhancers CE1 and CE2 …… 49
Figure 2.2: Vector map for E1b:Cherry vector …… 50
Figure 2.3: Enhancer 14 sequence and respective forward and reverse primer sequences …… 54
Figure 3.1: Detected PACAP/VIP receptor subtypes in cultured rat cardiac cells ... 58
Figure 3.2: *Vip* KO mice phenotype and survival estimates …… 62
Figure 3.3: Volcano plot and scatter plot of differentially upregulated epicardial genes after cryoinjury of the zebrafish heart …… 64
Figure 3.4: HCR detection of the co-localisation of *vipr2* and *tcf21* mRNA ……. 66
Figure 3.5: HCR detection of the co-localisation of *vipr2* and *myl7* mRNA …… 67
Figure 4.1: ATAC-seq accessibility map of the *vipr2* gene locus of the epicardial, myocardial and neural crest populations at 5dpf zebrafish embryos …… 73
Figure 4.2: CE1 and CE2 negative data throughout embryonic development …… 74
Figure 4.3: Characterising the spatiotemporal activity patterns of enhancer 14 … 75
Figure 4.4: HCR investigation into enhancer 14 activity and *vipr2* expression overlap with *tcf21* epicardial marker …… 77
Figure 4.5: HCR investigation into enhancer 14 activity and *vipr2* expression overlap with *myl7* myocardial marker …… 78
Figure 4.6: Schematic of the TFBS-mutation strategy to test for upstream TF-enhancer regulation …… 80
Figure 4.7: Schematic demonstrating workflow for TFBS-mutation strategy ……. 81
Figure 4.8: *Meis1* binding was found to be needed for enhancer 14 activity through the TFBS-mutation strategy …… 82
Figure 4.9: Graphs and representative images from spinning disc and imaris image analysis for the different TFBS-mutated enhancer-reporter conditions …… 85
Figure 4.10: Expression of *vipr2* and the individual TF expression in scRNA-seq dataset …… 88
Figure 5.1: Improved HCR specificity with *in situ* HCR v3.0 with split-initiator probes …… 98

List of abbreviations

3C - chromatin conformation capture

4C - circular chromatin conformation capture

4-HT - 4-hydroxytamoxifen

Ach - acetylcholine

Ac RNA - activator RNA

ALPM - anterior lateral plate mesoderm

Amhc/ myh6 - atrial-specific atrial myosin heavy chain

ATAC-seq - assay for transposase accessible chromatin with high throughput sequencing

AVN - atrioventricular node

avcPE – atrioventricular boundary

BAC - bacterial artificial chromosome

Baf - Brg1 associated factor

BirA - biotin ligase

Bp - base-pair

bHLH - basic helix-loop-helix

Bmp - bone morphogenic protein

BRE - TFIIB recognition element

bZIP - basic leucine zipper

cAMP - cyclic adenosine monophosphate

C/EBP - CAAT/ Enhancer Binding Protein

CCS - cardiac conduction system

cDNA - complementary DNA

CHF - congestive heart failure

ChIP-seq - chromatin immunoprecipitation sequencing

CNS - central nervous system

CR - conserved region

CREEs - cardiomyocyte regeneration enhancer elements

CTCF - CCCTC-binding factor

cTnT - cardiac troponin T

DPE - downstream promoter element

Dpf - days post-fertilization

Dpi - days post-injury

Ds - dissociation elements

ECR - evolutionarily conserved region

EMSA - electrophoretic mobility shift assay

EMT - epithelial to mesenchymal transition

EPDC - epicardium-derived cell

EPO - erythropoietin

eRNA - Enhancer RNA

Fgf - fibroblast growth factor

GCPR - G-protein-coupled receptor

Grh - grainyhead transcription factor

GO - Gene Ontology

GRN - gene regulatory network

GTF - general transcription factors

HAT - histone acetyltransferases

HCR - hybridization chain reaction

Hh - hedgehog signalling

Hic1/2 - hypermethylated in cancer 1 and 2

Hltf - helicase-like transcription factor

HOX - homeobox factor

Hpf - hours post-fertilization

HTH - helix-turn-helix

IF - immunofluorescence

Igf2 - insulin-like growth factor 2

Inr - initiator

IR - ischaemia-reperfusion

ISH - in-situ hybridisation

Kb - kilobase

KO - knock-out

LAD - descending coronary artery

LB - lysogeny broth

LHT - linear heart tube

lncRNA - long non-coding RNA

Meis1 - myeloid ecotropic viral integration site 1 homolog

MEF - mouse embryonic fibroblast

MI - myocardial infarction

Mpf - months post-fertilization

MS222 - tricaine

Myl7 - myosin light chain 7

NO - Nitric Oxide

NOS - Nitric Oxide Synthase

NPR-C - natriuretic peptide clearance receptors

OCT - oligonucleotide capture technology

PACAP - pituitary adenyl cyclase-activating protein

PAGODA - pathway and gene set overdispersion analysis

PBS - phosphate-buffered saline

PCNA - proliferating cell nuclear antigen

PEO - proepicardial organ

PFA - paraformaldehyde

PIC - pre-initiation complex

PIL - personal license

PKA - protein kinase A

Plb - phospholamban

PNS - peripheral nervous system

Pol II - RNA polymerase II

PPL - project license

PRC - polycomb repressive complex

PTU - phenylthiourea

qPCR - quantitative polymerase chain reaction

RA - retinoic acid

Raldh2 - retinaldehyde dehydrogenase 2

RNA-seq - RNA-sequencing

re-ChIP - sequential chromatin immunoprecipitation

RFP - red fluorescent protein

RT - room temperature

RXRα - retinoid X receptor α

SA - sinoatrial

SAN - sino-atrial node

siRNA - small interfering RNA

SSC - sodium chloride sodium citrate

SSCT - sodium chloride sodium citrate triton

Stat- signal transducer and activator of transcription

SWI/SNF - switch/sucrose fermentable

T-β4 - thymosin-β4

TBP - TATA-binding protein

TESS - transcription element search system

TF - transcription factors

TFBS - transcription factor binding sites

Tnnt3 - cardiac troponin T3

TREE - tissue regeneration enhancer element

TSS - transcriptional start site

Vmhc - ventricular myosin heavy chain

Vipr2 - vaso-intestinal peptide receptor 2

vpPE - venous pole of the heart

VSD - ventricular-septal defects

Wpf - weeks post-fertilization

WT – wild-type

YSL - yolk syncytial layer

Chapter 1: General introduction

1.1 Cardiovascular disease background

The human heart is widely believed to be a post-mitotic organ with limited potential for regeneration. Following myocardial infarction (MI), inflammatory-mediated repair causes collagen-rich, non-contractile fibrotic scarring and subsequent pathological remodelling of the heart leading to congestive heart failure (CHF) (Houser & Margulies, 2003; Nadal-Ginard et al., 2003; Pfeffer & Braunwald, 1990). Heart transplantation is the only cure to this condition, but the scarcity of donor hearts and immunological disparity (Sheldon & Poulton, 2006) leaves most patients without viable options. Current medication-mediated CHF management (AHA, 2016) is only enough to prolong the onset of CHF and does not provide a permanent solution. To date, there is still no clinically viable cell therapy for the replacement of the infarcted tissue; although several candidate cell-types and clinical-trials have been conducted (Chugh et al., 2012; Laflamme & Murry, 2011; Menasché et al., 2008). Clinical improvements observed in various cell therapy trials hold promise, but experimental findings pose pertinent fundamental issues of increased arrhythmogenicity (Fouts et al., 2006), and insubstantial contribution to myocardial restoration due to meagre transplant survival (Chimenti et al., 2010). Instead, marginally improved clinical outcome can be attributed to paracrine effects of transplanted cells leading to increased neovascularization within the injured myocardium (Malliaras & Marbán, 2011). Hence, there is an unmet need to establish alternative approaches whereby the focus is redirected to understanding pathways to augment the activation of existing, tractable sources of *in vivo* cell-types within the injured heart.

1.2 The zebrafish as a model of heart regeneration

The zebrafish, *Danio rerio*, is a powerful model in which to study cardiac development and regeneration. During cardiac development, the zebrafish heart develops within 24 hours post-fertilization (hpf). Interestingly, the embryos can survive without cardiovascular circulation up to 10dpf, which allows for the study of otherwise lethal cardiac congenital malformations in the zebrafish model as opposed to other vertebrate models. Other advantages of the zebrafish includes: ease of handling due to external fertilization of embryos; propensity for laying large batches of embryos (>100 embryos at weekly intervals); reaches adulthood with the ability to breed within 3 months post-fertilization (mpf); adult zebrafish may reproduce all-year-round in controlled laboratory conditions; embryo transparency which allows for visualisation of organ development; a fully sequenced genome and ease of fluorescent reporter transgenesis.

Unlike mammals, the zebrafish has remarkable regenerative abilities which persist into adulthood, making it an attractive model to study the cellular and molecular mechanisms underlying regeneration. In seminal work investigating cardiac regeneration, adult zebrafish were found to fully regenerate up to 20% (González-Rosa et al., 2011; Poss et al., 2002) of lost heart tissue, with no loss of cardiac function. This was found to be achieved by proliferation of pre-existing adult cardiomyocytes, through the dedifferentiation and disassembly of the highly-ordered contraction-related sarcomeric structures to an immature/naïve state as characterized by a reactivation of cardiac developmental markers *gata4, hand2, nkx2.5*, G2/M cell-cycle regulator *plk1* (Jopling et al., 2010; Kikuchi et al., 2010), replication-related proliferating cell nuclear antigen (*PCNA*) (Lepilina et al., 2006), as well as detachment from the tight cell-to-cell conformation to facilitate proliferation. These events are closely coupled with the reactivation of the resident epicardial population which provides mitogenic factors, such as retinaldehyde dehydrogenase 2 (RALDH2*)* and

fibroblast growth factor (FGF) that facilitate cardiomyocyte proliferation as well as a source of epicardium-derived cell (EPDC) which contribute to the revascularization of the heart (Lepilina et al., 2006). In parallel, there is an acute clotting response elicited following injury, which constitutes a fibrin-based clot with minor deposits of collagen, followed by an inflammatory response dominated by neutrophils and monocytes/macrophages which subsequently leads to the gradual dissolution of the necrotic tissue shrinking the cardiac lesion for repopulation by proliferating cardiomyocytes, and new coronary blood vessels through angiogenesis (Marín-Juez et al., 2016). The inflammatory and regeneration process leads to the complete restoration of the working myocardium within 60 days, with the length of time varying accordingly to the size of the injury. However, the ventricular resection surgical method, which was used in the initial zebrafish heart regeneration studies does not compare well to mammalian MI. The resection method, although efficient, does not lead to widespread ischemic cell death and lacks the extent of scarring and ensuing fibrosis associated with mammalian MI. The current alternative, which better recapitulates the mammalian post-MI response is the cryo-injury method (González-Rosa et al., 2011). This technique utilises a copper filament-cooled in liquid nitrogen and immediately applied to the ventricular apex of the sedated zebrafish. The cryo-injury induces widespread cell death at the foci of cardiac injury resulting in the formation of a fibrotic scar (González-Rosa et al., 2011).

The zebrafish has also been utilised as a model organism to study dynamic chromatin accessibility and epigenetic changes in genomic regulatory regions underlying the injury and regeneration response. A recent study utilizing the zebrafish model uncovered a class of enhancers termed tissue regeneration enhancer elements (TREEs) (Kang et al., 2016a). These enhancer elements were found to respond to heart and fin injury, by regulating regeneration- specific expression of injury-response genes, not previously expressed in

development, in a spatiotemporal manner.

1.2.1 Early morphogenetic processes in the formation of the early zebrafish heart

In vertebrates, the heart is the first organ to form during development. The mammalian heart consists of four chambers, whereas the zebrafish heart is a two-chambered organ. However, despite the marked anatomical differences, the zebrafish heart exhibits several evolutionary similarities in cellular and molecular processes, which makes the model amenable for developmental studies (Stainier et al., 1993)

At the blastula stage of the zebrafish embryo, cardiac development begins with the initial specification of early cardiac progenitor cells of both the ventricular and atrial fate, located on either side of the lateral marginal zone at 5hpf. The early ventricular, atrial and endocardial progenitors are specified at this stage (Staudt & Stainier, 2012). Several signalling pathways have been implicated in regulating the number of cardiomyocyte progenitors in early embryonic development, namely, *Fgf*, *Wnt*, bone morphogenic protein (*Bmp*), retinoic acid (*RA*) and hedgehog (*Hh*) signalling. During early development, the WNT pathway plays an interesting temporal role, where the signalling pathway increases cardiac progenitor numbers in the blastula. However, the continued *Wnt* expression decreases cardiac progenitor numbers after the onset of gastrulation (Dohn & Waxman, 2012). RA signalling restricts and controls the size of the cardiac progenitor pools, establishing the balance of both cardiac and non-cardiac cell identities in the early embryo. *Hh* signalling was also demonstrated to play a critical role in promoting cardiac precursor specification for progenitors of both chambers. The overexpression of *Hh* causes an increase in progenitor numbers, and the knockdown of HH signalling results in the concomitant reduction of progenitors (Thomas et al., 2008). Similarly, FGF signalling affects cardiac progenitor specifications for both ventricular and atrial precursors (Marques et al., 2008). Lastly, BMP signalling was found to impact the atrial progenitor specification

directly, whereas the ventricular progenitors are less susceptible (Marques & Yelon, 2009).

During the gastrulation process, the primordial cells of the embryo involutes resulting in the formation of the three germ layers of the embryo. The early cardiac progenitor cells begin to migrate dorsally towards the mid-line to the posterior half of the anterior lateral plate mesoderm (ALPM) arriving at the 6-9 somite stage. The region is marked by early cardiac differentiation transcription factors *hand2* and *gata4* expression (Yelon, 2001). Several signalling pathways facilitate the migration process. Apelin, a signalling peptide ligand involved in the migration process, is expressed in the mid-line, while its receptor Agtrl1b is expressed in the marginal zone. Apelin/ Agtrl1b signalling gradient has been demonstrated to direct the migratory process of the progenitor cells towards the ALPM (Staudt & Stainier, 2012). Furthermore, HAND2, aside from its role in cardiac progenitor differentiation, was found to play an important role in the regulation of fibronectin deposition at the interface between the yolk syncytial layer (YSL) and the myocardial progenitors to enable their migration towards the midline (Yelon, 2001).

In the zebrafish model, the initiation of the cardiogenic differentiation begins at the 12-15-somite stage, with the expression of key cardiomyocytic markers such as *nkx2.5*, myosin light chain 7 (*myl7*) and ventricular myosin heavy chain (*vmhc*) in the medially positioned ventricular progenitor pool (Yelon et al., 1999). The next wave of cardiogenic differentiation follows with the slightly delayed expression of the atrial-specific atrial myosin heavy chain (*amhc/ myh6*), instead of the ventricle-specific *vmhc*, in the laterally positioned atrial progenitor pool. Similar to amniotes, a combination of GATA transcription factors, namely, GATA4, GATA5 and GATA6, promotes cardiac cell specification cooperatively with the Brg1 associated factor (BAF) (Lou et al., 2011). Morpholino-mediated knockdown of both Gata5 and Gata6 leads to the overt disappearance of *nkx2.5*-positive cardiac progenitors. The overexpression of *Gata5* and the

Baf complex component, *Smarcd3b*, results in the increase in the cell-autonomous specification of cardiac progenitors by driving all non-cardiac progenitor cells overexpressing *Gata5* and *Smarcd3b* to migrate to the APLM cardiogenic regions and re-specify to a cardiac fate, thereby increasing cardiac progenitor numbers (Lou et al., 2011). *Hand2* is also known to play a critical role in cardiac progenitor specification. *Hand2* is found to be expressed in all cardiomyocytic progenitors, especially in ventricular precursors. The factor is also needed to maintain expression of the T-box transcription factor, *tbx5*, which has been shown to synergistically promote cardiomyocyte differentiation with *nkx2.5*, an essential marker of the cardiomyocytic lineage. The knockdown of *hand2* gives rise to a separate formation of a population of *nkx2.5*-positive, *myl7*-negative population, showing the functional role of *hand2* in the differentiation of *nkx2.5*-expressing cells into cardiomyocytes (Yelon et al., 2000).

At the 16-somite stage, the endocardial progenitor population leads the critical migration process from the ALPM medially towards the embryonic midline, whereas the migration of the differentiating ventricular and atrial populations, which are situated in between the overlying endodermal precursors and underlying YSL, lagged behind slightly (Staudt & Stainier, 2012). Similar to the role of the endodermal progenitors in leading the cardiac progenitor migration to the ALPM, mutants with endoderm formation defects result in cardia bifida, showing the critical function of the endodermal progenitors in the migration process (Fukui et al., 2014). The bilateral progenitor populations differentiate, migrates and fuses at the embryonic mid-line by the 18-somite stage to form a shallow cone-shaped cardiac disc (Kimmel et al., 1995). The cardiac disc is comprised of, the endocardial population lining the central lumen, ventricular progenitors at the dorsal apex and the atrial progenitors at the periphery of the structure (Bakkers, 2011). The nascent conical structure extends, and the apex of the cone angles to towards the posteriorly right side, thus changing

the axis of the cardiac cone from dorsal-ventral, to the anterior-posterior plane. The cardiac cone proliferates and extends along the anterior- posterior axis and lengthens into the structure known as the linear heart tube (LHT) by 28hpf (Bakkers, 2011). The LHT is made up of cardiac progenitor cells from the first heart field (FHF). The second heart field (SHF) progenitors undergo differentiation later than the FHF counterparts and contribute to the two ends of the developing LHT, giving rise to the inflow and outflow tracts of the forming zebrafish heart, respectively (Felker et al., 2018). The LHT finally undergoes asymmetrical cardiac looping starting from 36hpf, where the ventricle is displaced towards the right, and concurrent chamber ballooning causes chamber formation in the zebrafish embryo (Bakkers, 2011).

1.2.2 Overview of the post-looping cardiac morphogenesis

The adult zebrafish ventricle is recognised to contain three distinct myocardial layers, formed by proliferation of dominant cardiomyocyte clones during cardiac morphogenesis (figure 1.1). This cardiomyocytic clonal expansion was demonstrated by multicolour clonal analysis with the Brainbow1.0L construct under the control of the *cmlc2* Cre-driver (Gupta & Poss, 2012). Brainbow drives expression of three spectrally distinct fluorescent reporters following inducible Cre-expression, with the red fluorescent protein (RFP) being the default expression marker pre-recombination. 4-Hydroxytamoxifen (4-HT) mediated recombination was induced from 2dpf in the looping heart, at a time point when the zebrafish chamber wall consists of a single-cardiomyocyte thickness and is hypothesised to be responsible for subsequent cardiomyocyte proliferation, and the initiation of trabeculation beginning at 3dpf. Histological examinations of the developing zebrafish heart at 10dpf revealed the maintenance of a single cardiomyocyte layer, referred to as the primordial layer, which was demonstrated to be retained throughout subsequent developmental stages. The primordial layer was shown to give rise to the innermost cardiac

trabecular layer by delaminating and subsequently seeding at another site within the inner myocardial chamber and then perpetuating trabecular growth at the secondary site. Thus, the trabeculae layer consists of cardiomyocytes from different clonal origins from the primordial cardiomyocyte layer (Gupta & Poss, 2012). The late-emerging outermost myocardial layer known as the cortical muscle layer, emerges from 6 weeks post-fertilization (wpf) externally to the primordial layer, as clonally related cardiomyocytes which first appears at the base of the ventricle. By 8wpf, the cortical layer clones proliferate and extend from the base of the ventricle, usually reaching half of the total ventricular surface area. At 10wpf, the clonally-related cardiomyocytes from the developing cortical layer converge with each other; and by 90dpf, the entire outer surface of the ventricle has been covered by the multi-cell thick cortical muscle layer consisting of a few large cortical clones (Gupta & Poss, 2012).

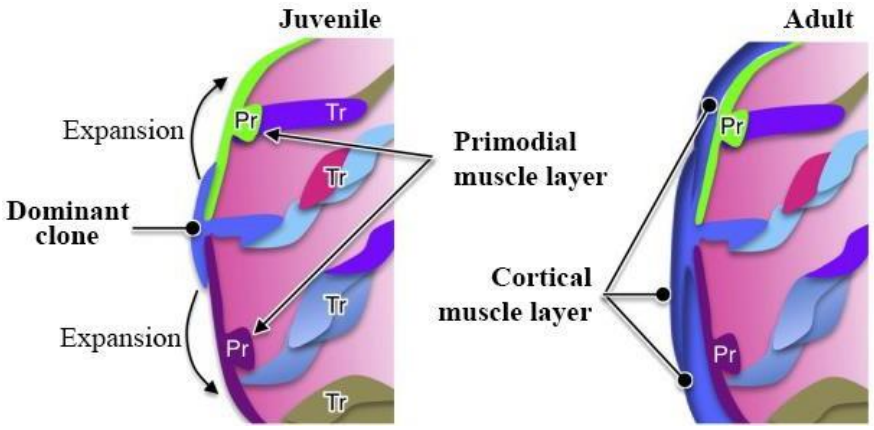


Figure 1.1: Development of the three myocardial muscle layers in the zebrafish ventricle

Tr, trabecular muscle; Pr, primordial muscle. In the juvenile zebrafish heart, the primordial muscle layer is maintained as a single-layer of cardiomyocytes in the ventricular wall. From 3dpf, dominant clones from the primordial layer would delaminate and seed elsewhere in the ventricle chamber to give rise to the trabecular muscle layer. In the adult zebrafish heart, the multi-cell thick cortical muscle develops from dominant primordial muscle clones and is situated externally to the primordial layer which is retained as a single cardiomyocyte layer throughout all life stages. (Figure adapted from (Kikuchi, 2014))

1.2.3 Overview of the zebrafish cardiac conduction system

The cardiac conduction system (CCS), is an organisation of a specialised network of cardiomyocytes which generates and perpetuates uniform contraction of the heart to transport blood around the body. The rhythmic contraction spontaneously originates from a group of pacemaker cells, known as the sino-atrial node (SAN). The electrical signal then radiates through the atrium, depolarising the cardiomyocytes and leads to atrial contraction. The electrical impulse spreads to the atrioventricular canal until reaching another specialised group of conduction cells which forms the atrioventricular (AV) node, where the signal is delayed, to maintain the relaxation of the ventricle while being filled with blood (Poon et al., 2016). After the AV node delay, the electrical signal enters the ventricular fast conduction system and propagates rapidly through the His-Purkinje fibres to the apex and throughout the ventricle, resulting in ventricular contraction. The series of contractions constitutes a heartbeat, which perpetuates the apex-to-base contraction required for ejecting blood from the ventricle to the body. Although morphologically different, the cardiac conduction system of the 2-chambered zebrafish heart has been demonstrated to be similar to the amniote 4-chambered heart via electrophysiology and optogenetics analyses (Arrenberg etal., 2010; Sedmera et al., 2003).

The cardiac conduction is first active at 24hpf, indicating differentiation and functionality of the pacemaker cells even at the LHT stage. At this early stage, the electrical impulse is initiated from the venous pole and travels through the LHT with peristaltic cardiomyocytic contractions. The AV conduction delay is first observed at 40hpf, and by 48hpf, calcium imaging and electrophysiology is able to distinguish between the atrial, ventricular and AV conduction cells via unique conduction signatures (Poon et al., 2016).

1.2.4 The formation and role of the epicardium during cardiac development

The epicardium is a mesothelial cell layer that covers the heart and originates from a transient organisation of coelomic cells known as the proepicardial organ (PEO), a structure which is highly conserved in vertebrates. During early development, cells from the PEO undergo epithelial to mesenchymal transition (EMT), proliferate and migrate to envelop the looping heart tube at E8.5-9.5 (Balmer et al., 2014) in mouse, forming a single epicardial cell-layer mid-gestation. In the mouse and zebrafish, the PEO originates from two sites after initiation of heart tube looping, proximal to the venous pole of the heart (vpPE) and near the atrioventricular boundary (avcPE) (Peralta et al.,2014). The epicardial derived cells (EPDC) then undergo further rounds of EMT, delaminating, gaining migratory capabilities and contributing to the developing heart by differentiating into multiple functional cell-types such as endothelial cells, smooth muscle cells, adipose tissue, vascular smooth muscle cells and cardiac fibroblasts (Dettman et al., 1998). The wide range of differentiation capability is due to the heterogeneity of epicardial cells, identified by characteristic markers such as *wt1*, *tbx18*, *tcf21*, *scx*, *sema3D* during development (Masters & Riley, 2014).

The epicardium is a prominent source of trophic signals to the underlying myocardium, promoting cell growth and coronary vessel development (Masters & Riley, 2014). Epicardial RA signalling is known to be critical in heart morphogenesis. Heterozygous null mice which lack the retinoid X receptor α (RXRα), an essential receptor for RA signalling, results in epicardial detachment, hypoplastic myocardium and embryonic lethality mid-gestation. Interestingly, the effect was found to be epicardial specific through knock-out (KO) studies, where *Rxrα* deletion in epicardial population recapitulated *Rxrα* null effect, whereas *Rxrα* deletion in the myocardial population resulted in unimpaired heart development (Merki et al., 2005). The importance of RA signalling for cardiac

development was directly explored with in-vitro chicken heart explant studies where myocardial proliferation was induced in explants containing epicardial cells (Van Wijk et al., 2009).

RA signalling is also implicated in organ-level crosstalk, as RA secreted by the epicardium binds to receptors on the liver to produce a hormone, erythropoietin (EPO). EPO, in turn, binds to EPO receptors on the epicardium, thus activating expression of Insulin-like Growth Factor 2 (*igf2*) which is an essential cardiovascular mitogen leading to myocardial proliferation and expansion and a normal cardiac development. Therefore, the literature strongly supports that the epicardium plays an essential role during cardiac development both as a centre of paracrine activity which provides proliferative cues for the underlying myocardium, as well as being a source of cardiovascular supportive cells such as vascular smooth muscle cells (VSMC), cardiac fibroblasts, endothelial cells through EMT of EPDCs (Carmona et al., 2010; Masters & Riley, 2014).

1.3 Role of the epicardium in regeneration

The epicardium has been identified as a viable resident source of multipotent progenitors contributing to adult cardiovascular regeneration (Smart & Riley, 2012). In the adult heart following MI, the ischaemic insult triggers the re-emergence of foetal gene expression in the myocardium (Porrello et al., 2011), and an epicardial-wide reactivation of *wt1*, *tcf21*, *tbx18* and *raldh2* foetal genes (Masters & Riley, 2014). The reactivation of the epicardium induces EPDC proliferation, enveloping the injury site and provides regenerative cues to the myocardium (Zhou et al., 2011). Studies in zebrafish have shown that ablation of the epicardium disrupts heart regeneration (Wang et al., 2015) and in mice, pre-MI treatment with thymosin-β4 (T-β4) enhanced re-activation of the epicardium to drive adult EPDC differentiation towards vascular lineages (Smart et al., 2013), as well as a low number of de novo cardiomyocytes (Smart et al., 2013). Other than zebrafish, neonatal mice were

found to retain cardiac regeneration capacity, as shown by the robust replacement of tissue following amputation of 15-20% of the neonatal ventricle or ligation of the lateral anterior descending coronary artery (LAD). However, this remarkable ability is lost within the first week of life (mouse post-natal day 7), in which the cardiomyocytes become more mature, less proliferative and are refractory to epicardial-secreted factors (Chen et al., 2002; Porrello et al., 2011). The epicardium itself is known to enter a state of dormancy during this neonatal period and this is thought to contribute to the loss of regenerative capacity.

A seminal study demonstrated the importance of the epicardium in zebrafish cardiac regeneration, where they utilized *tcf21*-nitroreductase to ablate 45% of the epicardial cells by 7 days post-amputation. This led to a concomitant decrease in cardiomyocyte proliferation by 33%, as well as marked reduction of the re-vascularisation and re-muscularization of the injured adult heart as compared to experimental controls without epicardial ablation (Wang et al., 2015). The epicardium seemingly recapitulates its role during cardiac development during the cardiac regeneration process. Thus, the quiescent adult epicardium harbours innate regenerative potential, which may be realised by augmenting the reactivated developmental gene programmes. Therefore, by studying the initial molecular events, pathways and regulatory networks of the epicardium that occur during normal cardiac development, we may uncover novel molecular stimuli and regulatory pathways that may be targeted for re-activation following heart injury to stimulate the regenerative process.

1.4 Overview of the eukaryotic transcriptional process

Transcription of genetic information encoded in DNA constitutes the first step in genetic expression, which consists of three distinct phases. The first phase, transcriptional initiation, is the stage where the pre-initiation complex (PIC) is recruited and bound to the promoter, a DNA sequence which is usually located immediately adjacent to the

transcriptional start site (TSS) for the commencement of subsequent transcription initiation. The second stage is known as the elongation phase, which constitutes the reading of the DNA sequences and lengthening of the RNA strand by the addition of nucleotides complementary to the DNA template. Lastly, the transcriptional termination phase marks the end of the transcription process, where the PIC reads the stop codon resulting in stoppage of the PIC, the release of RNA polymerase II (Pol II) from the DNA template and the completed messenger RNA molecule detaches from the RNA polymerase complex (Nechaev & Adelman, 2011).

Three main groups of factors are involved in the precise process of transcription. The general transcription factors (GTF), activators and co-activators, together with Pol II, which collectively constitutes the basal transcription machinery. The classical GTPs consists of TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH which, when bound together systematically, form the PIC that recruits Pol II to the TSS for transcription initiation (Smale & Kadonaga, 2003). Activators are a class of sequence-specific DNA-binding proteins, with classic examples such as zinc-finger proteins, homeobox (HOX) transcription factors (TF), basic helix-loop-helix (bHLH) containing proteins, basic leucine zipper (bZIP) proteins, ETS factors, fork-head (FOX) and POU DNA-binding domain-containing proteins which are essential for the stimulation of transcription. Activators are hypothesised to stimulate PIC assembly through direct interactions with specific components of the transcriptional machinery such as GTPs and co-activator proteins, as well as recruiting chromatin remodelling proteins which facilitate PIC assembly by enabling accessibility of surrounding chromatin regions. Co-activators generally functions similarly to activators in their function in promoting transcription. However, they do not typically display DNA binding properties but are mainly recruited by direct protein-protein interaction with activator proteins. Some examples of co-activators include histone

acetyltransferases (HAT), histone methyltransferases, chromatin remodelling factors and mediator complexes. The current view of the field is that most of the regulation involving protein-coding gene expression is controlled at the level of transcriptional initiation (Maston et al., 2006).

1.5 Cis-regulatory elements in gene regulation

1.5.1 Proximal promoter and core promoter elements

The promoter is a cis-regulatory element which directly interacts with the basal transcriptional machinery facilitating transcriptional initiation. The promoter generally contains multiple binding sites for activators of transcription, as well as motifs which function as docking sites for the assembly of the PIC (Smale & Kadonaga, 2003). The core promoter describes a DNA sequence containing one or more core promoter elements which are able to function as a transcriptional initiation site either independently or in synergy with other core promoter elements.

The TATA box was first discovered from sequence comparison of 5' flanking sequences of protein coding-genes in several model organisms, showing that the site was highly evolutionarily conserved. The TATA sequence can be found throughout the genome, 25-30 base-pairs (bp) upstream of the TSS in many protein-coding genes. The TATA box element functions as a binding site for the general transcription factor, TFIID, also known as the TATA-binding protein (TBP) (Maston et al., 2006). Other than facilitating TFIID binding, the TATA box also influences the orientation of TFIID polar binding leading to the correctly oriented PIC, thereby directly affecting the directionality of transcription. However, to this end, distal activators have been demonstrated to enhance the polarity of TFIID binding in the determination of transcriptional directionality.

Another core promoter element that can be found commonly in eukaryotic genomes is the

initiator (Inr) element. Similar to the TATA box, the Inr element was found to function independently as a binding site for TFIID which determines the start site of RNA synthesis in a promoter without a TATA box element. *In vitro* transcription experiments performed on nuclear extracts from HeLa cells revealed that Inr-mediated transcriptional activity was comparable to that of an isolated TATA box. Furthermore, the Inr element and the TATA box were found to work synergistically if the two elements are within 25-30bp of each other within the same promoter, where the Inr functions by synergistically increasing the efficiency of transcriptional initiation and the TATA box element acting as the primary site of PIC assembly (Javahery et al., 1994).

The downstream promoter element (DPE) is a downstream core promoter motif which is typically but not exclusively found in promoters lacking a TATA box. The DPE does not function independently, but rather, it works in conjunction with the Inr element, cooperatively forming a core recognition motif for the binding of TFIID (Burke & Kadonaga, 1996). In drosophila, it was demonstrated that the Inr-DPE cooperation is greatly reliant on the spacing between the two elements, where even the difference in a single nucleotide spacing results in the decrease in TFIID binding and transcriptional activity (Kutach & Kadonaga, 2000).

The B-recognition element (BRE), is a core promoter element, usually found flanking the TATA box with 5 nucleotides both upstream and downstream, known as the major and minor groove respectively (Nikolov et al., 1995). BRE may enhance transcription initiation by facilitating TFIIB binding to the minor groove. TFIIB contains a helix-turn-helix (HTH) motif which recognises and binds to BRE, thus stabilising the assembly of TFIIB at the promoter. Interestingly, while the TFIIB HTH motif is conserved in eukaryotes and archaeal systems, it is notably absent in yeast and plant systems, indicating that BRE may not contribute to gene regulation in these models (Deng & Roberts, 2005). TFIIB-BRE

interaction was found to stabilise the formation of TFIIB-TBP-BRE complex, as demonstrated via electrophoretic mobility shift assay (EMSA). TFIIB-BRE binding also enhances polar binding of TBP to the TATA box for transcriptional directionality (Nikolov & Burley, 1997). TFIIB-BRE interaction has been demonstrated to act as a negative regulator of transcription, where a construct containing mutated BRE element showed higher basal transcription than the wild-type sequence, both in basal transcription assay with HeLa nuclear extracts, as well as in HEK293 cells (Evans et al., 2001). However, in the presence of transcriptional activators, the innate level of transcription was found to be unaffected by TFIIB-BRE binding. In this regard, it has been proposed that BRE regulates the step from PIC formation to transcriptional elongation.

1.5.2 Silencer elements

Cis-regulatory regions known as silencers functions to repress transcription of a target gene. Silencers are made up of a collection of binding sites for repressive transcription factors and corepressors which perpetuates its repressive functions. Repressors exert their function through direct competition for a binding site, mediating the recruitment of chromatin remodelling factors for chromatin stability and repressive histone modifications, blocking the binding of a transcriptional activator, as well as the direct inhibition of the formation of the PIC (Maston et al., 2006). Silencers may act independent of the distance and orientation of their target gene, much like enhancers.

1.5.3 Insulator elements

Insulators elements are also known as boundary elements, which function to limit the regulatory actions of transcriptional activating or repressive elements and partitioning the genome into discrete functional domains in a position and orientation-dependant manner. Insulator elements, which are usually between 0.5-3.0 kilobase (kb) in length, exerts its function by either blocking promiscuous and unspecific enhancer-promoter interaction, as

well as to prevent the formation of heterochromatin across boundaries (Recillas-Targa et al., 2002). The only trans-acting factor known to perform insulator function is the CCCTC-binding factor (CTCF), which is responsible for the insulation between TADs. CTCF is a critical regulator of genome topology due to the specification of TADs. CTCF depletion from a TAD boundary was found to disrupt the insulation of the neighbouring TADs and could even lead to the neighbouring TADs to be encapsulated into a single domain. This encapsulation of adjoining TADs also impacts gene regulation and enhancer specificity (Nora et al., 2017).

However, evidence suggests that there may be other factors in play which affects insulator function. A publication studied the locus of *Firre*, a long non-coding RNA (lncRNA) on the X-chromosome which is conserved across both the human and mouse models. The studied locus is located at a TAD boundary with classical CTCF insulator enrichment with enriched CTCF binding across multiple cell types (Barutcu et al., 2018). The *Firre* locus was deleted in mouse embryonic fibroblast (MEF) cell-line, resulting in the complete loss of the dense CTCF binding at the locus at the TAD boundary, as shown by CTCF chromatin immunoprecipitation sequencing (ChIP-seq) experiments. Interestingly, the TAD boundary was found to be largely preserved despite the loss of the CTCF-rich TAD boundary as confirmed by Hi-C. However, the deletion led to the decrease in strength across the TAD boundary resulting in a marginally higher rate of unspecific interactions across the boundary as compared to wildtype controls.

1.5.4 Role of enhancers in gene regulation

The initiation of transcription can be performed by the minimal promoter, the site of PIC assembly; however, transcription is often weak without additional inputs from distal regulatory sequences such as enhancers. Enhancers are known to be the most abundant class of cis- regulatory factors within the genome. Enhancers are generally short regions of

DNA sequences, usually less than 1kb in length (Blackwood, 1998) and function to facilitate dynamic and precise patterns of gene expression. Enhancers have also been likened to a molecular computer, where specific molecular signals can be received and resulting in a spatiotemporal gene expression pattern. To this end, enhancers are greatly enriched in transcription factor binding sites (TFBS), which are short, generally 6 to 8bp DNA motifs, for the sequence- specific recognition of TFs. The enhancer region is first accessed by pioneer factors, which are TFs which harbour the ability to recruit chromatin remodelling complexes to facilitate the accessibility of DNA. The typically activated enhancer is in an open and accessible conformation, devoid of nucleosomes which are structural packaging units of inactive chromatin. Subsequently, through TF recruitment and binding of multiple TFs in a combinatorial manner on the nucleosome-free enhancer sequence, provides the input which activates the enhancer to facilitate the transcription process. Once activated by specific TF binding, the enhancer becomes physically associated with the promoter regions of their cognate genes via looping of chromatin to facilitate gene expression through the assembly of the pre- initiation complex at the promoter site (Pennacchio et al., 2013). Enhancers may regulate genes which are nearby or may exert their genetic influence over large genetic distances (~1 mega- base) (Visel et al., 2009) due to the native 3-dimensional conformation of chromatin (Spitz, 2016) or chromatin looping (Kulaeva et al., 2012). The unique ability to regulate proximal or distal genes in a spatiotemporal and tissue-specific manner (Blackwood, 1998; Lenardo & Baltimore, 1989) stems from the compartmentalisation of the genome into locally interacting regulatory environments where enhancers may interact within TADs with their cognate promoters.

Enhancers are also known to function in a modular fashion, where they contribute in a combinatorial and partly redundant manner. It is also possible that multiple enhancer regions may potentiate the differential spatiotemporal expression of the same gene under

different conditions and cellular processes including during, development (Perry et al., 2011) and tissue regeneration (Kang et al., 2016). Thus, the finding out the important upstream TF-enhancer interactions of active enhancers underlying cell or tissue-specific processes, would be instrumental in suggesting novel strategies in augmenting key biological processes.

1.6 Enhancer regulation of epicardial genes during development and regeneration

A previous study examined enhancers and GRN underlying epicardial development and after injury in mice (Huang et al., 2012). Since epicardial development is known to be highly conserved across most vertebrae species, it was hypothesized that conserved regulatory enhancer elements might also be present within conserved genomic sequences around the epicardial genes of interest. Sequence alignment and comparison were carried out between multiple epicardial gene loci (*wt1*, *raldh2*, *tcf21*, *tbx18*) between human, zebrafish and chicken, where a total of 39 conserved regions (CR) were identified. Transgenic mice harbouring each CR driving nLacZ and luciferase were created to confirm the enhancer-like activity of the identified, conserved regulatory elements within all the CR loci. Through screening, two CRs were identified to harbour cis-regulatory activity during epicardial development, namely, CR2 in the *Raldh2* locus and CR14 in the *Wt1* locus. The authors further examined the putative TFBS within the CR with the transcription element search system (TESS), to reveal multiple putative TFBS, which may underlie the enhancer activity. The most prominently repeated TFBS with 10 putative sites between the two identified CRs was the CAAT/ Enhancer Binding Protein (*C/EBP*), a pioneer TF which harbours the capability to open up the previously closed chromatin structure making the chromatin accessible to other TFs and co-activators thus facilitating the activation of the enhancer.

EMSA was employed to confirm the binding of C/EBP to nine out of ten of the identified

putative binding sites, and perturbation of these binding sites resulted in diminished enhancer activity, indicating the functional importance of *C/EBP* in regulating the identified enhancer activity. A tiled deletion-scanning approach further revealed the cooperativity of *Hox*, *Meis* and Grainyhead (*Grh*) TFs in establishing a transcriptional code for embryonic epicardial gene expression during development. Notably, both the enhancers were activated following cardiac injury by coronary artery ligation of adult mice, showing that at least some developmental enhancers are reactivated to drive embryonic gene expression in response to cardiac injury and the regenerative process.

Another study investigated the upstream gene regulation of *wt1,* especially on the epigenetic regulation of cis-regulatory elements underlying epicardial activation (Vieira et al., 2017). BRG1 is a subunit of the switch/sucrose fermentable (SWI/SNF) complex, which primarily functions in chromatin remodelling and affecting chromatin accessibility. *Brg1* was found to be expressed in the heart during development, overlapping the *wt1* expression domain. Evolutionarily conserved regions (ECR) were identified by sequence comparison between human and mouse genomes around the *wt1* locus, yielding seven ECRs which potentially constituted active regulatory elements. ChIP-seq for BRG1 was performed, which revealed that BRG1 is strongly associated with the *wt1* promoter and two other ECR identified elements. Furthermore, sequential ChIP (re-ChIP) was performed on BAF180, another member of SWI/SNF complex. The strong re-ChIP signal indicates that BAF180 associates with BRG1-enriched ECR elements, which confirmed the formation of the SWI/SNF chromatin remodelling complex. Small interfering RNA (siRNA) used to knock-down *Brg1* concomitantly decreased the *Wt1* and *Raldh2* expression in development, supporting that *Brg1* is required for *Wt1* expression. Further ChIP-seq experiments indicates that C/EBP enhancer binding protein forms a complex with BRG1 at the ECR sites, which serves to affect the chromatin conformation state, leading to increased

accessibility of the ECR cis-regulatory regions during development and regeneration. However, from the evidence presented, it would still be difficult to determine if all developmental epicardial enhancers are reactivated after injury, or whether there is an interplay between developmental and injury-specific enhancers to drive the injury response.

1.7 Identification of enhancers

1.7.1 Histone modifications, variants and co-activators

1.7.1.1 Epigenetic histone modifications

Histone modifications are covalent post-translational modifications which are enzymatically added or removed from histone subunits to epigenetically control the expression of genes by the alteration of chromatin structure. These modifications include phosphorylation, acetylation and methylation of histone subunits which makes up nucleosomes. Specific patterns of histone modifications have also been found to allude to functional enhancer elements; therefore, understanding the characteristics of each histone modification enables genome-wide interrogation of cis-regulatory regions with ChIP-seq.

H3K79me3 histone modification has previously been associated with gene repression (Zhang et al., 2015). However, other studies have also found this mark to be associated with active developmental enhancers in multiple cell types. Another well-known histone modification highly correlated with active enhancers is the H3K27ac mark. It has been demonstrated to represent increased levels of DNA accessibility at the locus, as well as increased proximal gene transcription; hence the detection of the H3K27ac mark has been used as a marker indicating active enhancers (Creyghton et al., 2010). H3K27me3 lies on the same H3 histone at the lysine 27 position; however, the trimethylation at the position has been associated with transcriptional silencing and a closed nucleosomal structure, closely

linked to the histone methyltransferase activity of polycomb repressive complex (PRC) (Ringrose & Paro, 2004). However, studies have also shown that bivalent chromatin with both H3K27me3 and H3K4me1 represents poised enhancers, as H3K4me1 facilitates the recruitment and stabilisation of the SWI/SNF nucleosome remodelling complex BAF to remodel nucleosomes on enhancer loci, facilitating enhancer activation (Rada-Iglesias et al., 2011). However, the detection of single chromatin modification (e.g. H3K27ac) has been found to be insufficient in identifying all active enhancers; hence, the combined detection of the poised and active enhancer modifications 3K4me1, and H3K27ac and H3K79me3 has been proposed as an improved modality to predict active enhancers.

1.7.1.2 H3.3 histone variant

H3.3 is a histone variant which differs from the canonical H3 by four amino acids. H3.3 was found to confer nucleosome hypermobility as well as nuclease and salt sensitivity. Studies of histone dynamics showed that canonical histone H3 was found to follow cell-cycle dependent incorporation into chromatin, as it was found that the administration of cell-cycle inhibitors blocked the deposition of the histone H3 (Ahmad & Henikoff, 2002). The variant H3.3 however, was unaffected by cell-cycle inhibition; hence, its deposition mechanism is deemed replication- independent. H3.3 density positively correlated with RNA pol-II activity, suggesting that H3.3 is incorporated during nucleosome-turnover of the pre-existing histones of the transcription process. The enrichment of H3.3 was also identified far upstream as well as downstream of the transcribed regions of genes, showing that H3.3 may mark cis-regulatory areas as well (Goldman et al., 2017). A recent publication studying cardiac regeneration in zebrafish found enrichment of H3.3 in enhancer regions near several upregulated cardiac genes during the cardiac injury healing response. However, they found that 78% of these H3.3-marked areas were not overlapping with the known activation histone modification H3K27ac, which was hypothesised to be

differential regulation during the regeneration event. They further showed that H3.3-associated enhancers, which they termed cardiomyocyte regeneration enhancer elements (CREEs) could be used to identify specific subsets of cardiomyocytes which are activated during regeneration (Goldman et al., 2017). However, the function of H3.3 is not without debate, as it was found that H3.3 also marks regions of telomeres which are unlikely to be actively transcribed (Szenker, Ray-Gallet, & Almouzni, 2011).

1.7.1.3 Co-activator: P300

P300 is a co-activator which is demonstrated to be associated with enhancers and promoters. Analysis of the protein structure of P300 showed that it contains a PHD domain which allows it to be recruited to and bind to chromatin. Additionally, P300 has a CREB-binding domain (KIX) which is a site for TF docking and protein-protein interaction, a bromodomain for chromatin binding and modulation of histone acetyltransferase activity, and lastly, a HAT domain which mediates a transfer of an acetyl molecule onto histones and proteins (Delvecchio et al., 2013). Genome-distribution study of P300 binding revealed a few interesting characteristics of P300 enrichment and pointed towards its close association with enhancers. Firstly, it was found that 75% of the P300 enrichment occurs at least 2.5kb away from known promoter sites, which is consistent with the diffuse and inconstant locations of enhancers in the genome relative to their cognate promoters (Heintzman et al., 2007). Analysis of the chromatin sequences near to the distal P300 sites shows clusters of TFBS in the surrounding vicinity (Blanchette et al., 2006). P300 enrichment at chromatin regions is also found to be enriched for the poised enhancer mark H3K4me1 (Gorkin et al., 2012). Lastly, P300 was found to acetylate the lysine residue on histone H3 at position 27, which, as previously discussed, is a histone modification used widely to predict active enhancers (Tie et al., 2009). Hence, the points above suggest the ability of P300 to predict at least a subset of enhancers (Heintzman et al., 2007; Witte et

al., 2015).

The methods for the detection of chromatin modifications and histone variants, although useful in predicting a subset of active enhancers, have several shortfalls: Firstly, despite many sources pointing towards the use of these modifications to identify enhancer regions, there are no agreed current standards on the combination of histone modifications that points out genome- wide enhancer localisation accurately and definitively. Secondly, the ChIP-seq technique requires large numbers of cells in the order of hundreds of thousands and, therefore, is not feasible for scarce cell populations (Miraldi et al., 2019).

1.7.2 Chromatin accessibility assays

1.7.2.1 DNase-seq

The packaging of chromatin is made up of DNA sequences intervals of 147bp wrapped around the octamer histone complex, forming a "beads on a string" conformation (Szerlong & Hansen, 2011). When DNA is tightly bound and wound around the histone complex, this structure of the nucleosome with relevant repressive histone marks represents a "closed" and condensed conformation of inactive genes, which safeguards against inappropriate TF binding and activation. When temporal expression of a gene is needed during development or specific cellular processes, nucleosomes will be displaced at various cis-regulatory sites and gene exons to facilitate chromatin accessibility.

Enhancer elements characteristically exhibit nuclease hypersensitivity, due to their nucleosome-free accessibility state, of which techniques such as DNAse-seq and variants such as MNase-seq and FAIRE-seq capitalises on to identify putative enhancers through the identification of open chromatin. These techniques utilise specific endonucleases, such as DNase-I, to catalyses cleavage events within DNase-I hypersensitivity sites (DHS) on chromatin, which correlate with nucleosome-free chromatin sequences (Weintraub &

Groudine, 1976), which may represent active enhancer regions. As such, a genome-wide accessibility map showing both nucleosome-bound and nucleosome-free sites can be generated. In addition, TF occupancy can be assessed whereby the presence of TF binding on chromatin protects the underlying chromatin from being cleaved by DNase thus leaving a "foot-print" of to enable the identification of binding motifs at the nucleotide level (Hesselberth et al., 2009).

1.7.7.2 ATAC-seq

Assay for Transposase Accessible Chromatin with high throughput sequencing (ATAC-seq) is a new and robust technique for mapping chromatin accessibility genome-wide by assaying for nucleosome-bound and nucleosome-free regions which requires only a minimal number of 500-50,000 cells (Buenrostro et al., 2013). The method consists of a simple two-step procedure. The first step involves genomic DNA fragmentation and simultaneous tagging named "tagmentation" which utilises a hyperactive prokaryotic transposase Tn5, to catalyse the integration of sequencing adaptors into regions of open chromatin in unfixed cell nuclei (Scharer et al., 2016). Steric hindrance from the nucleosome-bound heterochromatin sequences makes the transposing event less possible, thus creating a heterogeneous epigenomic map where nucleosome-bound sites and open chromatin can be distinguished. Tagmentation results in the preferential generation of DNA fragments at open chromatin regions, which can be amplified to construct a sequencing library for paired-end NGS subsequently. ATAC-seq thus reveals accessible regions of chromatin, such as the TSS, promoter regions, enhancer regions (Buenrostro et al., 2013). The sensitivity of ATAC-seq is suggested to be high enough to identify CTCF chromatin-binding factor footprint, comparable to standard techniques such as ChIP-seq and DNase-seq, which requires approximately three orders of magnitude more cells. Hence ATAC-seq would be ideal for the analysis of rare cells and subtypes where the starting material is in

short supply. Hence, the technique is able to circumvent the need for *ex vivo* expansion of cells which compromise on stage-specific epigenetic signatures due to the change in environmental cues and the lack of local and system cross-talk which may cause alterations in the histone modification state, which is not reflective of normal *in vivo* conditions. Another advantage of ATAC-seq is its ability to assay for both open regions and nucleosome occupation concurrently, broadening its usability in studying chromatin structure and gene regulation (Buenrostro et al., 2013). However, merely assessing chromatin accessibility state, cannot reveal whether these regions indeed harbour *in vivo* activity; hence, enhancer testing using reporter gene-based assays has to be carried out downstream of ATAC-seq.

1.7.3 Enhancer RNA

Enhancer RNA (eRNA) has been reported to be a viable modality to predict a subset of active enhancer activity (Ding et al., 2018). eRNAs are usually short, unspliced, non-polyadenylated RNA sequences which have no protein-coding ability, classified as a subclass of non-coding RNAs (ncRNA). The understanding of the *in vivo* biochemical functions of these transcripts is still limited and controversial. When they were first discovered, there were speculations that eRNAs may be transient by-products of RNA Pol-II transcription, which have no specific biochemical functions (Lam et al., 2014). Other findings and alternative viewpoints soon emerged, suggested that eRNAs may represent enhancer activity in the chromatin looping process through the recruitment of the cohesion protein which stabilises the enhancer-promoter linkage during transcription (Li et al., 2013). Only half of all active enhancers marked by the active chromatin mark H3K27ac were found to transcribe eRNAs. The genome-wide study of eRNA and gene regulation in mice suggests that enhancers exist in two states, non-eRNA-transcribing and eRNA-transcribing (Cheng et al., 2015). The eRNA-transcribing enhancers globally correspond to marked

increased expression of their cognate genes as compared to their non-transcribing counterparts across all tissues examined. Although looping events are more likely to occur at eRNA-transcribing enhancers, other evidence shows that eRNA is not an absolute prerequisite for chromatin looping events which still take place even when eRNA transcription is blocked with the small molecule flavopiridol (Hah et al., 2013). Hence, the use of enhancer RNA may be considered a viable modality, but not yet a robust method, to predicting active enhancers.

1.7.4 Chromatin conformation capture method and variants

When activated, an enhancer forms a chromatin loop and directly associates with its cognate promoter (Tolhuis et al., 2002). Chromatin conformation capture (3C) and technical variants exploit this transient chromatin association, as powerful biochemical techniques for determining which genes specific enhancers associate with directly and are focussed on capturing the physical interactions between different loci of chromatin at a given time-point, allowing sequence analysis of the elusive enhancer-promoter regulatory interactions underlying chromatin looping (de Laat & Dekker, 2012; Palstra et al., 2003). Chromatin conformation assays are established on the basis that proximity of physical interaction of chromatin can be covalently cross-linked by formalin, restriction enzyme fragmented and ligated, for downstream sequencing to determine locus identity (Dekker et al., 2002).

In the original 3C method, the ligated products would have their formaldehyde cross-link reversed to yield linearized ligated constructs of the interacting sequences, which are then evaluated by quantitative polymerase chain reaction (qPCR) using designed primers to determine the frequency of interaction. This requires insight into the interacting chromatin loci which underpins primer design for the qPCR, and as such 3C is only suitable for the study of an interaction between a single pair of genomic loci also informally known as

one-vs-one (Dekker, 2006). Subsequently, improvements were made to the 3C technique, to improve its usability as well as resolution. This gave rise to circular 3C (4C), also known informally as one-vs-many. 4C is an adaptation, whereby interacting chromatin are restriction enzyme digested and then circularized, while inverse PCR replaces conventional PCR as the amplification tool of choice as it effectively circumvents the limitation of conventional PCR in needing primers which are complementary to both interacting sequences. Oligonucleotide probes are used to capture previously unknown unique binding partners which would be sequenced. Other subsequent 3C adaptations such as 5C, Hi-C and ChIA-PET all improve the technique to achieve greater genomic coverage. However, the various methods may not be considered high-resolution or high-throughput due to the low number of informative interactions that can be generated (Sati & Cavalli, 2017).

Capture-C is a high-resolution, high-throughput variant of the 3C technique which combines 3C methodology, high-throughput sequencing and oligonucleotide capture technology (OCT), to analyse cis-regulatory element interactions at hundreds of gene loci and high resolution (Hughes et al., 2014). Firstly, the high-resolution 3C library is generated with a frequently cutting restriction enzyme DpnII using the 4C protocol, and the resulting circular DNA is sonicated into 300bp fragments where sequencing adapters would be ligated. Using OCT, probes may be multiplexed for each target region to capture hundreds of targeted loci in one single assay, and the interacting regions could be chosen without the need to consider the difference in chromatin structure, function or binding profiles. Due to the technique's high resolution, capture-C is touted to be able to identify hundreds to thousands of informative interactions from each OCT-targeted fragment (Zhao et al., 2006).

1.8 Phylogenetic foot-printing to identify putative cis-regulatory regions

With the advent of whole-genome sequencing, increasing availability of genomic sequence data and understanding about properties of cis-regulation led to algorithms which are designed to decipher the regulatory code. Phylogenetic footprinting is a computationally based method which analyses inter-species sequence conservation of non-coding sequences to shortlist putative evolutionary conserved cis-regulatory regions (Z. Zhang & Gerstein, 2003). The underlying principle behind phylogenetic footprinting is that evolutionary selective pressure causes non-functional elements to evolve faster than functional elements, leading to increased conservation. Hence, well-conserved sequences across species with evolutionary divergence serve as good candidates for functional regulatory regions. The standard phylogenetic footprinting method such as CLUSTALW constructs multiple genomic alignments of putative regulatory elements from several model organisms and identifies conserved sequences with high similarity (Chenna et al., 2003). Although generally non-coding sequence conservation is decreased in highly divergent species due to evolutionary distance, there are indeed striking examples. Several phylogenetic footprinting studies of highly evolutionarily divergent organisms successfully identified functional elements with *in vivo* function, such as the identification of a conserved retinoic acid response element within the enhancer of the *Hoxb1* homeobox gene in the mouse, chicken and pufferfish (*Fugu rubripes*) (Marshall et al., 1994).

However, there are some considerations when utilising this method. Firstly, the length of the region studied as the drawback of alignments is that functional elements could be shorter than the surrounding non-functional sequences, leading to noise which could cause misalignment and losing out on potentially useful conserved elements (Tompa, 2001). Secondly, evolution may lead to significant TF binding site degeneracy, which results in inexact binding motifs for TFs, posing a more significant challenge for the discrimination

of functional TFBS sequences from the non-functional background (Zandvakili et al., 2018).

1.9 Testing enhancer activity through transient transgenesis

Although the above-discussed methods are good predictors of cis-regulatory enhancers, resulting in the higher amount of putative cis-regulatory regions, they are all indirect measures of enhancer activity, and there is no guarantee that all identified genetic regions represent functional enhancers; hence there is still a need to validate these *in vivo*. ENCODE is a project undertaken by a consortium of international research groups, compiling genome-wide techniques such as ChIP-seq and DNase-seq, focussed on the identification of regulatory regions in the genome for the mouse, drosophila, c. elegans and human models (Bernstein et al., 2012). A systematic review on ENCODE predicted regulatory elements revealed that only 26% of the regions demonstrated enhancer functions *in vivo*, further showing the need for functional validation of putative enhancers (Kwasnieski et al., 2014).

Transgenesis is an effective method for testing enhancer activity *in vivo*. Although the technique is low-throughput, it provides direct evidence of the biological significance of the putative regulatory cis-regulatory regions. The enhancer-reporter system may be implemented by cloning the putative regulatory region upstream of a minimal promoter which drives a qualitative marker such as a fluorescent protein. In the zebrafish model, due to the transparency of the embryos throughout development, the fluorescent enhancer-reporter allows the effective visualisation and to characterise the spatiotemporal activity of the enhancer during development.

The Tol2-mediated transposition system is a well-established method of facilitating genomic transposition and has gained widespread use in the zebrafish community for the

generation of transgenic reporter lines (Ishibashi et al., 2013). However, the technique has inherent high variability in expression, which may be attributed to the non-specific background or mosaic expression of reporters, particularly in the F0 generation for transient sequence integration experiments (Chong-Morrison et al., 2018). Hence, enhancer reporter activity can only be reliably characterized in the F1 carrying the relevant genetic integrations.

The Ac/Ds system is a transposition system first identified in maize plants, as described by Barbara McClintock in 1950 (McClintock, 1950). The transposition system relies on two dissociation (Ds) genetic elements which flank the genetic sequence to be transposed, and an activator (Ac) transposase RNA which recognises the Ds sites and catalyses the transposition event. The Ac/Ds transposition system was further studied and optimised as a method to introduce transgene construct integration into the zebrafish genome, with the focus on the optimisation of germline propagation efficiency. The optimised system has since been adopted for several studies involving chemical-inducible transactivation systems (Emelyanov & Parinov, 2008; Kenyon et al., 2018) as well as mutagenesis gene trap screening (Quach et al., 2015). Recently, the Sauka-Spengler lab has optimised and validated the Ac/Ds system for its use in transient transgenesis *in vivo* in the zebrafish model (Chong-Morrison et al., 2018). Using a previously characterised strong *Pax3a* enhancer as well as a weak *Sox10* enhancer respectively cloned into the pVC-Ds-E1b:eGFP-Ds constructs with the enhancers controlling the expression of the enhanced green fluorescent protein (GFP) expression cassette, the authors compared the efficiency of the Ac/Ds system to the Tol2-mediated system on multiple fronts. It was firstly found that, for the clean expression pattern of the enhancer constructs, the Ac/Ds system required only 30pg of Ac/Ds vector as opposed to 150pg of Tol2 vector with 24pg of Ac or 50-90pg of Tol2 mRNA respectively, showing the Ac/Ds system harboured a higher somatic

integration efficiency. The higher integration efficiency of the Ac/Ds system also results in lower mosaicism in injected F0 embryos, due to the lower requirement of injected DNA constructs minimises ectopic activity resulting from episomal expression of non-integrated plasmids (Chong-Morrison et al., 2018). Cell toxicity could also be circumvented due to the smaller amount of Ac RNA needed for the physical integration process. The Ac/Ds transposition system was further demonstrated to be effective in enhancer reporters with either stronger or weaker enhancers. Altogether, the Ac/Ds system was proven to be an efficient transposition method for use in transient transgenesis experiments in zebrafish embryos for enhancer validation.

1.10 Thesis aims

1.10.1 Working hypothesis

Since the molecular events underlying epicardial gene activation and reactivation are mostly unknown, we are interested in identifying and studying cis-regulatory regions which are involved in epicardial development, as well as upstream gene regulatory networks (GRNs) to understand the molecular inputs into the formation of the epicardium and how its cellular derivatives contribute to the developing zebrafish heart. Thereafter, understanding how embryonic epicardial genes are regulated during development may be extrapolated to the reactivation/re-expression of embryonic gene programmes during adult cardiac regeneration. Ultimately, we aim to discover novel gene regulation or pathways which may be augmented in resident adult epicardial populations, for improving human cardiac regeneration.

1.10.2 Aims

1) Identification of epicardial genes of interest based on existing RNA-seq datasets
2) To characterise the expression pattern of the gene of interest

3) To verify putative enhancer activity as detected by ATAC-seq data in mosaic transient transgenesis
4) In-silico investigate TF-enhancer input upstream of confirmed enhancers
5) To perturb predicted binding sites required for TF-enhancer binding interactions *in vivo*, to elucidate regulatory activity upstream of enhancer activation

Chapter 2: Materials and methods

2.1 Zebrafish husbandry

The adult zebrafish were maintained in circulating water aquarium systems at 50 fish per litre of system water, with a controlled environment of 28.5°C, pH7.1-7.2, and a normal light/dark cycle consisting of 14 hours of light and 10 hours of darkness. The zebrafish were fed three times daily with a diet of protein pellets, flaked fish food and enriched brine shrimp. All collected zebrafish embryos were raised in 10cm Petri dishes at 28°C with the addition of 0.003% phenylthiourea (PTU) from 2dpf to prevent the development of melanin-producing pigmentation cells. All zebrafish procedures have been approved by the Research Ethics Committee of the University of Oxford, under project license (PPL) number: PDDE89C84 and personal license (PIL) number: I00833E3.

2.2 Zebrafish lines

Wild-type zebrafish lines used in the current study originates from Nigel Holder's lab at King's College London. The transgenic reporter lines described in this study, Tg(*enhancer 14:GFP*), Tg(*myl7:dsRed*), Tg(*tcf21*:*mCherry*) were generated in-house by Filipa Costa Simoes and Michael Weinberger. The transgenic lines were maintained as heterozygous carriers which were in-crossed or out-crossed when necessary.

2.3 Enhancer amplification and cloning

CE1 and CE2 putative enhancer sequences in the intronic regions of *vipr2* locus were isolated (Invitrogen Purelink genomic DNA purification kit) and amplified with PCR primers containing additional 5' and 3' homology-arms to the forward and reverse primers respectively. The DNA construct was then amplified with three-step PCR protocol (95°C for 3 minutes for one cycle, 95°C for 20 seconds, 60°C for 30 seconds, 72°C for 90 seconds for 5 cycles, 95°C for 20 seconds, 55°C for 30 seconds, 72°C for 90 seconds for

30 cycles and 72°C for 5 minutes for 1 cycle). The PCR products were run on a 1% agarose gel electrophoresis, and bands of the correct product size were excised under UV light and after that gel-purified with Gel and PCR purification kit (Machenery-Nagel) and quantified with NanoDrop2000 (Thermofisher Scientific).

Enhancer name	5’ Homology-arm (lower-case) and forward primer sequence (upper- case)	3’ Homology-arm (lower-case) and reverse primer sequence (upper-case)
CE1	tcgagtttacgtaccgctagTGAAAACAG CAGCAGAAGGG	tatcgccgcaagcttgctagCTTAATTGTG GAGCTGCGCA
CE2	tcgagtttacgtaccgctagACTTCGACTC TGAGGTGAACA	tatcgccgcaagcttgctagGCCAAACCA TCCAAAGCTAATTG

Figure 2.1: Primers for cloning putative enhancers CE1 and CE2

The table above shows the primers used to clone putative enhancers CE1 and CE2. Right column shows forward primer sequences, left column shows reverse primer sequences. Lower-case represents infusion arms sequence, upper-case represents complementary primer sequences for sequence of interest.

Enhancer-reporter vectors were generated with In-fusion HD cloning kit (Clontech). The PCR amplified enhancer sequences with corresponding 5’ and 3’ homology-arms were cloned into E1b-GFP or E1b-mCherry vectors upstream of the E1b minimal promoter using the NheI restriction site within the multiple cloning site by Gibson assembly. The E1b-GFP/Cherry vector contains various elements, such as the E1b minimal promoter sequence, NheI restriction site and infusion homology arms for site-specific insertion of enhancer sequences, ampicillin resistance gene for antibiotic selection of positive bacterial clones and flanked by 5’ and 3’ Ds transposable elements for Ac RNA-mediated transposition (figure 2.2). Chemically competent cells (Clontech) were transformed with the respective enhancer-reporter vector and plated onto ampicillin agarose plates overnight to select for successfully transformed bacterial colonies containing the ampicillin resistance gene. Individual colonies were picked and grown overnight in lysogeny broth

(LB), and the enhancer-reporter vectors were isolated using the Qiaprep Spin Mini-Prep kit (Qiagen). The isolated enhancer-reporter vector constructs were verified by PCR using primers flanking the MCS (figure 2.2) forward and reverse primers, sequence 13 and 14 respectively) for size-based determination of the insert by gel electrophoresis. The gel electrophoresis bands were gel purified by Wizard SV gel and PCR clean-up kit (Promega) and sent for Sanger sequencing and aligned to the enhancer sequence on snapgene for identification.

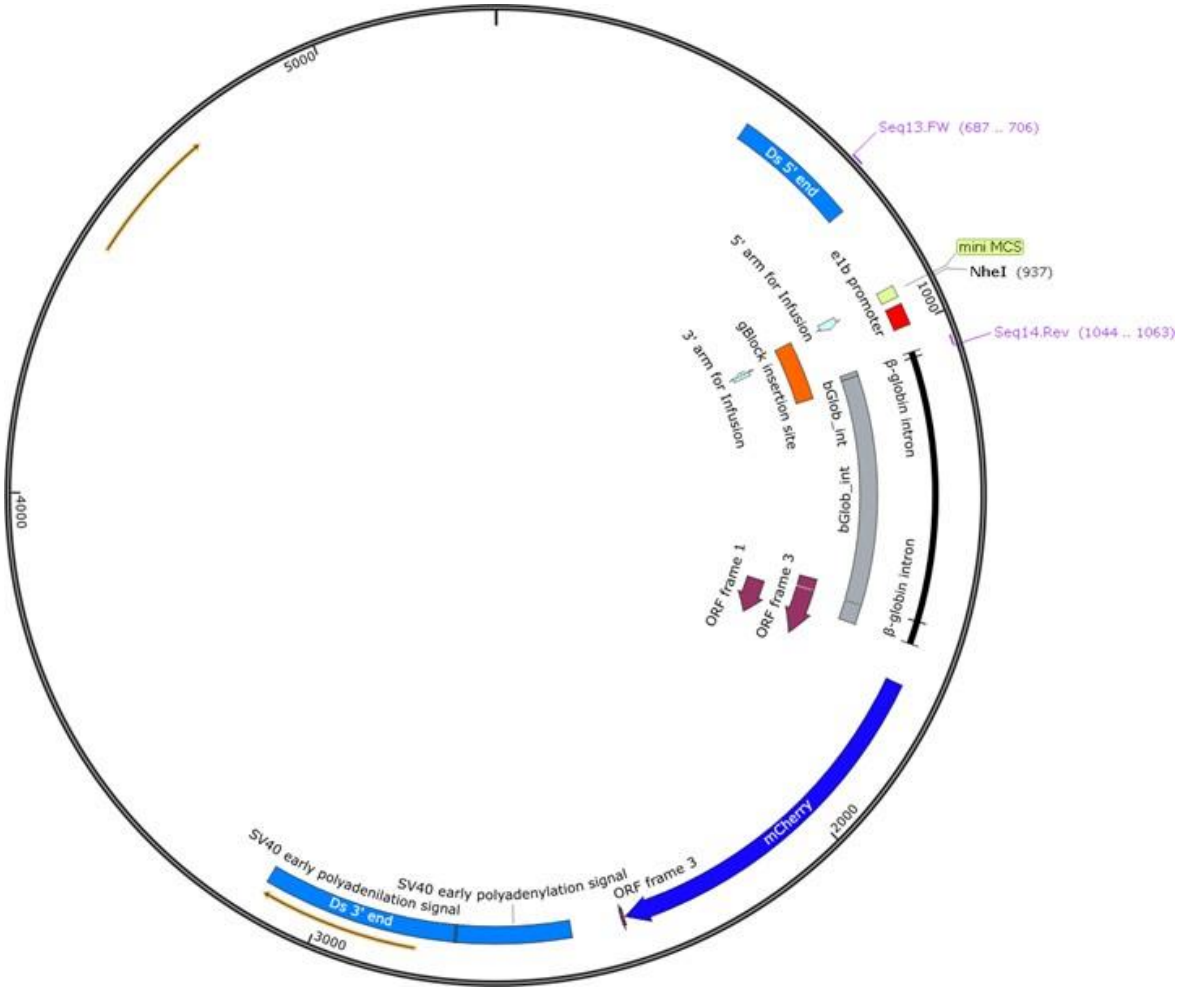


Figure 2.2: Vector map for E1b:Cherry vector

The vector construct contains a minimal E1b promoter (red), 5' and 3' homology arms for in-fusion cloning and the NheI restriction enzyme site for linearization of the plasmid vector and insertion of enhancer sequences. The vector contains the coding sequence for mCherry fluorophore, which serves as a fluorescent reporter for enhancer activity. A similar vector construct contains GFP instead of mCherry fluorophore for

cloning enhancer-reporters in spectrally distinct fluorophores. The enhancer sequence, E1b minimal promoter and mCherry fluorophore/ GFP are flanked by 5' and 3' Ds transposable elements for transgenesis into the zebrafish chromosome.

2.4 Transcription of Ac RNA for micro-injection procedure

Ac RNA was synthesised from the BamHI (NEB) linearised vector pAC-SP6, and in-vitro transcribed with the SP6 mMessage mMachine kit (Ambion). The transcribed Ac RNA was precipitated with 10M NH4Ac at a concentration of 0.3M with 100% ethanol at -80°C for 20 mins. The samples were centrifuged at 12,000RPM for 30 minutes at 4°C. RNA pellets collected were washed in 70% ethanol, air-dried at RT to remove remaining ethanol and finally resuspended in RNAse free water. The product was run on an RNAse free 1% agarose gel to ensure the transcription process has succeeded, and to confirm the size of the product. The samples were then quantified with the Nanodrop 2000 spectrophotometer (Thermofisher Scientific).

2.5 Transgenesis micro-injection procedure

About 12-16 hours before the microinjection experiment, zebrafish breeding pairs were set up in individual breeding tanks and separated by a tank divider. In the morning just before the microinjection procedure, the dividers were removed. The eggs were collected using net sieves from natural spawning and were aligned against the side of a fixed glass slide in a 10mm petri dish, ready for the microinjection procedure.

For the micro-injection of one-cell stage zebrafish embryos, the working solution containing 60ng/µl DNA vector constructs, 20 ng/µl Ac RNA and phenol red were prepared immediately before the injection procedure and loaded onto glass capillary needles. Prior to the microinjection, the diameter of a drop of the working solution expelled from the microinjector would be measured with the microscope eyepiece graticule to quantify the volume of working solution per injected embryo. The pressure of the injector

was adjusted to ensure 30pg of DNA vector and 8pg of Ac RNA would be reproducibly injected into each embryo for transgenesis. The injected embryos were observed for heart-specific fluorescence temporally until 5dpf and collected for staining, hybridisation chain reaction, or placed into the aquatics system to grow to adulthood.

2.6 Immunofluorescence and Imaging

Fertilised zebrafish eggs were collected from either natural spawning or after the microinjection procedure. The eggs were incubated at a constant 28.5°C in E3 embryo water, with the addition of PTU to block the formation of melanocytes from 2dpf, until the required experimental time-points. Zebrafish embryos collected for immunofluorescence (IF) staining were euthanised with 5g/L tricaine (MS222) and fixed with 4% paraformaldehyde (PFA) for 1 hour at room temperature (RT). The PFA-fixated embryos were permeabilized and blocked with PBTr (PBS with 0.5% Triton, 2% dimethyl-sulfoxide, 5% Goat serum) for 1-hour RT on the nutator. The samples were incubated at 4°C overnight on the nutator, in primary antibody solution (1 in 500 dilutions, chicken anti-GFP and rabbit anti-dsRed) with PBTr and 2.5% goat serum. The samples were washed in PBTr solution to remove excess primary antibody, 4x15 minutes at RT.

Subsequently, the samples were incubated at RT for two hours on the nutator in secondary antibody solution (1 in 1000 dilution, goat anti-chicken 488 and goat anti-rabbit 546 with PBTr and 2.5% goat serum, protected from light). The samples were washed 4x 15 minutes and subsequently overnight at RT, to remove the unbound excess secondary antibodies. The IF- stained samples were mounted on slides with VectaShield (H-1400, Vector Laboratories Inc) or 1% low melting-point agarose (Sigma-Aldrich) for subsequent imaging.

2.7 Hybridization chain reaction v3.0

Hybridization chain reaction (HCR) probes (version 3.0) were ordered from Molecular Technologies (Caltech). The staining procedure was performed according to the manufacturer's instructions. The zebrafish embryos maintained in embryo water at 28.5°C until 5dpf were collected and euthanised with 5g/L MS222. The samples were fixed with 4% PFA for 1 hour at RT, and dehydrated by graded increasing methanol phosphate-buffered saline (PBS) concentrations for 10 minutes each (25% methanol, 50% methanol, 75% methanol and 100% methanol solutions) and the dehydrated embryos were stored in 100% methanol at -20°C at least overnight (samples are stable in 100% methanol for up to 6 months at -20°C).

For the HCR procedure, the samples were rehydrated with 0.1% PBS-Triton and permeabilised with 20µg/ml proteinase K for 15 minutes at RT. The permeabilized samples were post-fixed in 4% PFA for 20 minutes at RT, and prehybridized with hybridisation buffer (30% formamide, 5X sodium chloride sodium citrate (SSC), 9mM citric acid (pH 6.0) 0.1% Tween-20, 50 µg/mL heparin, 1x Denhardt's solution, 10% dextran sulphate) at 37˚C for 30 minutes. The samples were subsequently incubated overnight at 37˚C with 2nM of HCR initiator probes in HCR probe hybridisation solution. Probes against GFP (amplifier B2), *vipr2* (amplifier B3), *tcf21* (amplifier B1), *myl7* (amplifier B5) were ordered directly from Molecular Instruments.

The samples were washed with HCR probe wash buffer (30% formamide, 5x SSC, 9mM Citric acid (pH 6.0), 0.1% Tween20, 50µg/ml heparin) to remove excess initiator probes. Fluorescent hairpins were snap-heated to 95˚C and allowed to cool down for 3-4 hours in the dark. The samples were incubated overnight at RT with 30nM fluorescent HCR hairpins in HCR hairpin amplification solution (protected from light). The samples were incubated with Hoechst reagent (1:1000, 5x SSC-T) for 45 minutes at RT and finally

washed with 5x sodium chloride sodium citrate triton (SSCT) (2x 5 minutes, 2x 30 minutes, 1x 5 minutes) to remove excess fluorescent hairpins. The fluorescent samples were mounted onto Superfrost glass slides using vectashield (vectorlabs) mounting medium.

2.8 Microscopy and image analysis

All high-resolution confocal images were taken on the Zeiss upright LSM780 confocal microscope, as well as the Zeiss Spinning Disc microscope.

CTGGCGCGTGATCAACTGGCCAATCATGGCATCTGTCATTGTGAGTATAACCTCACACCCGTACTTCTAAAC ACACAGACCAGCCTCATACTGTATGCATTATGTCAGGCAGGGAGGGATTCTGCCAGCAAAGCAGACGAGG GGATGTGCTGAGTCTCACAGACACTTTCCTGGATAAGACATGAATGCAGGCATGTCAGGAAGAGCAAGCA AACACGCTGTCACAAGCGCCTTTCTCAGCACACGTGCCAGAGCTCTCTCTCTCTTGTGCCATTTCCATTTTG CAACTCCTCACCTGCAGAAGATCAAATGTCTCTGGACAGCATGATGGCTATGTTTGACAGCTTTGCATCCA GGTCAAAATGTTCCTGTTACTCAGTCCTCCAGAGCCAAGAATGCAATAACAACAGAAAACGTCATAGCTTT TTTCTGCCAAACGTGCTTAGTTCTGGTAATCCACGTGTTTGGATCGCTCTTTCACAGAGCATTACATACCAA ATTGTCACTGGTAGAGGGAATCGCTGGCCAAACTGTTCTGTTTGTTTTGGTTTGAGGTTTTACGGCGGGATT AATTATGAGGAACAGAAGCAGGTGCCCTCTTTTGTTTTTACTATCTGTCTGATGGATTTAAGCAAGATGGAA ATTGAGGACGATCATTTACATTAGAGTTTACACGTCTGGCAGAATATATTTAAAACAACTTTTGGTGCAATT AAAGCATGCAGAGAAATAATTTTCAAATTTGGGGCTGCGCAGTATTGGTAAAAATCTGATATTGCAATATT CTTTTTTTCTG

Forward primer:	Reverse primer:
CAGAAAAAAAGAATATTGCAATATC	CTGGCGCGTGATCAACTGGCCAATC

Figure 2.3: Enhancer 14 sequence and respective forward and reverse primer sequences.

The figure above shows the entire 796bp (UCSC genome browser GRCz10/danRer10, located at chr24:27,258,181-27,258,976) enhancer 14 sequence as indicated by ATAC-seq results for *tcf21*+ epicardial cells at 5dpf. The corresponding forward and reverse primers used for enhancer 14 sequence isolation from zebrafish genomic DNA is as shown.

All confocal images were post-processed using the Zen software (Zeiss) or Imaris

(Bitplane) image analysis software. The high-resolution z-stack image is projected into a 3D-rendering for ease of visualisation and quantification using the Imaris image analysis software. With the mask creation tool, the different channels could be individually processed. Firstly, the GFP fluorescent (channel 2) areas were defined, and an object mask is created for the defined GFP fluorescent areas. Secondly, the cell nuclei were determined by DAPI staining (channel 1) with perimeters defining the size and shape of the cell nucleus (XY diameter = 4μm, z diameter = 6μm) which effectively creates masks for all the nuclei. Lastly, a heat map based on nucleus proximity to defined GFP fluorescence can be generated, and results can be exported to excel. Results would tally the number of nuclei proximal/overlapping with GFP fluorescence.

2.9 *In silico* binding site analysis and binding site perturbation

JASPAR binding motif database 2014 for used to predict the transcription factor binding sites for the identified enhancer 14 sequence (figure 2.3). The analysed region was defined by the area under the ATAC-seq peak, using a minimum similarity score of 87%. Using the *in silico* TFBS prediction data, putatively important TFBS families were identified using the following matrix, which has to satisfy at least 2 criteria to be considered:

1) The TF is expressed in the bulk epicardial RNA-seq dataset at 5dpf
2) Enrichment of the predicted TFBS in the enhancer sequence
3) The predicted location of TFBS in the enhancer sequence (sparsely located around the centre of under the ATAC-seq peak)
4) The length of predicted TFBS motif
5) The TF importance in cardiac development

For each putatively important TF family, sequence mutations were designed with the aid of position weight matrix (PWM) for each motif by designed point mutations for nucleotides

which are deemed more critical for the recognition by the TF. The TFBS-mutated enhancer was re-analysed with similar perimeters in JASPAR to confirm binding motif removal, as well as not introducing new TFBS inadvertently by the designed point mutations.

All designed TFBS-mutated sequence condition was ordered as double-stranded synthetic DNA sequences as gBlock gene fragments (Integrated DNA Technologies). The addition of flanking (both 5’ and 3’) homology-arms on each end of the sequence so that they could be cloned into the e1b-GFP Ac/Ds vector using In-fusion cloning after NheI linearization of the plasmid.

Chapter 3: Identification of the *vipr2* gene as a putative regulator of epicardial development

3.1 Introduction

3.1.1 *In vivo* bio-tagging toolkit for the isolation of cells for RNA-seq

Transcriptomic analyses based on the sequencing of reverse-transcribed complementary DNA (cDNA) molecules serve to provide valuable insight into the plethora of ongoing, dynamic cellular processes within cells, many of which are essential for normal function. Bio-tagging is a binary, tissue-specific biotinylation system which can isolate particular cell populations in- vivo (Trinh et al., 2017). The process makes use of two bacterial artificial chromosome (BAC) recombined transgenic zebrafish lines, the biotin ligase (BirA) driver line and the Avi effector line. In the driver line, BirA is expressed under the control of a tissue-specific promoter (*tcf21* promoter) as well as a 2A-separated mCherry fluorophore. In the effector line, β-actin promoter drives the expression of Avi-tagged RANGAP nuclear envelope protein and a cerulean fluorophore which enables the isolation of nuclei targeted by the cell-specific driver-line. The BirA-driver and Avi-effector transgenic lines were crossed, resulting in double-positive embryos with mCherry and cerulean labelled cells/nuclei. In the double-positive cells, tissue- specific expression of BirA catalyses an *in vivo* biotinylation event where a biotin group is tagged onto the Avi-tagged RANGAP protein. Magnetic streptavidin beads were used to efficiently pull-down the biotin-labelled nuclei for further processing (Trinh et al., 2017).

3.1.2 VIP and VIPR2: evolutionarily conserved neurotransmitter system

VIPR2, also known as VPAC2 or VP2 is a G-protein-coupled receptor (GCPR) of the secretin family which includes receptors for other members of the family 2 GCPRs, such

as growth hormone-releasing factor, glucagon, calcitonin and parathyroid hormone receptors (Henning & Sawmiller, 2001). In the mouse and human models, *Vipr2* shares 50% sequence homology with *Vipr1*, a sub-functionalized receptor which shares both similar and exclusive functions and receptor localisations. VIPR1 is found to be widely distributed in the central nervous system (CNS), as well as the liver, intestines, lungs, leukocytes and blood vessels of the pancreas, intestines; whereas VIPR2 is mainly detected in the CNS, aortic endothelium, skeletal muscle, adrenal cortex, renal medulla as well as the heart (Henning & Sawmiller, 2001). However, RT-PCR results of cultured rat cardiomyocytes and cardiac non-myocytes show that *Vipr2* and *Pac1* are the two main receptor transcripts expressed in cardiac cells, whereas *Vipr1* transcripts were not detected (figure 3.1) (Sano et al., 2002).

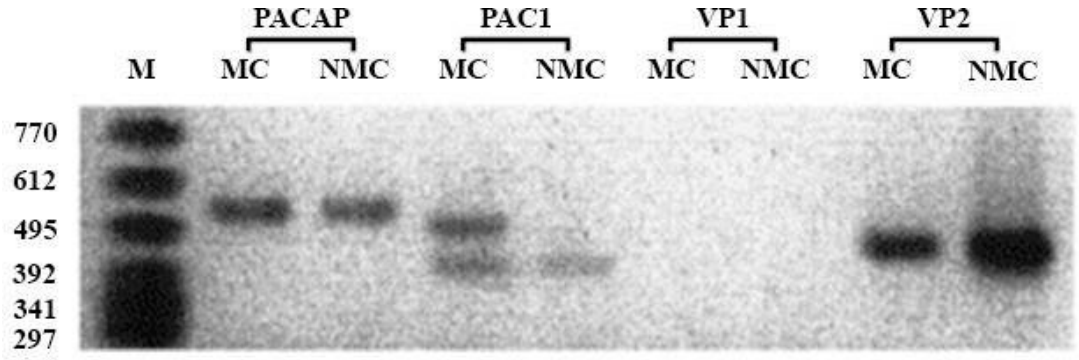


Figure 3.1: Detected pituitary adenyl cyclase-activating protein (PACAP) and vaso-intestinal peptide (VIP) receptor subtypes in cultured rat cardiac cells

RT-PCR detection of three PACAP/VIP receptor subtypes in cultured rat cardiac myocytes (MC) and nonmyocytes (NMC). M, PAC1, VP1 and VP2 represent DNA ladder, PAC1 receptor, Vipr1 receptor and Vipr2 receptor, respectively. *Vipr2* transcripts were detected as the main VIP receptor isoform present in the myocytes and non- myocytes. (Figure adapted from (Sano et al., 2002))

Two main peptide ligands, VIP and PACAP signal through binding with VIPR1 and VIPR2. In addition, PACAP signals through PACAP-specific receptors PAC1, PAC2 and PAC3. While PACAP can bind efficiently to VIPR1 and VIPR2, the VIP ligand binds to

the PAC receptors with a thousand times lower affinity than to its cognate receptors (Harmar et al., 2012). VIP and PACAP signalling through their cognate receptors have also been found to have mostly similar but non-redundant, biological effects. VIP, which functions as the primary ligand for both VIPR1 and VIPR2, is a small neurotransmitter peptide synthesized in neuronal cell bodies of postganglionic neurons and stored within large (100nm) dense-core vesicles in pre-synaptic nerve cell endings of the CNS and peripheral nervous system (PNS) where it functions primarily as a non-adrenergic and non-cholinergic neuromodulator of many tissue types (Lefebvre et al., 1995). Interestingly, the amino-acid sequence of the *Vip* gene is identical throughout many mammalian species such as humans, rat, dog, cow and sheep, and shows a high level of conservation in non-mammalian model systems such as chicken, frog, goldfish and cod (Henning & Sawmiller, 2001), only differing at four to five positions as compared to the human sequence. As *Vip* was found to be highly evolutionarily conserved, it might allude to essential functionality. Indeed, VIP was found to regulate many critical biological processes in various animal models, such as brain development (Wu et al., 2006), circadian rhythmicity (Hannibal et al., 2011), gut motility (Gregersen et al., 1991), secretion of pepsinogen, regulation of the release of both pros- and anti-inflammatory cytokines from T-helper cells (Goetzl et al., 2001) as well as potent cardiovascular effects (Henning & Sawmiller, 2001).

3.2.3 VIP and VIPR2: cardiovascular effects

VIP was initially found to be a potent vasodilator which facilitates vascular smooth muscle cell relaxation when bound *to* VIPR2. The GCPR cascade activates cyclic adenosine monophosphate (cAMP) and then the secondary activator, protein kinase A (PKA), which in turn translocate to the sarcoplasmic reticulum and phosphorylates phospholamban (Plb) which allows increased sequestration of Ca^{2+} by the sarcoplasmic reticulum leading to

decreased activation (Murray, 1990). *cAMP*, on the other hand, decreases myosin light chain kinase's affinity with the Ca^{2+}-calmodulin complex, leading to reduced myosin phosphorylation and decreases the actin- myosin affinity, resulting in smooth muscle relaxation and the vasodilation of blood vessels.

VIP-immunoreactive nerve fibres innervate many systemic and pulmonary blood vessels and have been indicated to be 50-100 times more effective, on a molar basis, than the canonical neuropeptide, acetylcholine (Ach). Aside from the vasodilatory effects, VIP-Vipr2 signalling has also shown to have anti-oxidative effects which are mediated through the release of Nitric Oxide (NO) via activation of the Nitric Oxide Synthase (NOS) by VIP activation of natriuretic peptide clearance receptors (NPR-C) (Henning & Sawmiller, 2001).

In the mouse and rat models, VIP-immunoreactive nerve fibres innervate the heart via both the parasympathetic (vagus) and sympathetic cardiac nerves, which supplies the sinoatrial as well as the atrioventricular node (AVN). In the SA node, VIP stimulation from the vagus nerve induces cAMP and activation of the hyperpolarisation-activated pacemaker I_f current in myocardial cells, which in turn causes rapid diastolic depolarization and a concomitant increase in heart rate (Accili et al, 1996). High frequency cardiac vagal nerve stimulation results in the endogenous release of VIP, which concomitantly upregulates coronary artery blood flow up to 62% in canines which have had their aortic pressure and heart rate controlled extrinsically, in addition to the administration of atropine and propranolol to block the muscarinic and β-adrenergic receptors (Feliciano & Henning, 1998). Blood flow returns to its average value about 30 minutes after cessation of vagal nerve stimulation, reflecting the slow speed of lymphatic clearing of endogenously released VIP. When VIP is administered directly to the SA node in the experimental

canines, the heart rate increases up to 37% over 40 seconds.

Furthermore, *Vipr2* transcripts have been detected via RT-PCR of mRNA transcripts in rat ventricular cardiomyocytes. VIP-induced stimulation increases PKA-dependant phosphorylation of the SERCA channel, thereby affecting the L-type calcium current and facilitating increased calcium release from the sarcoplasmic reticulum as part of calcium-induced calcium release (CICR) leading to an increase in the rate and strength of cardiac contraction (Dvoráková, 2005). *Vip* KO in mice results in decreased survival, which begins at 5mpf and resulting in 100% mortality at 20 months (figure 3.2a). *Vip* KO mice exhibits reduced body weight, asthma, increased production of proinflammatory cytokines and progressive right ventricular hypertrophy due to pulmonary hypertension (figure 3.2b). The *Vip* KO cardiac phenotype was also attributed to the dysregulation of key cardiac genes such as Ca^{2+} modulator calsequestrin, lumican, alpha tropomyosin, desmin and cardiac actin, together resulting in aberrant calcium release, biventricular dilated cardiomyopathy and increased mortality (Szema et al., 2015).

In the canine model of heart failure and human patients diagnosed with cardiomyopathy, VIP concentration in the myocardium is markedly decreased, by up to 50% (Bersell et al., 2009). During heart failure, VIP receptor density in the myocardium also decreases by up to 62%, resulting in a concomitant decrease of a similar scale in the myocardial contractile response (Henning & Sawmiller, 2001). Although receptor density is decreased, the remaining receptors were found to have increased affinity for the VIP ligand. This phenomenon indicates that the changes in VIP concentrations or VIP receptors may play a role in the pathogenesis of heart failure and hypertension.

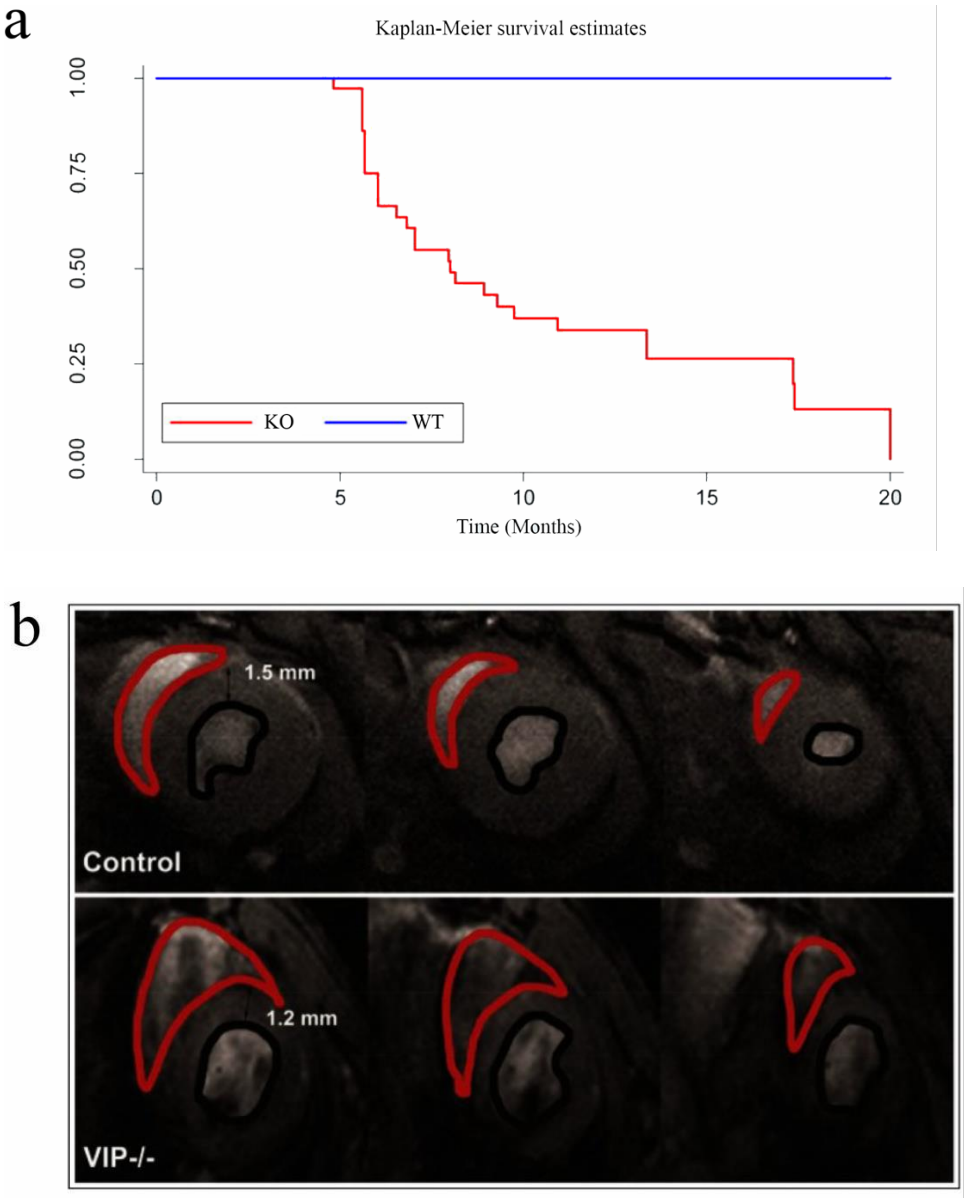


Figure 3.2: *Vip* KO mice phenotype and survival estimates

(a) Increased mortality can be observed in *Vip* KO mice as compared to wild-type mice. Progressive right ventricular hypertrophy was associated with the increasing age. (b) End-diastolic multi-slice magnetic resonance imaging images of mice hearts in the coronal plane orientation. *Vip* KO mice heart (bottom panel) and control mice heart (top panel), circled red areas represents the right ventricular area. Right ventricular dilation can be observed in the *Vip* KO mice hearts, with the RV area (red circle) approaching the size of the left ventricles (black circle). (Figure adapted from (Szema et al., 2015))

This chapter describes the selection of *vipr2* as a novel candidate regulator of epicardial development and applies the HCR technique, which is a novel in situ hybridisation

technique for the detection of target RNA via sequence-specific hybridisation to DNA probes, followed by self-assembly of tethered fluorescent DNA nanostructures in which the fluorescent signal is amplified. The aim is to demonstrate *vipr2* mRNA expression pattern during zebrafish development as an initial understanding of its possible function. Interestingly, although much work has been done on VIP-VIPR2 signalling effects in other model organisms (Henning & Sawmiller, 2001), little is known about *vip* and *vipr2* expression and role in the zebrafish.

3.2 Results

3.2.1 Differential expression analysis to identify genes of interest

The lab has previously generated RNA-sequencing (RNA-seq) data from *tcf21+* epicardial cells, during development and after adult heart injury and obtained by biotagging and nuclear pull-down, as the major subset of epicardial cells. The double-positive embryos were collected, and the hearts were isolated and lysed. Magnetic streptavidin beads were used to bind to the biotin-labelled nucleus, from which the nuclear RNA was subsequently isolated, and reverse transcribed to cDNA for sequencing (Chong-Morrison et al., 2018).

To compare epicardial gene expression during development and regeneration, three experimental groups were chosen, namely, 5dpf as the embryonic dataset, adult 3 days post-injury (dpi) and the corresponding adult sham control as the adult regeneration dataset. These time points were chosen to identify the known expression and re-expression of epicardial genes in each process during development and adult heart injury, respectively, at stages (5dpf; 3dpi) corresponding to epicardial cell expansion and activation. Computational analysis of the RNA-seq data yielded 376 genes (figure 3.3a) which were expressed in *tcf21+* epicardium during development and differentially upregulated in the 3dpi adult cryoinjury dataset as compared to the stage-matched sham control. A further 18

genes (figure 3.3b) were narrowed down for further study based on clean ATAC-seq signal around the respective gene loci for identification of putative enhancers, as well as their known or possible cardiovascular function.

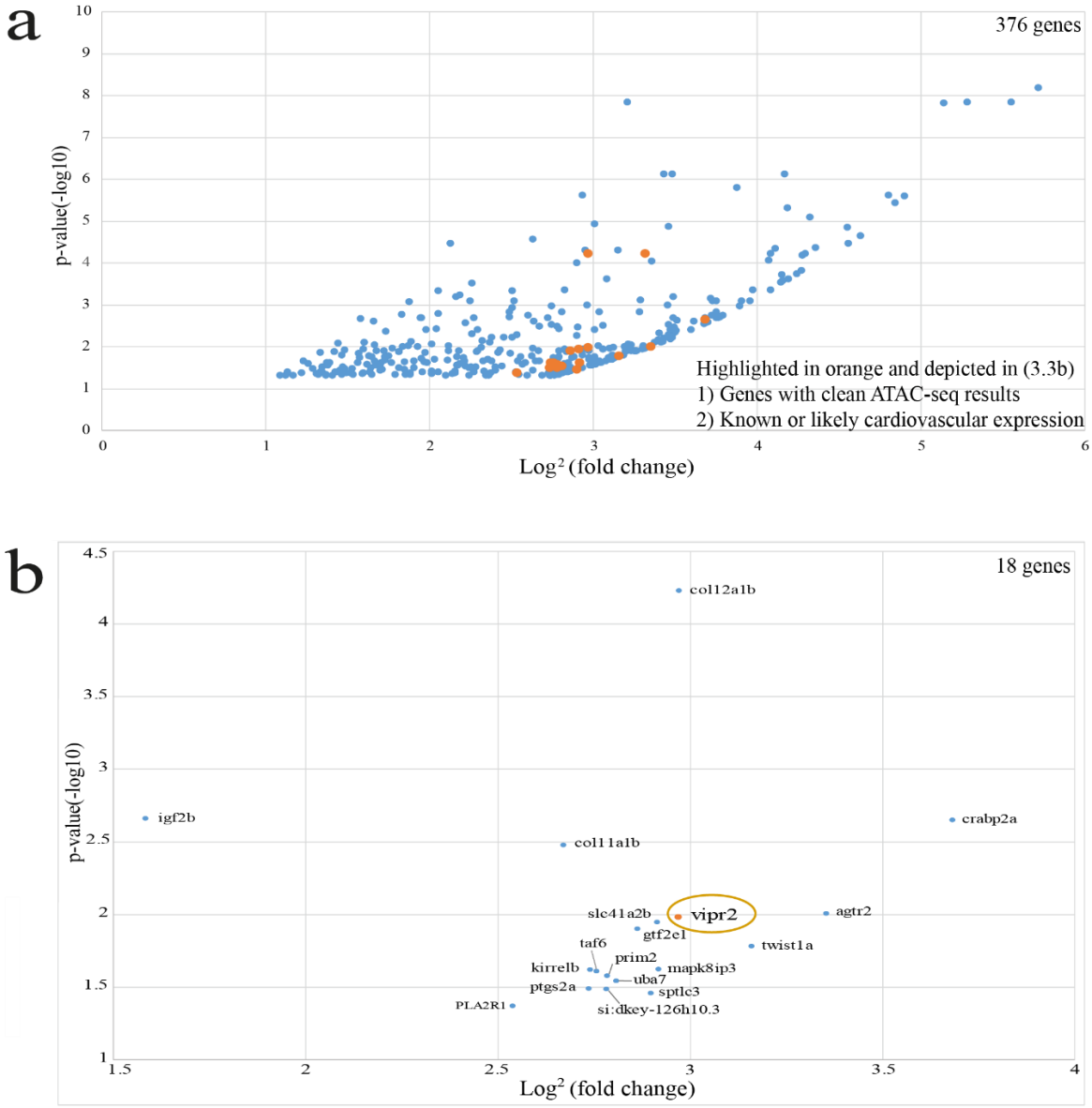


Figure 3.3: Volcano plot and scatter plot of differentially upregulated epicardial genes after cryoinjury of the zebrafish heart

(a) Volcano plot depicting 376 genes that were found to be significantly upregulated by differential expression analysis after cryoinjury as compared to sham controls, as well as being expressed in development. Gene expression data for differential analysis were from RNA-seq data previously produced by the lab. The 18 genes labelled in orange are depicted in figure 3.3b. **(b)** Scatter-plot is depicting 18 genes

which were selected from the 376 differentially upregulated genes for further characterization due to putative interesting function within the epicardium, as well as based on the clean signal from ATAC-seq data for putative enhancer calling. The gene of interest in this report, *vipr2* gene is circled in orange.

Vipr2 was an interesting target for study mainly due to its novel implication in the epicardium during development, and differential upregulation following cryoinjury in the zebrafish heart indicating a role for VIP or PACAP signalling during the cardiac regenerative process. It is especially interesting that *Vipr2* has been demonstrated to be upregulated in the canine and human myocardium during pathogenic cardiac hypertrophy, suggesting that *vipr2* may have an important role in this process. Having identified *vipr2* transcript through RNA-sequencing of biotagged population of *tcf21+* epicardial cells, this is the first report of *vipr2* expression in the epicardial cell layer during development.

3.2.2 Developmental expression of *vipr2* in *tcf21+* epicardial cells and *myl7+* cardiomyocytes

To assess the *vipr2* expression pattern and validate the developmental RNA-seq data, HCR was performed to probe for endogenously expressed *vipr2* mRNA in wild-type 5dpf zebrafish embryos. DNA probes against *vipr2* and *tcf21* mRNA were multiplexed, and fluorophore hairpins were designated to be spectrally apart to avoid bleed-through. HCR staining revealed endogenous mRNA expression as punctate fluorescent hairpin-tethered mRNA within or surrounding DAPI-stained nuclei. The expression pattern for *vipr2* was observed to be sparsely scattered in cardiac cell nuclei across the cross-section of the heart (figure 3.4a-c). *Vipr2* expression was also seen to be localised in greater density in the ventricular cardiomyocytes, as compared to the atrial cardiomyocytes in the developing zebrafish heart. *Tcf21* mRNA, as shown by punctate fluorescence, can be observed surrounding DAPI-stained nuclei of a single layer of cells constituting the outer lining of the heart. It was also observed that a small population of *tcf21* expressing epicardial cells

co-express *vipr2* mRNA (figure 3.4a-c).

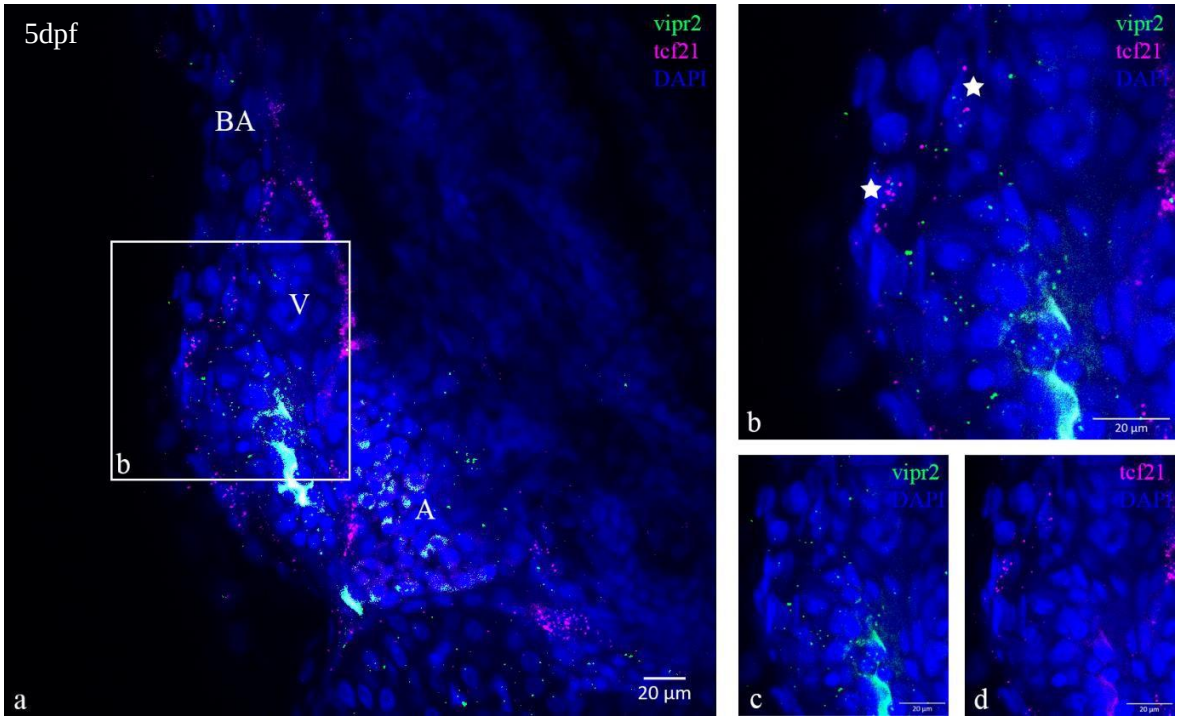


Figure 3.4: HCR detection of the co-localisation of *vipr2* and *tcf21* mRNA

(a) Cross-section of the embryonic zebrafish heart, with multiplexed HCR detection of mRNA in wild-type zebrafish at 5dpf. *Vipr2* mRNA is shown in green, and *tcf21* mRNA in magenta. The white box shows the area where the high magnification inset b (depicted on the right and two images bottom right labelled c and d) is taken from. **(b)** From the punctate fluorescent spots of both *tcf21* and *vipr2* can be seen co-localizing (white asterisk), showing *vipr2* expression in a small number of epicardial cells. **(c)** Punctate fluorescent spots showing *vipr2* mRNA on a DAPI background **(d)** Punctate fluorescent spots corresponding to *tcf21* mRNA expression in epicardial cells. V=ventricle, A=atrium, BA=bulbus arteriosus. Scale-bar=20µm.

To determine *vipr2* expression in the cardiomyocyte population, multiplexed HCR was also performed with DNA probes against *myl7*, a commonly used marker for the detection of cardiomyocytes, as well as *vipr2*, on wild-type 5dpf zebrafish embryos (figure 3.5a-c). *Myl7*-expressing cardiomyocytes were organised into a 2-3 cell thick wall constituting the myocardium of the developing zebrafish ventricle, where the trabeculation program has initiated. Sparsely located *vipr2* mRNA-tagged fluorescence pattern was observed in DAPI-stained nuclei which co-localised with *myl7*, across most cardiomyocytes that make

up the early zebrafish heart.

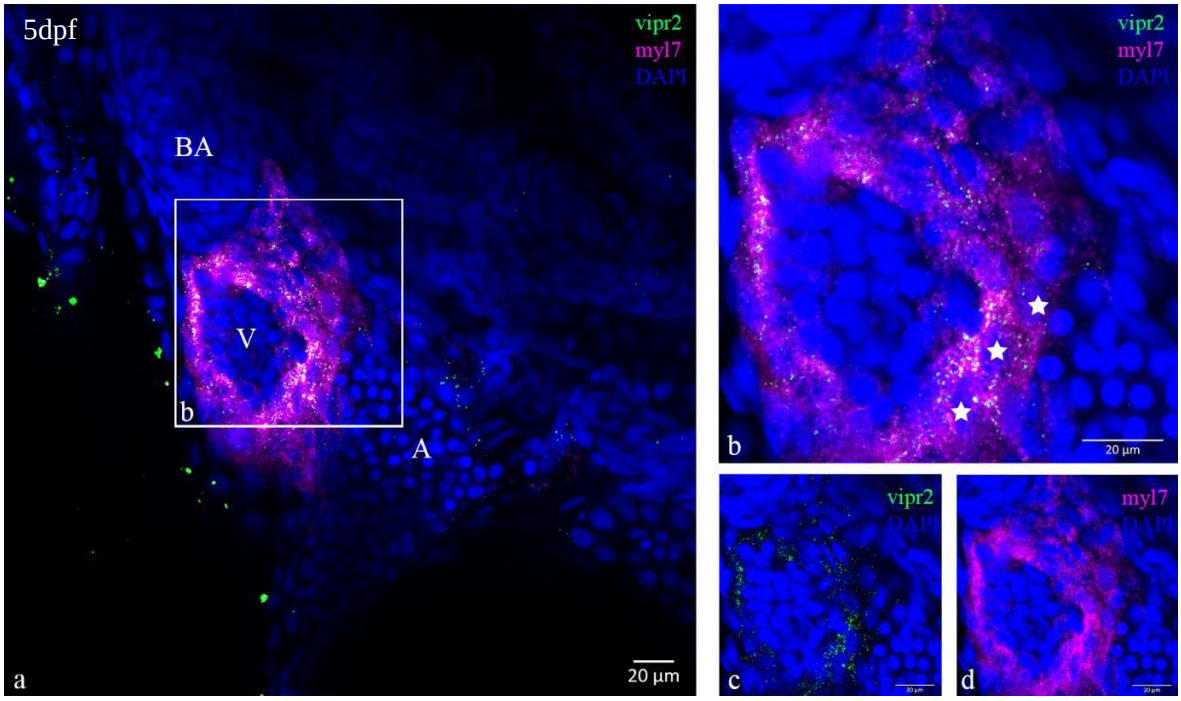


Figure 3.5 HCR detection of the co-localisation of *vipr2* and *myl7* mRNA

(a) Cross-section of the heart, with HCR detection of nascent mRNA in wild-type zebrafish at 5dpf. *Vipr2* mRNA is shown in green and *myl7* mRNA in magenta. High magnification inset is labelled b. **(b)** High magnification inset showing punctate spots of *vipr2* (green) fluorescence co-localising with *myl7* (magenta) cardiomyocytes (white asterisk). **(c)** Punctate fluorescent spots showing *vipr2* mRNA on a DAPI background **(d)** Punctate fluorescent spots corresponding to *myl7* mRNA expression in cardiomyocytes. V=ventricle, A=atrium, BA=bulbus arteriosus. Scale-bar=20µm.

3.3 Discussion

Although cardiac innervation and the cardiovascular regulatory role of the VIP-Vipr2 system has been described in detail in various model organisms, little is known about the *vipr2* orthologue in zebrafish. In this study, *vipr2* was found to be expressed in *tcf21+* epicardial cells (RNA-seq) during zebrafish development at 5dpf, and then differentially expressed post- cardiac injury at 3dpi in the regenerating adult zebrafish heart. *Vipr2* proved to be an interesting target for study due to a potentially novel role in epicardial cells during development, as well as a potential role during the injury and regenerative process that

might reflect its reported role in maintaining cardiac contractility elsewhere (Henning & Sawmiller, 2001).

However, as no previous publication has shown the expression pattern for *vipr2* in zebrafish, it was essential to determine where the *vipr2* gene is expressed in the cardiovascular system. This would allow us to understand firstly, whether there are differences in *vipr2* distribution between model organisms and secondly, to ascertain RNA-seq results, as well as the potential function of *vipr2* in the zebrafish. HCR was chosen for this purpose, as firstly, there are no commercially available antibodies with cross-reactivity in zebrafish for immunostaining. Secondly, HCR enables multiplexing of target mRNA by the use of different complementary DNA probes to be able to visualise co-expression of genes, and lastly, HCR is a sensitive technique which can amplify the signal of lowly expressed gene transcripts as compared to regular in-situ hybridisation (ISH) staining.

HCR fluorescence revealed that *vipr2* is indeed expressed in the developing zebrafish heart at 5dpf (figures 3.4a-c, 3.5a-c). In other model organisms, VIPR2 is known to be the primary VIP receptor perpetuating neuro-modulatory cardiovascular effects. In our RNA-seq dataset of *tcf21+* and *myl7+* cell populations, the closely related *vipr1* was not expressed at 5dpf nor differentially upregulated after cryoinjury, confirming that *vipr2* is indeed the main VIP receptor within the zebrafish cardiovascular system. The receptor expression was found to be expressed in most cardiomyocytes and a rare population of epicardial cells, corroborating with the RNA-seq data. The co-expression of *vipr2* with *tcf21* suggests a role for the peptide receptor in a subset of epicardial cells during heart development and co-expression with *myl7* in most cardiomyocytes indicates functional importance in cardiomyocyte physiology in the zebrafish.

Chapter 4: Genome-wide accessibility assays, verification of putative enhancers and elucidation of TF-enhancer interactions in *vipr2* expression.

4.1 Introduction

The various putative enhancers from various enhancer screens have to be tested for enhancer function *in vivo*. The putative enhancer sequences would be firstly picked, and via designed PCR experiments their genomic DNA could be obtained. The putative enhancer genomic DNA would be cloned into a fluorescent enhancer-reporter vector, which would serve to report on enhancer activity domains in transient experiments. If the putative enhancer shows enhancer-like activity, the ensuing activity pattern of the enhancer could be characterised via establishing a stable transgenic line.

4.1.1 *In silico* analysis of putative TFBS on enhancer sequences

To elucidate the upstream regulation of a target gene during development, cis-regulatory regions could be interrogated for TFBS within the enhancer locus, which may influence enhancer activity. Putative enhancer sequences were analysed using the online TFBS prediction tool JASPAR (Bryne et al., 2008), to identify all potential TFBS within the enhancer sequence. Next, putatively important TFBS were then screened for their influence on enhancer activity by transient transgenics utilising the enhancer sequence with designed point mutations aimed to perturb TF binding. In order to find out the likely importance of each putatively interacting TF family, I followed the lab's approached as detailed in chapter 2.9. A previous publication investigated significant TF-enhancer interactions across 47 cell-types, and through statistical clustering algorithms of top enriched TF-enhancer interactions in each cell type, classified six distinct positional binding- site

patterns (Grossman et al., 2018). TFs which were essential for cell-type-specific activity was found to be likely clustered sparsely at the centre of the nucleosome-free region (clusters 2 and 3) where the motif grammar shows likely TF-TF interactions underlying enhancer activity. This was used as one of the parameters in the determination of putatively important TF families.

4.1.2 Background of putatively important TFBS and corresponding TFs

Four significant TFBS were identified *in silico* as consensus sequences for binding of *meis1*, *hltf*, *hic1/2* and *stat1/6*, which are introduced below:

MEIS1 is a member of the HOX family of TFs, which are known to play crucial roles during development. MEIS1 is known to be important in cardiovascular development, involved in outflow tract (OFT) septation (Dupays et al., 2015) and endothelial cell development (Minehata et al., 2008), as well as regulation of the postnatal cardiomyocyte cell-cycle (Muralidhar & Sadek, 2016). KO of *meis1* in embryonic zebrafish leads to severe cardiac phenotypes of vascular defects and the failure to undergo heart looping, leading to embryonic lethality. The KO of *meis1* in the embryonic mouse model leads to the reduced sarcomeric formation and the increase in mono-nucleated cardiomyocytes as opposed to bi-nucleated cardiomyocytes.

HLTF is a member of the SNF/SWI family with ATP-dependant chromatin remodelling activity. *Hltf* is known to be expressed in the mouse model at E8.5-E9.5 corresponding to the formation of the linear heart tube (LHT) (Helmer et al., 2013). Studies performed in the mouse model has shown that the TF plays a variety of roles during cardiac development, in regulating transcription, remodelling of chromatin as well as coordinating DNA damage repair. *Hltf* has been found to directly regulate essential cardiac genes during cardiac development, such as, *gata4* which regulates proepicardium outgrowth, *wt1b* which is a

crucial factor in epicardial development, as well as *hif1a* which is a regulator of collagen hydroxylation and folding; all processes which are essential for the normal development of the embryonic heart. *Hltf* also plays an indispensable role in regulating the expression of key elements in centrosome maturation which is critical in G2/M cell cycle progression (Helmer et al., 2013). Therefore, the KO of *Hltf* during development leads to downregulation of cell cycle, epicardial, cardiomyocyte and collagen biosynthesis genes which are essential for the normal functioning of the heart. The KO of *Hltf* is semi-lethal, with 75% of neonatal mice pups dying within 12-24 hours of birth and a gross deviation in the shape of the heart due to reduced epicardial EMT, decreased angiogenesis, impaired collagen biosynthesis and the elevated apoptosis of cardiomyocytes.

HIC1 and HIC2 are zinc finger and BTB domain-containing TFs. A BTB/POZ domain is identified at the N-terminus of the factor, which mediates transcriptional repressive function and the GLDLSKK/R motif interacts with CTBP1 co-repressive factor (Dykes et al., 2014). In the mouse model, *Hic2* cardiac expression can be detected as early as E8.0 in the LHT, until E15.5 where expression can be observed throughout the myocardium. *Hic1* can be detected from E9.5 co-expressed with *Hic2* as well as the transcriptional co-repressive factor, Ctbp1, throughout the myocardium. HIC2 was found to negatively regulate cardiac troponin T3 (*Tnnt3*) which is the fast-skeletal isoform of cardiac troponin T (*cTnT*), as well as the positive regulation of the inhibitory SMAD6 protein, which decreases *Bmp10* production (Dykes et al., 2014). *Hic2* conditional KO (n = 8) with an *nkx2.5* Cre driver does not result in embryonic lethality; however, the phenotype results in cardiac malformation due to hypoplasia of the myocardium, apparent in 40% of studied embryos, as well as ventricular-septal defects (VSD) in 80% of the mouse embryos at E15.5. *Hic2* conditional KO with the *Mesp1* Cre driver leads to mutations within the anterior mesoderm where 26% lethality of *Hic2*-floxed embryos by E13.5 and 78% by P21.

STAT is a family of TFs which mediate many critical biological processes such as proliferation, differentiation and apoptosis. The STAT pathways have been shown to be essential to the myocardial response to various cardiac insults, such as myocardial infarction, pressure overload, hyperglycaemia and ischaemia-reperfusion (IR) injury (Dykes et al., 2014). STAT1 is known to play a pro-apoptotic role in IR injury via induction of effectors such as caspase-1.

4.1.3 *In silico* analysis of TFBS within an enhancer

To study whether the identified TFBS were indeed functional and essential for enhancer activity, point mutations were designed at the crucial bases for TF recognition within the TFBS sequence as determined by position weight matrixes. PWMs are a statistically derived representation of TFBS motifs of biological sequences, which are often depicted as graphical sequence logos. PWMs are constructed based on the frequency of occurrence of each nucleotide at each position within the sequence (Stormo, 2013). The probability of each nucleotide appearing at each specific location would then be normalized by dividing nucleotide count at each position by the entire number of sequences considered. PWMs assumes statistical independence between individual nucleotide positions which appear in the sequence pattern as the probability of each nucleotide appearing at any specific position is calculated independently of other positions.

4.2 Results

4.2.1 Characterizing spatio-temporal enhancer-reporter activity and generation of the enhancer 14 stable transgenic line

ATAC-seq was previously performed in the lab on bulk *tcf21+* epicardial cells at 5dpf, to determine accessible chromatin peaks for the identification of putative enhancers. Three putative enhancers (CE1, CE2 and enhancer 14) were identified from the ATAC-seq

results from 5dpf *tcf21+* epicardial cells during development (figure 4.1, track 1); all showed significant peaks in the embryonic 5dpf ATAC-seq track. In addition, putative enhancer CE2 and enhancer 14 both showed chromatin accessibility peaks corresponding to the same locus in the *myl7* dataset at 5dpf, indicating putative enhancer activity in myocardial cells in addition to the epicardial cells. To test the putative enhancers for activity, their respective sequences were isolated and cloned upstream of a minimal promoter E1b driving GFP. The plasmid was injected into one-cell stage embryos and imaged for fluorescence, indicating enhancer activity. For CE1 and CE2, clear ATAC-seq peaks were found in the 5dpf *tcf21* track and both *tcf21* and *myl7* tracks respectively. However, monitoring of both CE1 and CE2-E1b-mCherry zebrafish embryos did not reveal any cardiac-specific fluorescence throughout development (figure 4.2). These results suggest that although the chromatin loci may be open, they did not indicate bona fide enhancer activity during development.

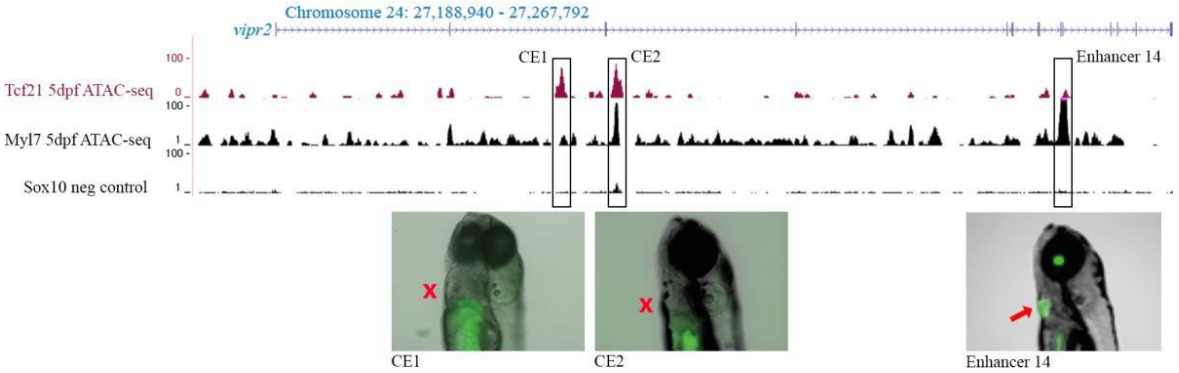


Figure 4.1: ATAC-seq accessibility map of the *vipr2* gene locus of the epicardial, myocardial and neural crest populations at 5dpf zebrafish embryos

The figure shows a UCSC genome browser view of ATAC-seq results for biotagged *tcf21+* epicardial cells (track 1), biotagged *myl7+* cardiomyocytes (track 2) and biotagged *sox10+* neural crest cells as a non-cardiac marker (track 3) at 5dpf. Accessible chromatin as seen as peaks from the *vipr2* non-coding regions and identified to be putative enhancers (marked by black boxes). CE1 putative enhancer shows a peak in *tcf21+* epicardial (track 1) cells, but no peak in *myl7* or *sox10* populations, however CE1 putative enhancer did not show heart-specific fluorescence when microinjected into zebrafish embryos in transient (shown in the

representative embryo at 5dpf). CE2 putative enhancer shows a peak for both *tcf21+* epicardial (track 1) and *myl7+* cardiomyocytic (track2) cells, as well as a small peak in the *sox10+* (track 3) populations, however CE2 putative enhancer, did not show heart-specific fluorescence when microinjected into zebrafish embryos in transient (shown in representative embryo at 5dpf). Enhancer 14 putative enhancer shows a small peak in the *tcf21+* (track 1) epicardial cells, and a big peak in the *myl7+* (track 2) cardiomyocytic cells and no peak in the *sox10+* neural crest cells. Positive enhancer 14 is labelled with a red arrow pointing to the heart-specific fluorescence in the 5dpf embryo. CE1 and CE2 regions which did not show heart-specific fluorescence is marked with a red cross.

The lab has previously ascertained the enhancer activity of enhancer 14, a sequence present within the intron of the *vipr2* gene (figure 4.1), through the use of transgenic report assay and the Tg(*enhancer14:GFP*) enhancer-reporter stable line has been generated by the lab previously. To characterise stable enhancer 14 activity, I monitored the in-crossed Tg(*enhancer14:GFP*) zebrafish embryos at various stages throughout development (figure 4.3a) to observe for spatiotemporal enhancer activity as indicated by GFP fluorescence. Spinning disc confocal microscopy revealed fluorescence throughout the embryo during early stages of embryogenesis, which gradually became restricted to the heart from 3dpf onwards to 5dpf (figure 4.3a).

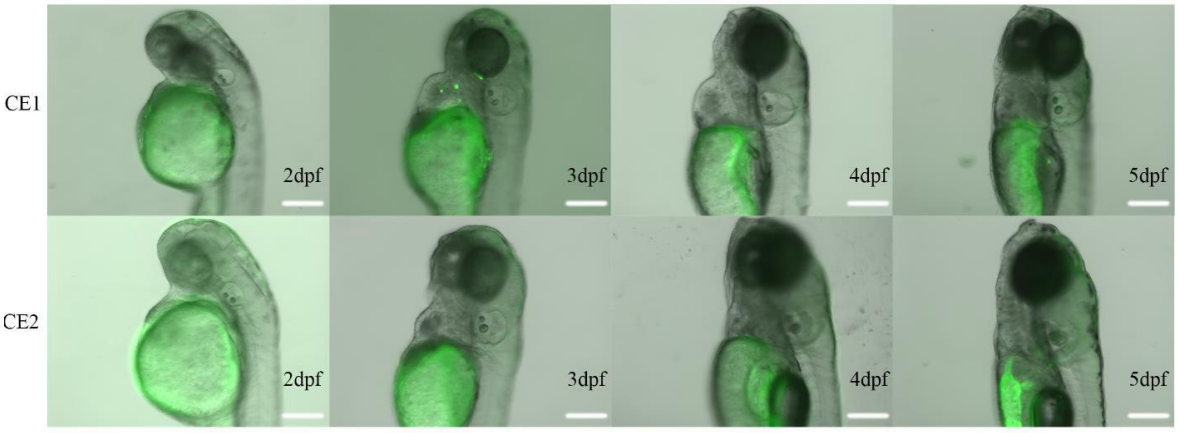


Figure 4.2 CE1 and CE2 negative data throughout embryonic development

The figure shows representative embryos of a transient experiment where CE1 and CE2 enhancer-reporter vectors were injected into one-cell stage embryos. CE1 (top row) and CE2 (bottom row),

throughout zebrafish embryonic development did not show any heart-specific fluorescence at any stage. CE1 zebrafish 3dpf had GFP expressing cells in the skin, but not in the heart. Scale-bar=150µm.

4.2.2 Immunofluorescence investigation of enhancer 14 expression

To investigate enhancer 14 activity in epicardial and myocardial cells, I crossed the Tg(*enhancer14:GFP*) stable line with in-house Tg(*tcf21:dsRed*) and Tg(*myl7:mCherry*) lines, which report epicardial and cardiomyocyte populations respectively. IF was performed on the double-positive embryos in the respective groups and imaged with the upright confocal microscope.

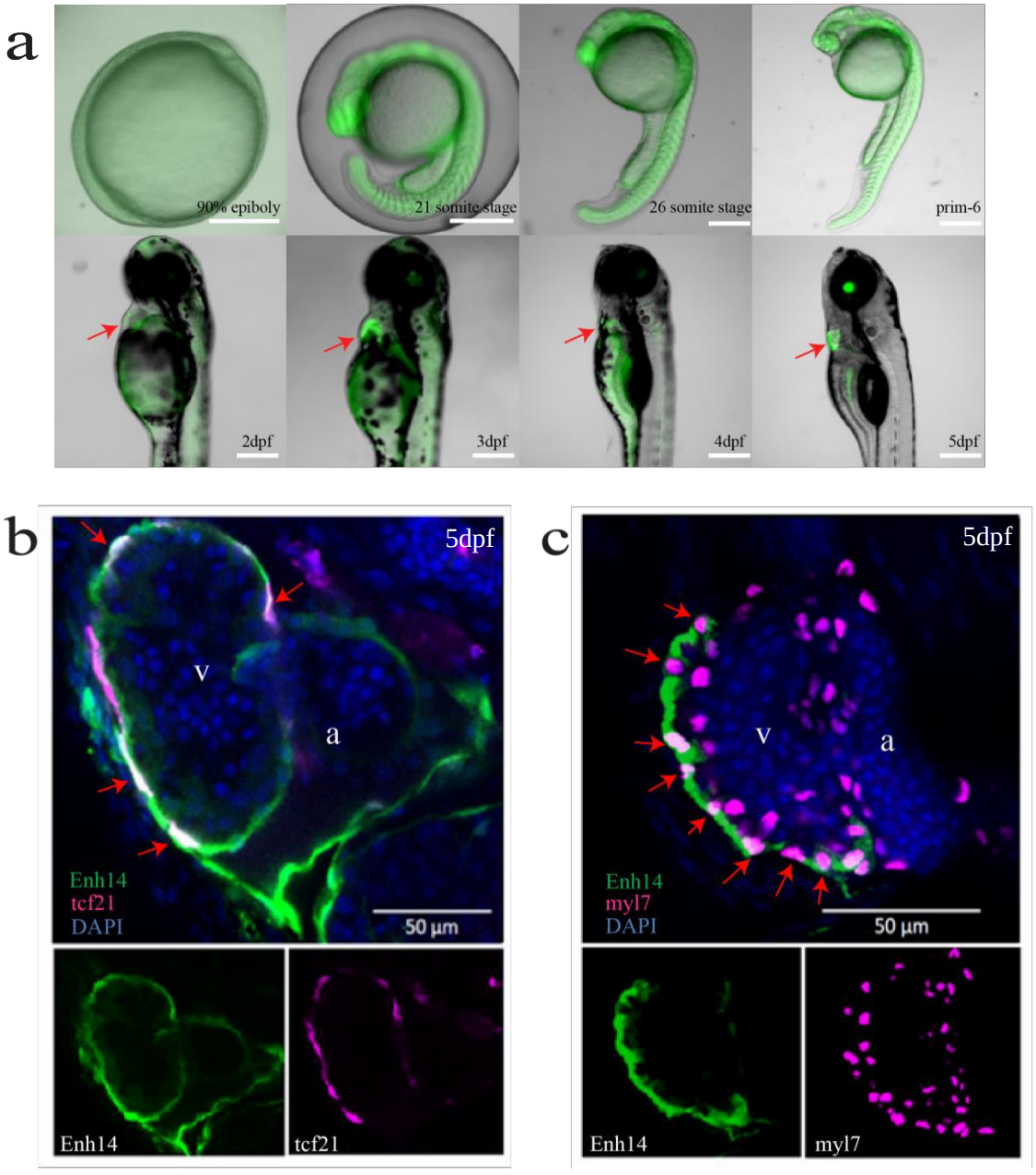


Figure 4.3: Characterising the spatiotemporal activity patterns of enhancer 14

(a) Tg(*Enhancer14:GFP*) enhancer-reporter stable line embryos throughout development from 90% epiboly stage to 5dpf. Heart specific fluorescence, showing enhancer 14 activity can be observed from 3dpf onwards to 5dpf. Red arrows show where the heart is located. Scale-bar=150µm **(b)** Tg(*Enhancer 14:GFP*; *tcf21:mCherry*) whole-mount embryonic zebrafish heart at 5dpf, stained for GFP and mCherry immunoreactivity with anti-GFP and anti-Cherry antibodies respectively, as well as DAPI (blue). GFP signal can be seen co-localising with *myl7+* myocardial cells (red arrows), showing that enhancer 14 directs GFP expression in epicardial cells during development. **(c)** Tg(*Enhancer 14:GFP*; *myl7:dsRed*) whole-mount embryonic zebrafish heart at 5dpf, stained for GFP and dsRed immunoreactivity with anti-GFP and anti-dsRed antibodies respectively, as well as DAPI (blue). GFP signal can be seen co-localising with *myl7+* myocardial cells (red arrows), showing that enhancer 14 directs GFP expression in cardiomyocytes during development. v=ventricle, a=atrium

GFP fluorescence was co-localised with dsRed fluorescence representing the *tcf21+* population of epicardial cells (figure 4.3b). Similarly, GFP fluorescence was also co-localised with the mCherry fluorescent signal of the *myl7+* cardiomyocytes (figure 4.3c). The IF results indicate that enhancer 14 is active in both epicardial and myocardial cell populations during embryogenesis at 5dpf.

4.2.3 Determining activity of enhancer 14 against *vipr2* expression domain

Next, multiplexed HCR was employed to monitor enhancer 14-reporter expression against the endogenous expression domains of *vipr2*. The Tg(*enhancer14:GFP*) line was in-crossed, and embryos were collected and fixed at 5dpf in accordance to the HCR protocol. Two experimental groups were carried out with DNA probes against *tcf21*, *vipr2* and *gfp* in the first experimental group (figure 4.4), and probes against *myl7*, *vipr2* and *gfp* in the other (figure 4.5), to analyse epicardial and myocardial cell expression, respectively.

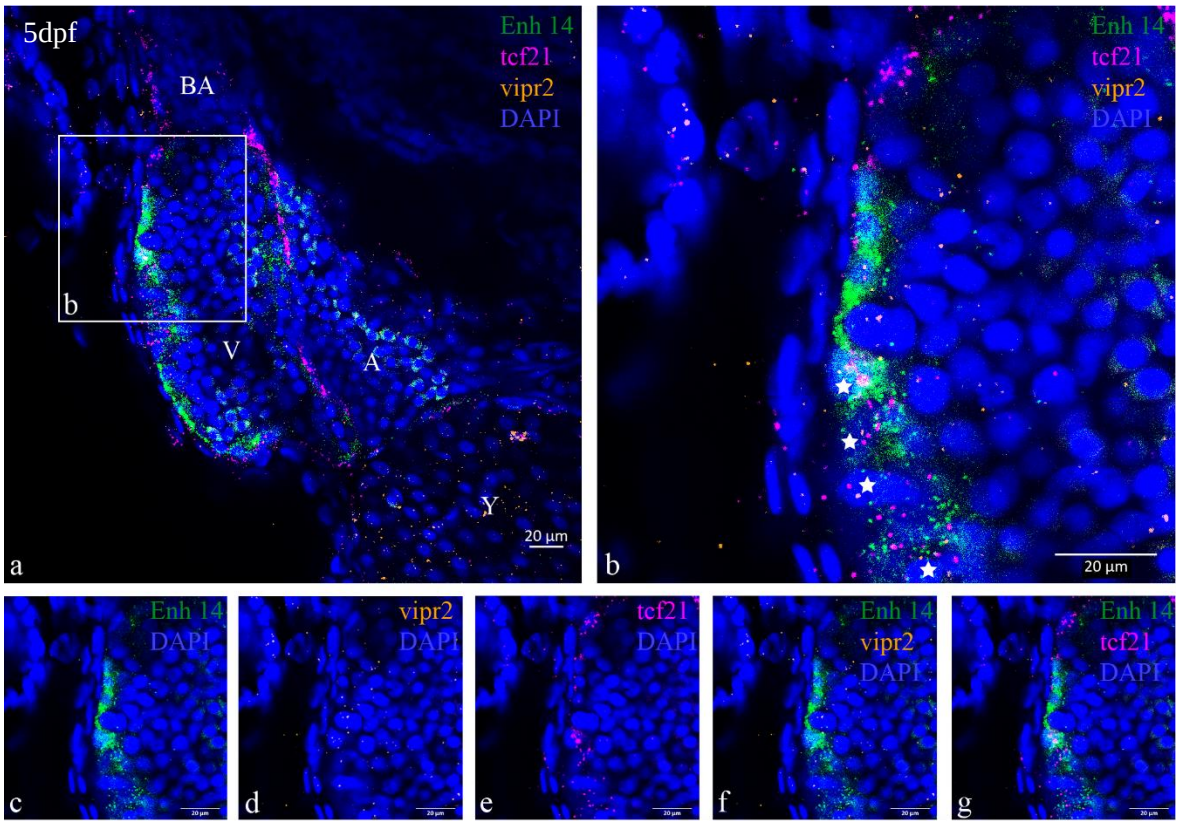


Figure 4.4: HCR investigation into enhancer 14 activity and *vipr2* expression overlap with *tcf21* epicardial marker

(a) Cross-section of the embryonic zebrafish heart, with multiplexed HCR to compare the activity pattern of enhancer 14 (green), expression of *vipr2* (orange) and *tcf21* (magenta). High-magnification inset labelled as b (panels c-g depicts single/dual channels from high-magnification inset b). **(b)** Observations were made of a small population of cells with overlapping *tcf21* (magenta), *vipr2* (orange) and enhancer 14 (green) expression (white asterisks). In all the cells with *vipr2* and *tcf21* co-expression, GFP expression indicating enhancer 14 activity may also be seen. **(c)** GFP (enhancer 14) is expressed in a single-celled layer of the heart adjacent to the epicardial *tcf21*+ lining. **(d)** *Vipr2* (orange) expression may be seen around a 3-4 cell-thick area. **(e)** Expression of *tcf21* (magenta) may be seen at the single, outermost layer of the heart. **(f)** Enhancer 14 (green) colocalizes with a subset of the total *vipr2* expressing cells. No cells were observed to express only GFP (enhancer 14) and not show *vipr2* co-expression. *Vipr2* mRNA-only expression without GFP can be observed in the cardiomyocytes closer to the ventricular chamber of the myocardium. **(g)** Enhancer 14 (green) can be seen co-localizing with a subset of *tcf21* (magenta) expressing cells. V=ventricle, A=atrium, BA=bulbus arteriosus, Y=yolk. Scale-bar=20µm.

In the high-resolution confocal images, the GFP expression representing enhancer 14

activity can be seen expressed in a subset of epicardial cells where GFP can be seen co-expressed with *tcf21* and *vipr2* (figure 4.4a, c, g). A large population of *tcf21+* epicardial cells do not express neither *vipr2* nor GFP (enhancer 14). However, there was no observation of an epicardial cell expressing GFP but not expressing *vipr2* mRNA. In the myocardial population, the majority of *myl7+* cells co-express *vipr2* mRNA; however, only a subset of *myl7+* myocytes co-express *vipr2* and GFP localised predominantly within a single-cell layer at the outer ventricular myocardium, whereas the cardiomyocytes present within the ventricle do not express GFP but express *vipr2* mRNA (figure 4.5a, e, g). This is speculated to be marking a specific subset of the myocardium, known as the primordial cardiomyocyte layer and would be discussed further below.

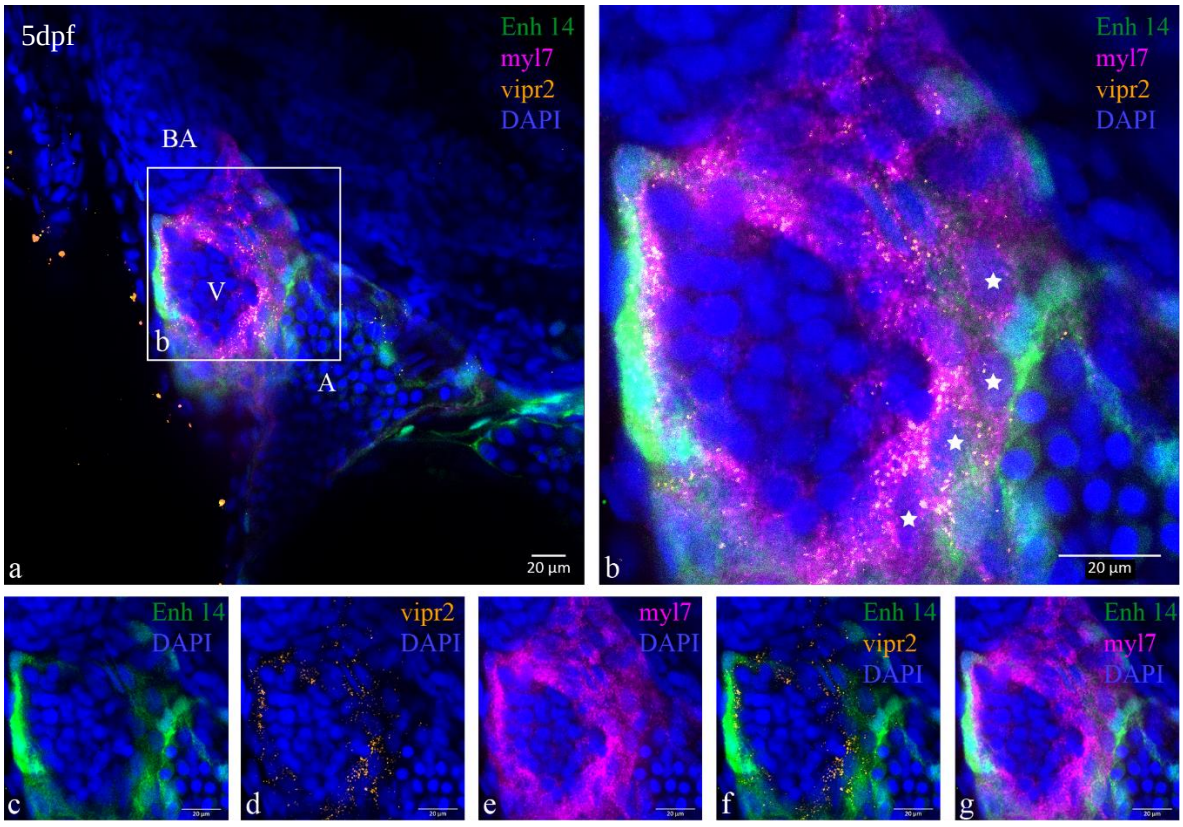


Figure 4.5: HCR investigation into enhancer 14 activity and *vipr2* expression overlap with *myl7* myocardial marker

(a) Multiplexed HCR to compare the activity pattern of enhancer 14 (green), expression of *vipr2* (orange) and *myl7* (magenta). High-magnification inset labelled as b (panels c-g depicts single/dual channels from high-

magnification inset b). **(b)** Observation can be made of cells which co-express all *tcf21*, *myl7* and GFP (white asterisks). HCR results show that while *vipr2* can be observed mostly co-expressed with *myl7*, enhancer 14 activity only partially overlaps both *myl7* and *vipr2*- expression domains. GFP (enhancer 14) is expressed in a single-cell layer of the heart overlapping with *myl7* expressing cardiomyocytes. *Vipr2* expression may be seen mostly co-expressing with *myl7* expression in the same cells. **(c)** GFP (enhancer 14) is expressed in a single-cell layer of the heart. **(d)** Expression of *vipr2* may be seen around a 3-6 cell thick area **(e)** Expression of *myl7* may be seen around a 3-6 cell thick area around the ventricle marking the myocardium. **(f)** Enhancer 14 (green) can be observed overlapping *vipr2* expressing cells in a single-layer of cells. No cells were observed to express only GFP (enhancer 14) and not show *vipr2* co-expression. **(g)** Enhancer 14 (green) can be seen co-localized with *myl7* expression at the single and outermost layer of the myocardium. V=ventricle, A=atrium, BA= bulbus arteriosus, Y=yolk. Scale-bar=20μm.

4.2.4 PWM-based TFBS mutation strategy to elucidate important TF-enhancer interactions

To test for important TF-enhancer regulation, the enhancer sequence was analysed for TFBS using JASPAR, and a TFBS perturbation strategy based on PWM was used. PWM is a matrix which represents statistical representation of the variance in binding preferences of TFs for biological motifs at each specific position in the TFBS, through experimentally validated data from SELEX-seq experiments (DNA-protein binding analysis). TFBS motifs were then aligned with the expression of the relevant TFs within the existing *tcf21+* RNA-sequencing datasets to refine the list of candidate TF-regulation of the enhancer. TFBSs clustered around the core enhancer sequence showing a higher likelihood of cell-type-specific activity and containing multiple predicted binding sites within the enhancer locus and longer binding motif (>6bp) were criteria for the initial site selection. For enhancer 14, four TFBS and their corresponding TFs were identified as potential functional binding sites acting to regulate *vipr2* enhancer 14, namely *meis1*, *hltf*, *hic1/2* and *stat1/6* (figure 4.4).

Point mutations were designed for all four of the identified putatively important TFBS. All similar TFBS reoccurrences (e.g. all *meis1* sequence reoccurrences) within the enhancer

sequence would be perturbed by PWM-designed point mutations which are ultimately aimed at disrupting the recognition and binding of the TF in question within the enhancer sequence (figure 4.4). Subsequently, the TFBS-mutated sequences were inputted into JASPAR, to determine in-silico, whether the mutation would be sufficient to disrupt TF-recognition and to ensure that no new site was inadvertently created through the mutagenesis, as would be picked up by the JASPAR algorithm. For each of the mutated enhancer, conditions were designed, ordered as double- stranded DNA gene blocks with flanking arms and subsequently cloned into the e1b-GFP plasmid vector for all four of the identified putatively important TFBS.

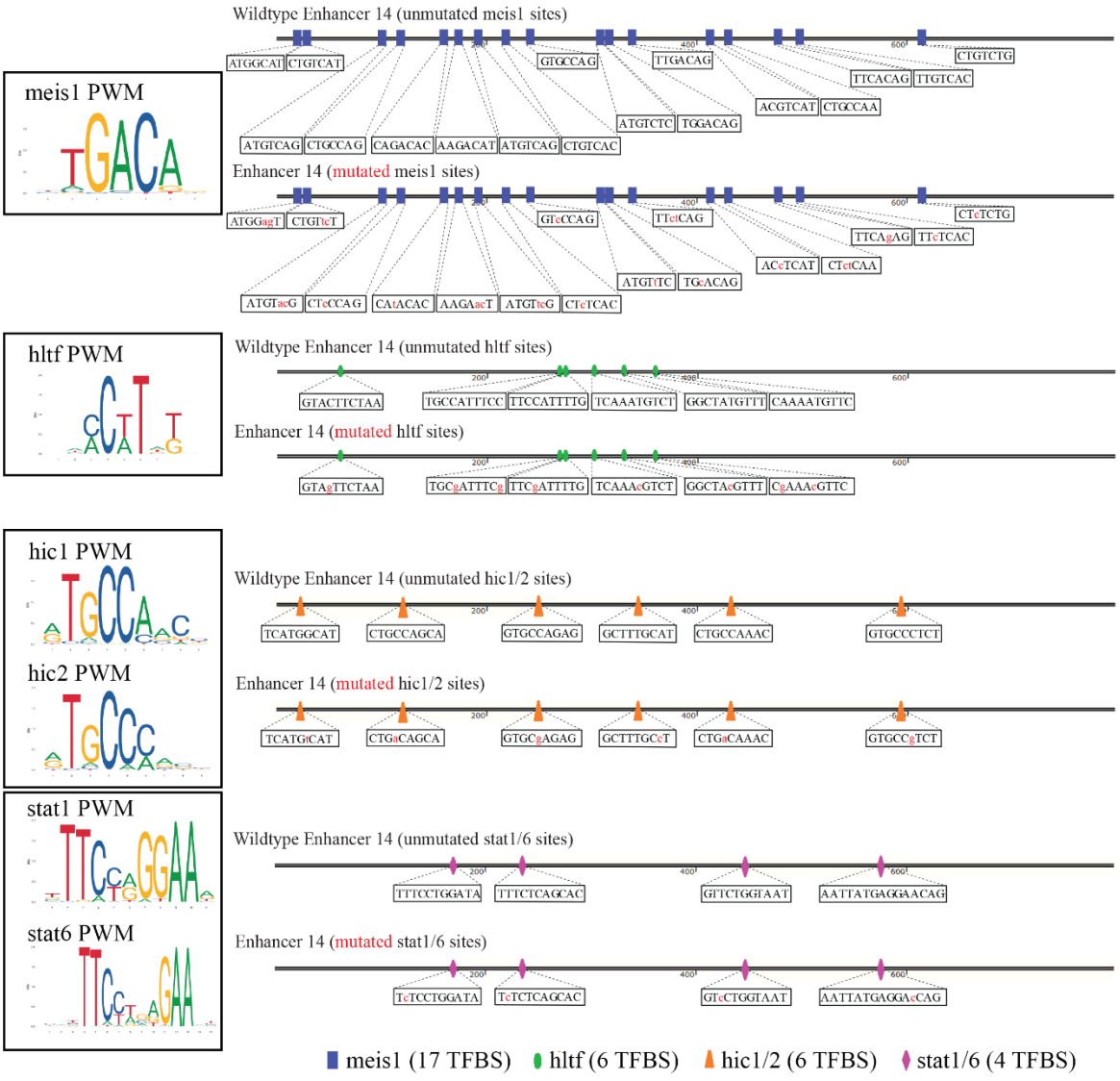


Figure 4.6: Schematic of the TFBS-mutation strategy to test for upstream TF-enhancer regulation

Schematic of the PWM for all four (*meis1/2, hltf, hic1/2, stat1/6*) putatively important TFBS for enhancer 14 activity (leftmost panel), followed by enhancer 14 depiction showing highlighted TFBS (refer to legend) as well as individual unmutated TFBS sequences as projected boxes. In sequence, *meis1* has 17 binding sites, *hltf* has 6 binding sites, *hic1/2* has 5 binding sites and *stat1/6* has 4 binding sites on enhancer 14 sequence as depicted. The paired corresponding enhancer depiction shows highlighted TFBS with individual mutated TFBS sequences as projected boxes, and the designed 1-2bp sequence substitution is indicated by lowercase lettering and red font.

4.2.5 Determination of TFBS-mutation on enhancer 14 activity

To determine whether the mutated TFBSs had any impact on enhancer activity, the respective TFBS-modified enhancer-reporter constructs (figure 4.6) were injected into one-cell stage zebrafish embryos, treated with 0.003% PTU and observed for heart-specific fluorescence temporally throughout early development until 5dpf and compared to the corresponding unmutated enhancer14 enhancer-reporter embryos as controls (figure 4.6, figure 4.7). The number of 5dpf embryos with heart-specific fluorescence was compared between the TFBS- perturbed group and control unmutated groups. Two-tailed unpaired student's t-test was used to analyse the data to determine whether the various TFBS-perturbation had resulted in a significant change in the absolute number of 5dpf embryos expressing heart-specific fluorescence.

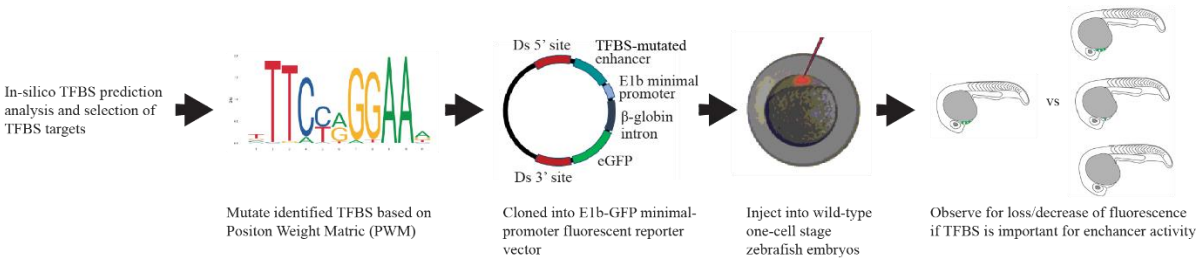


Figure 4.7: Schematic demonstrating workflow for TFBS-mutation strategy

Schematic showing the workflow to test for elucidating the important TFBS underlying enhancer activity.

TFBS identification was achieved by in-silico TFBS prediction software, then selected based on TF expression in RNA-seq dataset of the same time point, TFBS enrichment and location within enhancer locus, as well as the length of TFBS. Selected putatively important TFBS were mutated based on PWM to perturb specific TF recognition, followed by cloning of mutated enhancer sequences into E1b-GFP plasmid vector and injection into one-cell stage embryos to test for disruption of enhancer activity as compared to unmutated enhancer 14 enhancer-reporter vector to test for TF importance in enhancer activity as reported by loss/reduction of GFP expression.

In four rounds of micro-injections, the number of zebrafish with heart-specific fluorescence in wildtype (WT) unmutated enhancer 14-reporter compared to *meis1*-mutated enhancer-reporter injected embryos at 5dpf (figure 4.8), WT1 = 12/34 (34.9%) vs *meis*-mutated1 = 0/61 (0%), WT2 = 16/35 (45.7%) vs *meis*-mutated2 = 1/43 (2.3%), WT3 = 7/27 (25.9%) vs *meis*-mutated3 = 0/35 (0%), WT4 = 9/27 (33.3%) vs *meis*-mutated4 = 4/90 (4.4%) indicated that the *meis1* TFBS-perturbation had disrupted the enhancer's ability to drive GFP fluorescence, as reflected by the significant loss of heart-specific fluorescence from around 30-40% GFP+ embryos in the control wild-type enhancer-reporter condition to 0-4% GFP+ embryos in the mutant- enhancer-reporter condition (p = 0.0329) (figure 4.8); suggesting MEIS1 binding sites may be important for enhancer 14 functionality.

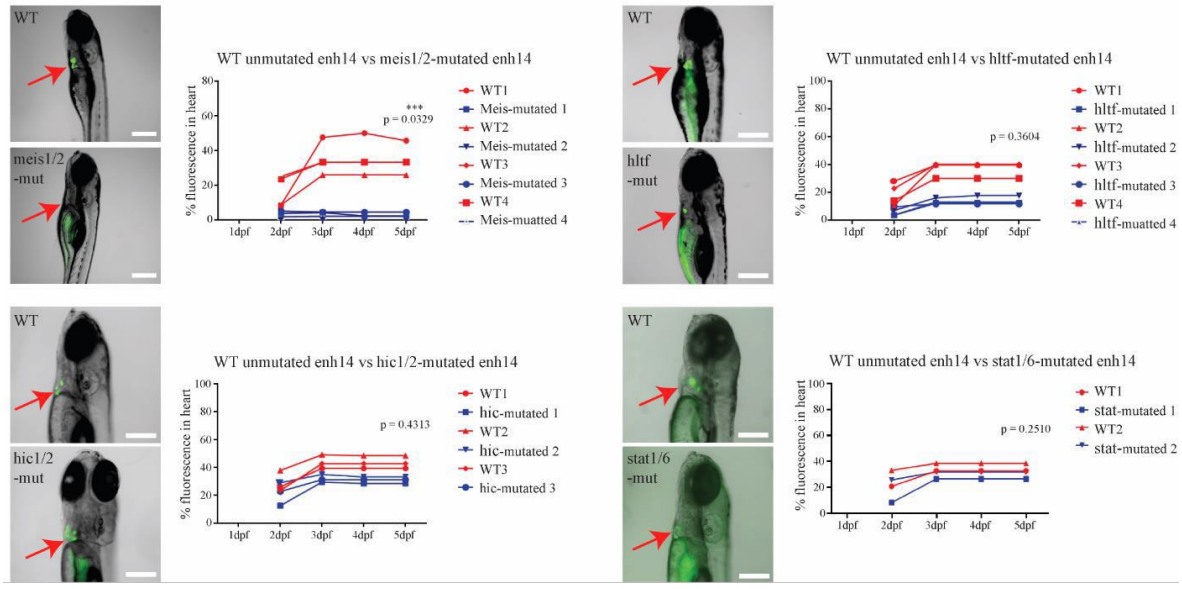


Figure 4.8: *Meis1* binding was found to be needed for enhancer 14 activity through the

TFBS-mutation strategy

Representative embryos depicting each of the injected mutated TFBS enhancer-reporter as compared to unmutated WT enhancer 14 enhancer-reporter injected embryos at 5dpf. The accompanying line graphs show temporal injection results for all injection rounds (e.g. WT experiment (exp) 1 versus *meis*-mutated exp 1 representing one injection round) for each of the mutated enhancer-reporters as compared to the unmutated enhancer-reporter vectors. *Meis1* TFBS-perturbation shows a significant decrease in numbers of embryos showing heart-specific fluorescence (as indicated by red arrows) at 5dpf as compared to the unmutated enhancer-reporter injected embryos. Significance was calculated with unpaired two-tailed t-test assuming equal variances was performed on the 5dpf percentage of embryos with heart fluorescence across the number of experiments in each condition. *The meis1*-mutation condition was found to be statistically significant with $p = 0.0329$. Scale- bar=150μm.

The other three TFBS-mutation experiments resulted in non-significant p-values, analysing *hltf*-mutated ($p = 0.3604$), *hic1/2*-mutated ($p = 0.4313$) and *stat1/6*-mutated ($p = 0.251$) (figure 4.8), enhancer-reporter expression versus wild-type enhancer controls, indicating that the TFBS-perturbations did not result in significantly reduced number of zebrafish embryos with heart-specific fluorescence at 5dpf. However, as the experiment only considers the absolute number of zebrafish embryos with heart-specific fluorescence but does not account for the possible difference in terms of the change in the intensity of heart-specific fluorescence, thus it would be premature to conclude that the conditions are not important for enhancer 14 activity.

4.2.6 High-resolution spinning disc imaging and Imaris analysis of *hltf*, *stat1/6* and *hic1/2* TFBS- mutation on enhancer 14 activity

The enhancer activity of the TFBS-perturbed enhancer 14 for the *hltf*, *hic1/2* and *stat1/6* conditions warranted further investigation to determine whether the mutation for the respective TFBS may cause any significant changes to the activity of the enhancer. Cloned E1b-GFP reporter plasmids under the control of the individual TFBS-perturbed enhancer

14 (figure 4.6) were injected into one-cell stage zebrafish embryos. In this experiment, the aim is to quantify the number of GFP fluorescent cells in the heart for all TFBS-mutated conditions, in the background of the Tg(*myl7:dsRed*) line, where dsRed fluorescence marks the cardiomyocytes.

Differences in *hic1/2*-mutated and *stat1/6*-mutated enhancer 14 activity as compared to that of the WT unmutated enhancer 14 were not significant (figure 4.9a, d), p = 0.55 and 0.073 respectively). This suggests that *hic1/2* and *stat1/6* TFs may not be required, or do not play a significant role in the activity of enhancer 14 during zebrafish cardiac development. However, it is interesting to note that while the *hic1/2*-mutated condition was not statistically significant, the condition had a mean fluorescent cell count of 43.7 (n = 30 embryos), which is slightly higher than that of the unmutated enhancer condition of 37.9 (n = 52 embryos), which may be attributed to the known repressive function of HIC1/2 (Dykes et al., 2014).In contrast, the *hltf*-mutated enhancer-reporter injected zebrafish embryos (figure 4.9a, d), revealed a significantly reduced number of GFP-expressing cells as compared to wild-type enhancer14-reporter injected embryos (p = 0.008), suggesting HLTF may be required for enhancer 14 activity during development.

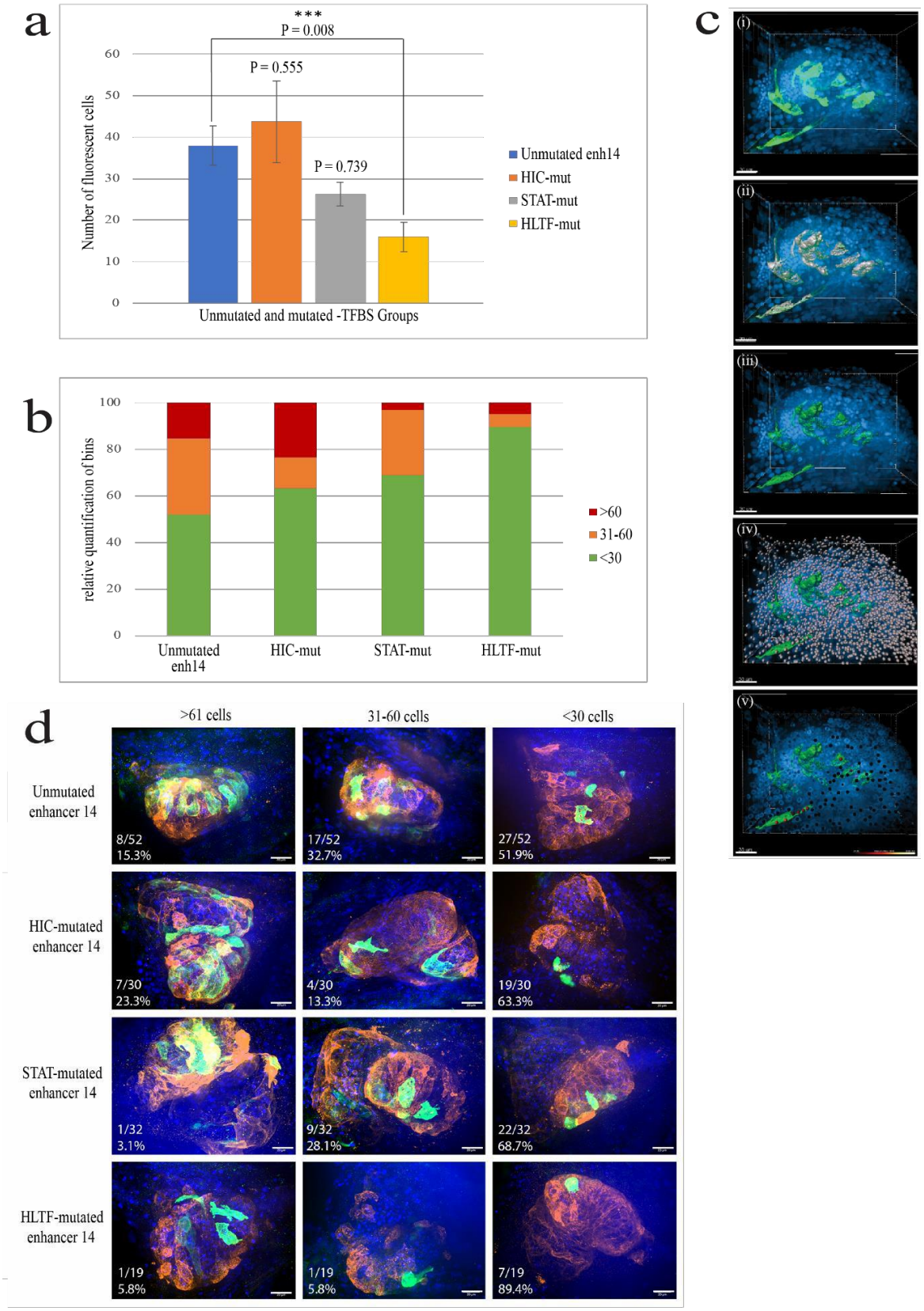


Figure 4.9: Graphs and representative images from spinning disc and Imaris image analysis for the different TFBS-mutated enhancer-reporter conditions

(a) Bar graphs showing data from spinning disc experiments, comparing the number of fluorescent cells in the

heart comparing unmutated enhancer 14 versus respective TFBS-modified enhancer-reporter vectors at 5dpf. P-value is calculated by comparing the number of fluorescent cells in each of the TFBS-mutated embryos with the number of fluorescent cells in the unmutated enhancer14 injected embryos using an unpaired two-tailed t-test, assuming equal variance. Error bar represents a 95% confidence interval. Then injections are performed in the background of Tg(*myl7:dsRed*) line with dsRed fluorescence marking the cardiomyocytic population. **(b)** Pictures showing Imaris analysis procedure. (i) The high-resolution z-stack images are projected to a 3D conformation (ii) GFP fluorescent (channel 2) areas are defined (iii) a mask is created for the defined GFP fluorescent areas (iv) the nucleus is defined by DAPI staining (channel 1) with perimeters defining the size and shape of the cell nucleus (xy diameter = 4µm, z diameter = 6µm) heat map based on nucleus proximity to defined GFP fluorescent areas. **(c)** Graphs showing relative quantification of the different numbers of embryos in each bin. The red area of the graphs shows percentage of cells with >60 fluorescent cells, orange area of the graphs show percentage of cells with 31-60 fluorescent cells and green area shows percentage of cells with <30 fluorescent cells. Scale-bar=30µm **(d)** Representative spinning disc images of the 5dpf embryonic zebrafish heart for each TFBS-mutated and wild-type unmutated condition used for Imaris imaging to calculate the number of fluorescent cells. Images are segregated by the individual conditions (rows) and their range of the numbers fluorescent cells detected (columns). The bottom left corner of each representative image shows the number of embryos with fluorescence in that respective range over total number of embryos in that condition, also presented as a percentage. Scale-bar=20µm.

4.3 Discussion

Enhancers are by far the most abundant class of cis-regulatory elements in the genome and may be identified by several different molecular techniques which exploit various characteristics associated with active enhancers.

In this study, ATAC-seq, a method for assaying for accessible chromatin was used as the initial step in identifying enhancer elements. The use of the ATAC-seq technique harbours several advantages as applied in this study. Firstly, due to the scarcity of *tcf21+* epicardial cells in the 5dpf zebrafish heart, methods such as DNase-seq which requires at least 1 to 10 million cells would be difficult/impossible to achieve as compared to ATAC-seq which only requires a minimal number of 500 to 50,000 cells for a comparable signal of

chromatin accessibility (Buenrostro et al., 2013). Furthermore, the ATAC-seq protocol preparation requires only 3 hours to complete, as compared to similar methods which may take days. After successful identification of accessible putative enhancers, enhancer activity has to be verified *in vivo* using transient transgenesis as chromatin accessibility does not equate to functionality.

Previously, the lab performed ATAC-seq on *tcf21+* epicardial cells at 5dpf, which enabled us to map the accessible chromatin region around *vipr2*. Two of three putative enhancer peaks, designated CE1 and CE2 when tested via transient transgenesis in enhancer-E1b-mCherry zebrafish embryos revealed no fluorescence throughout development. These results suggest that although the chromatin loci may be open in these regions, they do not regulate gene expression during zebrafish development. This could be due to a few reasons, firstly, the enhancers may be in a poised state and may not yet be active, due to the lack of appropriate TF-complement for activation. Secondly, it could also be likely that the putative enhancer regions may require their endogenous promoter sequences, as opposed to the use of E1b minimal promoter in this study, in order to exert their function in the spatiotemporal context. The other possibility is that the identified genetic locus may not be a regulatory region. ChIP-seq of acetylation/methylation states of histones would be useful to distinguish whether the open peak may be functional. The third accessible region identified, designated putative enhancer 14, was represented as a broad ATAC-seq peak in *myl7+* cardiomyocytes, and a smaller peak for *tcf21+* cells at 5dpf (figure 4.1), indicating that the putative enhancer, enhancer 14 may be active in cardiomyocytes and a sub-set of epicardial cells. This was corroborated by HCR staining (figure 4.4, figure 4.5) for enhancer 14 indicating expression in a small population of epicardial cells, as well as in the majority of cardiomyocytes. The small peaks observed in the *tcf21* ATAC-seq track (figure 4.1) may also be due to the fact that the accessibility peak from *vipr2*-expressing epicardial cells

may be masked by the major *tcf21+* population that is negative for *vipr2* (figure 4.1). Indeed, a recent study in our lab has explored the functional heterogeneity of the zebrafish epicardium during development using scRNA-seq and k-means clustering to detect functional subpopulations at 5dpf and identified three distinct epicardial subpopulations designated as 3 functionally distinct epicardial clusters, of which Epi1, Epi2, Epi3 were described. I leveraged on this existing scRNA-seq dataset and found that *vipr2*-expressing cells constitute a small population of the designated Epi3 cluster of the epicardium (figure 4.10) which correspond to a small population of epicardial cells which expresses *tcf21* exclusively with regards to other epicardial subpopulations. Interestingly, through the RNA-seq data, the Epi3 epicardial subpopulation was found to be enriched for leukocyte chemotaxis genes such as the chemokine CXCL12a which was found to function by recruiting ptprc/CD45+ haematopoietic cells during cardiac development. (Weinberger et al., 2018). The results indicate that *vipr2* may fulfil a specialised role in the small population of secretory epicardial cells.

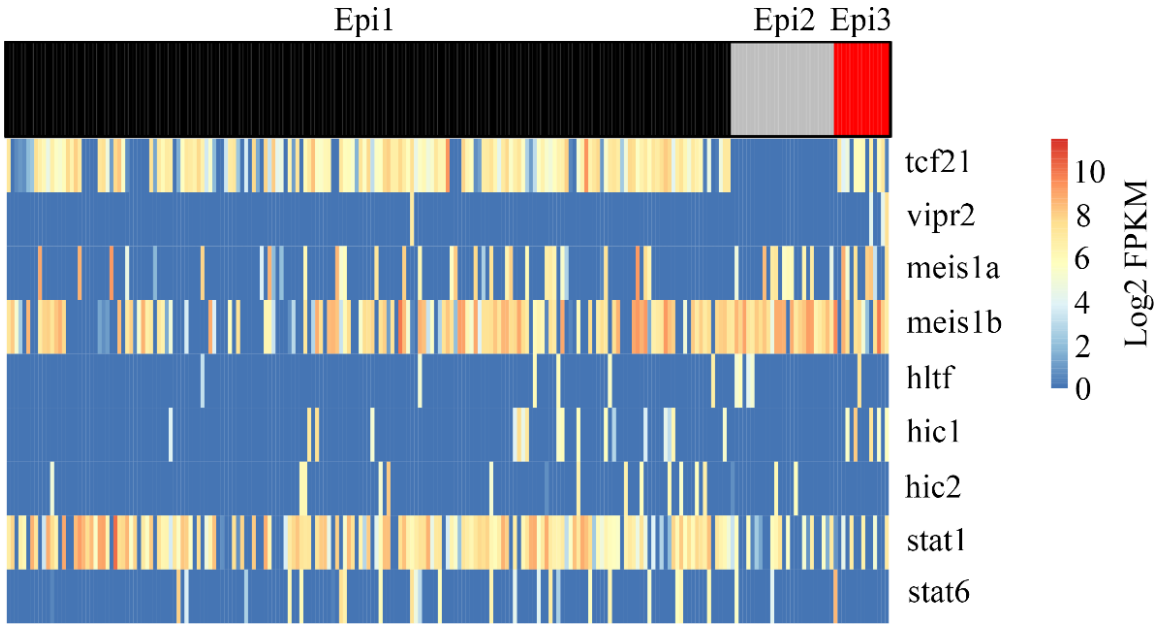


Figure 4.10: Expression of *vipr2* and the individual TF expression in scRNA-seq dataset

Marker gene expression of *vipr2*, *tcf21*, *meis1a/1b*, *hltf*, *hic1/2* and *stat1/6* (indicated directly on the right-side of heatmap) in the Epi1-3 clusters which were clustered in an unsupervised manner (columns). Colour-

coded cell cluster identity is shown at the top of the heatmap. Legend for the expression in the $\log_2$ FPKM scale is shown on the right-most side of the image. The image indicates that *vipr2* is expressed in a few epicardial cells in the Epi3 cluster, co-expressed with *tcf21*, *meis1a/b*, *hic1* and *stat1*. (scRNA-seq data from Michael Weinberger. Image partially adapted from (Weinberger et al., 2018))

Fluorescent microscopy images of the transgenic Tg(*enhancer 14:GFP*) line over time revealed generalised GFP-fluorescence throughout the developing embryo at 1-2dpf, indicating that the active enhancer might be ubiquitously active potentially indicating a requirement in multiple organ systems during early embryogenesis. From 3dpf, the GFP-fluorescence became restricted to the heart, suggesting a continued requirement for enhancer 14 activity in the heart, in both cardiomyocytes and a sub-population of epicardial cells.

HCR experiments on whole-mount 5dpf zebrafish embryos, multiplexing GFP (representing enhancer 14 activity), endogenous *vipr2*, *tcf21* and *myl7* expression confirmed the co- expression of enhancer 14 (GFP expression) and endogenously expressed *vipr2* in a subset of epicardial and myocardial cells. The results indicate that enhancer 14 does indeed partly recapitulate endogenous *vipr2* expression, further corroborated by observations from HCR staining images that all GFP expressing cells also co-express *vipr2*. As indicated by the ATAC- seq results, this co-expression pattern can be seen in a few *tcf21+* epicardial cells, while the majority of *tcf21+* cells within the epicardium do not appear to express either GFP or *vipr2* (figure 4.4a, b, g). The relative functional significance of this minor *vipr2+ tcf21+* population is unknown but would be interesting to follow-up in further studies. In the cardiomyocytic population, a large population of *myl7+* cardiomyocytes co-expressed both *vipr2* and GFP, while a subset of *myl7+ vipr2+* cardiomyocytes near to the inner trabecular/ endocardium region of the heart did not show GFP expression (figure 4.5a, b, g). This observation might be due to a few

possibilities. Firstly, there might be the existence of another unidentified *vipr2* enhancer which may be responsible for the spatiotemporal expression of *vipr2* in the inner trabecular subset of cardiomyocytes which is formed by 3dpf in the zebrafish heart (Gupta & Poss, 2012). Secondly, enhancer 14 may no longer be needed for *vipr2* expression for cardiomyocytes at the inner ventricle. Lastly, as an artefact of the HCR procedure due to insufficient permeabilisation (PKA step) or due to the size of the GFP probe, and related impaired tissue penetration. The third scenario is the least probable due to similar observations in multiple rounds of HCR, where PKA timing was adjusted from 20 minutes to 25 minutes. More putative enhancers need to be tested using transient transgenesis which together with enhancer 14, may fully recapitulate the entire expression pattern of *vipr2*. Capture-C, using an OCT probe designed against the *vipr2* promoter could be used to capture interacting sequences, which would reveal putative additional *vipr2* enhancers. It is interesting to note that GFP expression indicating enhancer 14 activity can be seen expressed by a single-layer of cells directly juxtaposed on the inner side of the *tcf21* expressing cells marking the epicardium (figure 4.3a). This single-layer of GFP expression can be seen co-localising with the outermost layer of *myl7* and *vipr2* co-expressing cells, while GFP expression was not seen in the cells of the inner ventricle, which do express *vipr2*. This may indeed indicate that enhancer 14 does not fully recapitulate the entire *vipr2* expression domain, revealing a possibility of other active *vipr2* enhancers. The single cardiomyocyte layer at 5dpf of which enhancer 14 was found to be active, may indeed allude to the primordial cardiomyocyte layer which constitutes the outermost single-cardiomyocyte layer of the myocardium during cardiac development (Gupta & Poss, 2012).

As enhancer 14 seems to be active in most *vipr2*-expressing cells in the heart, it would be interesting to find out which specific critical TF-enhancer interactions underlie its

activation. To assemble an in silico *vipr2* regulatory network via TF inputs upstream of the enhancer, an online database for TFBS recognition motifs was queried with the enhancer 14 sequence. Four TFs were identified to bind putatively and therefore might regulate enhancer 14 activity; hence their individual TFBS were mutated by 1-2bp and cloned into enhancer-reporter vectors to be tested using *in vivo* transient transgenesis in zebrafish embryos. Mutation of the binding site for *meis1* significantly reduced enhancer 14 activity, resulting in a reduction in cardiac expression in the majority of mutant enhancer-reporter embryos. In the remaining embryos showing heart-specific GFP expression, the area of GFP expression corresponding to the reduced number of GFP-expressing cells was grossly reduced as compared to the non-mutated enhancer driving GFP. These results indicate that the *meis1* TF may directly bind enhancer 14 to regulate some of the activity of *vipr2* during development.

Hltf binding site mutation on enhancer 14 was found to significantly reduce the number of GFP-expressing cells in the heart as demonstrated by TFBS-mutation (figure 4.9a), suggesting that *hltf* may be important for enhancer 14 activity during development. *Hltf* expression was detected to be expressed in bulk 5dpf *tcf21*+ RNA-seq results when the TFBS site was deemed to be putatively important. Intriguingly, while referencing to the scRNA-seq data of *tcf21* cells at the same timepoint, *hltf* was not found to be co-expressed in the *vipr2*- expressing epicardial cells at 5dpf, suggesting the RNA-seq signature of the small *vipr2*- expressing population may also have been masked by *hltf*-expressing *vipr2*-negative cells within the bulk RNA-seq dataset. However, as the mutation of the *hltf* TFBS resulted in a statistically significant reduction in enhancer activity, alternate explanations may include, *hltf* may be crucial for enhancer 14 activity during early development as a pioneer factor as it harbours chromatin remodelling functions. In mouse studies, it was found that most *hltf*-KO mice died between 12-24 hpf, showing the importance of the TF,

especially during early development (Helmer et al., 2013). Hence, it is possible that the expression of the *hltf* mRNA transcript may have already been downregulated at 5dpf in the zebrafish embryo. A second possibility is that, since *hltf* is not present in epicardial cells and cardiomyocytes at 5dpf, the reduction in enhancer 14 activity may be an artefact arising from the designed TFBS mutations. As a 1-2bp mutation was designed at all 6 putative *hltf* sites, one or more sites may have likely impacted the TF-enhancer interaction of another important TF which may have been overlooked by JASPAR analysis, resulting in the reduction of enhancer 14 activity. Lastly, mutating the *hltf* binding sites might have inadvertently resulted in non-specific binding of another TF, not detected by JASPAR analysis, which may have blocked the formation of other TF complexes important for enhancer 14 function. RNA-seq of *myl7+* cardiomyocytes would also reveal if *hltf* is present in the myocardial cells expressing *vipr2*, to determine the role of *hltf* in *vipr2* expression.

The *hic1/2* binding site mutated condition in enhancer 14, was not found to be statistically significant in terms of reducing enhancer 14 expression (figure 4.9a). However, through the results, *hic1/2* sites were indicated to be negatively modulating the activity of the active enhancer 14 under normal circumstances. The 6 designed point mutations caused an increase in enhancer activity which manifested as an increase in GFP-expressing cells in the mutant enhancer-reporter hearts. It can also be argued that the increased GFP-positive cells in the *hic1/2* mutated condition may be an artefact, as it was not statistically significant. Interestingly however, scRNA-seq data (figure 4.10) indicates that *hic1* may be co-expressed with *vipr2*. Hence, it could be interesting to study whether mutation of other motifs in the vicinity might result in a statistically significant increase in GFP-positive cells. It might also be interesting to analyse the enhancer 14 sequence further to potentially reveal other putative repressive TFBS. The presence of repressor motifs within enhancers

has been attributed to an evolutionary mechanism, whereby they are suggested to overcome enhancer robustness from the regulation of multiple partially-redundant enhancers, or to provide a mechanism to modulate enhancer activity in a spatiotemporal manner (Noon et al., 2016).

The lab has previously generated scRNA-seq data to assay for functional epicardial heterogeneity during development (Weinberger et al., 2018). The RNA-seq read count data was mapped in pathway and gene set overdispersion analysis (PAGODA) software, to test gene sets for coordinated expression pattern for the analysis of transcriptional heterogeneity. The developing epicardium at 5dpf was found to be functionally heterogeneous, containing three subpopulations of cells with different transcriptional profiles and associated Gene Ontology (GO) terms. The biggest epicardial cluster, Epi1, was found to mostly co-express *tbx18, tcf21* and *wt1b* epicardial markers, as well as being significantly enriched with GO terms associated with epithelial migration and cell adhesion. The Epi2 cluster was found to express the *tbx18* marker exclusively, and significantly enriched for genes associated with vasoconstriction and cell migration. The Epi3 cluster was found to be enriched for the epicardial *tcf21* marker, and significantly enriched for genes associated with leukocyte migration and axon extension. The *tcf21* and *vipr2* co-expressing cells were found to be classified under the Epi3 cluster. The Epi3 cluster was highly enriched for the expression of the chemokine *cxcl12a*, tight junction component *cldn11a* and *rxrα*, amongst other genes such as *ccl25b, bgnb, cntfr, nr1h5* and *nr2f2*. The Epi3 cluster also expresses *cxcl12a*, a paracrine chemo-attractant which was found to be essential for the recruitment of ptprc/CD45+ haematopoietic cells to the surface of the developing heart. The RXR*α*-mediated pathway in epicardial cells was found to affect the expression of downstream regulatory factors such as the FGF2 as well as canonical WNT signalling member, WNT9b. The epicardial-secreted paracrine factor FGF2 is well-known

as a crucial regulator of epicardial EMT, as well as a regulator of myocyte proliferation (Merki et al., 2005). The Epi3 cluster of epicardial cells harbours a secretory and immunomodulatory phenotype, which has been demonstrated functionally in development where the Epi3 population secretes the chemokine CXCL12a which attracts and retains ptprc/CD45+ leukocytes to the surface of the heart during development, as well as playing an important role in mediating the adult cardiac injury response (Balmer et al., 2014). The expression and potential role for the *vipr2* receptor in epicardial cells is novel and given that it appears localised to the *tcf21+* Epi3 subpopulation, I propose that VIPR2 may play a possible immunomodulatory role in epicardium consistent with the transcriptional profile of this cell cluster. Indeed, elsewhere VIPR2 been shown to possess immunomodulatory function in multiple cell types and to regulate the expression of multiple chemokines including those that are present within Epi3 (Goetzl et al., 2001).

Chapter 5: General discussion and future directions

5.1 Transcriptomic and genome-wide accessibility to elucidate cis-regulatory elements involved in epicardial gene expression

The epicardium is known to play an instrumental role in the facilitation of normal cardiac development by paracrine signalling and epicardial EMT, differentiation of supportive cardiovascular cell types and migration into the myocardium (Masters & Riley, 2014). In response to cardiac injury, the quiescent adult epicardium is the first cardiac cell-type to respond, by proliferating and enveloping the injured area while re-activating embryonic epicardial genes for cardiac regeneration (Cao & Poss, 2018). While the epicardial response is known to be important, we are indeed just beginning to characterise epicardial gene expression and gene regulation which precedes epicardial development and regeneration. The advent of high-resolution and high-throughput sequencing methods have indeed enabled unprecedented clarity into the gene regulation process and allowed the identification of chromatin interactions, as well as elucidating transcriptomes and chromatin accessibility at the single-cell level (Buenrostro et al., 2015; Patel et al., 2014).

The regulation of transcription efficiency exists mostly before transcription initiation, perpetuated by cis-regulatory enhancer elements which are known to drive gene transcription independent of orientation and distance to its target gene promoter. Enhancer elements contain an assortment of TFBS motifs which underlies its process-specific and spatiotemporal activity in facilitating cognate gene transcription. As discussed in chapter one, many genome-wide molecular techniques have been proposed to identify enhancer elements through evolutionary sequence conservation, chromatin accessibility and various histone modifications. Although each individual or combination of methods may have some success in enhancer identification, the robust discovery and identification of all

enhancers remain a significant challenge. In this study, I utilized the lab's previously generated differential expression analysis of RNA-seq data from biotagged *tcf21*-expressing epicardial population during development and regeneration as well as ATAC-seq data of genome-wide chromatin accessibility performed at 5dpf, to study the cis-regulation underlying the expression of *vipr2*, a GCPR for a pleiotropic neurotransmitter with novel implication within the developing epicardial cell population.

5.2 Characterising *vipr2* expression and activity patterns of putative enhancers

In multiple model organisms, VIP and PACAP, the two main neurotransmitter ligands exhibit overlapping pleiotropic functions such as during neural development, immunomodulatory effects, circadian rhythm control, as well as potent cardiovascular effects (Henning & Sawmiller, 2001). In the rat heart, VIPR2 is the main receptor present on both cardiomyocytes and non-myocytes to perpetuate CNS/PNS-mediated cardio-modulatory effects as well as vasodilative effects respectively. VIP/VIPR2 plays an important role in cardio-protection, as shown by the dysregulation of important cardiovascular genes resulting in pulmonary hypertension, aberrant calcium signalling, progressive biventricular cardiomyopathy and mortality (figure 3.2) (Szema et al., 2015).

Although the expression pattern of *vipr2* has been characterized in other model organisms, neither the molecular mechanisms of cis-regulation nor the expression pattern of *vipr2* have been described in the zebrafish model. I used the version 3.0 HCR with 'split-initiator' DNA probes to determine the expression pattern of *vipr2*, which requires both halves of the initiator DNA probe to bind to adjacent binding sites on the same mRNA transcript for the subsequent amplification of HCR signal through fluorescent hairpins (Choi et al., 2018). This system greatly improves the method's specificity and reduces nonspecific background signals. Initiator probes targeting the mRNA transcript of *vipr2* were multiplexed with DNA probes against *tcf21* and *myl7*, to mark the respective

epicardial and cardiomyocytic populations to determine *vipr2* expression in both these cell-types. Through IF and HCR staining, I found that v*ipr2* is co-expressed with *tcf21* in a few epicardial cells in the 5dpf zebrafish heart (figure 3.4, 3.5). This result is corroborated by previously generated scRNA-seq data from the lab, where *vipr2* was found to be expressed in a subset of an epicardial cell cluster Epi3 (figure 4.10). This is a specialized subpopulation of the heterogenous epicardial population which has been demonstrated to harbour a secretory and immunomodulatory phenotype. This suggests that *vipr2* may play an immunomodulatory role within this secretory subpopulation of epicardial cells during development.

Furthermore, *vipr2* was co-expressed with *myl7* in the majority of cardiomyocytes at 5dpf, suggesting a role in the forming myocardium (figure 3.5). In the rat heart, the small neurotransmitter *vip* signals through both parasympathetic and sympathetic nerves of the autonomic nervous system which innervates the heart through *vip*-reactive nerve fibres stemming from the vagus and sympathetic cardiac nerves to perpetuate chronotropic, inotropic and lusitropic effects (Henning & Sawmiller, 2001). Rat cardiomyocytes also express *vipr2*, which facilitates increased calcium extrusion and sequestration from the ryanodine and phospholamban channels of the sarcoplasmic reticulum, respectively. However, as the zebrafish conduction system has not been fully characterized, the role of VIP-Vipr2 signalling in the zebrafish remains to be determined, more experiments, including cell-specific loss of function, are needed to ascertain its functional role *in vivo*.

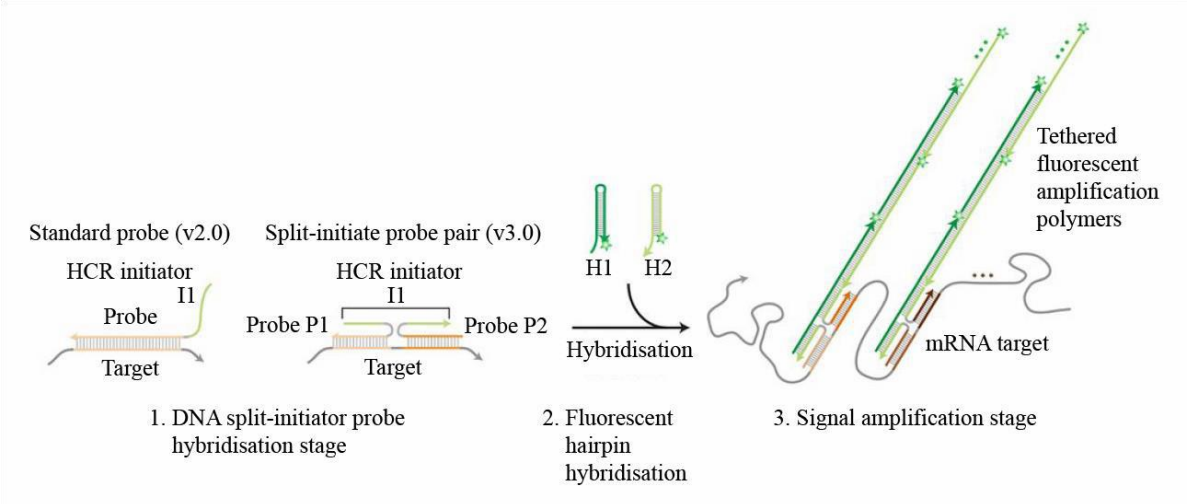


Figure 5.1: Improved HCR specificity with *in situ* HCR v3.0 with split-initiator probes

Schematic of HCR workflow. Stage 1 shows the difference between the standard probe (v2.0) which carries a full HCR initiator I1 and could lead to unspecific background fluorescence if it is bound non-specifically. Split-initiator DNA probes (v3.0) P1 and P2 each carry half of the HCR I1, and both probes have to co-bind to the target mRNA for the assembly of the full initiator I1; hence they would not lead to an amplified background if bound unspecifically, thus improving the method's specificity. Stage 2 shows the metastable fluorescent hairpins in their inactive state, through hybridisation with the I1 initiator element, the hairpins would become active. Stage 3 shows the signal amplification stage where the active fluorescent hairpins undergo and propagates hybridisation chain reaction, leading to the self-assembly of a tethered fluorescent amplification polymer to the mRNA target of interest. (Figure adapted from (Choi et al., 2018))

The lab previously generated ATAC-seq data from *tcf21*-positive epicardial cells at 5dpf, with *myl7*-positive cardiomyocytes, and *sox10* neural crest cells as a non-cardiac control mapped onto the UCSC genome browser (figure 4.1). Three small peaks representing accessible chromatin were identified in the 5dpf *tcf21* tracks and were tested for putative enhancer function. Recent advances in ATAC-seq has improved its resolution greatly, to enable its use at the single-cell level (Buenrostro et al., 2015). This modality would serve to increase resolution, especially for the analysis of rare subpopulations of cells such as the *vipr2* expressing epicardial cells. In the future, single-cell ATAC-seq could be performed on the sorted *vipr2+* population to check for any other accessible loci around *vipr2* which

may have been missed within the bulk *tcf21* ATAC-seq dataset.

The various identified putative enhancer regions were annotated to the nearest TSS due to the lack of data regarding chromatin organisation and interaction with the *vipr2* promoter. I have cloned and tested two putative enhancers of *vipr2*, CE1 and CE2, in transient transgenic zebrafish, using the previously described novel Ac/Ds reporter transgenesis system which ensures lower enhancer-reporter mosaicity making it suitable to be used in transient lines. However, the monitoring of the clutches of CE1 and CE2 enhancer-reporter injected embryos, revealed no heart-specific fluorescence throughout development.

Interestingly, with the CE2 sequence, the ATAC-seq track for both *myl7* and *sox10* revealed a small peak which may signify that CE2 may be ubiquitously accessible and may not represent a regulatory element (figure 4.1). ChIP-seq would be instrumental in determining histone states at both putative enhancer regions, which would inform on the presence of markers alluding to enhancer activity such as H3.3 or H3K4me1 and H3K27ac. The lack of reporter signal might also be attributed to abolished enhancer activity, due to the possible omission of important TFBS motifs at the edges of the enhancer while designing the cloning experiment. Selecting and cloning longer sequence may confirm whether the CE1 and CE2 regions may still constitute enhancer activity. Additionally, it might also be interesting to explore utilizing the *vipr2* endogenous promoter instead of the E1b-promoter when making the enhancer-reporter construct, as CE1 and CE2 putative enhancers may require their cognate promoter sequence in order to exert spatiotemporal regulation.

One of the putative enhancers, which was identified within the ninth intron of the *vipr2* gene, designated enhancer 14, was tested and shown to be active in transient lines. Therefore, I utilized the previously generated stable transgenic line, Tg(*enhancer 14:GFP*),

to monitor enhancer-driven GFP expression *in vivo*. I in-crossed the stable line and characterized the spatiotemporal activity pattern of enhancer 14 through fluorescent microscopy and showed that enhancer 14 is first expressed throughout the embryo from the 21-somite stage, indicating that the enhancer and by inference *vipr2* may be required in multiple cell types in the developing embryo. The enhancer activity became restricted to the heart from 3dpf, suggesting a heart-specific function at this key stage of zebrafish development (figure 4.3a).

To investigate which populations enhancer 14 is active in, I crossed the Tg(*enhancer 14: GFP*) stable line to the epicardial reporter line Tg(*tcf21:dsRed*) and myocardial reporter line Tg(*myl7: dsRed*). Immunostaining staining of the 5dpf embryos revealed that enhancer 14 is active in both epicardial and myocardial populations of the developing zebrafish heart, showing that enhancer 14 may indeed recapitulate the *vipr2* expression domain (figure 4.4, 4.5). Further investigation by HCR staining revealed that enhancer 14 activity fully recapitulated the expression profile of the *vipr2* gene the small population of *gfp+ tcf21+* epicardial cells (figure 4.4 a, b, g), as well as partially recapitulating the expression pattern of *vipr2* gene in a specific population of *myl7+* cardiomyocytes (figure 4.5a, b, g).

The *gfp* and *myl7* expressing cardiomyocytes were observed to be localized to a mono-layer of cardiomyocytes, situated in between the *tcf21+* epicardial cell population and the *myl7+ vipr2+* GFP- cardiomyocytes of the inner myocardium. I hypothesize that enhancer 14 may be active in the primordial cardiomyocyte layer, which is known to be maintained as a single- layer even into adulthood in zebrafish (Gupta & Poss, 2012), whereas the trabecular muscle layer which is directly juxtaposed on the inner side of the myocardium did not show enhancer 14 activity. The partial recapitulation of the entire *vipr2* expression pattern suggests there might be other yet unidentified developmental enhancers (s) acting on the *vipr2* promoter, which when combined with the activity profile of enhancer 14, fully

recapitulates the expression of *vipr2* within the total cardiomyocyte population. Capture-C could be used to identify further enhancers acting on *vipr2* promoter and identify all chromatin interactions.

5.3 Elucidating important TF-enhancer interactions underlying enhancer 14 activity

JASPAR, a database of TFBS recognition sites across many model organisms, was used to identify potentially informative sites which underlie enhancer 14 activity during zebrafish development. The selection criteria used was that the transcription factor would have to be expressed in the RNA-seq dataset, enrichment of homotypic motifs within the sequence, location of predicted TFBS within the sequence, favouring longer TFBS motifs, as well as previous publications which indicate TF importance in cardiac development. Selected TFBS would have to fulfil at least two or more of these criteria to be followed-up for further study. Through JASPAR analysis, the TFBS that were short-listed for further investigation included, *meis1*, *hltf*, *hic1/2*, *stat1/6*.

Another strategy that could be used to find putatively important TFBS is to utilize phylogenetic footprinting sequence alignment to find evolutionarily conserved motifs. Due to selective pressure, enhancers are more likely to be conserved than the non-functional sequences. This method has been used to determine important TFBS motifs and orthologous regulatory sequences across organisms (Neal et al., 2019). As discussed, motif degeneracy may pose certain challenges in sequence alignments; however, with the improvements in computational phylogenetic footprinting methods (Arnold et al., 2012; Lelieveld et al., 2016), this modality could also be another viable approach to potentially identifying the most important conserved TFBS motifs that underlie enhancer activity.

In order to elucidate whether the identified motifs had any direct implication on

modulating enhancer activity *in vivo*, a PWM-based enhancer mutation strategy was employed whereby all identified TFBS motifs for each TF family were mutated by 1-2 nucleotide substitutions for all four of the putatively important TFBS family motifs, so as to ensure the elimination of recognition whilst being careful to not create new binding sites inadvertently. Each of the mutated sequences containing TFBS-perturbed binding sites for all particular TF reoccurrence within the enhancer sequence was cloned into the Ac/Ds transgenesis vector as separate conditions, which was previously generated in the lab. Subsequently, I carried out the microinjection transgenesis experiments in one-cell stage zebrafish embryos, to compare the fluorescence driven by enhancer-reporters of the individual TFBS-perturbed conditions to the unmutated WT enhancer 14 control condition. In each clutch of microinjected embryos, there would be 50-90% of embryos which either did not show enhancer-driven fluorescence or non-specific fluorescence, especially in TFBS-perturbed conditions. This might be attributed to deleterious TFBS-perturbation effects, or various reasons such as the efficiency of transgenesis, the insertion site of the transgenic construct and intra-experimental user variability, suppression of vector expression due to local chromatin regulation as well as the inherent mosaicity of the transient system. It would be useful, therefore, to analyse the number and genomic integration sites from the transgenesis experiment, by using splinkerette PCR (spPCR) to analyse the exact genomic insertion site from the transgenesis experiment (Potter & Luo, 2010). This method utilizes restriction enzymes which digest genomic DNA producing sticky ends, and the addition of stable double-stranded oligonucleotide splinkerette hairpin loop containing designed compatible sticky ends for ligation to the restriction enzyme-digested genomic DNA fragment. Two rounds of nested PCR would be performed, which generates a product which contains genomic DNA between the transposable element and the restriction enzyme-digested genomic DNA site, which could be subsequently

sequenced to reveal the site of genomic integration. However, it would be counter-productive to perform this analysis for every clutch of microinjected embryos due to the time and costs involved, and the technique is only suggested to be used to map potentially complex transposon integrations (Potter & Luo, 2010).

The transgenic experiments could also be improved by the co-injection of another reporter which does not overlap with the activity domains of enhancer 14. For example, a transgenesis vector could be generated for a negative control such as the crystallin (*crybb1*) a constituent gene involved in lens development, driving expression of a spectrally distinct fluorophore. When co-injected together with the experimental vectors, the alternate reporter could serve as another internal control for the microinjections for the inherent mosaicity of the transient system. However, this method may be undermined by the known toxicity associated with microinjection, most notably when using transgenesis vectors. Hence, the DNA to be injected would need to be appropriately titrated to optimize the use of multiple reporters.

5.4 *Meis1* and *hltf* TFBS mutations abrogates enhancer 14 activity

Of the four TFBS-mutation experiments, *meis1* and *hltf* motif perturbation showed significant reduction in enhancer activity. *Meis1,* was the most enriched TFBS out of the four putatively important TFBS with 17 binding motifs on enhancer 14. The *meis1*-mutated binding sites, when introduced throughout the enhancer 14 sequence, showed the most significant disruption in enhancer-driven fluorescence, as compared to the unmutated enhancer 14 control. Only 1-5% of a clutch of *meis1*-mutated enhancer-reporter vector microinjected embryos exhibited any heart-specific enhancer-reporter driven fluorescence, as compared to 25-40% in the unmutated WT enhancer 14 control condition. Even in the minority of embryos which still expressed GFP in the heart, the levels of expression were significantly lower as compared to the unmutated condition (figure 4.8). This indicates that

the MEIS1 binding site is important and necessary in perpetuating enhancer 14 activity. In previous studies, morpholino KO of *meis1* during zebrafish development results in overt defects in blood vessel and heart development. The KO of *meis1* affects the *vegf* signalling pathway leading to defective cardiac looping during development. In the mouse model, *meis1* was also shown to be a regulator of the postnatal cardiomyocyte cell cycle (Muralidhar & Sadek, 2016), as well as an important regulator of sympathetic neuron development, which is also responsible for perpetuating VIP-VIPR2 cardiac innervation (Bouilloux et al., 2016).

Hltf had 6 putative binding sites on enhancer 14 with PWM-based mutation of 1-2 nucleotides per *hltf* binding motif. Unlike with *meis1*, mutation of HLTF binding sites resulted in a non-significant decrease, where 10-15% of a clutch of embryos exhibited heart-specific fluorescence as compared to 25-40% in the control unmutated condition (figure 4.6, $p = 0.3604$). However, I observed that the embryos of the *hltf*-mutated condition a reduced area of GFP fluorescence as compared to the unmutated control condition, which warranted further investigation. Hence, I adopted a high-resolution spinning disc imaging and image analysis approach to quantify fluorescent cell numbers in each of the conditions.

The use of spinning disc microscopy increased the acquisition rate of high-resolution z-stack images from embryos of all conditions, as compared to the conventional confocal imaging. Confocal laser scanning microscopy utilizes a system consisting of high-intensity excitation lasers and small apertures to scan samples using a point-by-point image acquisition method, which is designed to sequentially scan samples with a narrow excitation beam resulting in images with a high level of detail and low background signal at the expense of speed. In contrast, spinning disc microscopy utilizes a rotating disc with slit apertures which splits the excitation beam into a wide field of individual light beams which is able to traverse the entire image frame rapidly allowing accelerated image acquisition.

Thus, through spinning disc imaging and Imaris-based cell counting (figure 4.9c), I analysed all embryos for each mutated TFBS condition and performed two-tailed students t-test for the comparison of the fluorescent cell count in each of the TFBS-perturbed enhancer-reporter injected conditions versus the unmutated enhancer 14-reporter control.

Through the above-mentioned experimental pipeline, I discovered that the *hltf* TFBS-mutation resulted in a statistically significant decrease in heart-specific GFP-fluorescent cells as compared to the unmutated control (figure 4.9a). This indicates that HLTF may be necessary for the activity of enhancer 14. Interestingly, from bulk RNA-sequencing results, *hltf* was found to be expressed in *tcf21+* epicardial cells at 5dpf. However, from our scRNA-seq dataset, it was revealed that the *hltf* transcript was not expressed in the *vipr2*-expressing cells. This highlights an inherent drawback of bulk RNA-seq data; as distinct rare sub-populations of cells might exist within a large population of cells used to generate the bulk RNA-seq library, the variability in gene expression between such populations may not be detected. Hence, when studying rare subpopulations, bulk RNA-seq (and bulk ATAC-seq) must be used with caution. HLTF is known to function as an ATP-dependant chromatin remodelling protein and has been demonstrated to be expressed in the epicardium of the mouse at E8.5-9.5 (Helmer et al., 2013). I hypothesize that the enhancer 14 disruption from loss of *hltf*-binding sites may have been due to the downregulation of *hltf* transcripts by 5dpf, as expression profile of *hltf* in the embryonic mouse shows peak *hltf* expression at 12-24hpf. The transcript could be expressed earlier in zebrafish development which could be detected with RNA-seq of the *vipr2*-expressing cells at the zebrafish pim-25 to long-pec stage somite stage, which is equivalent to E8.5-9.5 in the mouse embryo. Furthermore, RNA-seq of cardiomyocytes may also inform on whether *hltf* is indeed expressed at 5dpf, which alludes to the TF having an effect on the cardiomyocytic population as opposed to the epicardial population.

Hic1/2 and *stat1/6* TFBS-mutated conditions did not result in a statistically significant change in GFP expression in the heart. HIC1/2 functions as a transcriptional repressor, cooperatively with CTBP1 during mouse development. *Hic2* can be detected as early as E8.0 in the mouse and becomes widely expressed throughout the myocardium by E15.5 (Dykes et al., 2014). It has been demonstrated to negatively regulate *Tnnt3,* the fast-skeletal muscle isoform of *cTnT,* as well as upregulating *Smad6,* which negatively regulates *Bmp10*. In this experiment, following *hic1/2*-binding site mutation there was trend towards an increase in the number of fluorescent cells in the heart as compared to the non-mutated enhancer 14 conditions, suggesting that *hic1/2* sites may function as repressive sites for enhancer 14 activity, although this was not found to be statistically significant.

In the future, tissue-specific *vipr2* KO zebrafish lines could be generated with *Cas-9* mediated deletion to observe for interesting phenotypes which would serve to elucidate its function during epicardial and myocardial development. If the *vipr2* KO line is viable, adult cryoinjury could be performed to study the *vipr2* KO phenotype associated with adult zebrafish cardiac regeneration, to determine if *vipr2* has a role to play as indicated by the adult 3dpi RNA-seq results. Single- cell RNA-seq could also be performed on the *vipr2*-expressing epicardial cell population, to determine their transcriptomic profiles as a guide towards VIPR2 function in the zebrafish heart. It also would be interesting to confirm, whether enhancer 14 is active in the intact zebrafish heart, as well as to characterize the enhancer 14 endogenous activity pattern in the adult zebrafish heart, to see if enhancer 14 is restricted to the primordial cardiomyocyte layer as I have hypothesized. Enhancer 14-KO line could also be generated to study the degree of effect of enhancer 14-KO for *vipr2* expression. More putative enhancers could be identified and tested using comparative genomics to look for evolutionary sequence conservation across multiple model organisms. This modality would serve to inform on previously unknown, important enhancer regions

which perpetuates conserved function during both development and cardiac injury; as well as possibly indicating enhancers which are responsible for perpetuating *vipr2* expression in the cardiomyocytes of the inner myocardium. Lastly, cardiac cryo-injury would also be performed for the Tg(*enhancer14:GFP*) transgenic fish as well as wild-type controls, in order to characterise domains of *vipr2* re-expression, as well as enhancer 14 reactivation regions.

5.5 Summary

In summary, this project focused on the novel characterisation of the developmental expression and upstream gene regulation of *vipr2*, a GPCR receptor, which was found to be expressed in a subset of epicardial cells and cardiomyocytes during heart development and differentially re- expressed as part of the epicardial response to cardiac injury. Through HCR detection of endogenously expressed RNA, I demonstrated the *vipr2* expression domain in the developing zebrafish heart and utilizing genome-wide accessibility and transcriptomic data, I have tested putative cis-regulatory regions which might be acting on the *vipr2* locus. I have also characterized the activity pattern of enhancer 14, a proven cis-regulatory element residing within the intronic regions of the *vipr2* gene, in a subset of epicardial and myocardial cells partially recapitulating endogenous *vipr2* expression. Notably, in the myocardium, enhancer 14 was found to be co-expressed with *vipr2* expression in a subset of cardiomyocytes which I hypothesized to be cardiomyocytes of the primordial muscle layer. *Vipr2* expression and enhancer 14 activity were also demonstrated in a rare subpopulation of epicardial cells which have been characterised by another study in the laboratory, based on scRNA-Seq profiling, as secretory and playing a role in modulating the immune cell migration during zebrafish heart development. Lastly, I carried out a TFBS-perturbation study on enhancer 14 to elucidate important TF-enhancer interactions underlying its activity and have shown that

MEIS1 and HLTF may function upstream of the cis-regulatory enhancer regulating *vipr2*. In conclusion, my study demonstrates that insights into the upstream regulation of potentially important genes functioning during heart development and in response to injury may be learnt from probing the cis-regulatory regions and testing candidate enhancers *in vivo*. This study would serve to build on the previously unknown gene regulation upstream of *vipr2,* as a novel pleiotropic GCPR which may play a role in the forming epicardium and myocardium during zebrafish heart development and regeneration.

References

Accili, E. A., Redaelli, G., & DiFrancesco, D. (1996). Activation of the hyperpolarization-activated current (if) in sino-atrial node myocytes of the rabbit by vasoactive intestinal peptide. *Pflugers Archiv European Journal of Physiology*.

AHA. (2016). Medications Used to Treat Heart Failure. Retrieved from American Heart Association website: http://www.heart.org/HEARTORG/Conditions/HeartFailure/TreatmentOptionsForHeartFailure/Medications-Used-to-Treat-Heart-Failure_UCM_306342_Article.jsp#.WC5CkC3hB1N

Ahmad, K., & Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Molecular Cell*, *9*(6), 1191–1200. https://doi.org/10.1016/S1097-2765(02)00542-7

Arnold, P., Erb, I., Pachkov, M., Molina, N., & Van Nimwegen, E. (2012). MotEvo: Integrated bayesian probabilistic methods for inferring regulatory sites and motifs on multiple alignments of DNA sequences. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btr695

Arrenberg, A. B., Stainier, D. Y. R., Baier, H., & Huisken, J. (2010). Optogenetic control of cardiac function. *Science*. https://doi.org/10.1126/science.1195929

Bakkers, J. (2011). Zebrafish as a model to study cardiac development and human cardiac disease. *Cardiovascular Research*. https://doi.org/10.1093/cvr/cvr098

Balmer, G. M., Bollini, S., Dubé, K. N., Martinez-Barbera, J. P., Williams, O., & Riley, P. R. (2014). Dynamic haematopoietic cell contribution to the developing and adult epicardium. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms5054

Barutcu, A. R., Maass, P. G., Lewandowski, J. P., Weiner, C. L., & Rinn, J. L. (2018). A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. *Nature Communications*. https://doi.org/10.1038/s41467-018-03614-0

Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., & Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*, 57–74. https://doi.org/10.1038/nature11247

Bersell, K., Arab, S., Haring, B., & Kühn, B. (2009). Neuregulin1/ErbB4 Signaling Induces Cardiomyocyte Proliferation and Repair of Heart Injury. *Cell*, *138*(2), 257–270. https://doi.org/10.1016/j.cell.2009.04.060

Blackwood, E. M. (1998). Going the Distance: A Current View of Enhancer Action. *Science*, *281*(5373), 60–63. https://doi.org/10.1126/science.281.5373.60

Blanchette, M., Bataille, A. R., Chen, X., Poitras, C., Laganière, J., Lefèbvre, C., … Robert, F. (2006). Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. *Genome Research*, *16*(5), 656–668. https://doi.org/10.1101/gr.4866006

Bouilloux, F., Thireau, J., Ventéo, S., Farah, C., Karam, S., Dauvilliers, Y., … Marmigère, F. (2016). Loss of the transcription factor Meis1 prevents sympathetic neurons target-field innervation and increases susceptibility to sudden cardiac death. *ELife*. https://doi.org/10.7554/eLife.11627

Bryne, J. C., Valen, E., Tang, M. H. E., Marstrand, T., Winther, O., Da piedade, I., … Sandelin, A. (2008). JASPAR, the open access database of transcription factor-binding profiles: New content and tools in the 2008 update. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkm955

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open

chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., … Greenleaf, W. J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*. https://doi.org/10.1038/nature14590

Burke, T. W., & Kadonaga, J. T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes and Development*. https://doi.org/10.1101/gad.10.6.711

Cao, J., & Poss, K. D. (2018). The epicardium as a hub for heart regeneration. *Nature Reviews Cardiology*. https://doi.org/10.1038/s41569-018-0046-4

Carmona, R., Guadix, J. A., Cano, E., Ruiz-Villalba, A., Portillo-Sánchez, V., Pérez-Pomares, J. M., & Muñoz-Chápuli, R. (2010). The embryonic epicardium: An essential element of cardiac development. *Journal of Cellular and Molecular Medicine*. https://doi.org/10.1111/j.1582-4934.2010.01088.x

Chen, T. H. P., Chang, T. C., Kang, J. O., Choudhary, B., Makita, T., Tran, C. M., … Sucov, H. M. (2002). Epicardial induction of fetal cardiomyocyte proliferation via a retinoic acid-inducible trophic factor. *Developmental Biology*. https://doi.org/10.1006/dbio.2002.0796

Cheng, J. H., Pan, D. Z. C., Tsai, Z. T. Y., & Tsai, H. K. (2015). Genome-wide analysis of enhancer RNA in gene regulation across 12 mouse tissues. *Scientific Reports*, *5*. https://doi.org/10.1038/srep12648

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., & Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkg500

Chimenti, I., Smith, R. R., Li, T. S., Gerstenblith, G., Messina, E., Giacomello, A., & Marbán, E. (2010). Relative roles of direct regeneration versus paracrine effects of human cardiosphere-derived cells transplanted into infarcted mice. *Circulation Research*, *106*(5), 971–980. https://doi.org/10.1161/CIRCRESAHA.109.210682

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., … Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Development (Cambridge)*. https://doi.org/10.1242/dev.165753

Chong-Morrison, V., Simoes, F. C., Senanayake, U., Carroll, D. S., Riley, P. R., & Sauka-Spengler, T. (2018). Re-purposing Ac/Ds transgenic system for CRISPR/dCas9 modulation of enhancers and non-coding RNAs in zebrafish. *BioRxiv*. https://doi.org/10.1101/450684

Chugh, A. R., Beache, G. M., Loughran, J. H., Mewton, N., Elmore, J. B., Kajstura, J., … Bolli, R. (2012). Administration of cardiac stem cells in patients with ischemic cardiomyopathy: The SCIPIO trial: Surgical aspects and interim analysis of myocardial function and viability by magnetic resonance. *Circulation*, *126*(11 SUPPL.1). https://doi.org/10.1161/CIRCULATIONAHA.112.092627

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., … Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1016071107

de Laat, W., & Dekker, J. (2012). 3C-based technologies to study the shape of the genome. *Methods*, *58*(3), 189–191. https://doi.org/10.1016/j.ymeth.2012.11.005

Dekker, J. (2006). The three "C" s of chromosome conformation capture: Controls, controls, controls. *Nature Methods*, *3*(1), 17–21. https://doi.org/10.1038/nmeth823

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome

conformation. *Science*. https://doi.org/10.1126/science.1067799
Delvecchio, M., Gaucher, J., Aguilar-Gurrieri, C., Ortega, E., & Panne, D. (2013). Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. *Nature Structural and Molecular Biology*, *20*(9), 1040–1046. https://doi.org/10.1038/nsmb.2642
Deng, W., & Roberts, S. G. E. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes and Development*. https://doi.org/10.1101/gad.342405
Dettman, R. W., Denetclaw, W., Ordahl, C. P., & Bristow, J. (1998). Common Epicardial Origin of Coronary Vascular Smooth Muscle, Perivascular Fibroblasts, and Intermyocardial Fibroblasts in the Avian Heart. *Developmental Biology*, *193*(2), 169–181. https://doi.org/10.1006/dbio.1997.8801
Ding, M., Liu, Y., Liao, X., Zhan, H., Liu, Y., & Huang, W. (2018). Enhancer RNAs (eRNAs): New insights into gene transcription and disease treatment. *Journal of Cancer*. https://doi.org/10.7150/jca.25829
Dohn, T. E., & Waxman, J. S. (2012). Distinct phases of Wnt/β-catenin signaling direct cardiomyocyte formation in zebrafish. *Developmental Biology*. https://doi.org/10.1016/j.ydbio.2011.10.032
Dupays, L., Shang, C., Wilson, R., Kotecha, S., Wood, S., Towers, N., & Mohun, T. (2015). Sequential Binding of MEIS1 and NKX2-5 on the Popdc2 Gene: A Mechanism for Spatiotemporal Regulation of Enhancers during Cardiogenesis. *Cell Reports*. https://doi.org/10.1016/j.celrep.2015.08.065
Dvoráková, M. C. hottov. (2005). Cardioprotective role of the VIP signaling system. *Timely Topics in Medicine. Cardiovascular Diseases*.
Dykes, I. M., Van Bueren, K. L., Ashmore, R. J., Floss, T., Wurst, W., Szumska, D., … Scambler, P. J. (2014). HIC2 is a novel dosage-dependent regulator of cardiac development located within the distal 22q11 deletion syndrome region. *Circulation Research*. https://doi.org/10.1161/CIRCRESAHA.115.303300
Emelyanov, A., & Parinov, S. (2008). Mifepristone-inducible LexPR system to drive and control gene expression in transgenic zebrafish. *Developmental Biology*. https://doi.org/10.1016/j.ydbio.2008.04.042
Evans, R., Fairley, J. A., & Roberts, S. G. E. (2001). Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes and Development*. https://doi.org/10.1101/gad.206901
Feliciano, L., & Henning, R. J. (1998). Vagal nerve stimulation during muscarinic and beta-adrenergic blockade causes significant coronary artery dilation. *Journal of the Autonomic Nervous System*. https://doi.org/10.1016/S0165-1838(97)00109-4
Felker, A., Prummel, K. D., Merks, A. M., Mickoleit, M., Brombacher, E. C., Huisken, J., … Mosimann, C. (2018). Continuous addition of progenitors forms the cardiac ventricle in zebrafish. *Nature Communications*. https://doi.org/10.1038/s41467-018-04402-6
Fouts, K., Fernandes, B., Mal, N., Liu, J., & Laurita, K. R. (2006). Electrophysiological consequence of skeletal myoblast transplantation in normal and infarcted canine myocardium. *Heart Rhythm*, *3*(4), 452–461. https://doi.org/10.1016/j.hrthm.2005.12.016
Fukui, H., Terai, K., Nakajima, H., Chiba, A., Fukuhara, S., & Mochizuki, N. (2014). S1P-Yap1 signaling regulates endoderm formation required for cardiac precursor cell migration in zebrafish. *Developmental Cell*. https://doi.org/10.1016/j.devcel.2014.08.014
Goetzl, E. J., Voice, J. K., Shen, S., Dorsam, G., Kong, Y., West, K. M., … Harmar, A. J.

(2001). Enhanced delayed-type hypersensitivity and diminished immediate-type hypersensitivity in mice lacking the inducible VPAC2 receptor for vasoactive intestinal peptide. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.241503798

Goldman, J. A., Kuzu, G., Lee, N., Karasik, J., Gemberling, M., Foglia, M. J., … Poss, K. D. (2017). Resolving Heart Regeneration by Replacement Histone Profiling. *Developmental Cell*. https://doi.org/10.1016/j.devcel.2017.01.013

González-Rosa, J. M., Martín, V., Peralta, M., Torres, M., & Mercader, N. (2011). Extensive scar formation and regression during heart regeneration after cryoinjury in zebrafish. *Development*. https://doi.org/10.1242/dev.060897

Gorkin, D. U., Lee, D., Reed, X., Fletez-Brant, C., Bessling, S. L., Loftus, S. K., … McCallion, A. S. (2012). Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. *Genome Research*, *22*(11), 2290–2301. https://doi.org/10.1101/gr.139360.112

Gregersen, H., Dall, F. H., Jørgensen, C. S., Jensen, S. L., & Ahrén, B. (1991). Effect of vasoactive intestinal peptide on duodenal motility in the isolated perfused porcine pancreaticoduodenal block. *Digestive Diseases (Basel, Switzerland)*. https://doi.org/10.1159/000171328

Grossman, S. R., Engreitz, J., Ray, J. P., Nguyen, T. H., Hacohen, N., & Lander, E. S. (2018). Positional specificity of different transcription factor classes within enhancers. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1804663115

Gupta, V., & Poss, K. D. (2012). Clonally dominant cardiomyocytes direct heart morphogenesis. *Nature*. https://doi.org/10.1038/nature11045

Hah, N., Murakami, S., Nagari, A., Danko, C. G., & Lee Kraus, W. (2013). Enhancer transcripts mark active estrogen receptor binding sites. *Genome Research*, *23*(8), 1210–1223. https://doi.org/10.1101/gr.152306.112

Hannibal, J., Hsiung, H. M., & Fahrenkrug, J. (2011). Temporal phasing of locomotor activity, heart rate rhythmicity, and core body temperature is disrupted in VIP receptor 2-deficient mice. *American Journal of Physiology - Regulatory Integrative and Comparative Physiology*. https://doi.org/10.1152/ajpregu.00599.2010

Harmar, A. J., Fahrenkrug, J., Gozes, I., Laburthe, M., May, V., Pisegna, J. R., … Said, S. I. (2012). Pharmacology and functions of receptors for vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide: IUPHAR Review 1. *British Journal of Pharmacology*. https://doi.org/10.1111/j.1476-5381.2012.01871.x

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., … Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, *39*(3), 311–318. https://doi.org/10.1038/ng1966

Helmer, R. A., Martínez-Zaguilán, R., Dertien, J. S., Fulford, C., Foreman, O., Peiris, V., & Chilton, B. S. (2013). Helicase-like transcription factor (Hltf) regulates G2/M transition, Wt1/Gata4/Hif-1a cardiac transcription networks, and collagen biogenesis. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0080461

Henning, R. J., & Sawmiller, D. R. (2001). Vasoactive intestinal peptide: Cardiovascular effects. *Cardiovascular Research*. https://doi.org/10.1016/S0008-6363(00)00229-7

Hesselberth, J. R., Chen, X., Zhang, Z., Sabo, P. J., Sandstrom, R., Reynolds, A. P., … Stamatoyannopoulos, J. A. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nature Methods*, *6*(4), 283–289. https://doi.org/10.1038/nmeth.1313

Houser, S. R., & Margulies, K. B. (2003). Is depressed myocyte contractility centrally

involved in heart failure? *Circulation Research*, Vol. 92, pp. 350–358. https://doi.org/10.1161/01.RES.0000060027.40275.A6

Huang, G. N., Thatcher, J. E., McAnally, J., Kong, Y., Qi, X., Tan, W., … Olson, E. N. (2012). C/EBP transcription factors mediate epicardial activation during heart development and injury. *Science*. https://doi.org/10.1126/science.1229765

Hughes, J. R., Roberts, N., Mcgowan, S., Hay, D., Giannoulatou, E., Lynch, M., … Higgs, D. R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics*. https://doi.org/10.1038/ng.2871

Ishibashi, M., Mechaly, A. S., Becker, T. S., & Rinkwitz, S. (2013). Using zebrafish transgenesis to test human genomic sequences for specific enhancer activity. *Methods*. https://doi.org/10.1016/j.ymeth.2013.03.018

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., & Smale, S. T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.14.1.116

Jopling, C., Sleep, E., Raya, M., Martí, M., Raya, A., & Belmonte, J. C. I. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature*, *464*(7288), 606–609. https://doi.org/10.1038/nature08899

Kang, J., Hu, J., Karra, R., Dickson, A. L., Tornini, V. A., Nachtrab, G., … Poss, K. D. (2016a). Modulation of tissue repair by regeneration enhancer elements. *Nature*, *532*(7598), 201–206. https://doi.org/10.1038/nature17644

Kang, J., Hu, J., Karra, R., Dickson, A. L., Tornini, V. A., Nachtrab, G., … Poss, K. D. (2016b). Modulation of tissue repair by regeneration enhancer elements. *Nature*, *532*(7598), 201–206. https://doi.org/10.1038/nature17644

Kenyon, A., Gavriouchkina, D., Zorman, J., Chong-Morrison, V., Napolitani, G., Cerundolo, V., & Sauka-Spengler, T. (2018). Generation of a double binary transgenic zebrafish model to study myeloid gene regulation in response to oncogene activation in melanocytes. *DMM Disease Models and Mechanisms*. https://doi.org/10.1242/dmm.030056

Kikuchi, K. (2014). Advances in understanding the mechanism of zebrafish heart regeneration. *Stem Cell Research*. https://doi.org/10.1016/j.scr.2014.07.003

Kikuchi, K., Holdway, J. E., Werdich, A. A., Anderson, R. M., Fang, Y., Egnaczyk, G. F., … Poss, K. D. (2010). Primary contribution to zebrafish heart regeneration by gata4+ cardiomyocytes. *Nature*. https://doi.org/10.1038/nature08804

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., & Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Developmental Dynamics*. https://doi.org/10.1002/aja.1002030302

Kulaeva, O. I., Nizovtseva, E. V., Polikanov, Y. S., Ulianov, S. V., & Studitsky, V. M. (2012). Distant Activation of Transcription: Mechanisms of Enhancer Action. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.01127-12

Kutach, A. K., & Kadonaga, J. T. (2000). The Downstream Promoter Element DPE Appears To Be as Widely Used as the TATA Box in Drosophila Core Promoters. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.20.13.4754-4764.2000

Kwasnieski, J. C., Fiore, C., Chaudhari, H. G., & Cohen, B. A. (2014). High-throughput functional testing of ENCODE segmentation predictions. *Genome Research*, *24*(10), 1595–1602. https://doi.org/10.1101/gr.173518.114

Laflamme, M. A., & Murry, C. E. (2011). Heart regeneration. *Nature*, Vol. 473, pp. 326–335. https://doi.org/10.1038/nature10147

Lam, M. T. Y., Li, W., Rosenfeld, M. G., & Glass, C. K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends in Biochemical Sciences*, Vol. 39, pp. 170–182. https://doi.org/10.1016/j.tibs.2014.02.007

Lefebvre, R. A., Smits, G. J. M., & Timmermans, J. -P. (1995). Study of NO and VIP as non-adrenergic non-cholinergic neurotransmitters in the pig gastric fundus. *British Journal of Pharmacology*. https://doi.org/10.1111/j.1476-5381.1995.tb16406.x

Lelieveld, S. H., Schütte, J., Dijkstra, M. J. J., Bawono, P., Kinston, S. J., Göttgens, B., … Bonzanni, N. (2016). ConBind: Motif-aware cross-species alignment for the identification of functional transcription factor binding sites. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv1518

Lenardo, M. J., & Baltimore, D. (1989). NF-κB: A pleiotropic mediator of inducible and tissue-specific gene control. *Cell*. https://doi.org/10.1016/0092-8674(89)90833-7

Lepilina, A., Coon, A. N., Kikuchi, K., Holdway, J. E., Roberts, R. W., Burns, C. G., & Poss, K. D. (2006). A Dynamic Epicardial Injury Response Supports Progenitor Cell Activity during Zebrafish Heart Regeneration. *Cell*, *127*(3), 607–619. https://doi.org/10.1016/j.cell.2006.08.052

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., … Rosenfeld, M. G. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, *498*(7455), 516–520. https://doi.org/10.1038/nature12210

Lou, X., Deshwar, A. R., Crump, J. G., & Scott, I. C. (2011). Smarcd3b and Gata5 promote a cardiac progenitor fate in the zebrafish embryo. *Development*. https://doi.org/10.1242/dev.064279

Malliaras, K., & Marbán, E. (2011). Cardiac cell therapy: Where weve been, where we are, and where we should be headed. *British Medical Bulletin*, *98*(1), 161–185. https://doi.org/10.1093/bmb/ldr018

Marín-Juez, R., Marass, M., Gauvrit, S., Rossi, A., Lai, S. L., Materna, S. C., … Stainier, D. Y. R. (2016). Fast revascularization of the injured area is essential to support zebrafish heart regeneration. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1605431113

Marques, S. R., Lee, Y., Poss, K. D., & Yelon, D. (2008). Reiterative roles for FGF signaling in the establishment of size and proportion of the zebrafish heart. *Developmental Biology*. https://doi.org/10.1016/j.ydbio.2008.06.033

Marques, S. R., & Yelon, D. (2009). Differential requirement for BMP signaling in atrial and ventricular lineages establishes cardiac chamber proportionality. *Developmental Biology*. https://doi.org/10.1016/j.ydbio.2009.02.010

Marshall, H., Studer, M., Pöpperl, H., Aparicio, S., Kuroiwa, A., Brenner, S., & Krumlauf, R. (1994). A conserved retinoic acid response element required for early expression of the homeobox gene Hoxb-1. *Nature*. https://doi.org/10.1038/370567a0

Masters, M., & Riley, P. R. (2014). The epicardium signals the way towards heart regeneration. *Stem Cell Research*. https://doi.org/10.1016/j.scr.2014.04.007

Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics*. https://doi.org/10.1146/annurev.genom.7.080505.115623

McCLINTOCK, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.36.6.344

Menasché, P., Alfieri, O., Janssens, S., McKenna, W., Reichenspurner, H., Trinquart, L., … Hagège, A. A. (2008). The myoblast autologous grafting in ischemic cardiomyopathy (MAGIC) trial: First randomized placebo-controlled study of myoblast transplantation. *Circulation*, *117*(9), 1189–1200. https://doi.org/10.1161/CIRCULATIONAHA.107.734103

Merki, E., Zamora, M., Raya, A., Kawakami, Y., Wang, J., Zhang, X., … Ruiz-Lozano, P. (2005). Epicardial retinoid X receptor α is required for myocardial growth and

coronary artery formation. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0504343102

Minehata, K. ichi, Kawahara, A., & Suzuki, T. (2008). meis1 regulates the development of endothelial cells in zebrafish. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2008.07.075

Miraldi, E. R., Pokrovskii, M., Watters, A., Castro, D. M., De Veaux, N., Hall, J. A., … Bonneau, R. (2019). Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells. *Genome Research*. https://doi.org/10.1101/gr.238253.118

Muralidhar, S. A., & Sadek, H. A. (2016). Meis1 regulates postnatal cardiomyocyte cell cycle arrest. In *Etiology and Morphogenesis of Congenital Heart Disease: From Gene Function and Cellular Interaction to Morphology*. https://doi.org/10.1007/978-4-431-54628-3_11

Murray, K. J. (1990). Cyclic AMP and mechanisms of vasolidation. *Pharmacology and Therapeutics*. https://doi.org/10.1016/0163-7258(90)90060-F

Nadal-Ginard, B., Kajstura, J., Leri, A., & Anversa, P. (2003). Myocyte death, growth, and regeneration in cardiac hypertrophy and failure. *Circulation Research*, Vol. 92, pp. 139–150. https://doi.org/10.1161/01.RES.0000053618.86362.DF

Neal, A., Nornes, S., Payne, S., Wallace, M. D., Fritzsche, M., Louphrasitthiphol, P., … De Val, S. (2019). Venous identity requires BMP signalling through ALK3. *Nature Communications*. https://doi.org/10.1038/s41467-019-08315-w

Nechaev, S., & Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2010.11.001

Nikolov, D. B., & Burley, S. K. (1997). RNA polymerase II transcription initiation: A structural view. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.94.1.15

Nikolov, Dimitar B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., … Burley, S. K. (1995). Crystal structure of a TFIIB–TBP–TATA-element ternary complex. *Nature*. https://doi.org/10.1038/377119a0

Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., … Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. https://doi.org/10.1016/j.cell.2017.05.004

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., & De Laat, W. (2003). The β-globin nuclear compartment in development and erythroid differentiation. *Nature Genetics*, Vol. 35, pp. 190–194. https://doi.org/10.1038/ng1244

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., … Bernstein, B. E. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. https://doi.org/10.1126/science.1254257

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: five essential questions. *Nature Reviews Genetics, 14*(4), 288–295. https://doi.org/10.1038/nrg3458

Peralta, M., González-Rosa, J. M., Marques, I. J., & Mercader, N. (2014). The epicardium in the embryonic and adult zebrafish. *Journal of Developmental Biology*. https://doi.org/10.3390/jdb2020101

Perry, M. W., Boettiger, A. N., & Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proceedings of the National Academy of Sciences, 108*(33), 13570–13575.

https://doi.org/10.1073/pnas.1109873108
Pfeffer, M. A., & Braunwald, E. (1990). Ventricular remodeling after myocardial infarction. Experimental observations and clinical implications. *Circulation*, *81*(4), 1161–1172. https://doi.org/10.1161/01.CIR.81.4.1161
Poon, K. L., Liebling, M., Kondrychyn, I., Brand, T., & Korzh, V. (2016). Development of the cardiac conduction system in zebrafish. *Gene Expression Patterns*. https://doi.org/10.1016/j.gep.2016.08.003
Porrello, E. R., Mahmoud, A. I., Simpson, E., Hill, J. A., Richardson, J. A., Olson, E. N., & Sadek, H. A. (2011). Transient regenerative potential of the neonatal mouse heart. *Science (New York, N.Y.)*, *331*(6020), 1078–1080. https://doi.org/10.1126/science.1200708
Poss, K. D., Wilson, L. G., & Keating, M. T. (2002). Heart regeneration in zebrafish. *Science*, *298*(5601), 2188–2190. https://doi.org/10.1126/science.1077857
Potter, C. J., & Luo, L. (2010). Splinkerette PCR for mapping transposable elements in Drosophila. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0010168
Preger-Ben Noon, E., Davis, F. P., & Stern, D. L. (2016). Evolved Repression Overcomes Enhancer Robustness. *Developmental Cell*. https://doi.org/10.1016/j.devcel.2016.10.010
Quach, H. N. B., Tao, S., Vrljicak, P., Joshi, A., Ruan, H., Sukumaran, R., … Sampath, K. (2015). A multifunctional mutagenesis system for analysis of gene function in zebrafish. *G3: Genes, Genomes, Genetics*. https://doi.org/10.1534/g3.114.015842
Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. https://doi.org/10.1038/nature09692
Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., … Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken β-globin insulator are separable activities. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.102179399
Ringrose, L., & Paro, R. (2004). Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins. *Annual Review of Genetics*. https://doi.org/10.1146/annurev.genet.38.072902.091907
Sano, H., Miyata, A., Horio, T., Nishikimi, T., Matsuo, H., & Kangawa, K. (2002). The effect of pituitary adenylate cyclase activating polypeptide on cultured rat cardiocytes as a cardioprotective factor. *Regulatory Peptides*. https://doi.org/10.1016/S0167-0115(02)00193-3
Sati, S., & Cavalli, G. (2017). Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma*. https://doi.org/10.1007/s00412-016-0593-6
Scharer, C. D., Blalock, E. L., Barwick, B. G., Haines, R. R., Wei, C., Sanz, I., & Boss, J. M. (2016). ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naïve SLE B cells. *Scientific Reports*, *6*. https://doi.org/10.1038/srep27030
Sedmera, D., Reckova, M., DeAlmeida, A., Sedmerova, M., Biermann, M., Volejnik, J., … Thompson, R. P. (2003). Functional and morphological evidence for a ventricular conduction system in zebrafish and Xenopus hearts. *American Journal of Physiology - Heart and Circulatory Physiology*.
Sheldon, S., & Poulton, K. (2006). HLA typing and its influence on organ transplantation. *Methods in Molecular Biology (Clifton, N.J.)*, *333*, 157–174. https://doi.org/10.1385/1-59745-049-9:157

Smale, S. T., & Kadonaga, J. T. (2003). The RNA Polymerase II Core Promoter. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev.biochem.72.121801.161520

Smart, N., Dubé, K. N., & Riley, P. R. (2013). Epicardial progenitor cells in cardiac regeneration and neovascularisation. *Vascular Pharmacology*. https://doi.org/10.1016/j.vph.2012.08.001

Smart, N., & Riley, P. R. (2012). The epicardium as a candidate for heart regeneration. *Future Cardiology*, *8*(1), 53–69. https://doi.org/10.2217/fca.11.87

Spitz, F. (2016). Gene regulation at a distance: From remote enhancers to 3D regulatory ensembles. *Seminars in Cell and Developmental Biology*. https://doi.org/10.1016/j.semcdb.2016.06.017

Stainier, D. Y. R., Lee, R. K., & Fishman, M. C. (1993). Cardiovascular development in the zebrafish: I. Myocardial fate map and heart tube formation. *Development*.

Staudt, D., & Stainier, D. (2012). Uncovering the Molecular and Cellular Mechanisms of Heart Development Using the Zebrafish. *Annual Review of Genetics*. https://doi.org/10.1146/annurev-genet-110711-155646

Stormo, G. D. (2013). Modeling the specificity of protein-DNA interactions. *Quantitative Biology*. https://doi.org/10.1007/s40484-013-0012-4

Szema, A. M., Dang, S., & Li, J. C. (2015). Emerging novel therapies for heart failure. *Clinical Medicine Insights: Cardiology*. https://doi.org/10.4137/CMC.S29735

Szenker, E., Ray-Gallet, D., & Almouzni, G. (2011). The double face of the histone variant H3.3. *Cell Research*, Vol. 21, pp. 421–434. https://doi.org/10.1038/cr.2011.14

Szerlong, H. J., & Hansen, J. C. (2011). Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire*, *89*(1), 24–34. https://doi.org/10.1139/O10-139

Thomas, N. A., Koudijs, M., van Eeden, F. J. M., Joyner, A. L., & Yelon, D. (2008). Hedgehog signaling plays a cell-autonomous role in maximizing cardiac developmental potential. *Development*. https://doi.org/10.1242/dev.024083

Tie, F., Banerjee, R., Stratton, C. A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., … Harte, P. J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development*, *136*(18), 3131–3141. https://doi.org/10.1242/dev.037127

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active β-globin locus. *Molecular Cell*, *10*(6), 1453–1465. https://doi.org/10.1016/S1097-2765(02)00781-5

Tompa, M. (2001). Identifying functional elements by comparative DNA sequence analysis. *Genome Research*. https://doi.org/10.1101/gr.197101

Trinh, L. A., Chong-Morrison, V., Gavriouchkina, D., Hochgreb-Hägele, T., Senanayake, U., Fraser, S. E., & Sauka-Spengler, T. (2017). Biotagging of Specific Cell Populations in Zebrafish Reveals Gene Regulatory Logic Encoded in the Nuclear Transcriptome. *Cell Reports*. https://doi.org/10.1016/j.celrep.2017.03.045

Van Wijk, B., Van Den Berg, G., Abu-Issa, R., Barnett, P., Van Der Velden, S., Schmidt, M., … Van Den Hoff, M. J. B. (2009). Epicardium and myocardium separate from a common precursor pool by crosstalk between bone morphogenetic protein- and fibroblast growth factor-signaling pathways. *Circulation Research*. https://doi.org/10.1161/CIRCRESAHA.109.203083

Vieira, J. M., Howard, S., Villa Del Campo, C., Bollini, S., Dubé, K. N., Masters, M., … Riley, P. R. (2017). BRG1-SWI/SNF-dependent regulation of the Wt1 transcriptional landscape mediates epicardial activity during heart development and disease. *Nature Communications*. https://doi.org/10.1038/ncomms16034

Visel, A., Rubin, E. M., & Pennacchio, L. A. (2009). Genomic views of distant-acting enhancers. *Nature*, *461*(7261), 199–205. https://doi.org/10.1038/nature08451

Wang, J., Cao, J., Dickson, A. L., & Poss, K. D. (2015). Epicardial regeneration is guided by cardiac outflow tract and Hedgehog signalling. *Nature*. https://doi.org/10.1038/nature14325

Weinberger, M., Simões, F. C., Patient, R., Sauka-Spengler, T., & Riley, P. R. (2018). Functional heterogeneity within the developing zebrafish epicardium. *BioRxiv*. https://doi.org/10.1101/460394

Weintraub, H., & Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. *Science*, *193*(4256), 848–856. https://doi.org/10.1126/science.948749

Witte, S., Bradley, A., Enright, A. J., & Muljo, S. A. (2015). High-density P300 enhancers control cell state transitions. *BMC Genomics*, *16*(1), 903. https://doi.org/10.1186/s12864-015-1905-6

Wu, S., Adams, B. A., Fradinger, E. A., & Sherwood, N. M. (2006). Role of two genes encoding PACAP in early brain development in zebrafish. *Annals of the New York Academy of Sciences*. https://doi.org/10.1196/annals.1317.091

Yelon, D. (2001). Cardiac patterning and morphogenesis in zebrafish. *Developmental Dynamics*. https://doi.org/10.1002/dvdy.1243

Yelon, D., Horne, S. A., & Stainier, D. Y. R. (1999). Restricted expression of cardiac myosin genes reveals regulated aspects of heart tube assembly in zebrafish. *Developmental Biology*. https://doi.org/10.1006/dbio.1999.9406

Yelon, D., Ticho, B., Halpern, M. E., Ruvinsky, I., Ho, R. K., Silver, L. M., & Stainier, D. Y. R. (2000). The bHLH transcription factor Hand2 plays parallel roles in zebrafish heart and pectoral fin development. *Development*.

Zandvakili, A., Campbell, I., Gutzwiller, L. M., Weirauch, M. T., & Gebelein, B. (2018). Degenerate Pax2 and Senseless binding motifs improve detection of low-affinity sites required for enhancer specificity. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1007289

Zhang, T., Cooper, S., & Brockdorff, N. (2015). The interplay of histone modifications – writers that read. *EMBO Reports*. https://doi.org/10.15252/embr.201540945

Zhang, Z., & Gerstein, M. (2003). Of mice and men: Phylogenetic footprinting aids the discovery of regulatory elements. *Journal of Biology*.

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., … Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics*. https://doi.org/10.1038/ng1891

Zhou, B., Honor, L. B., He, H., Qing, M., Oh, J. H., Butterfield, C., … Pu, W. T. (2011). Adult mouse epicardium modulates myocardial injury by secreting paracrine factors. *Journal of Clinical Investigation*. https://doi.org/10.1172/JCI45529