Functional analysis of cardiac specific *cis*-regulatory elements involved in heart development and disease

Julian Stolper

ORCID 0000-0001-7233-2686

Doctor of Philosophy

March 2021

Faculty of Medicine, Dentistry and Health Sciences

Department of Paediatrics

Submitted in total fulfilment for the degree of

Doctor of Philosophy

Summary

Birth defects are the leading cause of infant death and the most common are congenital heart defects (CHD). In the majority of cases, the underlying genetic cause is unknown. There is increasing evidence that variants in non-coding DNA within the functional *cis*-regulatory elements (CREs) controlling gene expression can cause CHDs. However, little is known about the human CREs that are crucial for cardiovascular development.

To address this knowledge gap, we bioinformatically identified cardiovascular CREs based on the occupancy of the CRE by the homeodomain protein NKX2-5 and cardiac chromatin histone modifications. This search defined several putative CREs within the loci of genes specifically implicated in heart development and disease. We show that human regulatory elements are functionally conserved and capable of driving reporter transgene expression *in vivo* throughout the developing cardiovascular system of medaka (*Oryzias latipes*). Using heart muscle cells and endothelial cells derived from human embryonic stem cells in which these CREs were deleted, we identified molecular perturbations in the gene regulatory network controlling vascular development. Among the CREs identified, we show that the deletion of the human *enFLT1* enhancer *in vitro*, a highly conserved enhancer within the *FLT1* locus, using endothelial tube formation assays did not disrupt vessel formation. However, deletion of the endogenous *enFLT1* enhancer in medaka results in blood vessel malformations, indicating an important role for this regulatory element *in vivo*.

These findings highlight the crucial role of human regulatory elements and their function as a regulator and buffer of transcriptional regulation in cardiovascular development. Our studies further support the hypothesis that perturbations in human CREs are important contributors of developmental defects.

General Declaration

I declare that this contains no material that has been accepted for any other degree or diploma at any university or equivalent institution. To the best of my knowledge and belief, this thesis compromises only my original work towards the degree of Doctor of Philosophy and no material previously published or written by any other person except where due referenced in the preface and text of this thesis.

This thesis consists of not more than 100,000 words. The ideas, development and writing of all chapters in this thesis were the principal responsibility of myself, the candidate, under the supervision and guidance or A/Prof. David Elliott, A/Prof. Mirana Ramialison and A/Prof. Enzo Porrello.

Julian Stolper

March 2021

Acknowledgements

I would like to start by expressing my sincerest gratitude to my supervisors, A/Prof. David Elliott, A/Prof. Mirana Ramialison and A/Prof. Enzo Porrello, as well as Jr. Prof. Lázaro Centanin and Prof. Jochen Wittbrodt. Thank you for letting me carry out this crazy idea of an intercontinental, interprofessional PhD project.

Dave, thank you for giving me the freedom to venture out and make my own mistakes and successes, and for the constant support you provided in and out of work. Thank you for the many laughs, it has been a blast! Mirana, the support and empathy you showed me throughout these years is beyond anything I could ask for. I am honoured to have been your first PhD student. Enzo, thank you for directing my focus to the big picture, for the great suggestions you have provided and the generosity you have shown me throughout my time in the lab. To Lázaro and Jochen, I am truly grateful to you for taking me in and enabling me to continue my project in the lab. Thank you for giving me creative freedom and endless support throughout my time in Germany. You have always treated me kindly and made me feel welcome. I am also very grateful for the support and input from my advisory committee members Prof. Christine Wells, Prof. Ed Stanley and A/Prof. Shireen Lamande. I would also like to thank the amazing current and previous members in the Porrello/Elliott, Ramialison and Centanin/Wittbrodt Labs. You have made my PhD experience a truly wonderful one. I would like to thank Evangelyn, Kathy and Ling for your generosity, help and wisdom as well as your sense of humour and to be able to deal with my jokes and banter. Thank you to the Porrello/Elliott lab, specifically Menno, Francesca, Céline, Adam, James, Dad, Emma, Beth, Ravi, Hayley, Neda and Holly, for providing valuable input, making me laugh, challenging me and finally for bringing me back to reality when I needed it. I am incredibly thankful for the warm and comfortable environment you created in the lab. Furthermore, my sincere thanks to the Blood and Stem Cell groups, namely Tanya, Freya, Jasna, Rita, Koula and Kat. I will not forget the laughs we had during lunch and coffee breaks. To all of the current and previous Ramialison lab members and fellow PhD students Gulrez, Lisa and Sebastian, thank you for making my time at Monash an enjoyable one. I appreciate your help in the past couple of years. Also, thank you Jeannette, for your constant support throughout the past years and for never letting me go hungry while being in your presence.

I would also like to take this opportunity to thank all members of the COS-labs In Heidelberg that have helped me to grow scientifically as well as personally. Specifically, I'd like to thank the members of the Centanin/Wittbrodt labs: Ali, Isa, Karen, Liz and Omar. The time in our little box at the end of the corridor was lively, full of joy and singing and I am grateful for your friendship and dedication. Silvi, Naima and Maike, thank you for being there for me whenever I needed to talk, vent or laugh. I miss all of you!

Finally, I would like to show my love and gratitude to my family, especially my parents and Denis. Thank you for always believing in me and supporting me from afar. It has been an incredible journey and I am glad that you have walked it alongside me. This would have never been possible without you.

Table of Contents

List of Tables ..... *viii*

List of Figures ..... *ix*

List of Abbreviations ..... *x*

1 Chapter 1: Introduction *1*

1.2 Heart Development and Congenital Heart Disease **1**

1.2.1 Mammalian cardiovascular development 1

1.2.2 Teleost cardiovascular development 2

1.2.3 Cardiac networks and its perturbations 3

1.3 Genome Wide Association Studies **4**

1.4 Enhancers **5**

1.4.1 Evolution of Enhancer Elements – New and old 6

1.4.2 Motifs and Transcription Factor Binding 10

1.4.3 Conservation of enhancer elements 12

1.4.4 Additional effect of Active Enhancers 14

1.4.5 Enhancer Redundancy 16

1.5 Assessing Enhancer Activity **17**

1.5.1 Genomic Landscapes - DNA accessibility and Histone modification 18

1.5.2 Assessing enhancer transcripts and their function 20

1.5.3 Enhancer sufficiency 21

1.5.4 Enhancer necessity 23

1.6 Enhanceropathies **24**

1.6.1 Variants in enhancer sequences 25

1.7 Regulatory elements in cardiac disease **26**

1.8 Aims **29**

2 Chapter 2: Material and Methods *30*

2.1 Bioinformatic mining and enhancer prediction **30**

2.2 Routine methods *in vivo* **31**

2.2.1 Fish maintenance and egg collection 31

2.2.2 Mating and microinjection 31

2.3 Routine methods *in vitro* **32**

2.3.1 Culturing of human pluripotent stem cells 32

2.3.2 Polymerase Chain Reaction (PCR) 32
2.3.3 Quantitative Polymerase Chain Reaction (qPCR) 35
2.3.4 Gel electrophoresis 36
2.3.5 Transformation 36
2.3.6 Plasmid preparation 37
2.3.7 Flow cytometry 38

2.4 Transgenesis in medaka 39
2.4.1 Enhancer expression vector 39
2.4.2 Transgenesis via *Tol2* and SceI 39

2.5 CRISPR/Cas9 gene editing 40
2.5.1 sgRNA target prediction 40
2.5.2 Generation of sgRNA for *in vivo* gene editing 40
2.5.3 CRISPR/Cas9 injection 41
2.5.4 Generating sgRNA for *in vitro* gene editing 41
2.5.5 Transfection of cell lines 42
2.5.6 Colony picking and PCR screening 43
2.5.7 Single Nucleotide Polymorphism (SNP) analysis and Sequencing 43

2.6 Differentiation Assays 43
2.6.1 Cardiomyocyte differentiation 43
2.6.2 Endothelial cell differentiation 44

2.7 Functional assays 45
2.7.1 Cardioexcyte 45
2.7.2 Angiogenesis assay 45
2.7.3 Wound healing assay 45

2.8 Immunofluorescence and confocal microscopy 46
2.8.1 *in vivo* whole mount immunostaining and confocal microscopy 46
2.8.2 *in vitro* immuno staining and confocal microscopy 46
2.8.3 *in vitro* image analysis and statistics 47

2.9 RNA sequencing and analysis 48

2.10 Chromatin Confirmation Capture (Hi-C) analysis 48

3 Chapter 3: Regulatory Elements – Prediction and Validation 49

3.2 Background 49

3.3 Results 51
3.3.1 Predicting primate specific *cis*-regulatory elements 51

3.3.2 Human cardiovascular specific elements are able to drive gene expression in medaka 54
3.3.3 Ultra-conserved cardiac specific regulatory elements 58

3.4 Discussion 60

4 Chapter 4: enFLT1 – a regulator of gene expression during angiogenesis 64

4.2 Background 64

4.3 Results 66

4.3.1 The human enhancer of *FLT1* is able to drive cardiovascular gene expression *in vivo* 66
4.3.2 Deletion of *enFLT1* perturbs cardiovascular development *in vivo* 68
4.3.3 Enhancer of *FLT1* is essential for pathways involved in blood vessel morphogenesis 70
4.3.4 Deletion of *enFLT1* does not impair angiogenesis *in vitro* 77

4.4 Discussion 81

5 Chapter 5: cis-Regulation in human cardiomyocytes 86

5.1 Background: 86

5.2 Results 87

5.2.1 Investigating the role of cardiac specific CREs *in vivo & in vitro* 87
5.2.2 Chromatin interaction reveals targets of *enC5ORF35*, *enKCNH2* and *enSMG6* 90
5.2.3 Putative cardiac specific CREs regulate gene expression in human cardiomyocytes 92

5.3 Discussion 95

6 Chapter 6: Final Discussion 101

6.1 Identifying putative *cis*-regulatory elements 101

6.2 Enhancer redundancy and robustness 102

6.3 The relevance of enhancer in human disease 105

6.4 Future perspectives 106

7 References 107

8 Appendices 128

8.2 Supplementary data 128

8.3 Research output 131

8.3.1 Understanding *cis*-regulatory elements in cardiovascular development 131
8.3.2 Understanding cardiac biology in the fish model system 132
8.3.3 Understanding cardiac transcription factors and their role in CHD 132
8.3.4 Understanding and developing skillsets for the medaka as a model stystem 133

8.3.5 Understanding enhancer grammar and TFBS composition 134

8.4 COVID-19 related impacts 134

List of Tables

Table 2-1: Human pluripotent stem cell lines used in this thesis 32

Table 2-2: Oligos for *in vivo* enhancer transgenesis assay 33

Table 2-3: CRISPR/Cas9 gene editing screening oligos 34

Table 2-4: Probes for real time PCR-based gene expression profiling 36

Table 2-5: Plasmids containing human RE for *In vivo* enhancer assay 37

Table 2-6: sgRNA sequences used for *in vivo* gene editing 40

Table 2-7: sgRNA sequences used for *in vitro* gene editing 41

Table 2-8: List of anitbodies used in IHC, FACS and angiogenesis assays 47

Table 3-1: putative REs used for transgenesis assays i*n vivo* 53

Table 3-2: Non-coding REs ultra-conserved from fish to human 59

Table 8-1: Uniquely differentially expressed genes in *ΔenFLT1* 129

List of Figures

Fig. 1-1: Mammalian heart development. 2
Fig. 1-2: Models of regulatory element evolution. 8
Fig. 1-3: Models of enhancer-transcription factor binding. 11
Fig. 1-4: Models of transcription activation via regulatory elements. 15
Fig. 3-1: Predicting primate specific and cardiac specific *cis*-regulatory elements. 51
Fig. 3-2: Schematic representation of the Medaka life cycle. 54
Fig. 3-3: Human enhancers drive GFP expression in medaka. 56
Fig. 3-4: Bioinformatic pipeline to identify ultra-conserved cardiac RE. 58
Fig. 4-1: The *enFLT1* regulatory element efficiently drives expression in the medaka cardiovascular system. 66
Fig. 4-2: Functional analysis of *enflt1* during angiogenesis *in vivo*. 68
Fig. 4-3. Genetic modification of *FLT1* locus to study the role of *enFLT1* in human endothelial cells. 71
Fig. 4-4: Expression of endothelial specific makers is not perturbed in *ΔenFLT1* and *Δex1FLT* pluripotent stem cell derived endothelial cells. 72
Fig. 4-5: *enFLT1* is required for regulation of *FLT1* dependent transcriptional pathways. 73
Fig. 4-6: *enFLT1* regulates alternative pathways involved in blood vessel development. 75
Fig. 4-7: Functional analysis of *enFLT1* during angiogenesis. 78
Fig. 4-8: Altered expression pattern of *FLT1* in *ΔenFLT1* cardiac and chromatin arrangement at the *FLT1* locus in pluripotent stem cell derived cardiomyocytes and HUVEC cells. 79
Fig. 5-1: Chromosomal localization of putative cardiac specific *cis*-REs. 88
Fig. 5-2:Functional analysis of *enppargc1a in vivo*. 89
Fig. 5-3: Chromatin capture analysis of putative REs. 90
Fig. 5-4: Enhancers for *C5orf38*, *FLT1*, *KCNH2*, *PPARGC1A* and *SMG6* are not essential for cardiac differentiation. 93
Fig. 5-5: Expression analysis of RE knockout in cardiomyocytes. 94
Fig. 8-1: Expression of FLT1 does not seem to be perturbed in *ΔenFLT1* and *Δex1FLT* pluripotent stem cell derived endothelial cells. 128
Fig. 8-2: Genotyping and sequencing of *Δenflt1* medaka and *ΔenFLT1* human embryonic stem cells. 129
Fig. 8-3: Electrophysiological analysis of RE knockout cardiomyocytes. 131

List of Abbreviations

AMD	Age-related macular degeneration
APC	Allophycocyanin
bFGF	Basic fibroblast growth factor
BMP	Bone morphogenic protein
Bp	Base pair
BSA	Bovine serum albumin
BX-C	Bithorax Complex
CAGE	Cap Analysis of Gene Expression
Cas9	CRISPR associated protein 9
CBP	CREB Binding Protein
CHD	Congenital heart disease
ChIP-seq	Chromatin Immunoprecipitation sequencing
CHIR	CHIR99021
CRE	Cis-regulatory elements
CRISPR	Clustered regularly interspaced short palindromic repeats
D	Day
DHS	DNaseI hypersensitive sites
DNA	Deoxyribonucleic acid
E	Embryonic (day)
ECM	Extracellular matrix
EDTA	Ethylenediamine-tetra-acetic acid
EGM	Endothelial Growth Medium
ERM	Embryo-rearing medium
eRNA	enhancer RNA
ERV	Endogenous retrovirus
ETS	ETS Proto-Oncogene
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum

FGF8/10	Fibroblast Growth Factor 8/10
FLT1	Fms Related Receptor Tyrosine Kinase 1
FOXH1	Forkhead Box H1
GATA	GATA Binding Protein
GFP	Green fluorescent protein
GRO-seq	Global run-on sequencing
GWAS	Genome wide association studies
HAND2	Heart And Neural Crest Derivatives Expressed 2
hESC	Human embryonic stem cell
Hi-C	Chromosome conformation capture
HUVEC	Human umbilical vein endothelial cell
INF	Interferon
IWR	WNT pathway inhibitor
KDR	Vascular endothelial growth factor receptor 2
KO	Knockout
LTR	Long terminal repeats
MEF	Mouse embryonic fibroblast
MEF2C	Myocyte Enhancer Factor 2C
mg	Milligram
mM	Millimolar
MPRA	massively parallel reporter assay
NET-seq	Native elongation sequencing
ng	Nanogram
NGS	New generation sequencing
NKX2-5	NK2 Homeobox 5
PCR	Polymerase chain reaction
PE	Phycocerythrin
PFA	Paraformaldehyde
PI	Propidium Iodide
PITX2	Paired Like Homeodomain 2

PTM	Post-translational modifications
RNA	Ribonucleic acid
RT	Room temperature
S2E	Stripe 2 enhancer
SAN	Sinoatrial node
SCF	Stem cell factor
SCN5a/10A	Sodium Voltage-Gated Channel Alpha Subunit 5/10
sgRNA	Single-guide RNA
SHH	Sonic Hedgehog Signaling Molecule
SNP	Single nucleotide polymorphism
STARR-seq	Self-transcribing active regulatory region sequencing
TAD	Topologically associating domain
TBE	TRIS-Borat-EDTA Buffer
TBX5	T-box Transcription Factor 5
TE	transposable elements
TF	Transcription factor
TFBS	Transcription factor binding site
TSS	Transcription start site
TT-seq	Transient Transcriptome sequencing
UTX	Lysine Demethylase 6A
VEGF	Vascular Endothelial Growth Factor
VWF	Von Willebrand Factor
WNT	Wingless-Type MMTV Integration Site
ZRS	Zone of polarizing activity regulatory sequence
µg	Microgram
µl	Microliter
µM	Micromolar

1 Chapter 1: Introduction

1.2 Heart Development and Congenital Heart Disease

Heart development is a complex process, which when disrupted has serious consequences on the later life of the organism. Cardiac malformations can range from cardiac dysfunction to impaired pump function and death. Extensive studies in the mouse have been used to understand the molecular, cellular and morphological basis of mammalian heart development and disease (Krishnan *et al.*, 2014).

1.2.1 Mammalian cardiovascular development

During gastrulation, two separate cardiac progenitor cell populations that share a common progenitor are the earliest morphological markers of the developing heart. The primary heart field, also known as the cardiac crescent, is formed at embryonic day 7.5 (E7.5, human: day 15) in the anterior mesoderm (Tam *et al.*, 1997). It will give rise to left ventricle and to both atria. Subsequently, the secondary heart field is derived from the pharyngeal mesoderm directly medial and anterior to the primary heart field. This in turn will contribute to the formation of the right ventricle, the outflow tract and both atria as well. At E8 (human: day 20) cardiac progenitor cells from the cardiac crescent come together to form a linear heart tube that is then expanded by progenitor cells from the secondary heart field. This leads to the formation of the arterial and venous poles. The heart tube is subsequently looped, the first appearance of asymmetry in the embryo, at E8.5 (human: day 28) to form the primitive ventricles and atria. The venous poles are then brought anteriorly (E10.5, human: day 32) and the primitive chambers are separated by septation into atrium and ventricle and the valves begin formation (E15) (Garry and Olson, 2006; Koefoed *et al.*, 2014; Srivastava, 2006; Xin *et al.*, 2013) (*Fig. 1-1*).

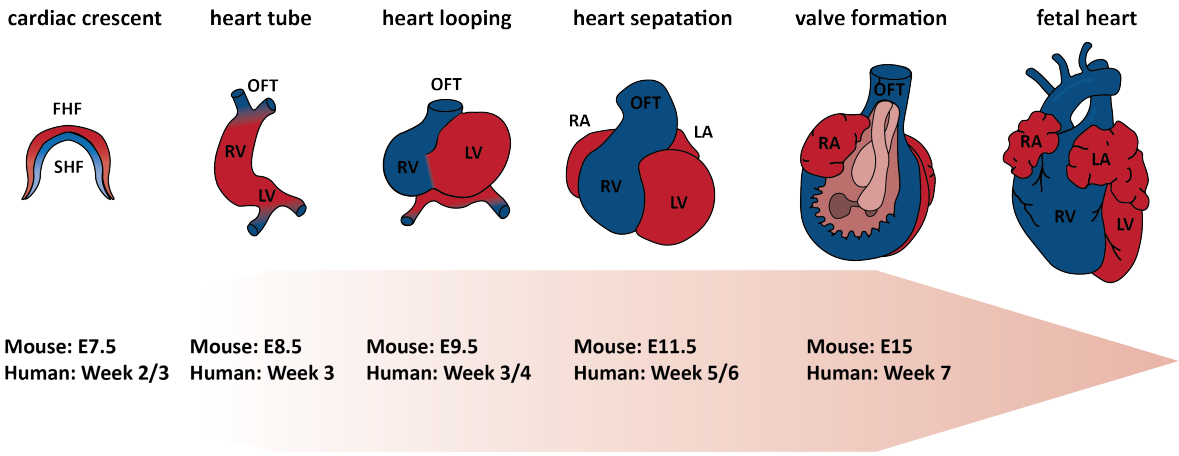


Fig. 1-1: Mammalian heart development. Mammalian heart development starts with the cardiac crescent divided with the primary or first heart field (FHF, red) and second heart field (SHF, blue). The heart then forms a heart tube which is subsequently looped to form primitive ventricles (right ventricle (RV), left ventricle (LV), atria (right atrium (RA), left atrium (LA) and outflow tract (OFT). The chambers are then separated into atrium and ventricle and the valves develop to then form the mammalian fetal heart.

Once the heart is fully matured, many different cell types contribute to maintain heart function (Litviňuková *et al.*, 2020; Pinto *et al.*, 2016; Quaife-Ryan *et al.*, 2017). Cardiomyocytes make up the myocardium, the muscular walls of the heart. The majority of cell types in the four-chambered heart are cardiac fibroblasts that will contribute to the epicardium, the outer layer of the heart (Xin *et al.*, 2013). The inner layer, the endocardium, consists of endothelial cells that form the blood vessels and cardiac valves. Smooth muscle cells maintain structural stability in coronary arteries and the in- and outflow vasculature. In order for the heart to pump efficiently, the electrical properties of specialized cardiomyocytes in the conduction system (pacemaker cells and Purkinje fibers) are essential. One special group of pacemaker cells contribute to the sinoatrial node (SAN) or sinus node. This structure is located in the right atrium and spontaneously generate electrical impulses that act as initial impulses to start the contraction of the heart (Buckingham *et al.*, 2005).

1.2.2 Teleost cardiovascular development

For a long time, beside mammalian models, fish model systems have played an essential role in understanding heart development and cardiovascular molecular networks. Teleost fish such as medaka (*Oryzias latipes*) and zebrafish (*Danio rerio*) are among the most studied models in this field (Lai *et al.*, 2017; Poon and Brand, 2013; Wittbrodt *et al.*, 2002). Although there are many differences between zebrafish and medaka, living conditions and anatomical structures as well as gene expression are similar and directly comparable (Furutani-Seiki and Wittbrodt, 2004; Lin *et al.*, 2016). Therefore, medaka has served as a suitable and complementary model

to zebrafish, which has been well established for the past 30 years (Hilgers and Schwarzer, 2019). In this study the medaka has been used as a model of choice due to its many advantages and applications in cardiovascular development, gene expression analysis, identification of non-coding *cis*-regulatory elements and genetic editing (Lin *et al.*, 2016). In brief, medaka heart development starts with the appearance of the tubular heart or heart anlage at stage 22 (1 day, 14h) underneath the head. Three hours later, the heart tube extends, and blood vessels begin to form on the yolk. At stage 24, the heart starts to beat and increases in pulsation until the onset of blood circulation at stage 25 (2 days, 2 h) while the dorsal aorta develops. The differentiation of the sinus venosus, atrium, ventricle and outflow tract are completed at stage 29. As the embryo grows, blood vessel development continues by branching of the arteries in the body trunk the gills and the tail. At stage 36 (6 days) heart development is fully completed by flexion of the atrio-ventricular region (Iwamatsu, 2004).

1.2.3 Cardiac networks and its perturbations

Signalling pathways that are essential for cardiogenesis are well established (McCulley and Black, 2012; Srivastava and Kwon, 2010). In order to correctly develop the heart, it requires key signalling molecules to be expressed in a precise spatio-temporal manner. The main protagonists such as bone morphogenic protein (BMP), WNT, sonic hedgehog (SHH) and FGFs initiate transcriptional activation of cardiac genes at the onset of heart development in the primary and secondary heart fields. These key regulators include factors that are essential for heart growth and patterning. In the primary heart field, T-box 5 (TBX5) and also in the secondary heart field, GATA4 and NK2 homeobox 5 (NKX2-5) are expressed and coordinate the transcriptional networks required for cardiogenesis. Other important factors that are restricted to the secondary heart field only, include ISL1, FOXH1, HAND2, MEF2C, FGF8/10 and PITX2 (Bruneau, 2013; Olson, 2006; Xin *et al.*, 2013). Any mis-regulation of this tightly controlled network can have severe impacts on cardiogenesis in the developing embryo and can lead to life threatening conditions like Holt-Oram Syndrome or tetralogy of Fallot (Edwards and Gelb, 2016).

Around 1% of babies are born each year with heart malformations (Virani *et al.*, 2020). Congenital heart disease (CHD) can be caused by many different genetic and environmental factors, such as infections, maternal drug use and maternal nutrient deficiency (Blue *et al.*, 2012). Therefore, CHDs are complex and identifying the underlying cause is often difficult. CHDs

themselves are varied and include a broad range of defects such as holes in the septum and thickening of the internal walls, abnormalities of the valves and defects in heart chambers and blood vessels. Symptoms can range from mild to lethal and often corrective heart surgery is the only treatment option (Blue *et al.*, 2012; Gurvitz *et al.*, 2016; Mendis *et al.*, 2011). Unfortunately, around 80% of the causes of CHD are still unknown and it remains challenging to determine the pathogenesis of CHD in order to work towards early detection of the disease (Edwards and Gelb, 2016). In order to investigate the genetic factors contributing to CHD, genomic approaches and genetic screens have been used to understand the genetic architecture of congenital heart diseases.

1.3 Genome Wide Association Studies

To understand the genetic contribution to a human phenotype on a large population based scale, Genome Wide Association Studies (GWAS) have been performed for many different diseases or biological traits (Agopian *et al.*, 2017; Wang and Wang, 2018). GWAS aim to identify genetic risk factors in diseases ranging from common to rare and complex defects, such as CHD. In these studies, genetic variants linked to the disease phenotype are identified by either Single nucleotide polymorphisms (SNP)-arrays or, increasingly, human whole genome sequencing. The aim is to identify genetic variants that can reliably be used to assign disease risk and to subsequently be used to investigate the underlying genetic mechanism of the disease. Furthermore, the clinical utility of GWAS approaches has enabled researchers to identify previously unknown genes and link them to specific diseases. As an example, through combination of association and genetic linkage the *Complement Factor H* gene was identified as an essential key player in the onset of age-related macular degeneration (AMD) by GWAS (Haines *et al.*, 2005). Not only do GWAS provide the background to understand disease and help with the development of possible therapies and prevention strategies on a large scale, but they also have the potential to tailor healthcare to individuals and their needs due to personal genetic testing (*i.e*. personalised medicine)(Bush and Moore, 2012).

The basis of all GWAS studies is to identify genetic variation in each patient in the form of SNPs. These single base-pair changes occur in a high frequency in the genome (Durbin *et al.*, 2010). In GWAS, SNPs are being used as markers to navigate through the genome. In the majority of

cases, SNPs do not have any impact on the genetic network in an organism. However, some SNPs can have mild to severe consequences on biological systems like transcriptional and translational regulation, mRNA stability and amino acid changes. In these cases, GWAS aim to connect disease causative SNPs in different individuals to a specific trait in comparison to a healthy control group. When analysing sequencing data, it has been shown that more abundant SNPs are less likely to be implicated in a genetic disease compared to rare ones due to evolutionary pressure (Kerem *et al.*, 1989). This is due to extremely rare variants that lead to severe alteration in protein sequence and integrity or transcriptional output, which in turn leads to disease. When looking at datasets that were acquired for GWAS, many studies still focus on coding regions when aiming to find a cause for a specific trait or disease. However, only 2% of the genome are protein-coding while 98% harbour non-coding DNA and RNA regulatory elements that are essential for transcriptional regulation. While regulatory regions are often neglected, a striking majority of SNPs (92%) fall in non-coding regions in the genome, underlining the importance of *cis*-regulatory elements in regard to development and disease (Corradin and Scacheri, 2014).

1.4 Enhancers

Enhancers are a class of *cis*-regulatory element (CREs) that usually span between 100 base pairs (bp) and 1000 bp (Schaffner, 2015). They have the capacity to regulate and activate transcription over long distances in the genome. In contrast to promoters, enhancers serve as transcriptional regulatory elements that vary in distance, location and their orientation relative to their target gene(s). This spatial separation means that genes are able to be regulated via multiple regulatory elements that in turn enable precise gene expression and transcriptional regulation. CRE redundancy and modularity allow the regulation of target genes in a spatio-temporal manner that facilitates cell-type specificity. Furthermore, CREs are also subject to natural selection during evolution, for example by harbouring variants that lead to genetic variation and by buffering the risk of lethality (Levine, 2010; Wittkopp and Kalay, 2012).

In recent years the study of gene regulation has been more accessible due to new screening technology, genome-editing techniques and newly acquired genome sequences. It facilitated the investigation of the evolution of *cis*-regulatory elements, the origin of enhancers, the role

of transcription factors in gene regulation and the role of genome folding during that process by comparative genomics (Long *et al.*, 2016).

1.4.1 Evolution of Enhancer Elements – New and old

Changes in regulatory elements themselves lead to a change in transcription activation because of their important role in gene regulation. Thus, this key role ensures these enhancers are also under extensive evolutionary pressure. It is important to look at the genomic background from which enhancers arose to elucidate mechanism of enhancer function. New regulatory elements can be established in many ways. The first example is genome duplication events as a genomic substrate for the evolution of new enhancers. Genome duplication is one of the key drivers providing the genomic diversity on which natural selection acts during evolution (Fig. 1-2a). With it, new loci arise that have the capability to evolve and diverge to then, in turn develop new functions and programs (Taylor and Raes, 2004). In many cases, after such events, many duplicated regions are often lost since there is no selective pressure on one of the paralogs. In terms of enhancers, a sub functionalisation of enhancer activity may occur (Fig. 1-2a'), meaning that expression domains of affected gene paralogs may shift to a different pattern. In addition, enhancers can drive a drastic change between the two paralogs, when already existing regulatory elements are deployed in new and different ways (Fig. 1-2a'') (Goode *et al.*, 2011; Lan and Pritchard, 2016). Therefore, enhancers can help to maintain both paralogs due to their effect on both copies.

A very prominent and important example for this is the whole genome duplication in teleost fish that occurred around 350 million years ago. This evolutionary milestone led to the creation and maintenance of many different genes and the divergence in function between different paralogs due to the evolutionary pressure on the gene regulatory network (Kleinjan *et al.*, 2008). However, new functions for enhancer elements is not solely reliant upon whole genome duplication. Duplication of smaller non-coding DNA sequences, though often lost due to selective pressure of genetic drift, can lead to the generation of additional regulatory elements. This can occur through change in function of established enhancer elements, or through previously non-regulatory sequences that develop into functioning enhancers (Rebeiz *et al.*, 2011). Repurposing of old enhancers, however, faces a lower evolutionary barrier since transcription factor binding sites (TFBS) are already present in the sequence permitting binding of transcription factors. In an extensive study, the transcription factor motif (*i.e.* the DNA

sequence recognised by a transcription factor) and grammar (arrangement of transcription factor binding sites) present within 1.3 million DNaseI hypersensitive sites (DHSs) in mouse and orthologous human genomes were analysed. This study revealed that the *cis*-regulatory network in these genomic compartments have been massively rewired and repurposed during evolution (Vierstra *et al.*, 2014). In turn, repurposing of the gene regulatory network leads to a divergent transcriptional output of target genes, enabling changes in morphology and gene expression. This has been demonstrated in the trichomes in *Drosophila* and the length of extremities in bats (Cretekos *et al.*, 2008; Frankel *et al.*, 2011).

Another group of regulatory elements involves enhancer exons, or eExons. These are enhancer elements that arose in protein coding regions – either in exons or overlapping exonic sequences - to regulate gene expression in certain tissues. For example, in the mouse liver 6% of the regulatory elements that have been identified were located in protein coding sequences which regulate liver-specific genes in close proximity (Birnbaum *et al.*, 2012). In addition, these enhancer exons can also display a dual function, which is an essential feature during genome evolution. This means that these “duons” can be translated into amino acid sequences and also display TF-binding capacity (Stergachis *et al.*, 2013). However, it remains to be elucidated if the enhancer function is subject to the same evolutionary pressure as the protein coding sequence (Agoglia and Fraser, 2016; Xing and He, 2015).

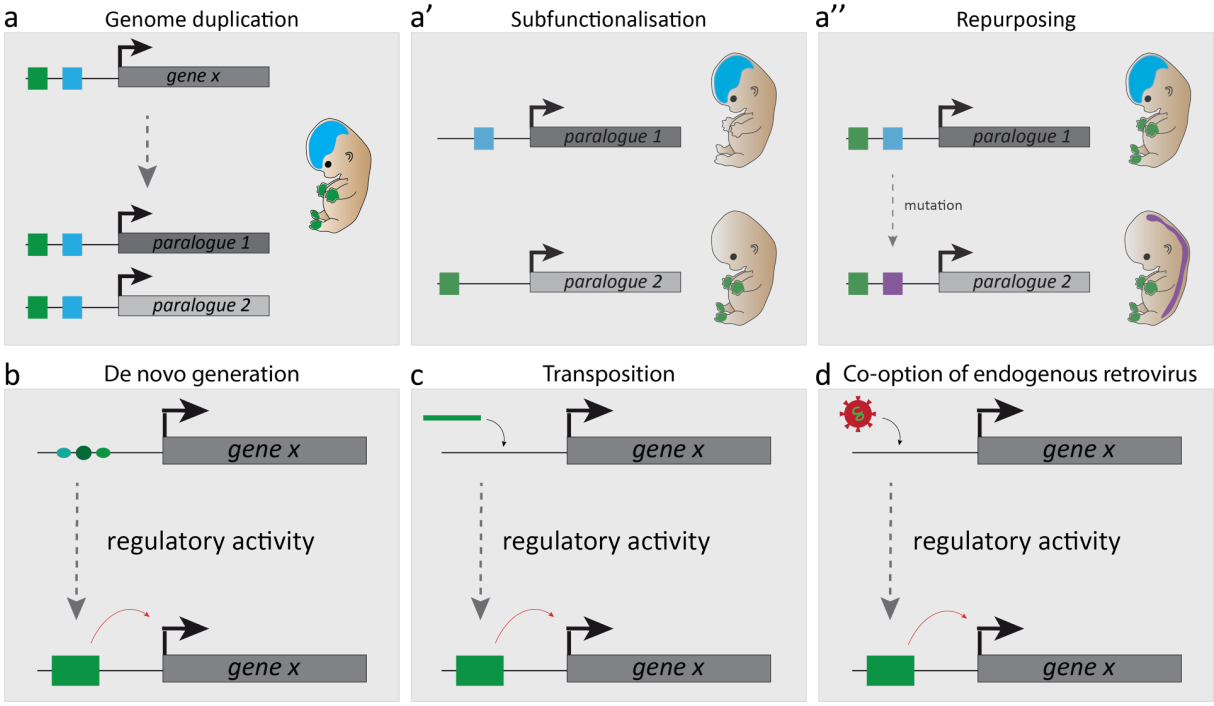


Fig. 1-2: Models of regulatory element evolution. With the occurrence of genome duplication, regulatory elements can be duplicated and evolve in a way to (a') sub-functionalise to a different pattern or (a'') get repurposed and drive gene expression in a different tissue or developmental stage. (c) Novel regulatory elements may arise through appearance of transcription factor binding sites or via genetic drift. (c) Transposable elements (TE) can lead to emergence of novel RE via repurposing of the regulatory sequence of the TE. (d) New RE may also emerge through exaptation of long terminal repeats of endogenous retrovirus (ERV) elements.

In addition, enhancer elements can also emerge *de novo* from non-regulatory DNA (Fig. 1-2b). When randomized sequences of synthetically generated "enhancers" were tested several were capable of driving expression in spatio-temporal manner (Smith *et al.*, 2013). This indicates that activity from regulatory elements can arise solely by random genetic drift. *in silico* approaches to model evolution of anterior-posterior (A-P) patterning *cis*-regulatory elements in *Drosophila* revealed neutral DNA sequences take within 0.5 to 10 million years to be incorporated as functional enhancers (Duque and Sinha, 2015). This modelling fits the *in vivo* observation of the evolution of enhancer elements within the *Drosophila* genus (Arnold *et al.*, 2014). Furthermore, this approach in mammalian systems supports the notion that the emergence of enhancers from non-regulatory and non-coding sequences generates diversity across the animal kingdom. As an example, it has been shown using comparative genomic of liver tissue of 20 different mammalian species, that 52% to 77% of regulatory elements emerged due to alteration of formally non-regulatory sequences. Moreover, analyses of the remaining sequences revealed that these alterations were caused primarily by transposable elements (TEs) (Villar *et al.*, 2015).

TEs are very common in almost all eukaryotic genomes and provide a whole new level of genetic regulation to the original regulatory network due to the fact that they carry regulatory sequences with them (Lander *et al.*, 2001). Since they are dispersed through the whole genome via transposon expansion they, use the endogenous gene regulatory machinery for their transcription (Feschotte, 2008). In addition, mutations in transposable elements can also lead to a tissue-specific shift in function that is then used for the endogenous regulatory network (Bourque *et al.*, 2008). Since this is the case, TEs thus play a crucial role in the evolution of enhancers and the rewiring and repurposing of non-regulatory sequences (Fig. 1-1c). Examples for such processes providing the molecular basis for evolution include sweet perception in primates (Ting *et al.*, 1992) and the evolution of the extraembryonic tissue boundaries between foetus and mother in eutherian mammals (Lynch *et al.*, 2011).

In fact, it has been shown that the origin of most of human and primate specific enhancer elements are transposable elements (Jacques *et al.*, 2013). One example for this are *Alu* elements, the most abundant and common TEs in the human genome. They make up around 10% of the DNA and consist of various enhancer motifs that, when altered through variants or other mutations, can acquire regulatory functions (Su *et al.*, 2014). From an evolutionary perspective, between human and chimpanzees many enhancers have changed their functional regulation since their last common ancestor, due to a functional shift of already pre-existing TEs (*e.g.* ERV1 elements in primate neural crest cells) (Prescott *et al.*, 2015). In addition, it has been shown in the human innate immune system that gene regulation can also occur via the mediation of long terminal repeats (LTR) (Chuong *et al.*, 2016). This supports the hypothesis that new regulatory elements emerged due to exaptation of long terminal repeats (LTRs) of endogenous retroviral (ERV) elements (Fig. 1-1d). These LTRs can gain new regulatory properties such as tissue specificity or induced and enhanced transcription of target genes (Chuong *et al.*, 2016). Since ERV arose from retroviral infections, further studies are needed to establish how strong the effect of such viral endogenization has been on the evolution of the regulatory network (Long *et al.*, 2016).

1.4.2 Motifs and Transcription Factor Binding

Enhancers are sequences in the genome that consist of multiple transcription factor (TF) binding sites. Typically, several TFs, which are in general cell-type and tissue specific proteins that facilitate the proper transduction of intrinsic and extrinsic signalling pathways, bind at enhancers. A major aspect of gene regulation is the structure of chromatin and the strong interaction of histone octamers with DNA. It is thought that the binding of multiple TFs can interfere with this interaction and enable binding of several DNA-binding proteins and co-factors to the regulatory element (Tillo *et al.*, 2010). In order for TFs to overcome the strong barrier to access the DNA, several scenarios can occur. "Direct cooperativity" involves the direct binding between TF before or while protein-DNA interaction occurs. "Indirect cooperativity" and "collaborative competition" on the other hand involves multiple TF that compete together with the histone octamer to bind to the DNA sequence that harbours the CRE (Miller and Widom, 2003). To predict the gene-expression output of these interactions, one can focus on the number of TF motifs. In general it can be said that a higher number of motifs is a good indication for a decrease in histone octamer binding and an increase on a transcriptional level (Erceg *et al.*, 2014; Smith *et al.*, 2013). Besides the cooperative binding and the collaborative competition, some TFs can bind to nucleosomal DNA and then initiate the activation of an enhancer. These TFs have been termed master regulators or pioneer factors, and it has been shown that these are crucial during cell lineage determination. The two-step mechanism of enhancer activation first involves the binding of the master regulator to the DNA, and then the recruitment of chromatin remodelling factors occurs. Subsequently, these factors then modify the surrounding histones via post-translational modifications such as acetylation and phosphorylation in order to permit transcription. Thus, the chromatin is in an open state and additional TFs and coactivator can then bind to the regulatory element (Zaret and Carroll, 2011).

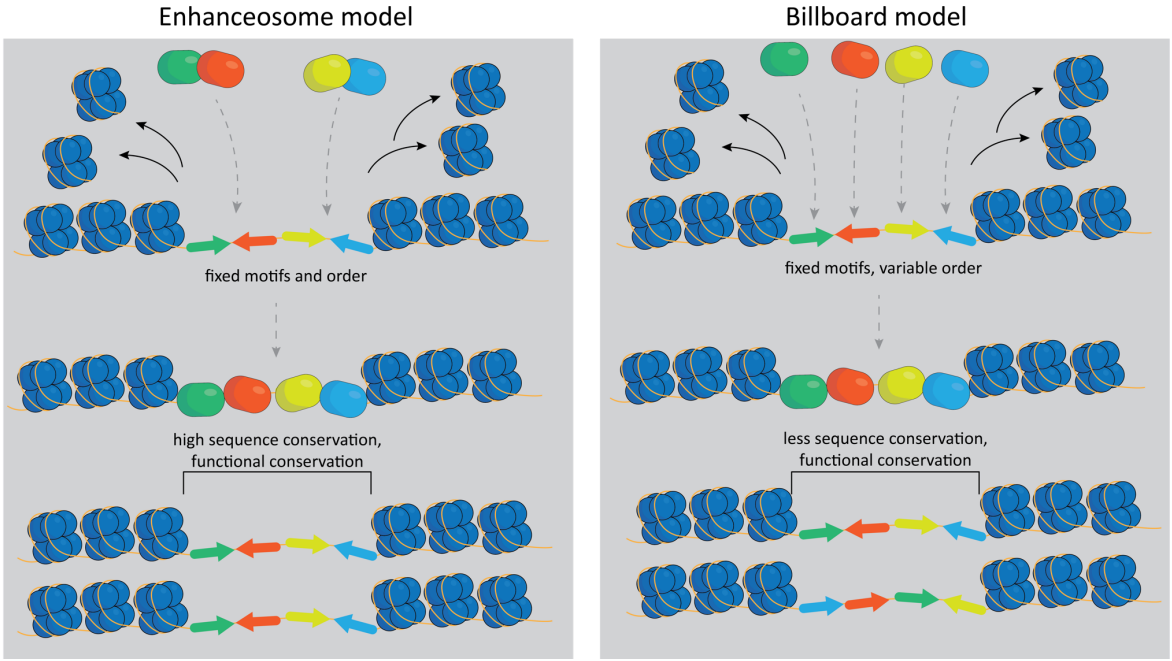


Fig. 1-3: Models of enhancer-transcription factor binding. Enhanceosome model (left) proposes fixed types and order of transcription factor binding motifs in regulatory elements. This results in a high sequence and functional conservation between species. The billboard model (right) describes fixed transcription factor binding motifs but a variable order, resulting in lower sequence conservation with the same functional conservation between species.

Enhancers affect the transcriptional output via a number of mechanisms. One example for this is the composition and organisation of the transcription factor motifs. The term enhancer "grammar" has been used to summarise the properties of motif composition within an enhancer. Many studies have aimed to unravel the mechanisms of how the organisation of motifs within an enhancer influences the activity and robustness of the system. This includes the type of motif, their binding affinity to various TFs, the amount and distribution within the enhancer, the orientation on the DNA strand, and the physical shape of the DNA helices. Although many models have been proposed that describe enhancer architecture, two main models have become established in the field. The "enhanceosome" model is based on a very organised and rigid motif grammar in the regulatory element (Fig. 1-3, left). Thus, the strict motif distribution leads to a very deep sequence conservation, which has a very distinct set of TFs that cooperatively bind to facilitate transcription (Long *et al.*, 2016). One of the most important examples of enhancer architecture underpinning this model, is the mammalian viral-inducible enhancer of interferon-β (INFβ) (Thanos and Maniatis, 1995).

The other widely accepted model is termed the "billboard" model. In this model, the distribution of TF binding motifs differs in number, orientation and order. This variability in TF motif organisation ultimately results, in comparison to the more stringent "enhanceosome" model, in less conserved enhancer sequences across species. However, the set of TFs does not change. Therefore, by relying on indirect TF cooperativity, in the "billboard" model a very specific set of TFs is deployed regardless of motif distribution (Fig. 1-3, right). It has been shown that these sites do contain variable binding motifs that in turn lead to suboptimal binding of the specific TFs. This feature may allow an increase in enhancer specificity and stabilizes the turnover of transcription factor motifs (Crocker *et al.*, 2015). As an example, when mapped to the consensus sequence, the GATA and ETS motifs in the *Otx-a* enhancer of *Ciona intestinalis*, show imperfect alignment. Interestingly, when these imperfections are optimised or the spacing is changed between the motifs, enhancer activity is increased and ectopic enhancer activity was observed due to a higher sensitivity for the TFs that bind to these domains (Farley *et al.*, 2015). While these theoretical models have been useful to guide functional experiments the emerging consensus in the field is that there is a continuum between the enhanceosome and the billboard models, consistent with the observed variation among motifs in terms of spacing and orientation within one regulatory element (Long *et al.*, 2016).

1.4.3 Conservation of enhancer elements

Sequence conservation of an enhancer implies that it is an important cog in the gene regulatory program. In more rigid and strict models of TF binding, any single nucleotide polymorphism or mutation might have dramatic effects on transcriptional regulation due to the lack of proper binding. Therefore, highly conserved enhancer sequences assure correct gene expression. While multiple studies have focused on such highly conserved enhancers, further work is required to establish if high sequence conservation is due to a distinct mechanistic requirement not needed at more evolutionarily flexible enhancers. The role of enhanceosomes could be to act as a genetic switch for correct gene expression upon cues from various signalling pathways (Spitz and Furlong, 2012). Other evidence for the importance of ultra-conserved sequences is the occurrence of clusters of such DNA sequences in close proximity to genes that encode transcription factors that are essential for development (Boffelli *et al.*, 2004; Dickel *et al.*, 2018; Letelier *et al.*, 2018). These clusters of enhancers are under such strict evolutionary pressure

that mutations in ultra-conserved regulatory elements can also have an effect on proper development and are also associated to disease. These effects can be very prominent, in the case of preaxial polydactyly for example, where transcription of the *Shh* gene is disrupted due to a mutation in its enhancer (Lettice *et al.*, 2003). Other effects can be very subtle and may not show any immediate impact on viability but can decrease fitness and, therefore, influence survival. The cluster of enhancers located close to the neuronal transcription factor *Arx* is such an example. It is one of the longest highly conserved sites in the genome and alteration in the enhancer sequence leads to neurological or growth abnormalities and thus brain defects (Dickel *et al.*, 2018).

Interestingly, ultra-conserved enhancers are still quite rare and more flexible enhancer sequences, as described in the billboard model, are much more common. Generally, enhancers that are functionally conserved among the animal kingdom often show very little sequence conservation (Wong *et al.*, 2020). Even *de novo* generation and the loss of regulatory elements is more frequent across evolution than highly conserved enhancer regions. It has been shown, that cardiac-specific regulatory elements from human and mouse with very little conservation in other animals, were able to drive heart specific expression in reporter assays performed in mouse embryos (Blow *et al.*, 2010; May *et al.*, 2012). An additional example is that poised enhancers identified in human embryonic stem cells (hESCs) were able to drive specific expression in various tissues at different stages during a reporter assay performed in zebrafish embryos, even when there was a lack of conservation of the identified enhancer sequences between the two species (Rada-Iglesias *et al.*, 2011).

A reason for the lack of sequence conservation but a maintenance of the initial function of the regulatory element could be the variability of transcription factor binding sites (TFBS). This compensation would then enable complete reshuffling of different motifs within one enhancer. One example for such an immense change in composition but maintained functional conservation is the *even-skipped stripe 2 enhancer* (*S2E*) in *Drosophila*, which still displays the same expression pattern of *even-skipped* in different species (Arnosti *et al.*, 1996; Ludwig *et al.*, 2000). Interestingly this is not only an effect of distantly related species, but such variability is also quite common for more closely related species. Therefore, functional conservation without sequence conservation can be found regularly and is a common event during evolution that enables a more flexible read out of enhancer motif grammar (Arnold *et al.*, 2014).

1.4.4 Additional effect of Active Enhancers

Gene regulatory networks are maintained through the modularity of enhancer elements in the genome, with each module regulating a spatio-temporal subset of the actual gene expression. Therefore, enhancers have to work together to maintain appropriate gene expression of the target gene. Essential developmental genes tend to be regulated by a cluster of modular enhancer elements that are very potent in their regulatory function, often termed super-enhancers (Hnisz *et al.*, 2013; Whyte *et al.*, 2013). There are several models that aim to explain how the regulatory network is established and how super-enhancers can act simultaneously without interfering with each other. Due to the nature of enhancers working in an autonomous, modular and additive manner, it is thought that regulatory elements are under higher selective pressure in comparison to non-regulatory sequences (Carroll, 2008). The existing data suggest that several enhancers can regulate the same locus in *cis* in different tissues and that enhancers in a cluster can be active all at the same time (Fig. 1-4) (Hay *et al.*, 2016; Maeda and Karch, 2011; Visel *et al.*, 2009). This combinatorial effect can differ depending on the discrete enhancers involved in the regulation of its target gene, resulting in variability of activity. Therefore, the actual gene expression pattern of a certain gene differs from the combined action of its regulating enhancer elements, and the deletion of modular enhancer elements in some cases can lead to a tissue-specific phenotype. It still remains to be elucidated on how common these non-additive regulatory networks are. Evidence in *Drosophila* suggests this type of regulation is fairly common and that around 37% of enhancers are being regulated by their own endogenous loci (Kvon *et al.*, 2014).

Since the discovery of the complex interactions of multiple enhancer elements in sea urchins (Yuh and Davidson, 1996; Yuh *et al.*, 1998), many other systems have been studied that show the multiplicative behaviour of enhancers (Stine *et al.*, 2011). Another important factor of the interplay between those regulatory elements is, beside the quantitative effect, the qualitive expression of target genes in a spatio-temporal manner. Multiple studies in *Drosophila* and mice demonstrate that the expression driven by a single enhancer does not always overlap with the expression pattern by all regulatory elements of a cluster are active all together. Thus, the combined action of multiple regulatory elements is necessary for appropriate spatio-temporal regulation of gene expression (Dib *et al.*, 2011; Dunipace *et al.*, 2011; Prazak *et al.*, 2010).

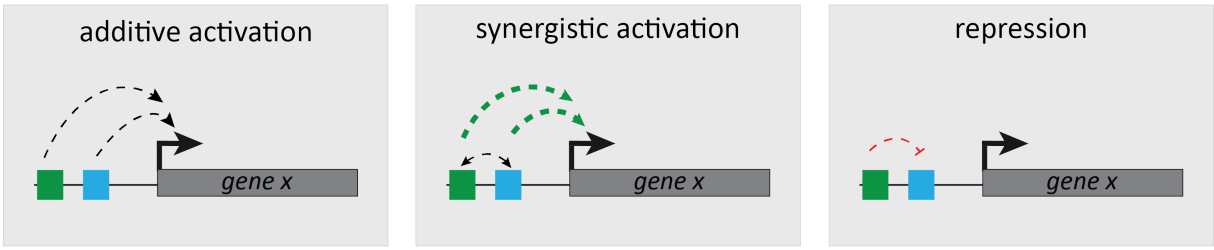


Fig. 1-4: Models of transcription activation via regulatory elements. Regulatory elements can act additively to drive gene expression of their target gene, work synergistically to enhance their function or repress each other.

Furthermore, regulatory unit interplay can also be repressive in nature to prevent ectopic expression of their targets (Fig. 1-4). It has been postulated that the regulatory elements themselves partially repress their neighbours due to the recruitment of long-range repressors (Dunipace *et al.*, 2011; Perry *et al.*, 2011). Examples for this include the correct gap-gene-expression patterns regulated by enhancers that are in close proximity (Perry *et al.*, 2011). In addition, it has been shown in mouse that the *Fgf8* expression is highly controlled due to repression mechanisms of regulatory elements (Marinić *et al.*, 2013).

Studies of the *Drosophila bithorax complex* (*BX-C*), which is important for the correct development of the anterior-posterior axis, show that a large sequence of around 300 kb contains regulatory elements that can be divided into nine subsegments. Each controls a subset of the *BX-C* homeotic genes to be appropriately expressed in each segment of the *Drosophil*a embryo (Maeda and Karch, 2011). When tested individually, not all enhancer elements in each subset will be actively mediating expression in the correct parasegment, but rather along the whole-body axis. Nonetheless, one or two in each domain will still drive the correct expression. These regulatory elements that show limited activity were called “initiator” elements due to their function in activating or silencing their neighbouring enhancers in the appropriate parasegmental pattern (Iampietro *et al.*, 2010; Mihaly *et al.*, 2006). In addition, it has been shown that initiator elements can also cause homeotic transformation when inserted into a different genomic context. Thus, the initiator element is able to regulate enhancer activity by itself. It remains to be determined if such a defined control mechanism exists in mammals. However, what has been described in mammals is that enhancers can also function to activate or repress activity at neighbouring regulatory elements. For example, an upstream enhancer element in close proximity to the *PU.1* locus in murine myeloid cells regulates gene expression

by an auto-regulatory loop with its downstream enhancer element through chromatin remodelling (Leddin *et al.*, 2011).

1.4.5 Enhancer Redundancy

When tackling the topic of enhancer activation, it is equally important to also consider redundant enhancers, *i.e*. those that fulfil the same, or similar, role in the regulatory network. These functionally redundant regulatory elements are termed "shadow enhancers" and are defined as secondary regulatory elements that are far from their target gene, with an activation mechanism that is overlapping with the primary enhancer of the gene (Hong *et al.*, 2008) to provide robustness in the network. However, having an overlapping activity is not true functional redundancy. Therefore, shadow enhancers may drive different spatio-temporal expression of the target gene and, thus, lead to subtle differences in the expression pattern. In addition, "shadow enhancers" are proposed to buffer gene expression from genetic changes by providing the capacity to cope with novel genetic variants (Frankel *et al.*, 2010). However, in stress conditions, these variants may trigger changes in gene expression (Frankel *et al.*, 2010). The precise mechanism by which enhancer elements target a given promoter remain to be elucidated. A number of models have been proposed including a competitive model where two individual enhancers compete to initiate transcription of the target gene and, thus, ensure a constant gene expression output (Long *et al.*, 2016). Strong redundant regulatory elements would then work in a sub-additively manner (the output of their combined activity is less than the predicted output), whereas weaker enhancers would function more additively due to a generally lower promoter competition. *in vivo* data on the regulation of *hunchback* and *snail* in *Drosophila* suggests that regulatory elements do function in a non-additive manner when competing for a shared promoter (Bothma *et al.*, 2015). With this mechanism, the robustness of certain enhancer-gene networks is due to the fact that the consequence of a novel disruptive variant is buffered by the action of the remaining functional enhancers. The remaining regulatory elements would subsequently adapt to the new requirements for transcriptional control and increase enhancer-promoter contacts accordingly. It is also possible that the competition model might explain the clear patterning and expression boundaries by recruiting long-range repressors (Long *et al.*, 2016). Such case was reported in the *knirps* and *Krüppel* loci,

where they act in an additive manner in one region but acquire repressive features in posterior regions to clearly restrict gene expression to a certain domain (El-Sherif and Levine, 2016).

1.5 Assessing Enhancer Activity

Advances in understanding enhancer biology has taught us a lot about the role of regulatory elements in gene expression and, more broadly, about evolution. Nevertheless, a number of gaps in our understanding of enhancer function remain. For example, while many regulatory elements have been defined, it remains unclear how each enhancer grammar of a specific enhancer influences the function of the CRE. A number of approaches have been made to synthetically generate enhancer elements to examine enhancer grammar, however, they have not been designed to mimic endogenous orientation or spacing of TFBs and also in regard to cooperative and individual binding strategies (Smith *et al.*, 2013). There are three major points that prove to be challenging when working with CREs. The first being putative enhancers are embedded within the genome. Thus, the sheer amount of DNA sequence makes it difficult to identify and analyse the functional ones (Long *et al.*, 2016). This is further illustrated by findings that synthetic enhancers with random sequences were able to function as regulatory elements (Smith *et al.*, 2013). Furthermore, the high number of enhancers that have similar or the same function (*i.e.* functional redundancy) make *in vivo* analysis technically challenging. Lastly, the environment or the context influences heavily the transcriptional output of the enhancer element, making the enhancer highly dependent to factors like chromatin structure, enhancer grammar, activity and spatio-temporal expression of target genes (Long *et al.*, 2016).

In order to assess enhancer activity and the influence on chromatin architecture on them, recent chromosomal conformation studies (*e.g*. Hi-C) have shown that most genomes of multicellular organism are organised in topologically associating domains (TADs) that enable proper transcriptional regulation over long distances. Interestingly, previous studies have also included high-resolution imaging to address the complex network of enhancer-promoter interactions providing a visual proof of spatial proximity between chromatin regions (Fabre *et al.*, 2015; Williamson *et al.*, 2016). To address these limitations, a number of technical developments have been utilised to dissect enhancer activity which are reviewed in detail in the following sections.

1.5.1 Genomic Landscapes - DNA accessibility and Histone modification

Next generation Sequencing (NGS) advances underlie investigations into the genetic control of development and disease. In the context of enhancer biology, NGS has been combined with Chromatin immunoprecipitation (ChIP), to shed light on many processes that involve protein-DNA interactions, including TF binding and histone modifications. Several histone marks are important to define enhancers and assess their state of activity: histone H3 Lys4 monomethylation (H3K4me1) labels *cis*-regulatory elements (*i.e.* enhancers), H3 Lys4 trimethylation (H3K4me3) is found at promoters, and H3 Lys27 acetylation (H3K27ac) marks active regulatory elements such as enhancers and promoters (Calo and Wysocka, 2013). Another important feature of regulatory elements is the structural accessibility for TF to bind to DNA, which is also widely used to predict enhancers within the genome.

The way DNA is compacted in the genome is in specific units called nucleosomes that consist of octamers of histones that wrap the DNA. When compacted, the DNA confirmation prevents TF and other proteins from binding and as a result the tightly packed nucleosome prevents transcription (Thåström *et al.*, 1999). However, the inhibition of enhancers by structural changes highly depends on the DNA sequence which is being packaged. Enhancers, in comparison to other genetic sequences, seem to consist of a specific sequence that have higher affinity of nucleosome positioning, which then in turn generates a "default off" state that needs to be overcome in order to activate transcription by TFs (Barozzi *et al.*, 2014). Transcriptional initiation requires enhancers to be activated by pioneer TFs that bind to the nucleosome and shift it to an open state. To asses accessible chromatin, a previous study has shown that the microccoal nuclease (MNase) is able to access DNA where the nucleosome has a lower binding affinity, highlighting open chromatin and a possible regulatory function (Mieczkowski *et al.*, 2016). Additional technologies which also help to investigate open chromatin include DNase I hypersensitive site sequencing (DNase-Seq) (Boyle *et al.*, 2008) and the more sensitive Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq) (Buenrostro *et al.*, 2013). Although DNA accessibility is a widely used tool to predict enhancer sequences, one has to keep in mind that also other regulatory elements such as insulators and promoters are also accessible regions in the chromatin (Boyle *et al.*, 2008). It is also important to note that open chromatin at enhancer elements does not necessarily reflect the state of activity of the

enhancer since inactive enhancers can also be accessible. Thus, it is essential to combine tools that identify open chromatin with other techniques to predict active *cis*-regulatory elements.

Post-translational modifications (PTMs) of histones are an important feature of chromatin remodelling and a widely used tool to identify enhancers and their state of activity. Important transcriptional regulators like P300/CBP and the mammalian H3K4 methyltransferases Mll3/4 catalyse these histone marks. PTMs can function in many ways, including the destabilisation of nucleosome stability (Neumann *et al.*, 2009; Pradeepa *et al.*, 2016) to promote opening of the DNA and the recruitment of proteins (Heintzman *et al.*, 2009; Shlyueva *et al.*, 2014). The histone mark H3K4me1 can be found at the majority of enhancers independently of their state of activity (Arnold *et al.*, 2013; Creyghton *et al.*, 2010). Although, H3K4me1 labels enhancers at any state, it is not essential for the activity of the regulatory element. As shown in flies and mammals, interfering with the function of the H3K4 methyltransferase in both systems had little effect on the transcriptional output of genes (Dorighi *et al.*, 2017; Rickels *et al.*, 2017). For example, a hyperactive mutant of the fly ortholog Trithorax-related (*Trr*) that showed increased levels of H3K4me1 had almost no significant impact on transcriptional regulation (Rickels *et al.*, 2017). However, in *Trr* knockout mice and flies, transcriptional regulation was heavily perturbed, resulting in non-viable offspring. This suggests while the protein itself is indeed required for proper gene regulation; its catalytic activity is not. The same approach was applied to UTX, an enzyme that recruits the acetyltransferase P300/CBP (Wang *et al.*, 2017). Interestingly, while the activity of P300/CBP is essential for transcriptional regulation, as shown in UTX deficient murine embryonic stem cells, heterochromatin formation was impaired altering transcriptional silencing (Santenard *et al.*, 2010), suggesting H3K2ac itself is not essential for the activation of regulatory elements (Pengelly *et al.*, 2013). Although used as a specific marker for active enhancers, not all active enhancers in *Drosophila* (Bonn *et al.*, 2012) and mice (Pradeepa *et al.*, 2016) have H3K27ac modifications. This could suggest that P300/CBP is not only targeting H3 Lys27 but also other residues and proteins that are not associated directly with histone modification (Edmunds and Mahadevan, 2004; Pradeepa *et al.*, 2016; Roe *et al.*, 2015).

1.5.2 Assessing enhancer transcripts and their function

As previously described, enhancers are found in exons and act as duons, displaying a dual function as gene transcript and enhancer (Stergachis *et al.*, 2013). In addition, there are also another family of enhancers, termed enhancer RNAs (eRNAs), that regulate gene expression after being transcribed into RNA. This has been shown in mammals, flies and nematodes via genome-wide analysis of transcriptional regulation (Bonn *et al.*, 2012; Chen *et al.*, 2013; Djebali *et al.*, 2012; Tuan *et al.*, 1992). Furthermore, enhancers can be transcribed bidirectionally (Andersson *et al.*, 2014) or unidirectionally (Koch *et al.*, 2011). These enhancer RNAs (eRNA) are subject to the same RNA processing as protein-coding transcripts. For example, some contain a polyA-tail to improve stability. However, the majority of eRNAs are short lived and are rapidly degraded (Schwalb *et al.*, 2016). Therefore, identification of eRNA is difficult due to the fast turnover rate of these eRNAs. In order to overcome these difficulties a number of new techniques have been developed such as inhibiting the formation of exosomes which the cell uses to degrade RNAs lacking a polyA tail. Alternative approaches include capturing premature RNA with techniques such as global run-on sequencing (GRO-seq) (Core *et al.*, 2008), START-seq (Scruggs *et al.*, 2015), native elongation sequencing (NET-seq) (Churchman and Weissman, 2011) and transient transcriptome sequencing (TT-seq) (Schwalb *et al.*, 2016). While many studies correlate eRNA expression with a functional output, not all transcribed regulatory elements have an enhancer function. It has been shown in Cap Analysis of Gene expression (CAGE) experiments, that only 70% of identified eRNAs actually functioned as *cis*-regulatory elements (Catarino and Stark, 2018). Consistent with data from enhancers, other regulatory elements are transcribed in a bidirectional manner such as insulators and other sequences in regions of open chromatin suggesting RNA from regulatory sequences may have important roles in gene regulation (Melgar *et al.*, 2011; Young *et al.*, 2017). If and when a regulatory element is transcribed into eRNA, depends on the sequence and location of the enhancer in the genome. For example, one prerequisite for eRNA expression is the existence of core promoter elements in the vicinity of the enhancer (Andersson *et al.*, 2014; Mikhaylichenko *et al.*, 2018). However, how this complex mechanism is regulated in a spatio-temporal manner in an organism remains to be elucidated. In addition, due to the co-occurrence and distinct function of promoter elements and enhancer elements, it is essential to carefully define the differences between the two regulatory elements in order to correctly predict enhancer activity in the genome (Arnold *et al.*, 2017; Catarino *et al.*, 2017). Therefore, a key definition

underpinning the enhancers biology is that they constitute DNA sequences that are capable of regulating and enhancing transcriptional output of their target genes in *cis* (Banerji *et al.*, 1981). It follows, therefore, that enhancer sequences must demonstrate the capacity to drive expression of a reporter gene(s) in a standardized manner. This framework for understanding enhancer biology gives rise to the question: is an enhancer necessary and/or sufficient for target gene expression?

1.5.3 Enhancer sufficiency

In order to assess the function of a putative regulatory element, it is important to test whether it is "sufficient" to drive expression of a target gene. This is generally done by taking the enhancer out of its endogenous environment and to determine if it functions in a heterogenous reporter assay. Such assays have a simple read out and remove all endogenous factors that could interfere with the function of the enhancer, which in turn, might lead to a cleaner comparison between different regulatory sequences. The regulatory output of the candidate regulatory element is assayed using a number of techniques including quantification of reporter gene expression directly via mRNA levels or indirect assays of the protein encoded by the reporter gene, usually measuring either enzymatic activity or fluorescence (Catarino and Stark, 2018). Deploying this classical approach the very first enhancer element was defined in 1981 (Banerji *et al.*, 1981), which led to the definition of enhancers as a DNA sequence that has a regulatory function and can trigger transcription of a target gene irrespective of distance, position and orientation (Banerji *et al.*, 1981; Catarino and Stark, 2018). Reporter assays continued to be used to identify regulatory elements, yet a disadvantage is that they are time consuming and are not amenable to high throughput screens. However, the emergence of next generation sequencing has facilitated the rapid identification and assessment of enhancers. For example, massively parallel reporter assays (MPRAs) can be used to identify putative candidates and their gene enhancing activity via genetic barcodes rather than reporter genes (Catarino and Stark, 2018). The abundance of these barcodes allows the determination of the activity of the regulatory element associated to this genetic marker. Therefore, this method allows a massive throughput and simultaneous testing of many different sequences (Kwasnieski *et al.*, 2012; Patwardhan *et al.*, 2012). Another example for a large-scale method to test enhancers is self-transcribing active regulatory region sequencing (STARR-seq) in *Drosophila* or mammalian

cells. The enhancers are being used in a positive feedback loop, where functional regulatory elements drive their own expression. This is achieved by fusing the regulatory elements to a transcription start site (TSS) that is upstream of the sequence. The transcript can then be used as its own barcode, due to its unique sequence. Therefore, millions of candidates can be tested simultaneously (Arnold *et al.*, 2013; Muerdter *et al.*, 2018). Such large-scale testing also enables interrogation of multiple enhancers at the DNA sequence level. Importantly this permits elucidation of the enhancer grammar and the composition of TF binding motifs found within regulatory elements driving common gene expression patterns. In turn, improved understanding of enhancer sequence composition enables the prediction of the susceptibility of regulatory elements to changes in the sequence, *e.g*. variants and SNPs that might alter the function. Thus, modern high-throughput enables a more detailed analysis of gene regulatory networks and their different components (Catarino and Stark, 2018).

In an *in vivo* situation, many enhancer functional assays are performed by integrating the putative enhancer sequence into the genome of the host to assess the spatio-temporal expression mediated by the enhancer. Classical approaches to bring foreign DNA into a host organism were either microinjection (Bessa *et al.*, 2009; Kothary *et al.*, 1989) or integration by viral vectors such as retroviruses and lentiviruses (Murtha *et al.*, 2014). However, when testing the sufficiency of enhancers to drive expression of a reporter gene, one has to keep in mind that random integration of the target sequence into the host DNA might have a major impact on the subsequent regulation. It has been shown, that genome integration is not entirely random but is usually favoured towards open chromatin and accessible regulatory regions (Bushman, 2003; Lander *et al.*, 2001). This bias can heavily influence the outcome of the enhancer assay. It can range from endogenous regulatory elements that act on the reporter genes, which then in turn lead to discrepancies in gene expression, to silencing of the inserted regulatory element. Nevertheless, many studies have been performed to identify and validate enhancer function *in vivo* during development in flies (Kvon *et al.*, 2014), zebrafish (Bessa *et al.*, 2009) and mice (Dickel *et al.*, 2016). In this thesis, we make use of the Japanese rice fish or medaka (*Oryzias latipes*) to assess cardiovascular specific CREs *in vivo*. Already established in the ENCODE project in 2012, medaka has been shown to be an effective model to study human regulatory elements and is well suited for genetic engineering and *in vivo* imaging (ENCODE Project Consortium, 2012; Letelier *et al.*, 2018; Lischik *et al.*, 2019; Stolper *et al.*, 2019)**.**

In order to fully assess the function of a regulatory element, testing for enhancer activity is not enough. It is important to determine whether a regulatory element is also capable of affecting transcription of its target gene in a negatively or positively. It has been shown that alterations of enhancer elements can lead to deregulation of gene expression and which can manifest as the pathological effects underlying diseases such as diabetes or cancer (Mansour *et al.*, 2014; Pasquali *et al.*, 2014; Pomerantz *et al.*, 2009). Indeed, GWASs have linked a large portion of disease-associated variants to noncoding regulatory sequences (Degner *et al.*, 2012; Karczewski *et al.*, 2013). Therefore, testing the function and influence of regulatory elements in the pathogenesis of disease requires precise genome editing, to alter specific base-pairs within the regulatory elements (discussed further in Enhanceropathies section below).

1.5.4 Enhancer necessity

Early approaches to perturb gene expression by altering enhancer sequences used insertional mutagenesis by transposable elements. Transposons such as "*sleeping beauty*" or "*Tol*2" randomly integrate and lead to genomic alterations (Kawakami and Shima, 1999; Kawakami *et al.*, 2017). For example, a key regulatory region of *decapentaplegic* (*dpp*) in *Drosophila* in 1990 (St Johnston *et al.*, 1990) was identified by insertional mutagenesis. This discovery enabled the dissection of a cluster of enhancers that govern the pattern of *dpp* expression (Blackman *et al.*, 1991). While the use of transposons leads to random alterations, targeted, precise genome editing is now widely used. A breakthrough in regard to genome editing was the discovery of the CRISPR/Cas9 system, a two-component mechanism that involves the Cas9 enzyme and special guide RNAs (gRNAs) (Hsu *et al*., 2014). The enzyme introduces DNA double strand breaks to specific sequences targeted to and defined by the sequence-complementary gRNAs, which then in turn allows genome editing such as insertions or deletions (Jinek *et al.*, 2012).

An excellent example of enhancer biology is provided by the dissection of the regulatory region of the *sonic hedgehog* (*shh*) gene in vertebrates, which leads to an absence of limbs when depleted (Sagai *et al.*, 2005). It has been a target of intense research and many studies have shown the importance of that region during evolution (Kvon *et al*., 2016; Letelier *et al*., 2018; Williamson *et al*., 2016). For example in mouse, when replaced with the orthologous enhancer sequence of *shh* of snakes, murine limb development is heavily impaired and leads to "serpentized" limbless mice (Kvon *et al.*, 2016). In addition, a recent study in medaka (*Oryzias*

latipes) has shown that the deletion of a single *shh* enhancer termed *ZRS* abolishes gene expression in the fins leading, in turn, to truncated pectoral and dorsal fin formation (Letelier *et al.*, 2018). Further examples of a disrupted CRE altering target gene expression and downstream biological processes include: i) the *Sox2* locus in embryonic stem cell differentiation (Li *et al.*, 2014; Zhou *et al.*, 2014), ii) the locus of the neuronal transcription factor *Arx* (Dickel *et al.*, 2018), iii) the locus of *Myc* in blood cells (Bahr *et al.*, 2018) and iv) the important cardiac genes in humans *KCNH2* (van den Boogaard *et al.*, 2019) and *TBX5* (Cunningham *et al.*, 2018). Taken together these studies provide ample evidence that CRISPR/Cas9 mediated genome editing provides a powerful approach to investigate enhancer biology. Many *in vivo* and *in vitro* approaches are now possible and, therefore, it has become increasingly easier to assess the importance of regulatory elements in development. Furthermore, in the disease context, it is now possible to functionally validate disease associated SNPs in non-coding regulatory sequences identified by GWAS.

1.6 Enhanceropathies

Alterations like mutations or single nucleotide polymorphisms in CREs can lead to disease (Villar *et al.*, 2020). Importantly, the notion underlying this genotype-phenotype correlation is that genetic variation in non-coding sequences are major drivers of phenotypic variation and this is supported by the segregation of functional SNPs with human disease (French and Edwards, 2020; Ward and Kellis, 2012). These groups of conditions that are caused by such deregulation in enhancer elements were termed enhanceropathies (Smith and Shilatifard, 2014).

High genome coverage arrays used in GWAS has led to an increasing number of risk variants linked to gene and tissue specific enhancer elements. As a result, it is estimated that a far greater number of SNPs impact *cis*-regulatory elements than protein-coding sequences. Two major themes have arisen from disease specific GWAS studies that identify CREs in linkage disequilibrium with the disease. The first is that sequences, which show histone marks that, indicate active enhancers (H3K4me1 & H3K27ac), are also highly enriched for genetic risk variants. In comparison, potentially pathogenic variants are less common in other *cis*-regulatory elements like promoters and insulators (Ernst *et al.*, 2011). The second theme is, that risk variants in enhancers which were associated to a specific disease are more likely to be enriched

in disease-relevant cell types (Ernst *et al.*, 2011). For instance in diabetes or colon cancer, risk variants were enriched in pancreatic islet enhancers or active enhancers in colon cancer cells respectively (Akhtar-Zaidi *et al.*, 2012; Pasquali *et al.*, 2014).

1.6.1 Variants in enhancer sequences

One of the earliest examples for disease causative variants that have been linked to altered gene regulation, are translocations in the *β-globin* gene locus that lead to thalassaemia. Although no mutations were detected in the protein coding sequence of the globin genes, the disease arose due to disruption of the interaction between the gene and enhancer element (Kleinjan and Lettice, 2008).

Whole-genome sequencing technology permits a deeper analysis of the effect of variants in CREs and an improved understanding of how variants manifest in human disease. Critically, mutations in regulatory elements can result in a broad spectrum of traits and display variable phenotypic penetrance ranging from Mendelian to small effects in complex diseases (Dickel *et al.*, 2016, 2018; French and Edwards, 2020; Kvon *et al.*, 2020; Pennacchio *et al.*, 2013). As a result of this complexity, it remains challenging to functionally characterise and validate putative disease-causative variants in enhancer sequences. Nevertheless, recent work has linked several genetic mutations in regulatory elements to certain human diseases that are segregated in a Mendelian manner (Villar *et al.*, 2020). Among one of the first identified examples is a mutation that causes pre-axial polydactyly in families. This variant is found in an enhancer that regulates the expression of the human *sonic hedgehog gene* (*SHH*). In addition, the same phenotype can be observed in patients that suffer from polydactyly due to a chromosomal translocation that results in a transfer of the enhancer to a new location which is too far removed to allow the enhancer to drive *SHH* expression (Lettice *et al.*, 2003). However, with respect to genotype-phenotype correlation, mutations in enhancers differ greatly from variants in protein-coding sequences. While mutation in coding sequences may affect a wider range of downstream effects, for example the processing and stability of RNA, protein folding, localisation in the cell and receptor-ligand interactions, non-coding sequences are mainly limited to *cis* effects during transcription (Sakabe *et al.*, 2012).

Due to the modularity of *cis*-regulatory elements, mutations in enhancer regions have different spatio-temporal impacts on the expression of the target gene. For example, mutations in the

protein-coding sequence of the human *TBX5* gene leads to Holt-Oram syndrome, which is characterised by cardiac and limb abnormalities. *TBX5* is a key player in early heart and forelimb formation and mutation of *TBX5* leads to impaired cardiac and forelimb development. However, specific CREs of *TBX5* are independently responsible for regulating limb and cardiac expression. Thus, mutations in the cardio-specific subset of enhancers will only lead to cardiac malformations while forelimb development is unaffected (Smemo *et al.*, 2012).

In contrast to loss-of-function mutations, which have been associated to the majority of identified enhancer variants, gain-of-function mutations may also occur and lead to a specific phenotype or disease. In Melanesians, the indigenous inhabitants of Melanesia, a variant has been identified that leads to the creation of a promoter-like functional enhancer sequence. This gain-of-function regulatory sequence is interfering with normal activation that in turn elevates expression of α-globin and leads to individuals suffering from α-thalassaemia (De Gobbi *et al.*, 2006).

The examples above demonstrate that regulatory variants contribute to the genetic basis of human diseases and malformations. However, it still remains challenging to identify disease-causing variants and assess their impact on a broader level. Therefore, it is essential to use new screening technologies and optimise them for a high-throughput manner. In addition, it is crucial to not only focus on mutations in enhancer sequences, but also to consider other types of *cis*-regulatory elements as well that can harbour variants, such as repressors, insulators and matrix attachment regions (Pennacchio *et al.*, 2013).

1.7 Regulatory elements in cardiac disease

Congenital heart defects (CHD) are complex diseases with a complex genetic aetiology. Progress in the molecular understanding of familial CHD have defined genetic factors as an important contributors to CHD pathogenesis (Richards and Garg, 2010). Since the majority of CHD associated SNPs lie in non-coding regions, such as enhancers or promoters, it's possible that the SNPs alter regulatory sequences such that TF binding is impaired and target gene expression is disrupted. In the context of heart development, highly conserved transcription factors are known to be critical regulators of heart development and perturbation of these TFs

results in CHD (McCulley and Black, 2012). However, the contribution of mutations within *cis*-regulatory elements to CHD remains relatively poorly defined.

A key foundational study was the identification of a regulatory variation in a *cis*-regulatory element of *TBX5*, a highly conserved transcription factor within the cardiac gene regulatory network, that contributes to heart defects in the Holt-Oram syndrome (heart-hand syndrome) (Smemo *et al.*, 2012). In this study three enhancers of *TBX5* were identified that are able to drive expression of the gene in the developing heart. Using mouse and zebrafish, *in vivo* modelling of a homozygous single-base-pair mutation in one enhancer located 90kb downstream of the gene revealed altered cardiac expression of *TBX5* which led to cardiac defects (Smemo *et al.*, 2012). Interestingly, another example of a CHD causative enhancer mutation reveals that it is not always the proximity to a gene that is an indication for the target of the regulatory element. This study found a common genetic variant in a regulatory element located in an intron of *SCN10A* resulted in the cardiac rhythm disorder Brugada syndrome. Through high-resolution 4C-seq analysis it was revealed that a cardiac enhancer in *SCN10A* is regulating the expression of the sodium channel-encoding gene *SCN5A* by enhancer-promoter interactions. Ultimately, variants in this enhancer sequence lead to reduced transcription levels of *SCN5A* and thus result in cardiac defects (van den Boogaard *et al.*, 2014; Man *et al.*, 2019). Another study focused on the essential cardiac transcription factor *HAND2*, for which several regulatory elements already have been identified (Charité *et al.*, 2001; McFadden *et al.*, 2000) within a super-enhancer displaying active histone marks (H3K27ac). In this study, a long non-coding RNA called *upperhand* (*Uph*) was shown to be essential to maintain the super-enhancer properties of the locus and transcription of *Hand2* via RNA polymerase II in mouse. Interfering with *Uph* transcription, results in depletion of *Hand2* expression leading to ventricular hypoplasia and embryonic lethality (Anderson *et al.*, 2016). In a *tour de force* in assessing heart enhancers in the mouse, more than 80,000 putative cardiac *cis*-regulatory elements were identified, highlighting the complex interplay of enhancers during both heart development and cardiac disease such as hypertrophic cardiomyopathy (Dickel *et al.*, 2016). Two human enhancer sequences regulating *MYL2* and *MYH7* (cardiac myosin genes) were functionally tested using the orthologs in mouse. In an enhancer assay, both sequences were sufficient enough to drive strong expression of the reporter construct in the murine heart due to high

functional conservation in mammals. Although no strong phenotypes were observed when each enhancer sequence was deleted, expression levels of *Myl2* were decreased in the respective enhancer knockout line. Interestingly, when looking for cardiac dysfunction, cases of cardiomyocyte disarray and karyomegaly were observed, suggesting an important role for these enhancers in correct transcriptional output and proper homeostasis of the heart morphology (Dickel *et al*., 2016). Based on this high throughput identification of cardiac REs more studies are emerging which are beginning to define the cardiac CREs important in development and disease. These studies identified and characterized how essential REs are for the phenotypic robustness and the regulation of correct expression patterns either working as a single enhancer or in a redundancy network of multiple regulatory elements. (Arvanitis *et al.*, 2020; van den Boogaard *et al.*, 2019; Galang *et al.*, 2020; Montefiori *et al.*, 2018; Villar *et al.*, 2020; Waymack *et al.*, 2020). Despite this, there remains significant work to functionally assess the 80,000 regulatory regions suggested to coordinate and control heart development and function. The work of this thesis aims to contribute to these efforts to precisely define genomic elements that drive heart formation and form a platform to understand how genetic variation in REs results in cardiovascular disease.

1.8 Aims

In order to further understand the underlying genetic mechanism behind congenital heart disease (CHD), this thesis focuses on dissecting the role of enhancers involved in cardiovascular development and disease. We hypothesise that genetic alterations in cardiac *cis*-regulatory elements can cause 'enhanceropathies' resulting in CHD by altering and interfering with the transcriptional output of the target gene. In this thesis, we aim to:

1. Identify *cis*-regulatory elements that are involved in heart development and disease
2. Functionally validate putative cardiac enhancers *in vivo* using the fish medaka
3. Model CHD *in vitro* using cardiomyocytes derived from human embryonic stem cells.

2 Chapter 2: Material and Methods

2.1 Bioinformatic mining and enhancer prediction

The first set of putative cardiac *cis*-regulatory elements were obtained through publicly available ChIP-seq datasets of histone modification marks retrieved from the Human Roadmap Epigenome Project (http://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/) (Kundaje *et al.*, 2015). Additional datasets consisted of histone mark ChIP-seq (H3Kme1, H3K27ac, H3K4me3, obtained from adult heart tissue from marmoset) and transcription factor ChIP-seq of NKX2-5 obtained from cardiomyocytes derived from human embryonic stem cells (Anderson *et al.*, 2018). STAR (Dobin *et al.*, 2013) and MACS2 (Zhang *et al.*, 2008) was used to map reads to the genome. The datasets were collected and uploaded to the UCSC genome browser (https://genome.ucsc.edu/)(Rosenbloom *et al.*, 2015). Human cardiac specific enhancers were mapped to the marmoset genome using liftOver. The GREAT tool (McLean *et al.*, 2010) was used to link enhancer candidates to target genes based on proximity rules. This occurs via predicted and validated enhancer-gene interactions. In order to assign these regions to congenital heart disease (CHD), the NHGRI-EBI GWAS Catalog (Welter *et al.*, 2014) was used to link cardiac-specific regulatory elements to SNPs that have been associated to congenital defects (such as atrial fibrillation, dilated and hypertrophic cardiomyopathy, long QT Syndrome and ventricular fibrillation).

The second set of highly conserved putative cardiac specific regulatory elements were also obtained by filtering sequences bound by NKX2-5. These were then subsequently filtered for ultra-conserved sequences via MultiZ alignments and intersected with histone modification marks datasets of the Human Roadmap Epigenome Project (Kundaje *et al.*, 2015). The GREAT tool (McLean *et al.*, 2010) was used to link enhancer candidates to target genes based on proximity rules. This second pipeline analysis was performed by Mirana Ramialsion.

2.2 Routine methods *in vivo*

2.2.1 Fish maintenance and egg collection

All fish lines were maintained under standard recirculating aquaculture conditions. Briefly, day-night cycles were set to 14 h of light and 10h of darkness and the fish were fed three times per day. The fish facility of the Centre for Organismal Studies in Germany is under the supervision of the local representative of the Animal Welfare Agency and all experiments on Medaka (*Oryzias latipes*) were performed according to European Union animal welfare guidelines and national animal welfare standards (Tierschutzgesetz §11, Abs. 1, Nr.1, husbandry permit number 35-9185.64/BH Wittbrodt and 7 mutagenesis number G-206/09). For this study and for the first time, we established medaka as a new model organism in Australia and obtained all legal requirements for a breeding colony to conduct *in vivo* experiments on these fish (MARP-2016-055-BC). After obtaining all the ethics documentation we are currently housing 8 transgenic and 2 mutant lines some of which were generated in Australia at the Australian Regenerative Medicine Institute (ARMI).

In the morning, the eggs were separated from the abdomen of the female using a hook. They were subsequently placed in petri dishes containing 1x embryo rearing medium (ERM). Separation of the embryos and removal of egg filaments occurred by gently rolling the eggs or using forceps. Embryos were developed in plates stored at 28°C. Unfertilized eggs or dead embryos were removed. When necessary, sedation of embryos and hatchlings was performed via tricaine.

2.2.2 Mating and microinjection

Synchronized egg production was achieved by separating medaka males and females the day before injection. Prior to injection, males re-joined the females. After mating, eggs were collected and placed on ice to slow down the development of the embryos. The eggs were aligned in injection moulds (1.5% agarose in water) and covered with ice-cold 1x ERM. Injections were performed at the one-cell stage of the zygote, prior to duplication of the genome to achieve high efficiency in genetic modification. The injected volume consisted of approximately 10% of the size of the zygote.

2.3 Routine methods *in vitro*

2.3.1 Culturing of human pluripotent stem cells

For this study, all cell lines (Table 2-1) were generated using human embryonic stem cells (HES3) derived *NKX2-5*$^{GFP/W}$ (Elliott *et al.*, 2011). All cultures were kept at 37^{o}C in 5% carbon dioxide and 20% oxygen. The stem cells were passaged enzymatically with TrypLE^{TM} Express Enzyme (Gibco). The cells were split in different ratios depending on the application used. HES3 cells were co-cultured with irradiated mouse embryonic fibroblasts (MEFs) at various seeding densities to support and stabilize the growth of the stem cells. HES3 cells were cultured in DMEM/F12 media, with20% (v/v) knockout serum replacement, 10mM non-essential amino acids, 2mM L-glutamine, penicillin/streptomycin, 50mM 2β- mercaptoethanol and 10ng ml^{-1} bFGF (all culture media and supplements were purchased from Life Technologies).

Table 2-1: Human pluripotent stem cell lines used in this thesis

Cell lines	Background	Introduced deletion (GRCh37/hg19)
H3 Nkx2.5 GFP	H3, human, female	none
enCAV1	H3 NKX2.5 GFP, human, female	chr7:116185687-116186944
enCAV1	IPSC, PB001.1, human, male	chr7:116185687-116186944
enC5ORF38	H3 NKX2.5 GFP, human, female	chr5:3091055-3093392
enFLT1	H3 NKX2.5 GFP, human, female	chr13:28982472-28983114
enKCNH2	H3 NKX2.5 GFP, human, female	chr7:150665090-150667258
enPPARGC1A	H3 NKX2.5 GFP, human, female	chr4:23819438-23821139
enPPARGC1A	IPSC, PB001.1, human, male	chr4:23819438-23821139
enSMG6	H3 NKX2.5 GFP, human, female	chr17:2140769-2142644
mutFLT1_2	H3 NKX2.5 GFP, human, female	chr13:29068751-29069297
mutFLT1_Cas	H3 NKX2.5 GFP, human, female	unknown
mutIRX1	H3 NKX2.5 GFP, human, female	chr5:3595983-3596623

2.3.2 Polymerase Chain Reaction (PCR)

For the *in vivo* enhancer assay, primers were designed to contain a restriction enzyme overhang (AscI and SpeI (NEB)) with a non-palindromic sequence for the enzyme to attach. All standard

amplification and screening PCR reactions were performed with the Q5 High-Fidelity DNA Polymerase according to manufacturer's protocols (New England Biolabs). Primers used for PCR are listed in Table 2-2 and Table 2-3. Annealing temperature and elongation time were adjusted to the specific properties of the primers and DNA template.

Table 2-2: Oligos for *in vivo* enhancer transgenesis assay

Name	Sequence
PPARGC1A_for	AATTTGGCGCGCCTCCGCTACCCACAACTTTCC
PPARGC1A_rev	AATTTACTAGTTGCTTGGCGACTCTTACTGTC
KCNH2_for	AATTTGGCGCGCCAACAAGGACAAAAGACGGCG
KCNH2_rev	AATTTACTAGTGAGCTGTCTGAGATGGGAGC
FLT1_for	AATTTGGCGCGCCTTAGGGGGAGGGGAATGTGC
FLT1_rev	AATTTACTAGTCCTCCCTGCCATTGTACTTGG
SMG6_for	AATTTGGCGCGCCACTGACTGTTCCGCCTAACG
SMG6_rev	AATTTACTAGTTCTTTTAACTCTGTGCCCTCC
C5ORF38_for	AATTTGGCGCGCCGTGCTGCACCCATTAACTCG
C5ORF38_rev	AATTTACTAGTTTGTGCTGCTTCTCTCCACC
C5ORF38_1_for	AATTTGGCGCGCCTTTTCTTATCTTCACCTCTGAAAGC
C5ORF38_1_rev	AATTTACTAGTAGCGGCACAATCCTTTTATACC
C5ORF38_2_for	AATTTGGCGCGCCCAAGGTTGGCTTTGGTGTCC
C5ORF38_2_rev	AATTTACTAGTTCCCCCAATGGTAGAAACACC
CAV1_for	AATTTGGCGCGCCCAGTGCACAGGACAGTACCC
CAV1_rev	AATTTACTAGTTGTAGCTTTAGGCCATCCAGC
PAPPA2_for	AATTTGGCGCGCCTGCGAGTCATCTATTGAGTGG
PAPPA2_rev	AATTTACTAGTTGATGCCATGTGCCTCTACC
CXCL12_for	AATTTGGCGCGCCCAACCAAGGTGTGTGTGCC
CXCL12_rev	AATTTACTAGTCTCCCCAGTGTTTTCCCTGG
BUD13_for	AATTTGGCGCGCCACAGGCCCTGTTTACTAAGC
BUD13_rev	AATTTACTAGTACCCCCTCCAATTTCTCTGC
KCNJ2_1_for	AATTTGGCGCGCCCTGTAATCCCAGCTACCCCG
KCNJ2_1_rev	AATTTACTAGTCAGAGAATCCCTCGAACCCG

KCNJ2_2_for	AATTTGGCGCGCCCATGGCAACCAAAACCAGGG
KCNJ2_2_rev	AATTTACTAGTTCTTGCCTATCTACTTCATGCC
CXCR4_for	AATTTGGCGCGCCACCAGTCACAAACTTGGAAGC
CXCR4_rev	AATTTACTAGTTTTCTCCTCTTTCCTGGCCC
TFGBR2_for	AATTTGGCGCGCCTGGCCAAAGTGAAGTCTCTCC
TFGBR2_rev	AATTTACTAGTGAATTCCAGCAACCAAGAGGG
MGAT4D_for	AATTTGGCGCGCCGGGCAAAATAGTCCTTGCTTTG
MGAT4D_rev	AATTTACTAGTTGTGAAGTTTTTCAGCAGCATC
SLC8A1_for	AATTTGGCGCGCCGCTTCTGCTACACTCGATGC
SLC8A1_rev	AATTTACTAGTTCCAAGGTCTCACCAACCAG
LINC01163_for	AATTTGGCGCGCCAGGGAGAAAAGCCAGCTTCC
LINC01163_rev	AATTTACTAGTACGTGCTGGACCCTAAATAGC
LIPA_for	AATTTGGCGCGCCTCCACTTCCTCCCCCTCTTC
LIPA_rev	AATTTACTAGTCATGGTGCCCAGCCTCTG
TNNT2_for	AATTTGGCGCGCCGGATACGTGAACCTCCACAGG
TNNT2_rev	AATTTACTAGTCTTCCACTCTTCCTGCCCTC

Table 2-3: CRISPR/Cas9 gene editing screening oligos

	In vitro
FLT1_scr_for	GTTCACTTGGGGCTGTTGAC
FLT1_scr_rev	GGCTCCCACTTGGAATGTAG
FLT1_scr_in_for	CCGAACAACAAGAGGAAAGC
SMG6_scr_for	AGAGACGGGATTTCACCATGT
SMG6_scr_rev	TCAATTGTGTGCATGTTCCAT
SMG6_scr_in_for	ACATGTTTGGGGCTGGACTA
FLT1_mut_scr_for	CAACCCGACAGCTGGATAAG
FLT1_mut_scr_rev	GGGAGGAAGAAGAGGGTAGG
FLT1_mut_in_scr_for	CCTCACCTGTGAGAAGCAGA
IRX1_mut_scr_for	CCCACCGCGAGAGTAGCT
IRX1_mut_scr_rev	TTCTGCGGCCTCTCTGAATT
IRX1_mut_in_scr_for	CCAACTACAGCGCCTTCC
PPARGC1A_screen_for	AATTTGGCTAATTCCCTAAACC
PPARGC1A_screen_rev	CCCAACAAATTCTTTTTACAGC

PPARGC1A_inside_for	GTGTTCACACTTGGCACAGAAG
KCNH2_screen_for	AGAAGAAAATGGAGACCCTCTG
KCNH2_screen_rev	TATTTGCGAAGAGCTGTCTGAG
KCNH2_inside_for	TCCCTCTTCTACTGACCCTGAC
C5orf38_screen_for	GAGAAATCAAGGGATCTGGTTG
C5orf38_screen_rev	CTGGCATTTCTTGTTTTGAGTG
C5orf38_inside_for	CCAAGTTTCCATCTGGTAGGAG
CAV1_screen_for	TCCAGGTAGAAAAACCCTGTTG
CAV1_screen_rev	CATACTGGAGTAGGGTGGAACC
CAV1_inside_for	GACTGGGCAACAAGAATGAGAC
	In vivo
medFLT_screen_for	GGTCATAGGGGTGTGTGGAC
medFLT_screen_rev	GCACTTTGCCTGGGAAAACC
medPPARGC1A_screen_for	TTTTGCAGACCTGGTGAGGC
medPPARGC1A_screen_rev	AATGGAGATTGATCCGGGGG

2.3.3 Quantitative Polymerase Chain Reaction (qPCR)

RNA for qPCR was isolated using the Direct-zol™ RNA Miniprep Kit (Zymo Research) and was subsequently reverse transcribed into cDNA using SuperScript® IV Reverse Transcriptase (Thermo Fisher Scientific) according to manufacturer's instructions. RNA levels were measured with the NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific). Input RNA for the qPCR was adjusted to ensure equal amount of RNA across all samples. For the real time PCR-based gene expression profiling, the TaqMan™ Fast Advanced Master Mix (Thermo Fisher Scientific) was performed in 96-well plates according to the user guide and quantified in the ABI 7300 Real-Time PCR detection system (Applied Biosystems) under fast cycling conditions (Table 2-4). Gene expression levels were normalised to GAPDH and calculated using the ΔΔCT method.

Table 2-4: Probes for real time PCR-based gene expression profiling

Taqman-Probes	Gene	Company
Hs01052961_m1	*FLT1*	Thermo Fisher Scientific
Hs02758991_g1	*GAPDH*	Thermo Fisher Scientific
Hs01063093_g1	*IRX1*	Thermo Fisher Scientific
Hs00918382_m1	*KCNH2*	Thermo Fisher Scientific
Hs00173304_m1	*PPARGC1A*	Thermo Fisher Scientific
Hs01034045_m1	*SMG*	Thermo Fisher Scientific

2.3.4 Gel electrophoresis

Agarose gels for gel electrophoresis consisted of 1% agarose in TBE. As a reference for DNA fragment sizes, the GeneRuler™ (Thermo Fisher Scientific) was used and the samples were loaded with 6x loading dye. Subsequently the gels were stained with 2μg/ml SYBR safe in TBE and imaged with a UV light-based gel imager. To proceed with molecular cloning, desired fragments were cut out of the gel and were purified using the innuPREP Gel Extraction Kit according to the manufacturer's instructions.

2.3.5 Transformation

The bacterial strains used for all subcloning were either Mach1 T1 (Thermo Fisher Scientific) or NEBExpress Competent *E.coli* (NEB). *E. coli* were cultured in Luria-Bertani (LB) broth containing Tryptone, NaCl and yeast extract or LB-agar plates with the addition of bacteriological agar prior to autoclaving. Filter-sterilised ampicillin or kanamycin was added to cooled broth to a final concentration of 100 µg/ml. For transformations, 50μl of chemically competent cells were thawed on ice and mixed with 3μl of ligation mix for 20 min before heat shocking at 42 °C for 45 seconds. Bacterial cells were then incubated at 37°C in 300μl SOC outgrowth medium (NEB) for 1 hour. The cells were then plated on agar containing ampicillin or kanamycin, depending on the plasmid, and incubated at 37°C overnight.

2.3.6 Plasmid preparation

Inoculation of 2ml Mini-bacterial cultures containing LB medium and ampicillin (100µg/mL) was performed and cultures were incubated overnight at 37°C. The cells were spun down for 2 min at 14000 rpm and resuspended in 200μl of P1 buffer (resuspension buffer) by vortexing. Lysis was performed by adding 200μl of P2 buffer and mixed by inversion. Subsequently the lysis was stopped by adding neutralization buffer (P3). The DNA was precipitated with 600μl isopropanol and the sample was mixed by vortexing and centrifuged for 10min at full speed. The supernatant was discarded, and washing was performed by adding 200μl ethanol. Residual ethanol was removed, and the samples were dried at 37°C and subsequently eluted in 30μl H_2O. In addition, the Wizard® Plus SV Minipreps DNA Purification System (Promega) was used. Medium-scale bacterial cultures (50 ml of LB containing ampicillin) were grown overnight at 37°C. The purification protocol, either Qiagen or Promega, was performed according to the manufacturer's instructions. However, a few alterations were made to the protocol. The centrifugation steps were performed at 4000rpm at 4°C. Furthermore, instead of centrifuging after the neutralization (Qiagen Kit), the samples were filtered using filter paper and funnels. Regarding the elution step, the pellet was resuspended in 30μl of H_2O instead of elution buffer. The plasmids generated in this study are shown in Table 2-5 and insert sequences for the plasmids used for CRISPR/Cas9 gene editing in Table 2-6 and Table 2-7.

Table 2-5: Plasmids containing human RE for *In vivo* enhancer assay

Name	Vector	Vector Source
LINC0_SCP1::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
KCNJ2_SCP1::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
KCNJ2_2_SCP1::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
CAV1_SCP1::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
KCNH2_ SCP1::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
BUD13_ SCP1::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
SLC8A1_SCP1:::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
PPARGC1A_ SCP1:::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
PAPPA2_ SCP1:::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
MGAT4D_ SCP1:::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)

TGFBR2_ SCP1:::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
CXCL12_ SCP1:::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
LIPA_ SCP1:::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
C5ORF38_ SCP1:::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
SMG6_ SCP1:::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
FLT1_ SCP1:::mmGFP	Modified ZED	(Gutierrez-Triana *et al.*, 2015)
enCAV1_1_Cas9-2A-GFP	pX458	48138, Addgene
enCAV1_2_Cas9-2A-GFP	pX458	48138, Addgene
enFLT1_1_Cas9-2A-GFP	pX458	48138, Addgene
enFLT1_2__Cas9-2A-GFP	pX458	48138, Addgene
enSMG6_1__Cas9-2A-GFP	pX458	48138, Addgene
enSMG6_2_Cas9-2A-GFP	pX458	48138, Addgene
enC5ORF38_1_Cas9-2A-GFP	pX458	48138, Addgene
enC5ORF38_2_Cas9-2A-GFP	pX458	48138, Addgene
enPPARGC1A_1_Cas9-2A-GFP	pX458	48138, Addgene
enPPARGC1A_2_Cas9-2A-GFP	pX458	48138, Addgene
enKCNH2_1_Cas9-2A-GFP	pX458	48138, Addgene
enKCNH2_2_Cas9-2A-GFP	pX458	48138, Addgene
mutFLT1_1_Cas9-2A-GFP	pX458	48138, Addgene
mutFLT1_2_Cas9-2A-GFP	pX458	48138, Addgene
mutIRX1_1_Cas9-2A-GFP	pX458	48138, Addgene
mutIRX1_2_Cas9-2A-GFP	pX458	48138, Addgene
mutSMG6_1_Cas9-2A-GFP	pX458	48138, Addgene
mutSMG6_2_Cas9-2A-GFP	pX458	48138, Addgene

2.3.7 Flow cytometry

Cells were harvested and resuspended in FACS wash buffer (Phosphate-Buffered Saline Buffer + 2% (v/v) fetal calf serum), filtered through a 40µM cell strainer (Falcon, Corning) and spun down. The cell pellet was then directly resuspended in FACS wash buffer containing propidium iodide (PI). When further staining was required, cell pellets were resuspended in FACS wash buffer containing antibodies (anti human CD31-APC and CD34-PE/Cy7 (Biolegend)) at a

concentration of 1:400 (Table 2-8), and were incubated at room temperature for 10 min. Cells were then washed in FACS buffer before being resuspended in FACS/PI. Live cells were sorted via FACS based on side scatter, forward scatter, and PI positive cells were excluded. Initial FACS gates were set using IgG isotype controls (BD Biosciences). All sorting was performed using assisted sorting at the BD INFLUX cell sorter (Becton Dickinson).

2.4 Transgenesis in medaka

2.4.1 Enhancer expression vector

Identified human CREs were amplified by Q5 High-Fidelity DNA Polymerase (NEB, M0491) using primers containing an *AscI* and *SpeI* restriction enzyme (NEB) overhang. PCR products were subsequently digested and cloned into a modified ZED vector (Bessa *et al.*, 2009; Gutierrez-Triana *et al.*, 2015) containing a minimal promoter (SCP1, synthetic) (Juven-Gershon *et al.*, 2006) driving a green fluorescent protein (mmGFP) flanked by insulator sequences (HS4, chicken). Approximately 2µg of the PCR product and 5µg of vector DNA were digested with AscI and SpeI for 2h at 37°C in Cutsmart buffer. Subsequently, the digestion was analysed via gel electrophoresis. A 3-fold excess of the vector DNA was used to ligate the vector with the insert and incubated with 1µl T4 ligase and 1µl T4 ligase buffer over night at RT. All plasmids contained a *Tol2* (transposon) element and SceI (NEB) restriction sites for subsequent linearization *in vivo*.

2.4.2 Transgenesis via Tol2 and SceI

The reporter line *enFLT1:GFP* was generated via injection of 5ng/µl donor DNA and 10 ng/µl *Tol2* mRNA into one cell stage medaka embryos. For transgenesis using the Meganuclease SceI, the injection mix consisted of 0.5x Yamamoto buffer, 0.5x I-SceI buffer, 0.3 U/µl I-SceI enzyme and 10-20ng/µl of the donor constructs, ddH_2O ad 30µl.

Injected embryos were kept at 28˚C in ERM and scored for heart or endothelial specific GFP expression at 4 days post fertilisation (dpf) using a Leica stereomicroscope. Positive embryos were imaged using a confocal laser scanning microscope (Leica TCS SP8) and raised to adulthood to generate a stable line.

2.5 CRISPR/Cas9 gene editing

2.5.1 sgRNA target prediction

Suitable sgRNAs with low predicted off targets targeting endogenous medaka enhancers were designed using the CRISPR/Cas9 target online predictor (CCTop) (Stemmer *et al.*, 2015). A list of sgRNAs used in this thesis is included in Table 6.

2.5.2 Generation of sgRNA for in vivo gene editing

100 µm of the oligo (Table 2-6), containing T7 RNA polymerase overhangs, were annealed in annealing buffer (10mM Tris-HCl ph 7.5-, 30mM NaCl). The subsequent ligation into DR274 (Addgene®, #42250) (Hwang *et al.*, 2013) was performed for 10 min at room temperature, transformed into chemically competent bacteria and grown overnight as described previously. Bacterial cultures were grown in kanamycin (100 µg/mL). Digests, to test for correct integration of the oligos, were performed with BsaI (NEB). In order to extract the DNA, the QIAprep Miniprep kit from Qiagen was used. 10µg of plasmid DNA containing the insert was digested using DraI fast digest (Thermo Fisher Scientific) overnight at 37°C and purified via gel extraction (GelExtraction MinElute Qiagen). The sgRNAs were transcribed using the MEGAscript™ T7 Transcription Kit (Life Technologies), treated with 1µl of TurboDNase (Thermo Fisher Scientific) and cleaned up using the RNAeasy Mini Kit from Qiagen according to manufacturer's protocols.

Table 2-6: sgRNA sequences used for *in vivo* gene editing

Name	Sequence	Source
enFLT1_sgRNA1_oligo1	TAggCCAGACCCAACAGTGGACCC	
enFLT1_sgRNA1_oligo2	AAACGGGTCCACTGTTGGGTCTGG	
enFLT1_sgRNA2_oligo1	TAGGGCTTGAGAGGTATGTGCT	
enFLT1_sgRNA2_oligo2	AAACAGCACATACCTCTCAAGC	
enPPARGC1A_sgRNA1_oligo1	TAgGATGTAGGGCAAAAGCAGTA	
enPPARGC1A_sgRNA1_oligo2	AAACTACTGCTTTTGCCCTACAT	
enPPARGC1A_sgRNA2_oligo1	TAggAGGGGGAATGAGAGTAATGG	
enPPARGC1A_sgRNA2_oligo2	AAACCCATTACTCTCATTCCCCCT	

Oca2_sgRNA1	TTGCAGGAATCATTCTGTGT	(Lischik *et al.*, 2019)
Oca2_sgRNA2	GATCCAAGTGGAGCAGACTG	(Lischik *et al.*, 2019)

2.5.3 CRISPR/Cas9 injection

To generate medaka knock-out mutants, 15ng/µl of each sgRNA was injected together with 150ng/µl of *Cas9* mRNA. The *Cas9* mRNA (Addgene Plasmid # 43861) was transcribed with the Sp6 Transcription Kit according to manufacturer's protocols (Thermo Fisher).

2.5.4 Generating sgRNA for in vitro gene editing

Single guide RNAs (Table 2-7) were designed using the CRISPR/Cas9 target online predictor (CCTOP) (Stemmer *et al.*, 2015). The oligonucleotides were annealed and cloned in pSpCas9(BB)-2A-GFP, also px458 (Addgene® plasmid #48138) for *in vitro* experiments. The plasmids were linearized with the restriction enzymes BsbI (NEB).

Table 2-7: sgRNA sequences used for *in vitro* gene editing

Name	sequence
IRX1_mut_1_1	CACCGCCGCAGTCGGCGCGCGATTG
IRX1_mut_1_2	AAACCAATCGCGCGCCGACTGCGGC
IRX1_mut_2_1	CACC GAGGAACTAGAAAGGTCCGG
IRX1_mut_2_2	AAACCCGGACCTTTCTAGTTCCTC
FLT1_mut_1_1	CACCGTAGTTGCAGCGGGCACGCTT
FLT1_mut_1_2	AAACAAGCGTGCCCGCTGCAACTAC
FLT1_mut_2_1	CACCGTTATAAATCGCCCCCGCCCT
FLT1_mut_2_2	AAACAGGGCGGGGGCGATTTATAAC
SMG6_mut_1_1	CACCGTAGGTAGCCCCAACACAGTT
SMG6_mut_1_2	AAACAACTGTGTTGGGGCTACCTAC
SMG6_mut_2_1	CACCGATAATGATACAGTTCAGTGC
SMG6_mut_2_2	AAACGCACTGAACTGTATCATTATC
FLT1_1_1	CACCGTAAGGGCACAAGCCCTAGTA

FLT1_1_2	AAACTACTAGGGCTTGTGCCCTTAC
FLT1_2_1	CACCGACCTGAAACAACTTAATTT
FLT1_2_2	AAACAAATTAAGTTGTTTCAGGTC
SMG_1_1	CACCGTAGGTCTAGGCATTCTTCAA
SMG_1_2	AAACTTGAAGAATGCCTAGACCTAC
SMG_2_1	CACCGCTGAGACTGCTGTTTCTTCT
SMG_2_2	AAACAGAAGAAACAGCAGTCTCAGC
PPARGC1A_1_1	CACCGATAGGCTTTCCATTAGGG
PPARGC1A_1_2	AAACCCCTAATGGAAAGCCTATC
PPARGC1A_2_1	CACCGTAAATTCCCCACAGAGCT
PPARGC1A_2_2	AAACAGCTCTGTGGGGAATTTAC
KCNH2_1_1	CACCGTAAGCCCAAGAGGCTACC
KCNH2_1_2	AAACGGTAGCCTCTTGGGCTTAC
KCNH2_2_1	CACCGGAGTTTCAGGAACGCTTC
KCNH2_2_2	AAACGAAGCGTTCCTGAAACTCC
C5ORF38_1_1	CACCGGATAGAGAGCTGTTTCTG
C5ORF38_1_2	AAACCAGAAACAGCTCTCTATCC
C5ORF38_2_1	CACCGTTGCTAAAACAGCTTA
C5ORF38_2_2	AAACTAAGCTGTTTTAGCAAC
CAV1_1_1	CACCGAAACAACCTCTACCGTG
CAV1_1_2	AAACCACGGTAGAGGTTGTTTC
CAV1_2_1	CACCGGACAAATTTCATGATTC
CAV1_2_2	AAACGAATCATGAAATTTGTCC

2.5.5 Transfection of cell lines

For transfection using the Neon™ Transfection System (Thermo Fisher Scientiffic) hES cells ($1x10^6$) were resuspended in Buffer R and 2.5µg per plasmid containing the relevant sgRNA with a final volume of 100µl. Electroporation was carried out using 100 µl Neon® Tips according to manufacturer's instructions. Each nucleofection involved 2 pulses of 1050V for 30ms. Subsequently, 1 million cells were split across two 6cm plate (pre-plated with MEFs) containing hES media lacking penicillin/streptomycin. Cells were then grown for 3 days prior to being single

cell sorted into 96 well plates according to GFP expression. After two weeks colonies were visible at the bottom of the 96 well plate and were then genotyped via PCR screening. Expanded clones were then further validated via Sanger sequencing and karyotyped via SNP analysis.

2.5.6 Colony picking and PCR screening

Visible colonies were washed with PBS and extracted using 30µl TrypLE. 10µl of the cell suspension was transferred to wells of a 48 well plate and 20µl were transferred to PCR tubes containing 50µl TE buffer for subsequent DNA extraction. DNA was extracted by boiling the cells at 100˚C for 10 min. Screening PCRs were carried in 96-well plates. Two strategies were designed to determine if the knock-out was successful. Briefly, three primers were designed (Table 2-3) - two flanking the target site and one within the target site – to confirm the deletion was created. If the deletion was successful: a shift in band in the flanking PCR was visible and no band was detectable when using PCR primers within the target site. Successfully modified clones were expanded for karyotyping and Sanger sequencing.

2.5.7 Single Nucleotide Polymorphism (SNP) analysis and Sequencing

To sequence DNA plasmids generated via molecular cloning, samples were prepared and premixed with 600-1200ng plasmid DNA template or 10-100ng PCR amplicon and 3.3pmol primer in a total volume of 12µl. Sequencing was performed offsite at AGRF at the Peter MacCallum Cancer Centre in Melbourne. For Karyotyping, a T25 flask with 80 to 90% confluency was harvested and the cell pellet was submitted to the Victorian Clinical Genetic Services (VCGS) onsite to perform single nucleotide polymorphism (SNP) analysis. Clones in which abnormalities or abnormal karyotypes were detected, were discarded.

2.6 Differentiation Assays

2.6.1 Cardiomyocyte differentiation

Human pluripotent stem cells were differentiated into cardiomyocytes according to an adapted protocol previously published (Anderson *et al.*, 2018). As described above, cells were cultured

on MEFs and harvested using TrypLE™ Express and resuspended in RPMI supplemented with B27 without Vitamin A (Thermo Fisher Scientific, Gibco), 2mM L-glutamine and penicillin/streptomycin. The cells were plated on a 12-well plate (Corning), which was coated with Geltrex™ (Life Technologies) in a 1:100 ratio and grown for 24 hours. On day 0 of the cardiac differentiation protocol, cells were induced using medium supplemented with 12µM CHIR99021 (Tocris), 80 ng/ml Activin A (In Vitro Technologies), 50µg/ml Ascorbic Acid (Sigma). On day 1, the media was replaced with standard cardiac differentiation media containing 5µM IWR-1 (Sigma Aldrich) and 50µg/ml Ascorbic Acid. This media was refreshed on day 3 and day 5 of the cardiac differentiation. The cells were then maintained in standard cardiac differentiation media until harvested.

2.6.2 Endothelial cell differentiation

For the endothelial differentiation of human embryonic stem cells, an adapted cardiac differentiation protocol (Anderson *et al.*, 2018) was used. Cells were plated and resuspended as described above. On day 0, the standard cardiac differentiation medium contained 12µM CHIR99021, 80ng/ml Activin A and 50µg/ml Ascorbic Acid. For day 3 and day 5, the standard media was supplemented with 5µM IWR-1, 50µg/ml Ascorbic Acid, 30ng/ml VEGF (PeproTech) and 50ng/ml SCF (PeproTech). At day 6 the cells were harvested and processed for flow cytometry sorting. The suspension was stained for the endothelial specific cell surface markers CD31 and CD34 to sort for endothelial progenitors. Anti-human CD31-APC (BioLegend) and anti-human CD34-PE/Cy7 (BioLegend) were used in a 1:400 dilution. During sorting the endothelial fraction was isolated by selecting double positive CD31+ and CD34+ populations. These cells were then collected via centrifugation at 1500 RPMI for 3 mins and resuspended in endothelial growth medium, EGM-2MV Bulletkit (Lonza). The cells were plated on 6 wells plates pre-coated with 1µg/ml Fibronectin (Merck) at a cell seeding density of $0.5x10^6$ per well. Endothelial cells were kept for a maximum of 5 passages.

2.7 Functional assays

2.7.1 Cardioexcyte

To assess the electrophysiological properties of cardiomyocytes, the cells were analysed using the Cardioexycte Nanion platform (Nanion). Briefly, cardiomyocytes were washed with Versene and harvested using Trypsin (incubated for 15 min at 37˚C) and resuspended in 1:1 FBS/Versene. After centrifugation, cells were resuspended in RPMI B27 (plus insulin, Gibco) containing 5µM ROCK inhibitor Y-27632 (Ri) and 150,000 cells per well were plated on NSP-96, CardioExcyte 96 Sensor Plates -Stim (Nanion) pre-coated with Matrigel (1:600 in PBS). After 24 hours, prior to the start of the measurement the media was replaced with maturation media (Mills *et al.*, 2017) and the cells were calibrated overnight.

2.7.2 Angiogenesis assay

Angiogenesis assays were performed according to the protocol of Arnaoutova and Kleinman (Arnaoutova and Kleinman, 2010). Briefly, 40µl of Geltrex^{TM} was added to each well of a 96 well glass bottom plates (Corning). The plates were kept on ice while pipetting, subsequently centrifuged and left to set in the incubator at 37˚C for 30min. Endothelial cells were harvested and resuspended in complete growth factor EGM-2V media supplemented with anti-human CD31-APC and anti-human CD34-PE/Cy7 antibodies (1:400) (Table 2-8). 15,000 cells were plated per well. The vessel formation was captured for 48 hours. Confocal images were acquired with Yokagawa CellVoyager CV8000 high-throughput discovery system under 37˚C and 5% CO_2. Imaging and data processing were performed with the assistance of Holly Voges.

2.7.3 Wound healing assay

Wound healing assays were performed using the µ-Plate 24 well with 2 well silicone inserts (Ibidi). The wells of the silicone inserts were coated with Fibronectin prior to seeding the cells. Endothelial cells were harvested resuspended in EGM-2V, and 70µl containing 50,000 cells were plated per insert well. After 24 hours, the insert was removed using clean forceps, the cells were washed with PBS and fresh EGM-2V media was used for the duration of the imaging.

The imaging was performed with the Yokagawa Cell Voyager CV8000 over the course of 48 hours.

2.8 Immunofluorescence and confocal microscopy

2.8.1 in vivo whole mount immunostaining and confocal microscopy

Whole mount medaka hatchling immunostaining was performed by fixing the samples with 4% paraformaldehyde (PFA) overnight at 4˚C. All washing steps were performed using PBS containing 0.1% tween (PBS-T). After washing the samples were permeabilized with acetone for 10 min at -20˚C before rehydration with water. Blocking was performed with PBS-T containing 10% goat serum. All primary antibodies were incubated overnight at 4˚C. Secondary antibodies were incubated for 2 hours at RT. Three times 5min washing steps with PBS were performed between all steps.

Medaka hatchlings (sedated with tricaine when alive) were mounted in 1% low melting agarose on glass bottom MatTek dishes. High-resolution imaging was carried out using confocal laser scanning microscopes (Leica TCS SPE or Leica TCS SP8). Analysis and processing were performed using the ImageJ software.

2.8.2 in vitro immuno staining and confocal microscopy

Endothelial cells, cardiomyocytes or human embryonic stem cells were seeded on plastic coverslips (Thermo Fisher Scientific Nunc™ Thermanox™ Coverslips). After 24 hours, cells were washed with PBS and fixed and permeabilized in 4% paraformaldehyde containing 0.1% triton X-100 for 10 to 15 mins at room temperature. Cells were washed three times with PBS and subsequently blocked with 5% Bovine Serum Albumin (BSA) or goat serum diluted in TBS/PBS. Primary Antibodies (Table 2-8) were incubated overnight in PBS with goat serum at 4 ˚C. Concentration and ratios of antibodies were adjusted according to the specific antibody used. Incubation of the secondary antibody was performed for 2 hours at room temperature. The cells were washed with PBS and mounted on coverslips with Fluoromount-G®

(SouthernBiotech) prior to imaging. Immunofluorescence was detected using the LSM 780 (Zeiss) or the Yokagawa Cell Voyager CV8000.

Table 2-8: List of anitbodies used in IHC, FACS and angiogenesis assays

Antibody	Concentration	Company
Anit-mouse Alexa Fluor 594	*1:200*	A21145, Invitrogen
Anit-mouse Alexa Fluor 647	*1:200*	A21242, Invitrogen
Anti-Chicken Alexa Fluor 488	*1:200*	A11039, Invitrogen
Anti-GFP	*1:200*	Ab13970, Abcam
Anti-human alpha-Actinin	*1:200*	A7811, Sigma-Aldrich
Anti-human CD31-APC	*1:400*	303116, Biolegend
Anti-human CD34-PE/Cy7	*1:400*	343516, Biolegend
Anti-human FLT1-647	*1:200*	ab32152, Abcam
Anti-Human SIRPA-PE	*1:200*	323806, Biolegend
Anti-human Troponin T	*1:200*	AB45932, Abcam
Anti-human VEGFR2	*1:200*	359909, Biolegend
Anti-human VWF	*1:200*	AB6994, Abcam
Anti-Ki-67	*1:200*	9129SCell Signalling Tech.
Anti-mouse IgG1-APC CTRL	*1:400*	555751, BD Biosciences
Anti-mouse IgG1-PE CTRL	*1:400*	555749, BD Biosciences
Anti-rabbit Alexa Fluor 555	*1:200*	A21429, Invitrogen
Anti-rabbit Alexa Fluor 647	*1:200*	A21244, Invitrogen

2.8.3 in vitro image analysis and statistics

Maximum intensity projection (MIP) images were constructed from 15µm z slices (300µm total z distance), captured every 40 minutes for 48 hours. Images were analysed in CellPathfinder software based on CD34-APC intensity. Analysis algorithm was defined using skeleton function to generate the parameters vessel area, vessel length, branch count, and junction count. The Kruskal-Wallis test was used to evaluate differences in medians among three cell lines. If the Kruskal-Wallis test was significant, the Wilcoxon Rank Sum Test was used to assess the level of difference significance on a variable between two cell lines. The statistical analysis was

performed in the R statistical programming language. Image analysis and statistics were performed in collaboration with Holly Voges and Neda Rahmani Mehdiabadi

2.9 RNA sequencing and analysis

Three biological replicates with each three technical replicates (per cell line) of endothelial cells were harvested 12 days after differentiation initiation. RNA was extracted using the Direct-Zol RNA Miniprep Kit (Zymo research). Paired-end mRNA sequencing was performed at the Victorian Clinical Genetics Services (VCGS) using the NovaSeq 6000 System (Illumina) with a 2 x 150 bp read length. The fastq files were processed using the RNAsik pipeline (2018). The STAR aligner (Dobin *et al.*, 2013) was used to align reads to the GRCh38 Assembly. Aligned reads were assigned to features from the GRCh38 EnsEMBL Annotation (Yates *et al.*, 2019) using the featureCounts program from RsubRead (Liao *et al.*, 2019). Degust (2019) was used to perform and visualise differential expression analysis. Firstly, the first dimension of unwanted variation were removed from counts using RUVr routine from the RUVSeq R package (Risso *et al.*, 2014), k=1. Next, TMM normalisation and the quasi-likelihood test was performed using EdgeR's (Robinson *et al.*, 2009) standard workflow. Bulk RNA sequencing was performed at VCGS and the dataset was analysed by Michael See.

Salmon (Patro *et al.*, 2017) was used to quantify transcript isoforms abundance, and the differential abundance of *FLT1* was tested and visualised using the DRIMSeq R Package (Nowicka and Robinson, 2016). The statistical analysis using the DRIMSeq R Package, assumes the Dirichlet Multinomal model for each gene.

2.10 Chromatin Confirmation Capture (Hi-C) analysis

Publicly available Hi-C datasets (Dekker *et al.*, 2017; Rao *et al.*, 2014; Zhang *et al.*, 2019b) were analysed in collaboration with Gulrez Chahal and Alessandro Bertero (Bertero *et al.*, 2019).

3 Chapter 3: Regulatory Elements – Prediction and Validation

3.2 Background

Disruption of heart and major blood vessel formation during development results in congenital heart defects at birth which are a major cause of childhood mortality and morbidity (Virani *et al.*, 2020). The development of the heart is strictly regulated by a tight network of genetic components (Edwards and Gelb, 2016) which when disrupted perturb normal cardiogenesis leading to disease. Genome wide association studies (GWAS) have been used to identify genes associated with CHD and begin to dissect the complex genetic architecture underlying heart disease (Agopian *et al.*, 2017; Wang and Wang, 2018). Nevertheless, an established problem is associating disease phenotype to a given variant or single nucleotide polymorphism (SNP). Strikingly, the majority of SNPs associated with CHD are found in non-coding regions in the genome, highlighting the importance of *cis*-regulatory elements in developmental processes and disease (Chahal *et al.*, 2019; Corradin and Scacheri, 2014). These regulatory elements (REs), such as enhancers, promote gene expression in a spatio-temporal manner through the coordinated binding of specific transcription factors (TFs). However, since SNPs in such sequences can lead to disruptions in TF binding motifs and therefore have no impact on the protein sequence directly, it is challenging to pinpoint how sequence alterations result in cardiac and blood vessel defects. Aberrations in regulatory elements can lead to the perturbation of a gene regulatory network which, in turn, causes genes to be over- or underexpressed, even for multiple targets at once. This ultimately results in enhanceropathies, a group of diseases caused by mutations in regulatory elements (Smith and Shilatifard, 2014).

There is now mounting evidence that disrupted cardiovascular regulatory elements can impair heart development leading to disease (van den Boogaard *et al.*, 2019; Dickel *et al.*, 2016; May *et al.*, 2012; van Ouwerkerk *et al.*, 2019). A prerequisite for functional studies to understand the effect of non-coding SNPs, is the accurate identification of the regulatory regions in the human genome that are important for cardiac development and disease. Publicly available datasets relevant to the human genome such as regulatory element associated chromatin marks (ENCODE Project Consortium, 2012; Roadmap Epigenomics Consortium *et al.*, 2015), TF analyses (Anderson *et al.*, 2018) and chromatin capture experiments (Arvanitis *et al.*, 2020;

Bertero *et al.*, 2019; Montefiori *et al.*, 2018) provide valuable resources to define the human gene regulatory network in the heart. However, key challenges to identify non-coding elements relevant for disease remain. The search space is still large, for example, the current registry of human *cis*-regulatory elements in the Encode data set is comprised of 926,535 entries. Even if it were possible to predict REs that could be involved in CHD, it is crucial to have functional validation methods to determine both the sufficiency and necessity of a given human regulatory element for normal development and disease (Catarino and Stark, 2018).

To assess REs in the context of organogenesis, many studies focus on enhancer assays *in vivo* such as mice (Dickel *et al.*, 2016; Visel *et al.*, 2007), fish, such as zebrafish and medaka (Bessa *et al.*, 2009; Letelier *et al.*, 2018; Minhas *et al.*, 2019) and drosophila (Arnold *et al.*, 2013). Due to their fast, extra-uterine development, low generation time and relatively easy genetic editing (Korzh *et al.*, 2018), zebrafish and medaka have emerged as a powerful tool to study enhancer function during cardiac development. The teleost heart is structurally distinct from the mammalian heart (Bakkers, 2011), with the most prominent difference being the number of the heart chambers. While fish have only two heart chambers, the ventricle and the atrium, the mammalian heart consists of four chambers in total. Despite these differences, many developmental processes are conserved between fish and mammals (Nguyen *et al.*, 2008). Here, we used medaka (*Oryzias latipes*) as a model system to study the potential of human enhancers to drive gene expression in cardiovascular tissues. Compared to zebrafish, medaka has a smaller genome (800 Mb) and less genome duplication (Hilgers and Schwarzer, 2019), which is beneficial when studying the effect of exogenous or endogenous REs. A smaller genome can be easier to interpret due to a reduction in genetic redundancy, which in turn, can lead to more robust transient phenotypes when being altered. Previous studies using medaka have already shown their potential of using human and medaka REs *in vivo* (Dunham *et al.*, 2012; Williamson *et al.*, 2016). As a starting point in this study, we developed bioinformatic pipelines to predict human CREs which might be implicated in cardiovascular development and potentially novel candidates that are implicated in CHD and subsequently tested their potential to drive reporter gene expression in medaka.

3.3 Results

3.3.1 Predicting primate specific cis-regulatory elements

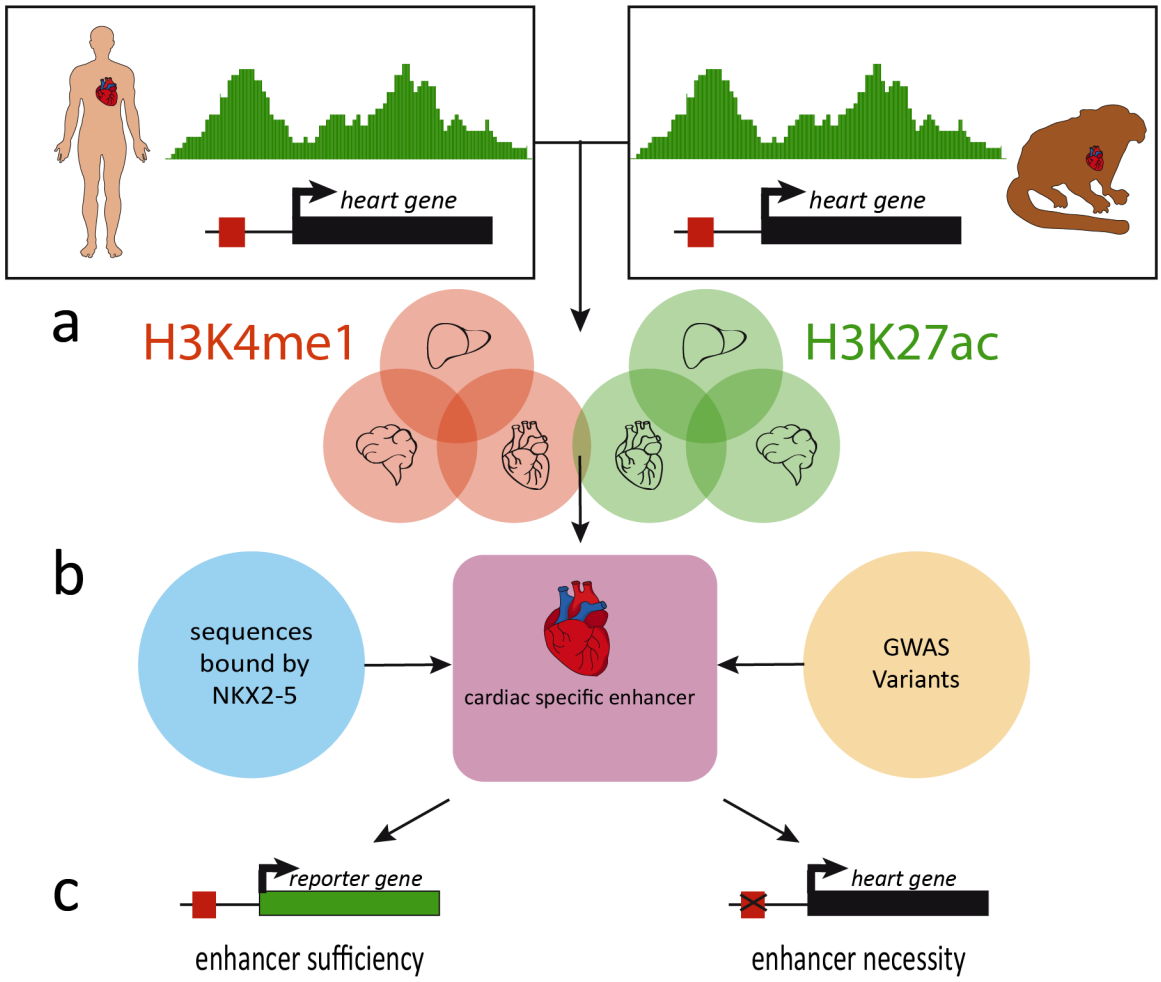


Fig. 3-1: Predicting primate specific and cardiac specific *cis*-regulatory elements. a) REs were identified based on histone marks (H3K4me1 and H3K27ac) in cardiac tissue of human and marmoset, b) NKX2-5 binding and variants associated to CHD by GWAS. c) Putative RE were tested for the ability to drive reporter gene expression and the necessity for endogenous target gene expression.

Identifying non-coding DNA sequences implemented in the development of the heart can be challenging since they do not follow a genetic code like protein coding genes (Catarino and Stark, 2018). In absence of such genetic code, sequence conservation is the best proxy to identify functionally relevant elements, but due to the low evolutionary pressure, these sequences are poorly conserved. This is especially noticeable between human and mouse, where only 3% of cardiac enhancers are conserved (May *et al.*, 2012). This can make it difficult

to identify functionally conserved mammalian RE when comparing human and mouse sequences.

In order to overcome this problem we applied different filters to our bioinformatic pipeline to uncover REs relevant in cardiac development and disease. In regards to the lack of conservation between human and mouse, previous work has made use of the close evolutionary relationship between the marmoset monkey (*Callithrix jacchus*) and humans and although closely related, these New World Monkeys have been shown to be sufficiently diverged from humans to provide the optimal evolutionary distance to identify primate specific REs (Wang *et al.*, 2007). Previously, various datasets were generated from marmoset tissues that included Chromatin-immunoprecipitation followed by deep-sequencing (ChIP-Seq) of biochemical modifications on lysine residues of the nucleosome component histone 3 (H3K4me1, H3K27ac and H3K4me3) that are indicative of enhancers, active and promoters respectively (unpublished data, performed by Markus Tondl and Michael Eichenlaub). In addition, we used human cardiac tissue specific histone marks which were absent in other tissues to filter for highly relevant REs involved in heart development. We found 13,627 human cardiac-specific enhancers via bioinformatic mining of publicly available H3K3me1, H3K27ac and K3K4me3 Chip-Seq datasets in human cardiac tissue (Roadmap Epigenomics Consortium *et al.*, 2015). We then used the conservation of histone marks between species as an additional filter for important cardiac REs. After mapping these regions to the marmoset genome, we obtained 7,898 sequences that were sequence conserved, out of which 2,833 regions also showed activity marks (H3K4me1) in marmoset. To provide further cardiac specificity, these primate and cardiac specific putative *cis*-regulatory elements were then intersected with the ChIP-Seq dataset of the transcription factor NKX2-5 of human cardiomyocytes (Anderson *et al.*, 2018), a known master regulator of heart development and highly implemented in cardiac disease. Lastly, to establish a link between these putative cardiac REs and heart disease, we made use of SNPs that were associated to CHD by genome wide association studies and used variants that were linked to different forms of cardiac diseases that can arise from developmental heart defects such as atrial fibrillation, dilated and hypertrophic cardiomyopathy and long QT syndrome (Hindorff *et al.*, 2009; Welter *et al.*, 2014). After applying these different filters, we identified a repertoire of 19 REs that potentially play an important role in heart development and disease, out of which 15 candidates were used for transgenesis assays (Fig. 3-1). This bioinformatic pipeline was

established together with Mark Drvodelic. We next proceeded to the functional assessment of these predicted REs *in vivo*.

Table 3-1: putative REs used for transgenesis assays i*n vivo*

CRE	Coordinates (human hg19 assembly			Location	Putative genes regulated by the RE
1	1	176565715	176574589	intronic	*PAPPA2*
2	2	40319558	40412896	intronic	*SLC8A1*
3	3	30695645	30712547	intronic	*TGFBR2*
4	4	23819370	23821538	intronic	*PPARGC1A*
5	4	141386524	141387822	intronic	*MGAT4D*
6	5	3090147	3093495	intergenic	*C5ORF38*
7	7	150664883	150667480	intronic	*KCNH2*
8	7	116166690	116200926	intronic	*CAV1*
9	10	44752539	44760041	intergenic	*CXCL12*
10	10	130365285	130370434	intergenic	*LINC01163*
11	10	90994596	91003229	intronic	*LIPA*
12	11	116611382	116613749	intergenic	*BUD13*
13	13	28980994	28984098	intronic	*FLT1*
14	17	68557471	68567891	intergenic	*KCNJ2*
15	17	2132301	2148779	intronic	*SMG6*

3.3.2 Human cardiovascular specific elements are able to drive gene expression in medaka

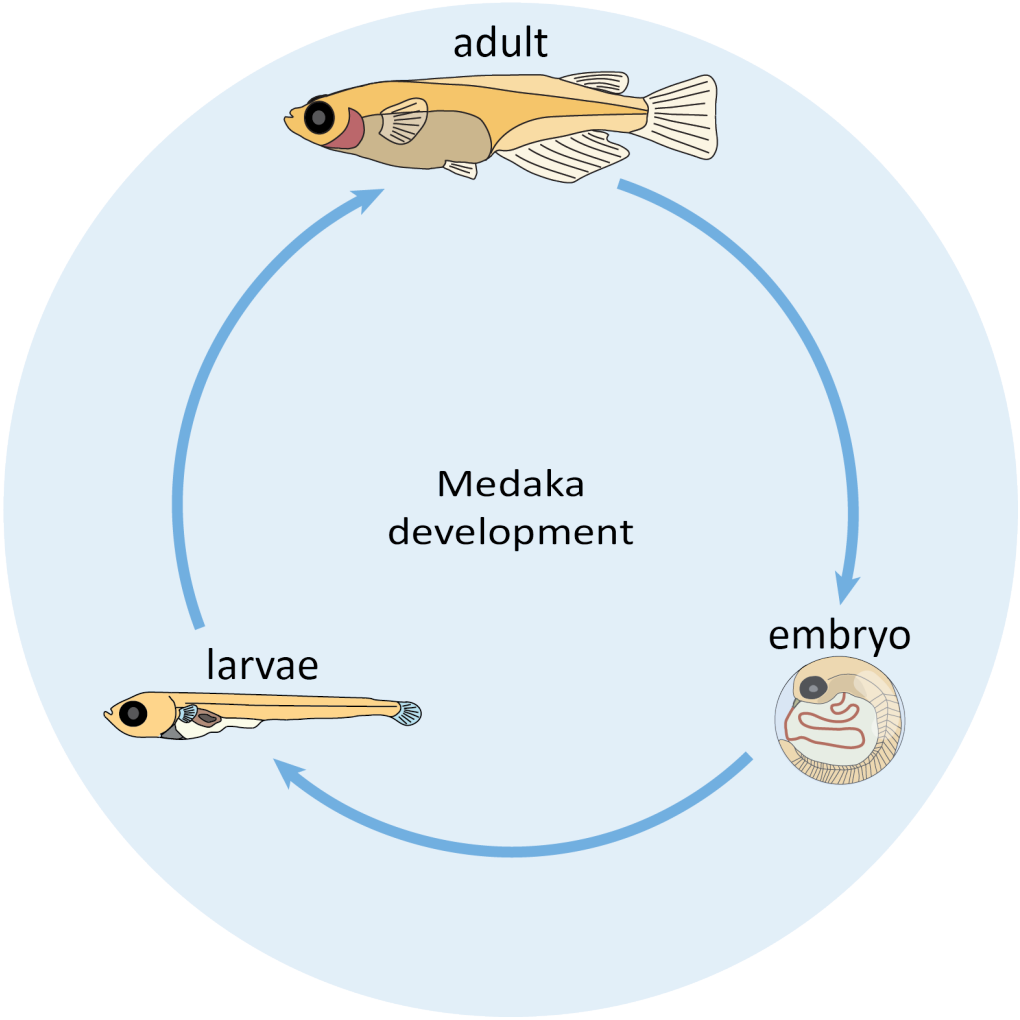


Fig. 3-2: Schematic representation of the Medaka life cycle. When kept at 28˚C, medaka embryos develop and hatch at 9dpf. At the point of hatching, the larvae are already able to feed and reach sexual maturation within 6 to 8 weeks (Iwamatsu, 2004).

In order to perform *in vivo* studies and assess regulatory function of human CREs during organogenesis, we used medaka as a model organism for the following reasons: 1) it has been successfully used to test the *in vivo* function of human enhancers (Dunham *et al.*, 2012), 2) it is a well- established model to study cardiac diseases (Gierten *et al.*, 2020; Lai *et al.*, 2019) and 3) it offers several advantages in efficient gene editing (Lischik *et al.*, 2019) and imaging during development (Stolper *et al.*, 2019). The initial set of enhancer assay experiments were performed in Germany. However, to conduct experiments in Australia, we imported medaka embryos and raised them to adulthood (Fig. 3-2) to establish the first medaka breeding colony in Australia.

Out of the 19 predicted cardiac-specific REs, 15 human regulatory elements (Table 3-1) were succesfully cloned into a modified ZED vector (Bessa *et al.*, 2009) containing a SCP1 promoter (Juven-Gershon *et al.*, 2006) (Fig. 3-3a) to assess their ability to drive reporter gene expression. The plasmids were linearized using the meganuclease SceI (NEB) and injected into the one-cell stage of medaka embryos. SceI enzyme will not cut in the host genome and, therefore, unwanted off target effects can be neglected (Thermes *et al.*, 2002). The developing embryos were screened for GFP at 4 days post fertilization (dpf). At this time point the heart in medaka has undergone heart looping and the development of the heart is almost complete (Iwamatsu, 2004). GFP expression was then monitored until around 9dpf, and the larvae were imaged directly after hatching. Non-specific GFP expression in the yolk syncytial layer (YSL) was observed in almost all injected embryos (data not shown). From 15 injected candidates, 3 human RE sequences associated to the genes *FLT1*, *SMG6* and *KCNJ2* were able to drive specific GFP expression in cardiac tissue, endothelial tissue and muscle tissue (Fig. 3-3b).

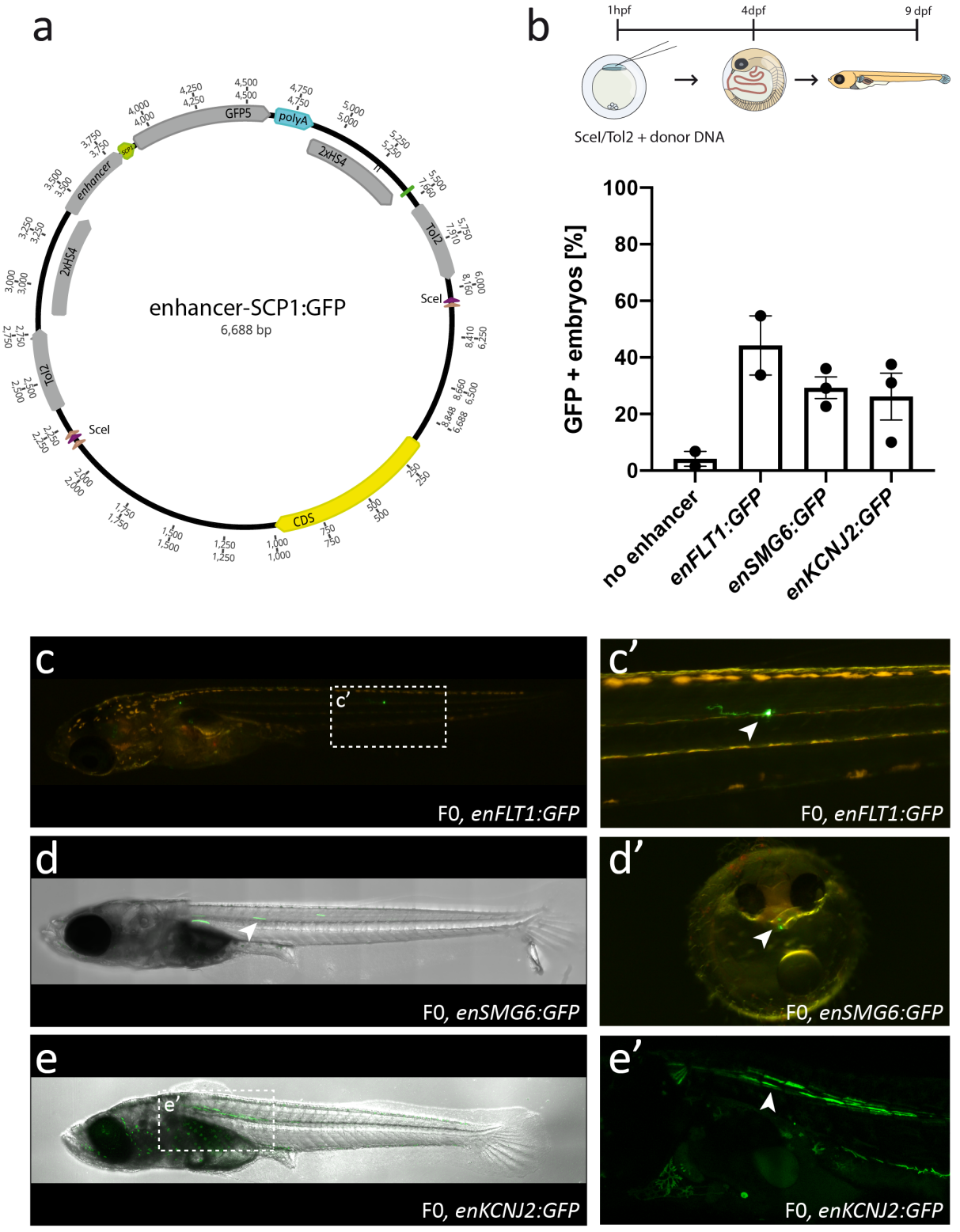


Fig. 3-3: Human enhancers drive GFP expression in medaka. a) Transgenesis assay plasmid containing *enFLT1*. Control vector had no RE sequence inserted. b) Plasmids were in injected in one cell stage medaka embryos. Occurrence of cardiovascular and muscle specific GFP expression was monitored until hatching. Embryos injected with the human regulatory elements for *FLT1* (43%), *SMG6* (30%) and *KCNJ2* (24) showed higher number of GFP positive embryos compared to injection with a control vector without a putative RE. c) The enhancer of *FLT1* drives GFP expression in c') endothelial cells (arrowhead). d) The human enhancer sequence of *SMG6* is able to drive expression of GFP in somites (arrowhead) and d') the heart (dorsal view of embryo, arrowhead). e) The expression pattern of the regulatory element of *KCNJ2* is restricted to e') somites (arrowhead). GFP expression in the heart was not observed. The yellow/orange fluorescence signal is non-specific autofluorescence of the pigments.

The enhancer sequence located in intron 10 of the human *FLT1* or *VEGFR1* (Vascular growth factor receptor 1) gene, consisting of 358 bases was able to drive GFP expression in endothelial and cardiac tissues (Fig. 3-3c and c'). Of 121 injected embryos, 52 embryos (43%) were positive for GFP expression. This is in agreement with previous studies that have characterised the zebrafish orthologue of this putative *FLT1* enhancer (Bussmann *et al.*, 2010), where the zebrafish enhancer drove expression of the reporter gene in blood vessels when fused together with the endogenous promoter of the *flt1* gene in zebrafish.

Injection of the enhancer sequence located in intron 9 of the human gene *SMG6* also resulted in GFP expression in cardiovascular tissue. 53 (30%) of 171 injected medaka embryos displayed reporter gene expression in blood vessels, heart and somites (Fig. 3-3d and d'). Interestingly, 12% of all injected embryos showed GFP expression at around 3 to 4dpf. In some cases, the GFP expression in the heart was not detected after hatching (9dpf), suggesting target gene expression might temporally regulated by this CRE.

The third candidate, a putative enhancer sequence of 1366 bp length, located downstream of the gene *KCNJ2* resulted in GFP expression in somites in the majority of embryos. Although, GFP expression was also observed in the endothelium, expression in the heart was not observed. Of 75 injected embryos, 18 (24%) fish were positive for GFP (Fig. 3-3e and e').

It is important to note, that microinjection into the 1-cell stage of medaka embryos will not lead to completely transgenic fish. While some cells will carry the construct, others will be lacking the transgene. Any injection will therefore result in mosaicism (F0 generation). To get a uniform expression of the construct, all GFP positive fish injected with the enhancer of *FLT1*, which we termed *enFLT1*, were raised and subsequently crossed to wildtype fish. Fish with germline transmission of the transgene then give rise to completely transgenic offspring (F1 generation). The spatio-temporal activity of the enhancer can therefore much more efficiently be analysed in the F1 generation. The functional analysis of *enFLT1* will be discussed in more detail in the next chapter.

3.3.3 *Ultra-conserved cardiac specific regulatory elements*

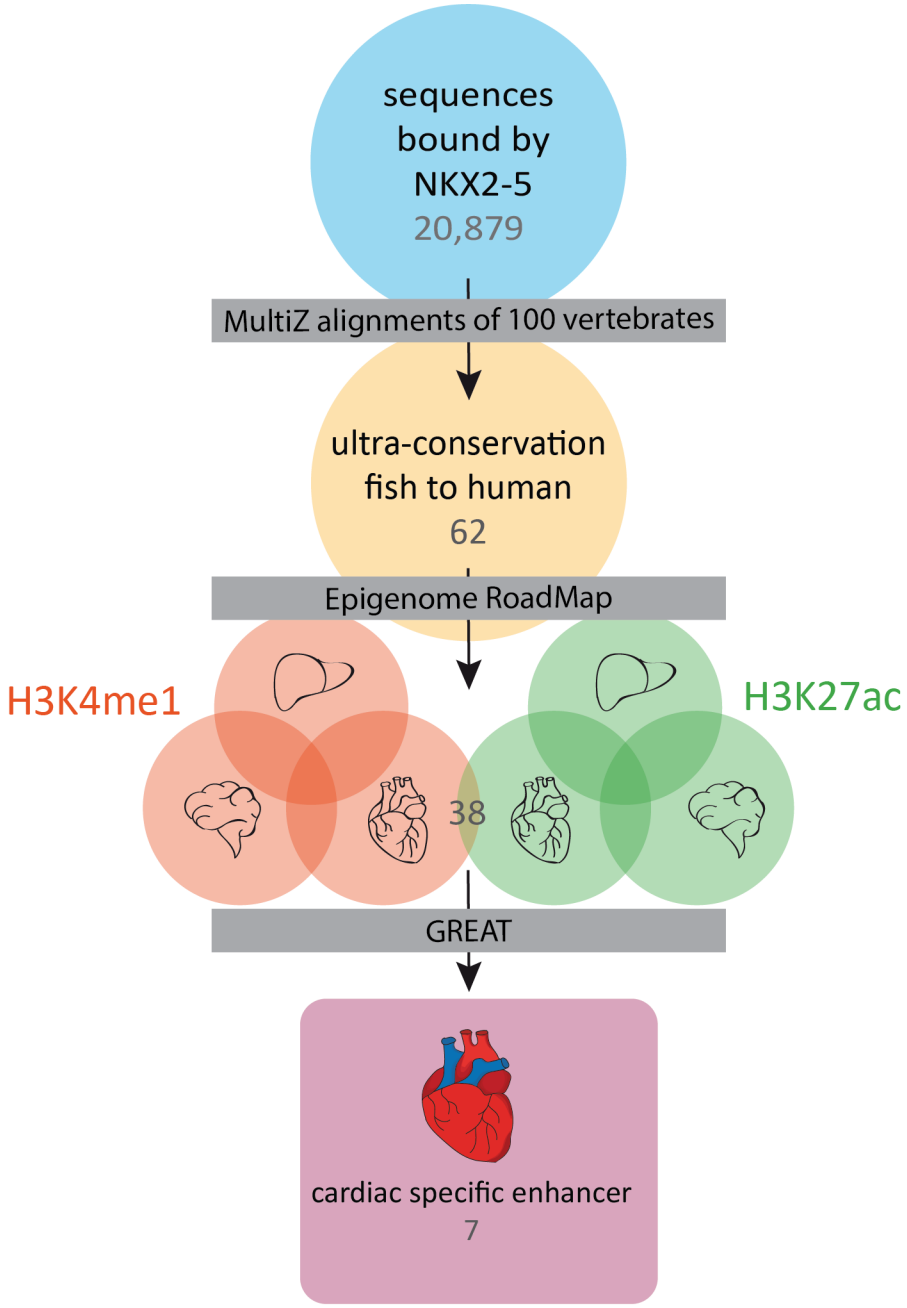


Fig. 3-4: Bioinformatic pipeline to identify ultra-conserved cardiac RE. REs were identified based on NKX2-5 binding, sequence conservation and active histone marks (H3K4me1 and H3K27ac). Grey boxes indicate software or datasets used for each step used (Material & Methods, section 2.1).

Our first round of functional validation indicated that 20% (3/15) of all tested human CRE injected in medaka showed expression in relevant tissues. Interestingly, the most promising candidate, *enFLT1*, showed high sequence conservation between human and fish. Therefore,

we hypothesised that ultra-conserved cardiac specific regulatory sequences would yield to a higher output when tested *in vivo*.

We therefore adjusted the bioinformatic pipeline to filter for ultra-conserved sequences that were directly bound by the "master regulator" of heart development NKX2-5 in human cardiomyocytes (Fig. 3-4). From all of the ChIP-seq experiments, 20,879 regions were identified to be directly bound by NKX2-5. We therefore filtered these regions for high sequence conservation and obtained 62 sequences which were ultra-conserved across 100 vertebrate species, from fish to human. Furthermore, in order to filter for sequences that were shown to be active REs, we also used publicly available datasets of histone modification marks as a measure to obtain active cardiac enhancers (Roadmap Epigenomics Consortium *et al.*, 2015). We identified 38 regions that showed histone marks for active enhancers H3K4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac). We further filtered these regions for those that could be associated with genes known to play a role in heart development. We identified 7 CREs associated with highly relevant genes expressed in the heart and deeply embedded in the genetic networks controlling heart development (Table 3-2).

Interestingly, the enhancer of *FLT1*, which was one of 2 REs identified in the first analysis that was deeply conserved from fish to human, was also identified after adjusting our bioinformatic pipeline. To then further understand the mechanism of regulatory elements involved in cardiovascular development and to illustrate the evidence supporting our hypothesis, we set out to further investigate the enhancer element located in the intron 10 of the gene *FLT1,* which we termed *enFLT1.*

Table 3-2: Non-coding REs ultra-conserved from fish to human

CRE	Coordinates (human hg19 assembly)			Location	Putative genes regulated by the RE
1	3	71573754	71574438	intronic	*FOXP1*
2	5	88129361	88130230	intronic	*MEF2C*
3	8	106341006	106341556	intronic	*FOG2*
4	13	28982241	28983259	intronic	*FLT1*

5	15	96808076	96808846	intronic	*COUP-TFII*
6	16	54540547	54541137	intergenic	*IRX5, IRX3*
7	21	38807473	38807875	intronic	*DYRK1A*

3.4 Discussion

Whole genome sequencing techniques enable fast and efficient screening of patients to identify genetic factors that are associated with disease. Regardless of the fact that the regulatory elements of the genome have been recognized as essential functional components in organ development and disease (Maurano *et al.*, 2012), most studies primarily focus on protein coding sequences that comprise less than 2% of the entire genome (UK10K Consortium *et al.*, 2015). In part, the focus has remained on protein-coding sequences because of the challenge in identifying functional non-coding sequences, due to missing genome annotations and the lack of tools to determine the exact penetrance of variants in regulatory sequences (UK10K Consortium *et al.*, 2015). However, when screening for the cause of congenital heart disease, it is crucial to focus on the non-coding sequences of the genome as SNPs that are associated to CHD are commonly found within these sequences (Corradin and Scacheri, 2014).

Since heart REs are poorly conserved between mice and human (May *et al.*, 2012), we aimed to identify primate specific REs that are involved in cardiovascular development and disease. We also aimed to link these sequences with known variants associated to CHD by genome wide association studies. Our output consisted of sequences that fulfilled all the other prerequisites (NKX2-5 binding, histone modification in human and marmosets) and overlapped with SNPs associated to CHD. This stringent approach may have limited the yield of positive candidate sequences and it is possible some important putative REs were missed in this analysis. We also did not take the linkage disequilibrium (LD) into consideration when filtering for important cardiac CREs. Our rationale for this strategy is that the leading disease associated SNP identified by GWAS is often not the causative SNP and, therefore, the LD may not directly correlate with causality. Often the LD region spanned distances of greater than 5 kilobases up and downstream of the associated SNP (Slatkin, 2008). Thus, including this information in our model

would have provided a less stringent filtering method, decreasing the number of false negatives.

In order to obtain a set of primate specific regulatory elements, we used adult marmoset heart tissue to perform ChiP-seq of histone modifications. Preferably, to obtain non-coding regulatory sequences implemented in heart development and the formation of CHD, fetal heart tissue should have been used for this analysis. However, Since the marmoset heart tissue was scavenged from another research group, we were limited to what kind of tissue we were able to obtain. By using sequences obtained from adult tissue, the dataset was most likely enriched for regions in the genome that are required for heart homeostasis and growth rather than development which might have also led to false negatives. However, we were able to identify 19 sequences which were primate specific with only 2 that showed ultra-conservation from fish to human.

All 15 sequences we obtained and tested, were linked to their target genes by GREAT based on proximity rules. This arbitrary association might not represent the true target genes with which these putative REs actually interacting. It can be assumed that an intronic CRE would interact with the promoter of the gene it is located in. However, it has been shown that intronic REs can also target neighbouring genes (van den Boogaard *et al.*, 2014). In order to assess what sequences are interacting with which genes, it is necessary to include datasets of chromatin capture experiments. This would provide a proof of physical interaction within the chromosome between RE and promoter. Often, these techniques have limited resolution and interactions below 40kb might be missed, which is a disadvantage when aiming to link intronic enhancers to genes they are located in. Modern prediction tools (Salameh *et al.*, 2020) and improved chromatin capture techniques will in time be able to address these problems.

Not all candidates that were obtained by this pipeline were expected to be restricted only to the heart. The heart consists of many different cell types such as heart muscle cells, fibroblasts and endothelial cells and whole heart datasets were used for the analysis. Therefore, some REs might be specific to a particular cell type within the heart, resulting in a cardiovascular rather than cardiac specificity. When assessing the specificity of these candidates, 13 out of 15 targets were associated to genes connected to cardiovascular developmental processes (van den Boogaard *et al.*, 2019; Das *et al.*, 2019; Guo *et al.*, 2019b; Nikpay *et al.*, 2020; Nistala and Savin, 2017; Peden *et al.*, 2011; Warren *et al.*, 2018; Xu *et al.*, 2018; Yuan and Braun, 2017). This

suggests, although missing out on some false negatives, that we were able to identify sequences which are very likely to be functional in heart development. In detail, the candidate REs, associated with the genes *SMG6*, *KCNJ2* and *FLT1*, which were able to drive reporter gene expression in medaka, were found to be regulators of cardiovascular development (Adler *et al.*, 2020; Krueger *et al.*, 2011; Wirtwein *et al.*, 2016).

The reason that only 20% of RE were able to drive reporter gene expression might indicate that a lack of sequence conservation could lead to a decreased transcriptional activity. Although conservation might be important to maintain reporter expression across species, it is not absolutely necessary. The REs of *SMG6* and *KCNJ2* proved to be only primate specific sequences. While the putative enhancer of *SMG6* was able to drive expression in the heart and muscle, expression of the reporter gene via the *KCNJ2* enhancer was only observed in somites. However, in order to define the exact tissue activity of these regulatory elements, co-staining with tissue specific markers for endothelial, muscle and cardiac cells need to be performed in the future. Given *SMG6*'s role as a mediator of embryonic stem cell differentiation through nonsense-mediated decay, it represents a promising candidate to study in the differentiation of cardiomyocytes from human embryonic stem cells (hESC). Even though *SMG6* does not seem to be obvious as an essential cardiovascular gene, it has been linked to coronary artery disease (van der Harst and Verweij, 2018) and the identified regulatory element might be crucial to facilitate differentiation from a naïve to a more specialised cardiovascular lineage.

Functional REs, *i.e.* able to drive gene expression, are not necessarily sequence conserved. However, we hypothesised that ultra-conservation plays an important factor when testing human RE in fish which ultimately might lead to a higher probability to initiate reporter gene expression. It remains to be investigated if REs which were not able to drive gene expression in medaka were missing crucial TFBS or were not functionally conserved from fish to humans and therefore not being recognized by the endogenous transcription machinery. In order to obtain a set of highly conserved cardiac specific CREs, we generated a new pipeline and obtained 7 sequences which were highly conserved from fish to human. These REs were also associated with genes implemented in cardiovascular development (Brady *et al.*, 2014; Chang *et al.*, 2013; Desjardins and Naya, 2016; Gaborit *et al.*, 2012; Hille *et al.*, 2016; Madonna and De Caterina, 2009; Al Turki *et al.*, 2014). Interestingly, this new set of REs also contained the previously

identified region of *enFLT1*, highlighting the likely importance of the putative enhancer. A detailed description of the enhancer and its function will be given in the next chapter.

Taken together, we were able to identify putative regulatory elements implicated in heart development via bioinformatic mining. In order to test the sequences efficiently, medaka proved to be an important tool to assess these sequences *in vivo*. One might argue that a greater number of non-conserved candidates would have been able to drive gene expression in a mammalian system as it has been successfully shown in other studies (Visel *et al.*, 2007). However, given the low conservation rate of REs between mouse and human, it needs to be determined if the non-conserved sequences identified here would have been more efficient in a murine background, which will be the focus of future work.

4 Chapter 4: *enFLT1* – a regulator of gene expression during angiogenesis

4.2 Background

Assessing the function of regulatory elements can be difficult due to the size of the genome and the complexity of gene regulation. To identify whether a sequence is truly able to regulate gene expression, enhancer necessity and sufficiency should be determined. Our initial set of cardiac specific REs has provided us with candidates that were able to drive gene expression in vivo. One of those candidates revealed an enhancer associated to the gene *FMS Related Receptor Tyosine Kinase 1* (*FLT1*) or *Vascular Endothelial Growth Factor Receptor 1* (Fong *et al.*, 1995). FLT1, together with VEGFR2, belongs to the VEGFR family and binds ligands such as VEGF-A, PIGF and VEGF-B and can be divided into four major domains (Shibuya *et al.*, 1990): an extracellular domain, a transmembrane domain, a tyrosine kinase domain and carboxy terminal region. FLT1 has important functions in angiogenesis, proliferation and differentiation and plays an important role in the formation of inflammatory diseases, cancer and atherosclerosis (Shibuya, 2006). FLT1 can be found in two major isoforms, a transmembrane bound protein and a soluble protein which is missing the transmembrane binding domain and kinase domain, but is still able to bind ligands such as VEGF-A. Therefore, FLT1 is proposed to act as an antiangiogenic factor by sequestering VEGF through ligand binding as well as a positive regulatory function by acting as a regulator of migration and proliferation via its kinase activity (Yang *et al.*, 2011). VEGFR2 and FLT1 are the main contributors of endothelial cell responses. Although FLT1 expression levels can vary in endothelial cells due to its negative regulation of VEGF-A, disruptions in FLT1 still lead to defects in blood vessel developments in many different organisms including humans. FLT1 mutants in mice and fish display an overgrowth and disrupted sprouting of blood vessels (Fong *et al.*, 1995; Krueger *et al.*, 2011) due to the lack of blood vessel sprout guidance (Nesmith *et al.*, 2017) and disruption of Notch signalling normally initiated by correct FLT1 activity (Wild *et al.*, 2017). However, the effect of impaired *FLT1* regulation via specific regulatory elements in the locus has yet to be elucidated.

In the context of a model of human endothelial cell biology, a number of key assays have been developed including, *in vitro* endothelial cell differentiation assays and angiogenesis assays to study human diseases (Arnaoutova and Kleinman, 2010). Pluripotent stem cells such as human embryonic stem cells and induced pluripotent stem cells (Takahashi *et al.*, 2007) have broadened the scope of experiments to dissect endothelial biology as they introduce the capacity to genetically modify regulatory elements and expand the number of cells available for downstream experiments. Therefore, PSCs provide a viable system for the genetic dissection of gene regulation during endothelial cell differentiation and angiogenesis. Importantly, well developed angiogenesis assays allow the investigation of cell proliferation, migration as well as the formation of three-dimensional structures blood-vessel like tubes *in vitro* (Nowak-Sliwinska *et al.*, 2018).

In this chapter we focus on a conserved genomic sequence referred to as *enFLT1*, which acts as an enhancer of *FLT1*. The key question addressed in this chapter is: how does perturbation of this element alter the transcriptional regulation of *FLT1*? Due to its sequence and functional conservation, the regulatory element was able to drive gene expression of a reporter gene in medaka (*Oryzias latipes*). The deletion of the enhancer element in endothelial cells derived from human embryonic stem cells revealed a molecular disruption which overlaps with *FLT1* gene loss-of-function suggesting an important role in the regulation of *FLT1* and other pathways downstream of *VEGFR1*. These data support the hypothesis that *enFLT1* is important for blood vessel morphogenesis and development. Thus, we define a highly conserved *FLT1* enhancer and provide evidence that this enhancer plays an evolutionarily important role in the development of the cardiovascular system.

4.3 Results

4.3.1 The human enhancer of FLT1 is able to drive cardiovascular gene expression in vivo

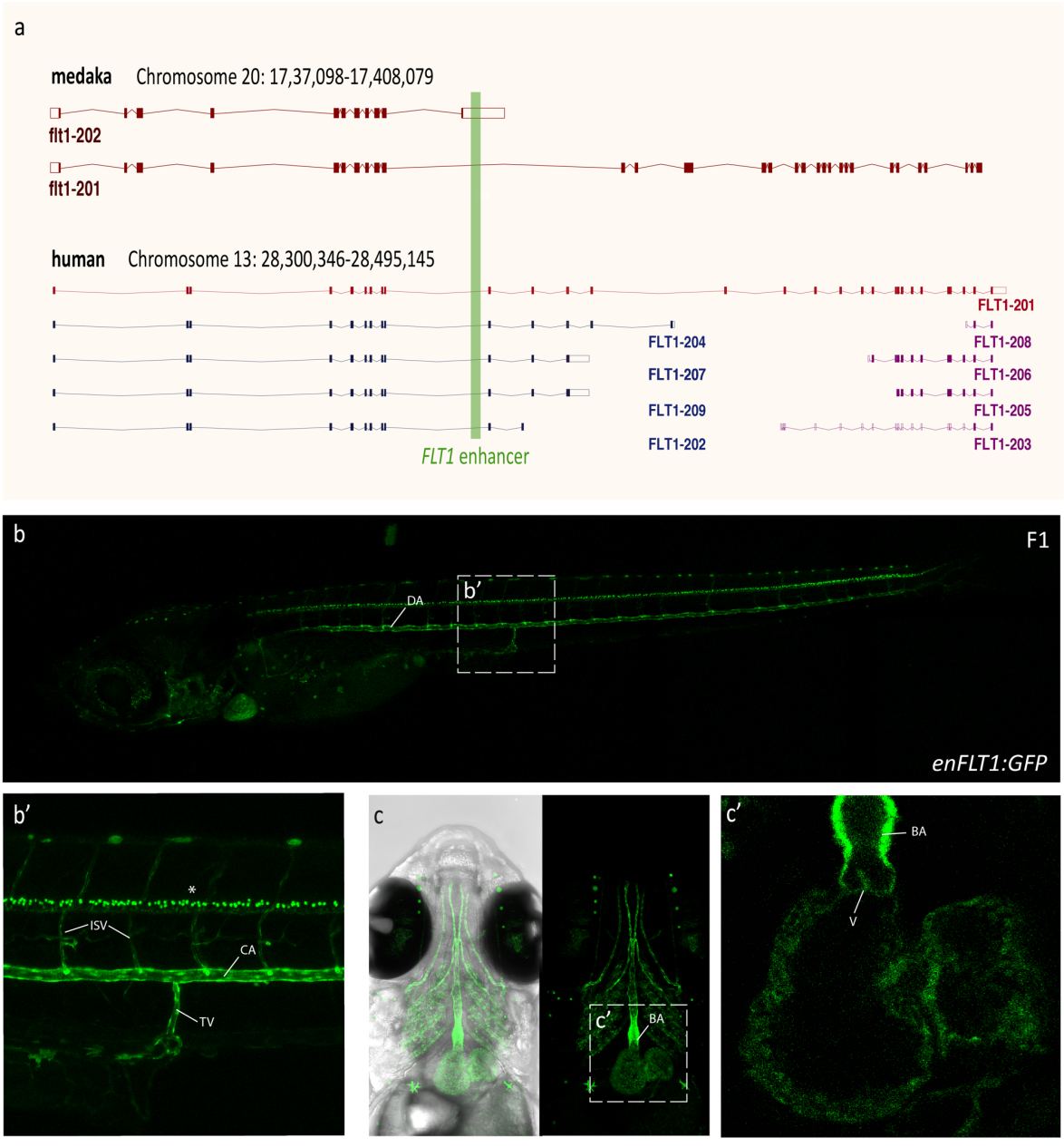


Fig. 4-1: The *enFLT1* regulatory element efficiently drives expression in the medaka cardiovascular system. a) Overview of the *FLT1* locus denoting the transcript isoforms in both humans (soluble isoforms in blue, transmembrane bound isoforms in red, other isoforms in purple) and medaka (soluble isoform: *flt1-202*, transmembrane bound isoform: *flt1-201*). The *enFLT1* RE used for *in vivo* transgenesis is highlighted in green. b) *enFLT1* drives stable reporter gene (green fluorescent protein (GFP)) expression in neurons (asterisk) and endothelium such as the dorsal aorta (DA) b' intersegmental vessels (ISV), the transverse vessel and caudal artery c) the bulbus arteriosus (BA) c') and valves (V) in the heart.

Enhancer reporter assays are an important tool to understand enhancer activity. In addition to validating the sufficiency of the enhancer sequence to drive expression of a reporter gene via a minimal promoter, we can also learn about the spatio-temporal regulation driven by the combination of transcription factors that bind the enhancer sequence. The ultra-conserved intronic enhancer of *FLT1*, *enFLT1*, is located in intron 10 in humans and in medaka, spanning both the soluble and transmembrane bound isoforms transcribed from the locus (Fig. 4-1a). As described previously, injected mosaic animals were insufficient to fully assess the transcriptional potential of *enFLT1*. Since all injected fish displayed GFP expression in a similar pattern, all GFP positive fish were crossed to wildtype fish to obtain a stable transgenic line (Fig. 4-1b) denoted *enFLT1:GFP*. Unfortunately, only one founder fish was identified with a successful germ line transmission of the construct. Characterisation of the *enFLT1:GFP* line revealed consistent expression in cardiac and endothelial tissues such as intersegmental vessels, dorsal aorta, caudal artery and transverse vessel (Fig. 4-1b&b') (Fujita et al., 2006), the outflow tract (Fig. 4-1c), blood vessels in the myocardium, the endocardium and the heart valves (Fig. 4-1c'). We also note GFP expression outside the vasculature which is consistent with reports of *flt1* in the zebrafish indicating that the enhancer may act in multiple tissues (Fig. 4-1b' asterisk) (Wild *et al.*, 2017).

This demonstrates that *enFLT1*, an ultra-conserved enhancer across all vertebrates, consisting of 358 bases is sufficient to drive GFP expression in endothelial and cardiac tissues *in vivo*.

4.3.2 *Deletion of enFLT1 perturbs cardiovascular development in vivo*

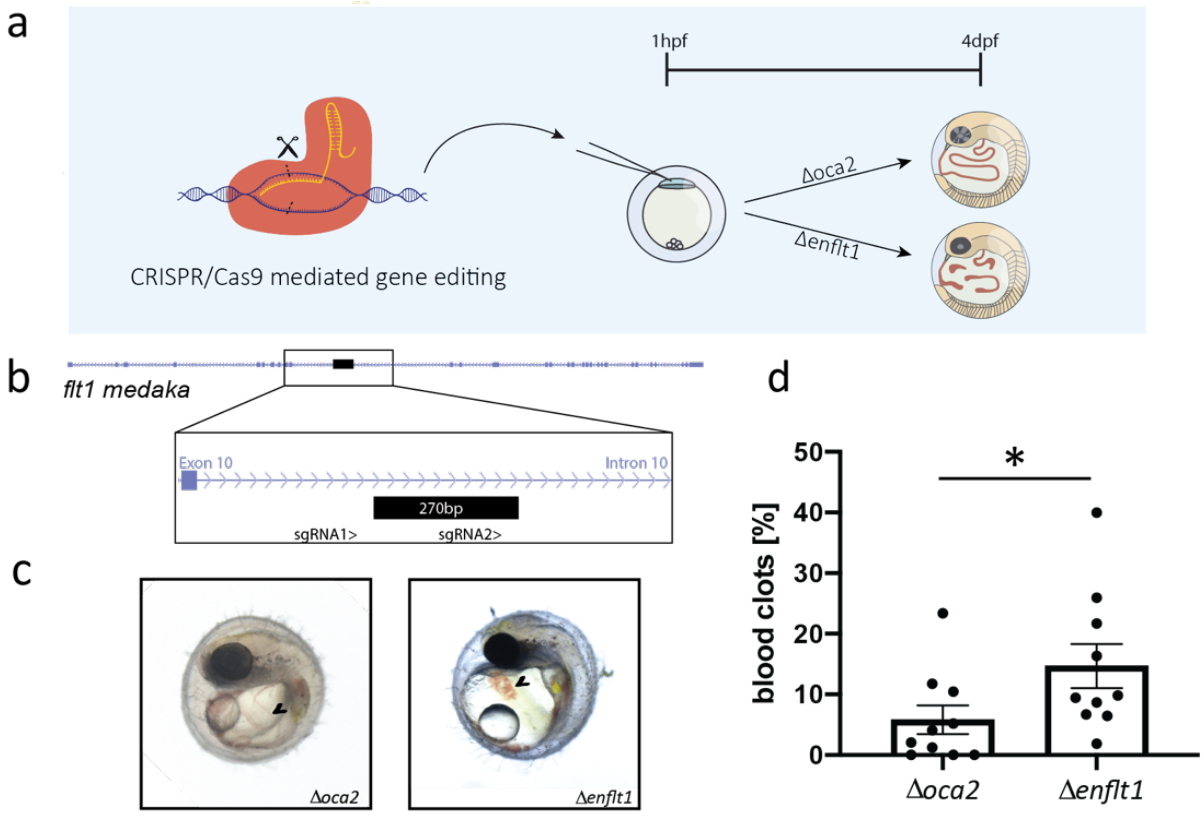


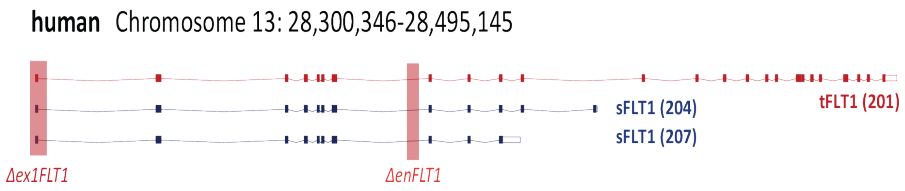
Fig. 4-2: Functional analysis of *enflt1* during angiogenesis *in vivo*. a) Overview of the experimental design. The endogenous enhancer of *flt1* was removed via CRISPR/Cas9. Targeting *oca2*, a gene responsible for the pigmentation in the eye, was used as a control. *b*) *flt1* locus and endogenous *enflt1* in medaka. 270bp were targeted with 2 sgRNAs flanking the sequence. c) Genetically modified medaka embryos targeting *oca2* as a control *(Δoca2*, left) or *enflt1* (*Δenflt1*, right). Vasculature on the yolk (arrowhead, left) was disrupted in *Δenflt1* and caused blood clots (arrowhead, right). d) Deletion of *enflt1* increased occurrence of blot clots (*Δenflt1*) compared to control injections targeting *oca2* (*Δoca2*) g) statistical analysis of blood clot occurrence. A significant difference (p= 0.0215) of blood clot numbers were observed upon deletion of *enflt1* (14.68%) compared to controls (5.81%).

To determine if loss of the conserved *FLT1* enhancer alters vasculogenesis *in vitro* we deleted the endogenous *enflt1* in medaka and assessed the effect on blood vessel development and morphology. In order to alter the regulatory sequence, which was located in intron 10 of the *tflt1* isoform and the 3'UTR of the *sflt1* isoform in medaka, we made use of the CRISPR/Cas9 system (Fig. 4-2a) to introduce a 270 bp deletion (Fig. 4-2b). The excision of the endogenous RE fragment in medaka (*enflt1*) was verified by PCR and injected embryos were analysed for defects in heart or blood vessel formation. To rule out an effect of the Cas9 enzyme or the process of microinjection itself on heart and blood vessel formation, controls were injected to target *oca2*, a gene important for pigment formation in the eye and body (Lischik *et al.*, 2019).

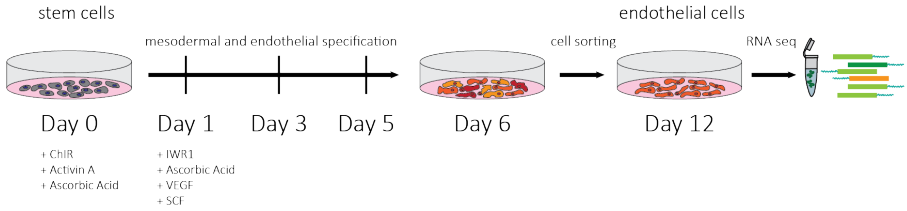
We observed blood clot formation and blood vessel disruptions when deleting *enflt1* in medaka (Fig. 4-2c). Interestingly, embryos injected with the sgRNA targeting *enflt1* resulted in a significant higher number of blood clots (14.68%) on the yolk compared to control embryos (5.81%) (p value=0.0215, Mann-Whitney test, one-tailed) (Fig. 4-2 d) suggesting a crucial role of this sequence in angiogenesis and vascularisation *in vivo* in fish.

4.3.3 *Enhancer of FLT1 is essential for pathways involved in blood vessel morphogenesis*

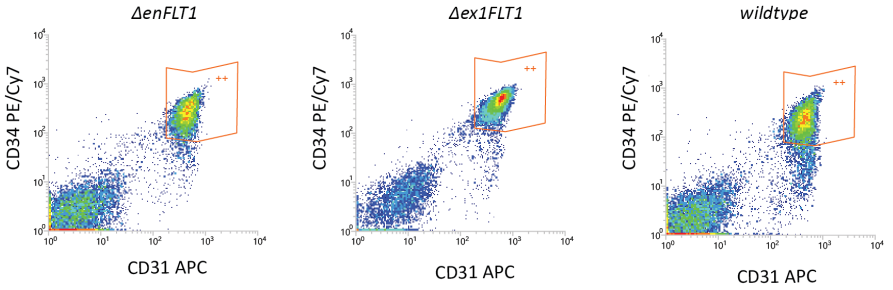
a



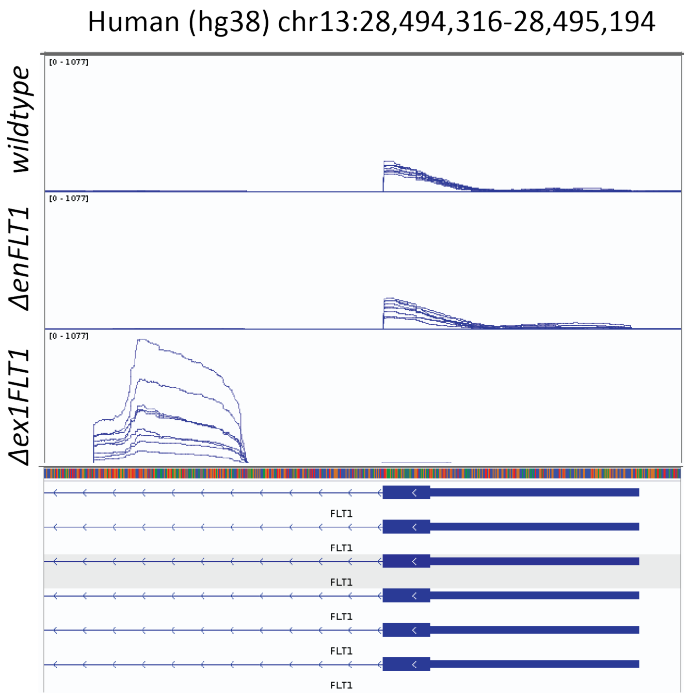
b



c



d



e

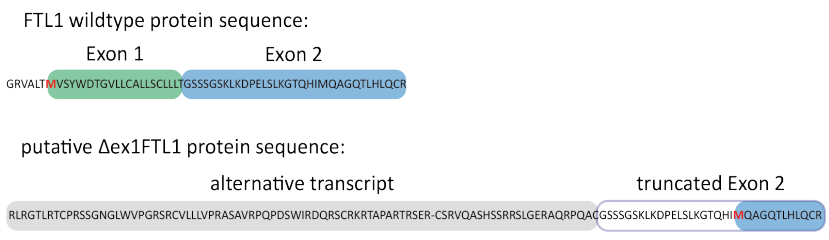


Fig. 4-3. Genetic modification of *FLT1* locus to study the role of *enFLT1* in human endothelial cells. a) Overview of the *transmembrane bound FLT1* (*tFLT1*) and *soluble FLT1* (*sFLT1*) isoforms. We generated a gene mutant (*Δex1FLT1*) and enhancer mutant (*ΔenFLT1*) cell line in human embryonic stem cells. b) human embryonic stem cells were differentiated for 6 days and sorted for CD31/CD34 expression. At day 12, endothelial cells were harvested for bulk RNA sequencing. c) FACS for CD31/CD34 positive cells (~11% of total amount of cells in all lines) at day 6 of the differentiation protocol in the three cell lines *ΔenFLT1*, Δex1FLT1 and wildtype. Only the double positive population was used for in vitro experiments. d) overview of sequencing reads of exon 1 of *FLT1* in wildtype, *ΔenFLT1* and *Δex1FLT1* endothelial cells. Expression of exon 1 has been disrupted in *Δex1FLT1* and transcription starts at an alternative start site. e) Predicted protein sequence of *FLT1* in wildtype and *Δex1FLT1* in endothelial cells. Alternative Start codon in Exon 2 might lead to a truncated protein in *Δex1FLT1*. Predicted translation start site (M) is labelled in red. Amino acid sequence prior to the translation start site are not predicted to encode any protein.

To understand the effect of the *FLT1* enhancer on the gene regulatory network and its function in humans, we deleted the RE via CRISPR/Cas9 mediated gene editing in human embryonic stem cells (hESC) (background line: H3, *NKX2-5*$^{eGFP/wt}$ (Elliott *et al.*, 2011)) to generate the *ΔenFLT1* line. In order to identify *enFLT1* specific pathways, we generated a *FLT1* null cell line by deleting exon 1, denoted *Δex1FLT1*, to serve as a control to allow us to identify the endothelial biological processes that are *FLT1*-dependent (red bars, Fig. 4-3a). We subsequently differentiated these cell lines into endothelial cells to assess the effect of enhancer and gene knockout during differentiation and angiogenesis. To understand the transcriptional consequences of deleting *enFLT1* we performed RNA sequencing on wildtype, *ΔenFLT1* and *Δex1FLT1* endothelial cells derived from hESC after 12 days of culturing (Fig. 4-3b). As part of the differentiation assay, we sorted for endothelial cells with cell type specific transmembrane markers CD31 and CD34 (Pusztaszeri *et al.*, 2006) to maintain pure endothelial cell cultures throughout this process. All differentiation assays resulted in a CD31 and CD34 double positive population at comparable proportions (~11% of total amount of cells) of total cell number across all three genotypes (Fig. 4-3c). Further, cells from all three genotypes maintained these markers while in culture.

At the molecular level, deletion of the enhancer had only a minor, not significant reduction on *FLT1* mRNA expression while the deletion of exon 1 of *FLT1* resulted in an alternative transcription start site (Fig. 4-3d) which in turn resulted in a truncated transcript and potentially protein in endothelial cells (Fig. 4-3e).

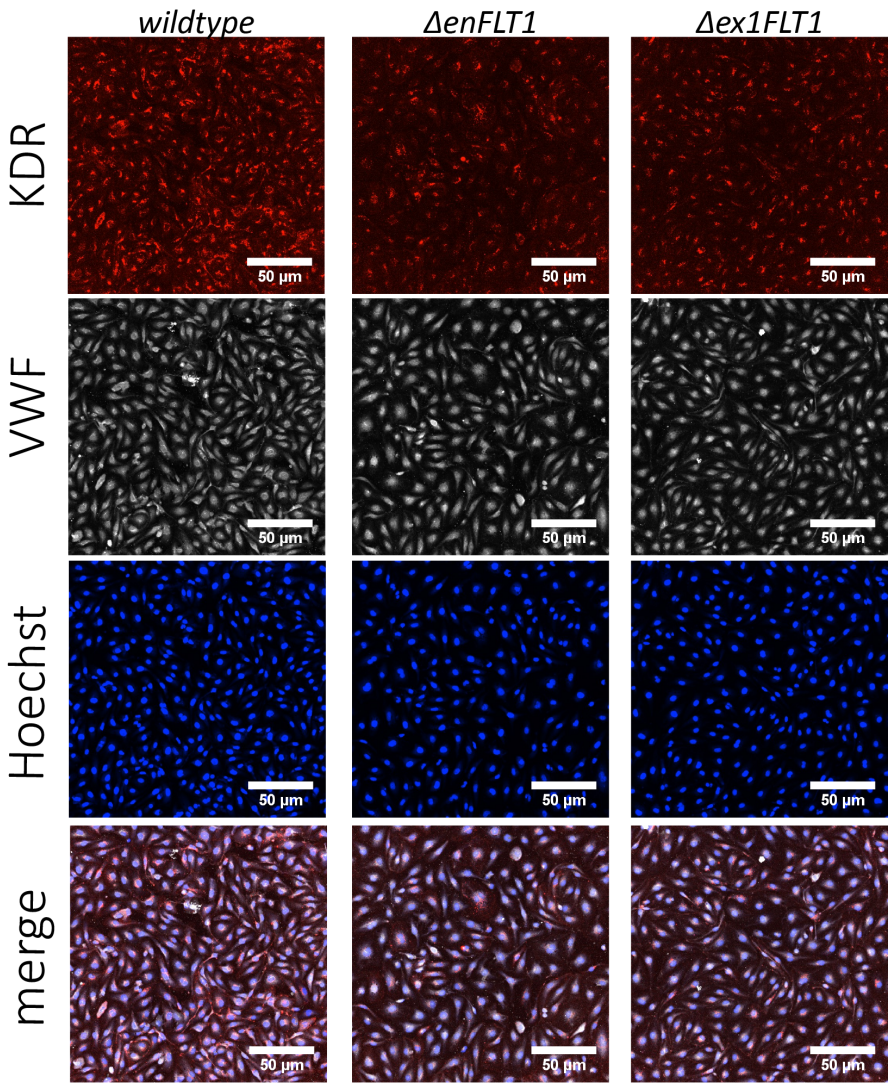


Fig. 4-4: Expression of endothelial specific makers is not perturbed in *ΔenFLT1* and *Δex1FLT* pluripotent stem cell derived endothelial cells. Immunohistochemistry of endothelial cell markers KDR (red) and VWF (grey) in *wildtype*, *ΔenFLT1* and *Δex1FLT* endothelial cells at d14 derived from sorted CD34/CD31 double positive endothelial cells. Nuclei were labelled with Hoechst (blue).

In order to characterise the endothelial cell lines *wildtype*, *ΔenFLT1* and *Δex1FLT1* morphologically, we performed immunohistochemistry (IHC) with known endothelial cell markers such as *KDR*, encoding VEGFR2, and VWF (Pusztaszeri *et al.*, 2006). We also aimed to assess the presence of FLT1 in all cell lines, however the FLT1 antibody did not produce a reliable signal in IHC (Fig. 8-1). Staining of transmembrane bound KDR and VWF was consistent across all three genotypes indicating endothelial cell differentiation did not require FLT1 (Fig. 4-4).

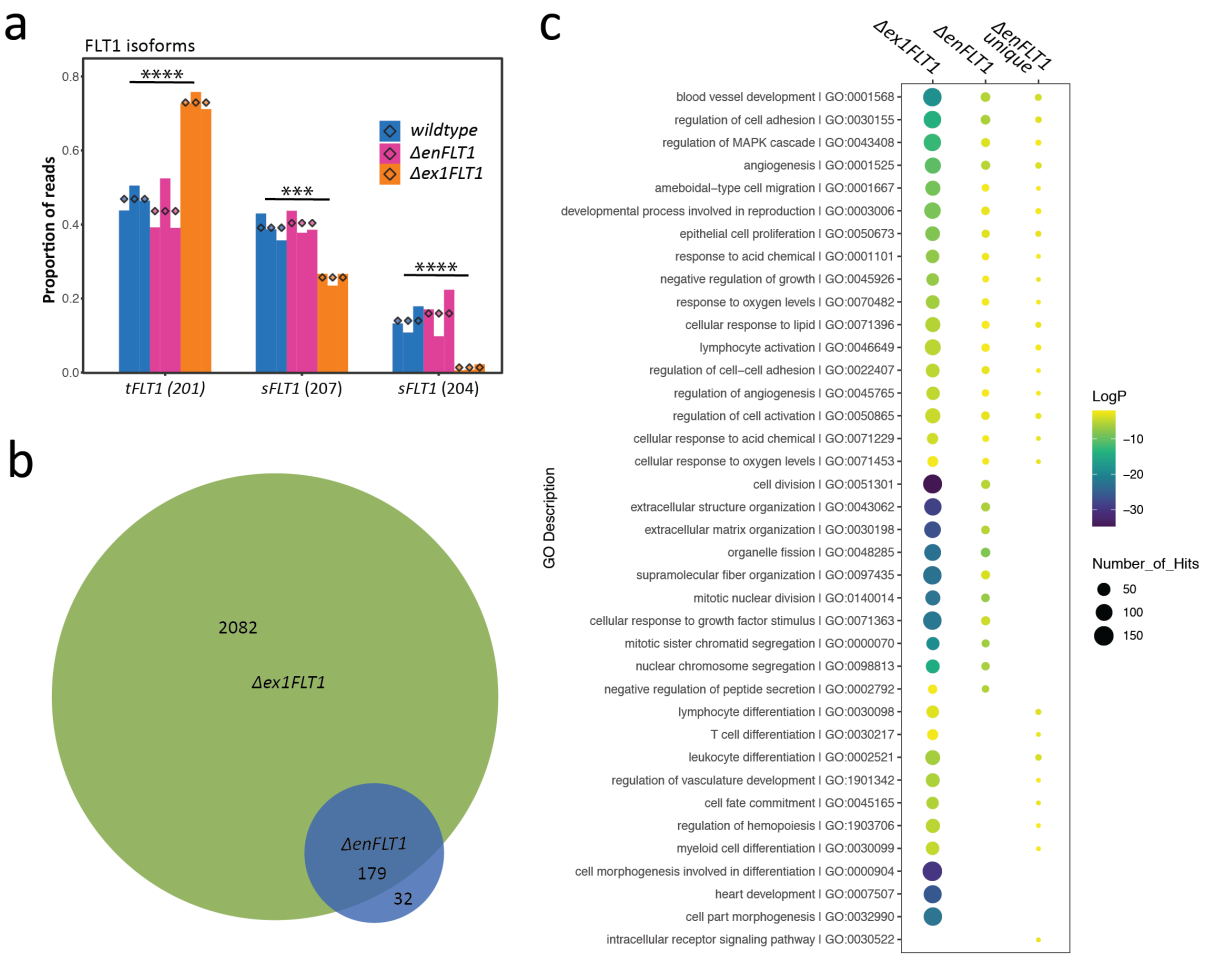


Fig. 4-5: *enFLT1* is required for regulation of *FLT1* dependent transcriptional pathways. a) Transcript reads of *transmembrane bound FLT1* (*tFLT1*) and *soluble FLT1* (*sFLT1*) isoforms in wildtype, *ΔenFLT1*, *Δex1FLT1* assessed by bulk RNA-seq (DRIMseq, Dirichlet Multinomial model). b) Venn diagram of differentially expressed genes (DEGs) (compared to wildtype) in *ΔenFLT1* and *Δex1FLT1*. The majority of DEGs in *ΔenFLT1* overlaps with genes in *Δex1FLT1*. c) Gene ontology of DEGs in *Δex1FLT1*, *ΔenFLT1* and the 32 genes unique to enhancer mutant. p-values for tFLT1-201: 3.252e-12, sFLT1-207: 1.173e-12 and *sFLT1-204*: 1.651e-4.

On a transcriptional level, when assessing the expression of different *FLT1* isoforms in *ΔenFLT1*, we could neither detect a significant decrease in *transmembrane bound FLT1* (*tFLT1*) expression (201) nor any difference in expression levels of any of the *soluble FLT1* (*sFLT1*) isoforms (204/207), suggesting a more redundant, subtle role of the CRE in *FLT1* regulation. However, deletion of exon 1 of *FLT1* significantly altered expression of both transmembrane and soluble isoforms (Fig. 4-5 a). Furthermore, we identified 211 differentially expressed genes (DEGs) as a result of deleting the regulatory element compared to cells with the intact wildtype *enFLT1*. The majority of these DEGs (179) overlapped with the 2261 identified genes which were differentially expressed in the human *FLT1* gene mutant, suggesting an essential role of the enhancer in the FLT1 pathway (Fig. 4-5b). Furthermore, gene ontology (GO) analysis of

DEGs in endothelial cells lacking the *FLT1* gene demonstrated that the main pathways affected were associated with blood vessel development, cell differentiation, cell division, heart development and extra cellular matrix organisation consistent with known biological processes regulated by *FLT1* (Chen *et al.*, 2010; Fong *et al.*, 1995). The genes shared by gene mutant and enhancer mutant human cell lines were involved in pathways such as organelle fission and extracellular structure organisation and other regulatory processes of extracellular matrix (ECM) composition. Pathways identified to be unique to differentially expressed genes in *ΔenFLT1* are associated with angiogenesis and blood vessel morphogenesis GO terms, suggesting a role of *enFLT1* in alternative pathways in blood vessel development distinct from the FLT1 pathway (Fig. 4-5c).

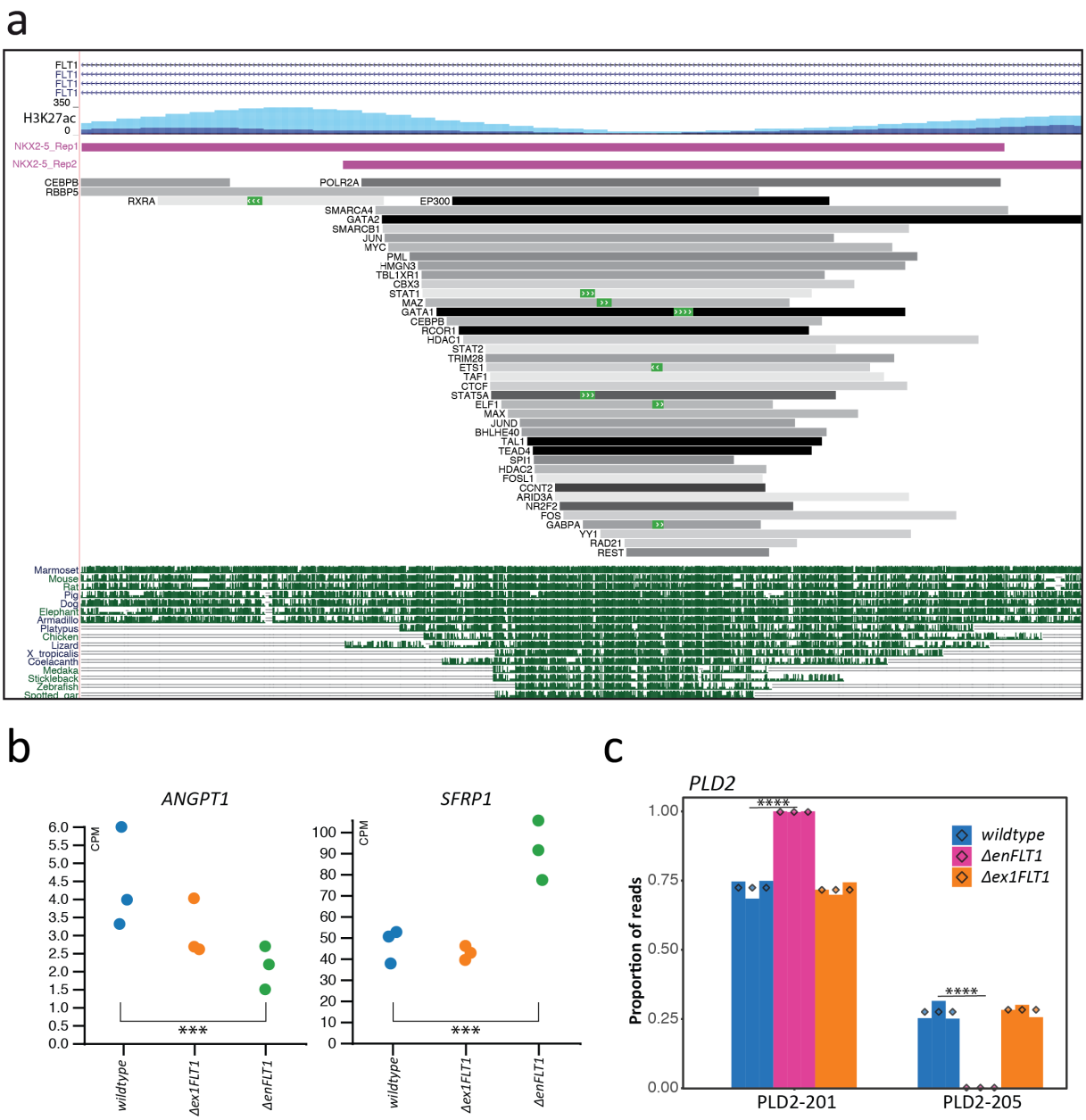


Fig. 4-6: *enFLT1* regulates alternative pathways involved in blood vessel development. a) Schematic of transcription factors that bind to *enFLT1*. Transcription factor binding overlaps with highly a conserved sequence across species and a valley between two H3K27ac activity marks suggesting open chromatin (UCSC Genome Browser). b) Two examples of differentially expressed genes involved in blood vessel development unique to *ΔenFLT1*. *ANGPT1* (left, p-value = 3.67e-4) and *SFRP1* (right, p-value = 1.81e-4). The statistical analysis was performed using RUVr package with k = 1). c) *PLD2* isoforms 201 and 205 (p-value = 9.405e-28) are affected by *enFLT1* deletion in endothelial cells (statistical analysis: DRIMseq, Dirichlet Multinomial model).

In order to understand how *enFLT1* might be regulating and buffering regulatory networks involved in blood vessel development, we focused on predicted transcription factor binding sites in the ultra-conserved sequence. In total, we identified 43 TFs important for gene regulation like p300 (Ghosh, 2020) and CTCF (Gomez-Velazquez *et al.*, 2017) as well as other cardiovascular regulators like GATA2 (Shi *et al.*, 2014) and MYC (Bywater *et al.*, 2020) (Fig. 4-6a).

Differentially expressed genes such as *ANGPT1* and *SFRP1* are examples for alternative regulators of blood vessel development distinct from VEGF (Fig. 4-6b). While we did not see any significant difference of expression levels in those genes between gene mutant and wildtype, we did identify that *ANGPT1* was significantly decreased in *ΔenFLT1* while *SFRP1* was overexpressed. Another example of how the enhancer might be acting separately to *FLT1* is *PLD2*, a gene association to ECM organisation, cell cycle control and transcriptional regulation (McDermott *et al.*, 2020). Previous studies have already identified *PLD2* as an essential contributor for endothelial growth (Ghim *et al.*, 2014) and that it is tightly linked to VEGF signalling (Lee *et al.*, 2016). The deletion of exon1 of *FLT1* had no effect of the expression of *PLD2* isoforms. However, the deletion of *enFLT1* altered expression of *PLD2* isoforms 201 and 205 significantly. This data provides a basis for further exploration of the role of specific transcription factors in the differential gene expression observed in *enFLT1* mutant endothelial cells.

4.3.4 *Deletion of enFLT1 does not impair angiogenesis in vitro*

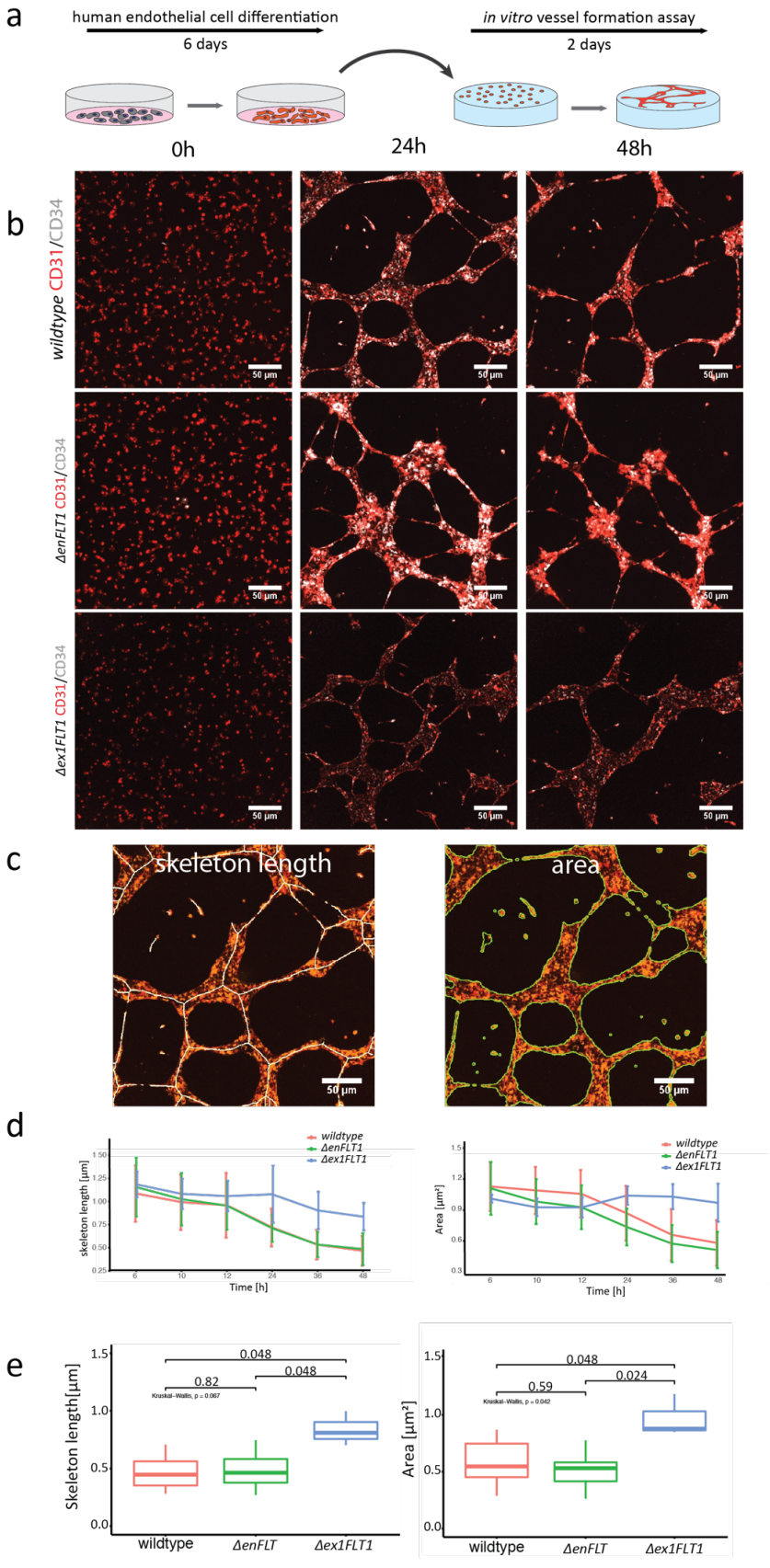


Fig. 4-7: Functional analysis of *enFLT1* during angiogenesis. a) Human embryonic stem cells were differentiated within 6 days and subsequently used for angiogenesis assays that were live imaged over 48 hours. b) Angiogenesis assays of wildtype, *ΔenFLT1* and *Δex1FLT1* cell lines at 0h, 24h, and 48h. All cell lines stayed positive for CD31 (red) and CD34 (grey) for 48h. *Δex1FLT1* vessels maintained stronger connections and collapsed over time forming thicker vessels (pictures representing the phenotype were taken from different differentiation assays and days). c) CellPathfinder analysis of vessels based on skeleton length and area. d) Statistical analyses of normalised skeleton length and area covered over the course of the experiment (Kruskal-Wallis test and Wilcoxon Rank Sum Test). e) *Δex1FLT1* cells displayed longer connections and covered a significant larger area compared to wildtype and *ΔenFLT1* cells. Image analysis and statistical analysis were performed in collaboration with Holly Voges and Neda Rahmani Mehdiabadi.

In a developing organism, endothelial cells form blood vessels during vasculogenesis and angiogenesis to establish a fully functional vascular system. Similar processes can be recapitulated *in vitro* using endothelial cells derived from pluripotent stem cells. To assess the potential of *ΔenFLT1* and *Δex1FLT1* endothelial cells to form tubes *in vitro*, we performed angiogenesis assays (Arnaoutova and Kleinman, 2010). For connections to form, cells were plated on a thick gel matrix, labelled with anti-human CD31 and anti-human CD34 antibodies and imaged over the course of 48 hours (Fig. 4-7a). While *wildtype* (*NKX2-5*$^{eGFP/wt}$) and *ΔenFLT1* endothelial cells formed tubes within 10h and then stopped and coalesced, we observed a continuation of angiogenesis and proliferation in *Δex1FLT1* cells until 48h (Fig. 4-7b). To assess difference in tube formation potential across genotypes, we analysed skeleton length, area, junction count and branch count over the course of 48h (Fig. 4-7c). *ΔenFLT1* cells showed no significant difference compared to wildtype cells in any of the categories assessed. However, we did see a significant change in area and skeleton length when comparing *Δex1FLT1* with the other cell lines (Fig. 4-7d) with the strongest significant difference at 48h (Fig. 4-7e), indeed suggesting a disruption of the *FLT1* pathway during angiogenesis consistent with published data on *FLT1* knockdown in HUVECS and *Flt1*$^{-/-}$ mutant in endothelial cells derived from murine embryonic stem cells (Nesmith *et al.*, 2017). This data suggests that the enhancer of *FLT1* is not essential for endothelial tube formation *in vitro* and that other regulatory elements act to buffer *FLT1* expression from the loss of *enFLT1*. Nevertheless, given the altered transcriptional profile observed in *ΔenFLT1* endothelial cells it remains possible that this enhancer is required for normal development.

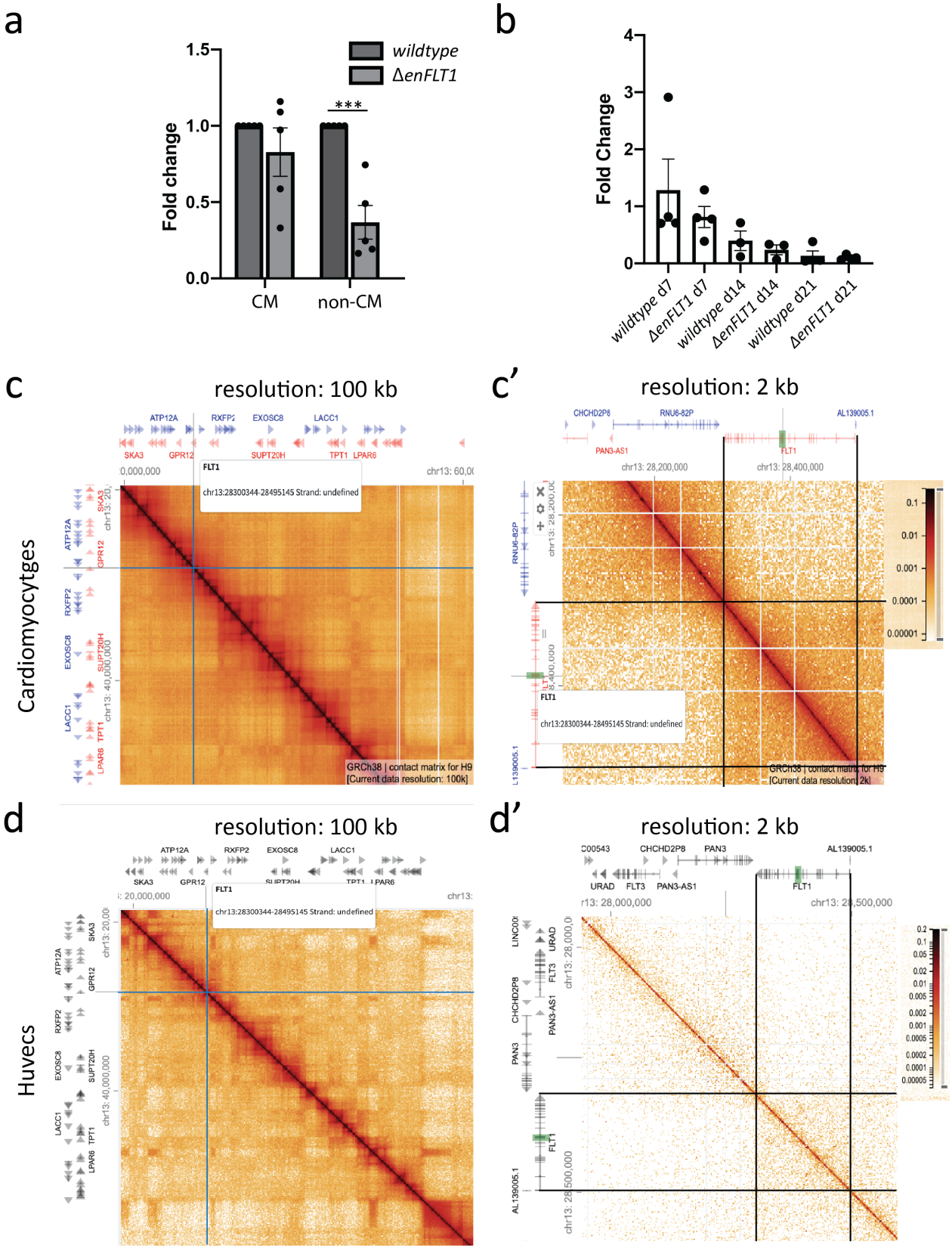


Fig. 4-8: Altered expression pattern of *FLT1* in *ΔenFLT1* cardiac and chromatin arrangement at the *FLT1* locus in pluripotent stem cell derived cardiomyocytes and HUVEC cells. a) *FLT1* expression is significantly lower in the non-cardiomyocyte population of a cardiac differentiation assay (p-value = 0.0004, t-test, two tailed). b) Temporal regulation of *FLT1* in wildtype and *ΔenFLT1* endothelial cells over the course of 21 days in culture. c) HiC interaction map of targets interacting with *enFLT1* in cardiomyocytes at 100kb resolution and c') 2kb resolution. d) HiC interaction map of targets interacting with *enFLT1* in human umbilical vein endothelial cells (HUVECS) at 100kb resolution and d') 2kb resolution. Blue and black lines indicate the

location of *FLT1* in 100kb and 2kb resolution respectively (Dekker *et al.*, 2017; Rao *et al.*, 2014; Zhang *et al.*, 2019b). Images were provided by Gulrez Chahal.

Many different cell types are required for the correct spatio-temporal establishment of the developing human heart. Since *enFLT1* has been shown to be implicated in cardiac development via the binding of the transcription factor NKX2-5 and other important cardiac regulators, we wanted to assess the effect of *enFLT1* deletion in cardiomyocytes and other cardiac cell types. We determined *FLT1* expression via real time quantitative PCR (qPCR) in the cardiomyocyte and non-cardiomyocyte fraction of a cardiomyocyte differentiation assay (described in the method section (Anderson *et al.*, 2018)) in *wildtype* and *ΔenFLT1* cells. We could not detect a significant difference in *FLT1* expression between wildtype and enhancer mutant in cardiomyocytes, as all cell lines resulted in 50 to 60% cardiomyocytes per assay. However, mRNA levels of *FLT1* were significantly lower in *ΔenFLT1* compared to *wildtype* in non-cardiomyocyte cells consisting of cardiac fibroblasts and endothelial cells (Fig. 4-8a). To assess temporal regulation of *FLT1* in endothelial cells, we also performed gene expression assays in *wildtype* and *ΔenFLT1* cells over the course of 21 days. *FLT1* expression levels decreased in both cell lines when cultured for 21 days. However, *FLT1* expression in *ΔenFLT1* was not significantly different compared to wildtype expression levels in endothelial cells when assessed at three different timepoints (7d, 14d and 21d in culture), supporting RNA sequencing data and the notion that *enFLT1* deletion does not have a strong effect on *FLT1* gene expression (Fig. 4-8b). To determine the interaction of *enFLT1* with other regions within the locus and thus assess how the RE might be regulating *FLT1* or neighbouring genes, we used publicly available datasets of chromatin confirmation capture experiments (Hi-C) (van Berkum *et al.*, 2010) in cardiomyocytes (Fig. 4-8c and c') and human umbilical vein endothelial cells (HUVECS) (Dekker *et al.*, 2017; Rao *et al.*, 2014; Zhang *et al.*, 2019b) (Fig. 4-8d and d'). Overall, there were more interactions detectable within the *FLT1* locus in cardiomyocytes compared to endothelial cells, suggesting that *enFLT1* interacts with the *FLT1* promoter in cardiac cells. We also overlapped this Hi-C data obtained from cardiomyocytes and HUVECS with differentially expressed genes in *ΔenFLT1* to determine if there was any direct physical interaction between the enhancer and any of the target genes. However, we could not find any evidence of such interaction, suggesting an indirect role of the enhancer on the transcriptional network.

4.4 Discussion

FLT1 is an important regulator of blood vessel development, cell proliferation, migration, differentiation and cell survival. Loss of FLT1 or perturbation in *FLT1* regulation has been implicated in blood vessel defects in human, mouse, and zebrafish (Fong *et al.*, 1995; Krueger *et al.*, 2011; Nesmith *et al.*, 2017). Although *FLT1* has been very well studied, its regulatory elements and transcriptional regulation is less well understood. However, a few studies have highlighted the importance of *cis*-regulatory elements of *FLT1* and suggested that disruptions in these non-coding regulators affect gene expression of *FLT1*, and its isoforms, and can ultimately lead to disease. (Kikas *et al.*, 2020; Menendez *et al.*, 2006; Owen *et al.*, 2014; Thomas *et al.*, 2010).

We aimed to functionally test regulatory elements which were linked to cardiovascular development and disease in humans and identified, among other regulatory elements, an enhancer in intron 10 of *FLT1*. This enhancer has been identified to be well conserved across species and the corresponding zebrafish sequence has been used in Bussmann *et al.* where it was shown that the endogenous zebrafish *flt1* enhancer in combination with the endogenous *flt1* promoter is able to drive reporter gene expression throughout the developing vasculature *in vivo* (Bussmann *et al.*, 2010). Here we demonstrate that the human *enFLT1* was also able to drive stable reporter gene expression in combination with a synthetic minimal promoter (SCP1) (Juven-Gershon *et al.*, 2006) in endothelial tissues, suggesting a functionally as well as sequence conserved mechanism of transcription factor binding and transcription initiation. Although GFP expression was detected in blood vessels, the heart and neurons, the exact cell types expressing the transgene in these tissues is unknown and needs to be further assessed. Future experiments such as angiograms and co-staining with myocardial, endothelial and neuronal markers will be performed to characterise their true identity within the cardiovascular system and other tissues.

Since the role of the regulatory element and connection to human disease has not been elucidated, we further investigated the regulatory potential of this human enhancer and determined the regulatory mechanism of *FLT1* and its effect on vessel formation *in vivo* and tube formation *in vitro*. *in vivo* knockout of the endogenous conserved enhancer in medaka led to an increase of blood vessel disruptions on the yolk resulting in blood clot formation, hinting

towards an important role of this sequence located in the 3'UTR of *sflt1* in vessel morphogenesis in fish. Therefore, deletion of the RE might lead to an altered expression or early degradation of the *sflt1* isoform, potentially causing a blood vessel collapse. Due to the function of *flt1* in the formation of the endocardium, heart development could also be impaired resulting in altered blood flow or coagulation abnormalities due to a reduction in cardiac function. However, whether the blood clot phenotype is due to a defective vasculature or altered haematopoiesis needs to be investigated molecularly and phenotypically in a homozygous knockout mutant line and will be part of future experiments. In human endothelial cells derived from pluripotent stem cells, the deletion of the enhancer had only a slight but no significant effect on expression levels of *FLT1*. In order to determine the true nature of the mutant *FLT1* transcript and FLT1 protein in *Δex1FLT1,* further studies are required to investigate the molecular phenotype. Protein analysis via western blotting of endothelial cell lysates, flow cytometry using reliable FLT1 antibodies and immunohistochemistry would provide more insights. Previous studies have shown that the deletion of *cis*-regulatory elements, despite their capacity to drive gene expression, may only lead to subtle effects on transcription levels or phenotype. This occurs due to the fact that redundant and shadow enhancers often act additively to provide genetic and phenotypic robustness during the development of an organism (van den Boogaard *et al.*, 2019; Cunningham *et al.*, 2018; Letelier *et al.*, 2018; Osterwalder *et al.*, 2018; Sarro *et al.*, 2018). Therefore, we can speculate that a compensatory mechanism exists which involves other regulatory elements in the locus to maintain a stable *FLT1* expression. Although we can be confident that this enhancer is regulating *FLT1*, we cannot rule out that it might be regulating other vasculature related genes in the vicinity by switching targets in a spatiotemporal manner. Additional data suggests (Dekker *et al.*, 2017; Rao *et al.*, 2014; Zhang *et al.*, 2019b) that this specific enhancer might be active in cardiac muscle cells and poised or inactive in endothelial cells (HUVECS). Strikingly, the evidence that there is an active promoter-enhancer interaction in cardiomyocytes and its lack in endothelial cells is inconsistent with our data. We could not determine a change in *FLT1* expression upon deletion of the enhancer. This could potentially be explained by the tight temporal regulation of REs and different maturation statuses of the cardiomyocytes that were investigated. Further expression analysis experiments investigating different time points of cardiomyocyte differentiation would help to provide more insights. We also did not see a change in *FLT1* expression in the enhancer mutant cell line when assessing mRNA levels in endothelial cells over the course of 21 days.

However, while the deletion of the enhancer element in cardiomyocytes did not lead to a significant decrease in *FLT1* expression, we did see a significant decrease of *FLT1* expression in the non-cardiomyocyte population of the differentiation assay. This could directly indicate that we are either dealing with two different endothelial cell populations in a cardiac differentiation assay compared to an endothelial cell assay, or that *enFLT1* is active in cardiac fibroblasts, another cell type present in the non-cardiomyocyte fraction, which could explain a decrease in *FLT1* expression upon deletion of the RE. It remains to be determined why we are able to detect an effect on a transcriptional level in the non-CM fraction. Future experiments looking at distinct cell types will clarify this finding. Despite no obvious effect on *FLT1* expression, RNA sequencing of *ΔFLT1* and *ΔenFLT1* revealed an overlap of 179 differentially expressed genes suggesting that the RE in intron 10 is indeed involved in the *FLT1* pathway, strengthening its role in providing robustness during vessel morphogenesis. In our angiogenesis assays, we see vessel collapse and a reduced rate of retraction of vessels in the human gene mutant cell line. These findings are consistent with previous studies where mutant cell lines and mouse models highlighted a clear role of *FLT1* during vessel interaction (Nesmith *et al.*, 2017). However, we were not able to observe any significant changes in the enhancer knockout in regard to angiogenesis *in vitro*. To exacerbate a potential phenotype of *ΔenFLT1*, future work could include investigating the role of this enhancer in induced pluripotent stem cell lines of patients with known congenital cardiovascular diseases. In addition to vessel formation assays, we also performed wound healing assays to determine differences in proliferation and migration in the enhancer and gene mutant compared to wildtype cell lines (data not shown). Unfortunately, these assays were technically challenging, and further optimisation is necessary to obtain quantifiable data. Additionally, further experiments are required to assess these differences using techniques such as flow cytometry or immunohistochemistry with proliferation markers. However, *in vitro* experiments might be insufficient to detect the subtle effects of the enhancer deletion. Alternatively, more *cis*-regulatory elements of the *FLT1* locus may need to be deleted to elucidate the function of each regulatory element and subsequent chromatin capture experiments are required to observe the resultant alterations in 3D chromosome architecture. Furthermore, an *in vivo* mammalian system may provide more insight into whether its function is necessary for organogenesis than a more artificial *in vitro* angiogenesis assay. In addition, deleting the regulatory element in a more permissive genetic background, for example disease specific lines with mutations in genes within the *FLT1* network, might lead to a more robust

phenotype in the vessel formation assays. Interestingly we do see an effect of the deletion of the RE in medaka *in vivo*. In fish, the conserved enhancer lies in the 3'UTR of the *sflt1* isoform, which might lead to premature degradation of the mRNA and therefore might have a stronger effect on vessel morphogenesis. It is interesting to note that in humans, the enhancer does not fall in any untranslated region of any of the isoforms and, thus, deletion does not interfere directly with mRNA stability or degradation.

Although the RE might not have a direct impact on *FLT1* expression, it is still implicated in the FLT1 pathway downstream and plays a role during blood vessel development. We also observed 33 differentially expressed genes that were restricted to the enhancer mutant of *FLT1*. This suggests an indirect role of the enhancer in parallel pathways to *FLT1* that are also involved in blood vessel morphogenesis and cardiovascular development via the regulation of the extracellular matrix composition. Among these genes, *ANGPT1* is involved in angiogenesis and disruptions in the gene have been associated with angioedema (Bafunno *et al.*, 2018; d'Apolito *et al.*, 2019).

As an antagonist of angiogenesis, FLT1 has been predicted to be a functional partner of ANGPT1 and both proteins share similar signalling cascades such as Ras signalling pathway and MAPK signalling pathway (https://string-db.org/network/9823.ENSSSCP00000009946). As a regulator of WNT signalling, another SFRP1 might be implemented in the WNT-depended suppression of angiogenesis (Stefater *et al.*, 2011) and be therefore a potential indirect partner of FLT1. However, future studies are required to support this hypothesis. Since we do see effects on isoforms expression of other genes such as *PLD2*, it is not clear how such changes in non-related pathways occur. The related function and implementation in the same pathways could suggest a likely regulatory function of the *FLT1* enhancer on those genes, at least indirectly via other intermediate factors. However, in order to fully understand the interaction network of the enhancer with other genes, a more detailed 3D chromatin interaction map of the enhancer and related REs is required to identify direct physical interaction partners.

In this chapter we have provided a functional and molecular basis to understand the complex regulation of the *FLT1* locus. Additionally, we have shown that enhancer necessity and sufficiency need to be separately assessed. This study describes the function of the enhancer of *FLT1* located in intron 10 and highlights how regulatory elements provide robust reinforcement of the transcriptional regulation of *FLT1*. In conclusion, we show that

perturbations in the sequence of a regulatory element of *FLT1* affects other pathways, such as the ANGPT1 pathway. Investigating the role of *enFLT1* in similar biological processes related to blood vessel development will ultimately assist in annotating variants that fall into this enhancer and link them to enhanceropathies.

5 Chapter 5: *cis*-Regulation in human cardiomyocytes

5.1 Background:

Congenital heart disease is one of the major causes of infant death in the world (Virani *et al.*, 2020). To study developmental defects in humans that present at birth is a challenging endeavour and mechanistic studies have relied upon model systems. Animal models, like mice and fish, are frequently used to address these questions on a developmental level. However, significant differences between these vertebrates and human can make drawing direct connections about disease pathogenesis problematic. In order to recapitulate pathogenic processes in a human background the field has increasingly turned to pluripotent stem cells to study disease mechanisms (Anderson *et al.*, 2018; Gifford *et al.*, 2019). Stem cell derived heart muscle cells or cardiomyocytes have been used to research human heart disease since the early 2000s (Kehat *et al.*, 2001; Mummery *et al.*, 2002). and the ability to make patient-specific iPSCs, along with improved ease of genetic engineering with CRISPR/Cas9, has opened up avenues to make human models of human disease (Takahashi *et al.*, 2007). The process of differentiation of pluripotent stem cells into cardiomyocytes first starts with the induction of cardiac mesoderm, followed by differentiation into cardiac progenitors and is finalised by the maturation of the heart muscle cells. A differentiation assay *in vitro* will contain different cell types like cardiomyocytes, smooth muscle cells and endothelial cells (Mummery *et al.*, 2012). In order to understand the mechanisms of heart disease modelled *in vitro*, cardiomyocytes can then be characterised functionally by morphology, electrophysiological measurements (Kehat *et al.*, 2002) and other physiological functions like Ca^{2+} signalling (Fearnley *et al.*, 2011).

In recent years, the function of non-coding *cis*-regulatory elements, and gene regulation more broadly, has become a significant focus in the cardiovascular field. For example, with the rise of genome wide association studies and disease causative variants being discovered in non-coding regions, the notion that even small disruptions in the gene regulatory network can lead to enhanceropathies has gained prominence (Villar *et al.*, 2020). Several studies have focused on trying to understand chromatin architecture and its changes during the process of differentiation from naïve to a more fate restricted state (Arvanitis *et al.*, 2020; Montefiori *et al.*, 2018). Interestingly, it has been shown that many cardiac specific genes shift from a

repressed to an active state, which correlates with the start of gene expression and the opening of chromatin (Bertero *et al.*, 2019) early in the differentiation process.

This chapter will focus on several putative cardiac specific *cis*-REs that were identified by our in-house bioinformatic pipeline and tested in cardiomyocytes derived from human pluripotent stem cells. We generated knock out mutants of cardiac enhancers to assess whether there was an effect on cardiomyocyte development and/or function. Enhancer knockouts were also analysed for cell morphology changes and for altered target gene expression. Chromatin confirmation capture analysis of these candidate regions revealed chromosome looping to possible targets and perturbations of the most promising sequences, *enC5ORF38* and *enKCNH2*, resulted in target gene expression changes. Furthermore, removal of an ultra-conserved sequence in the enhancer of *PPARGC1A* in medaka resulted in severe developmental defects and malformations. Thus, we identify and characterise cardiac specific REs and present evidence suggesting these regulatory sequences may function during cardiac specification and differentiation.

5.2 Results

5.2.1 Investigating the role of cardiac specific CREs in vivo & in vitro

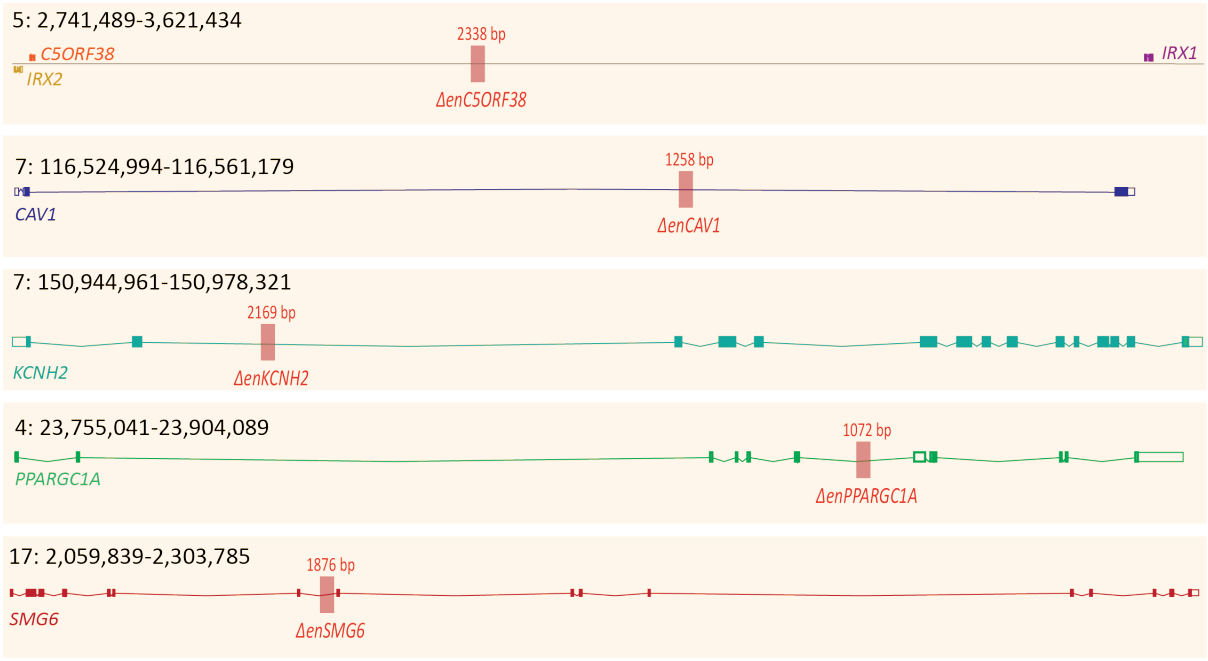


Fig. 5-1: Chromosomal localization of putative cardiac specific *cis*-REs. Five unique REs identified via bioinformatic mining and their putative target genes. REs are indicated by a red rectangle. Each RE was removed in hESCs via CRISPR/Cas9 gene editing to generate the enhancer knock-out lines *ΔenC5ORF38*, *ΔenCAV1*, *ΔenKCNH2*, *ΔenPPARGC1A* and *ΔenSMG6*.

To functionally characterize putative regulatory elements involved in cardiac development and disease, we generated enhancer knockout lines in human embryonic stem cells *(NKX2-5$^{eGFP/wt}$)*. The only intergenic candidate, a 2338 bp long putative regulatory element associated to the gene *C5ORF38* by GREAT was termed *enCORF38* due to its close proximity to the gene. This sequence was located in a gene desert, surrounded by cardiac specific Iroquois homeobox genes *IRX1* and *IRX2* (Kim *et al.*, 2012). All other sequences which were identified via bioinformatic mining, were located in intronic regions, like *enCAV1* which was located in intron 4 of the gene *CAV1*. We also identified a putative regulatory element in intron 2 in the cardiac specific gene *KCNH2*. Furthermore, we identified a 1072 bp putative enhancer in intron 9 of the gene *PPARGC1A* as well as a sequence located in intron 2 in *SMG6*. All putative regulatory elements were excised using the CRISPR/Cas9 system to generate the enhancer knockout lines, *ΔenC5ORF38*, *ΔenCAV1*, *ΔenKCNH2*, *ΔenPPARGC1A* and *ΔenSMG6* (Fig. 5-1).

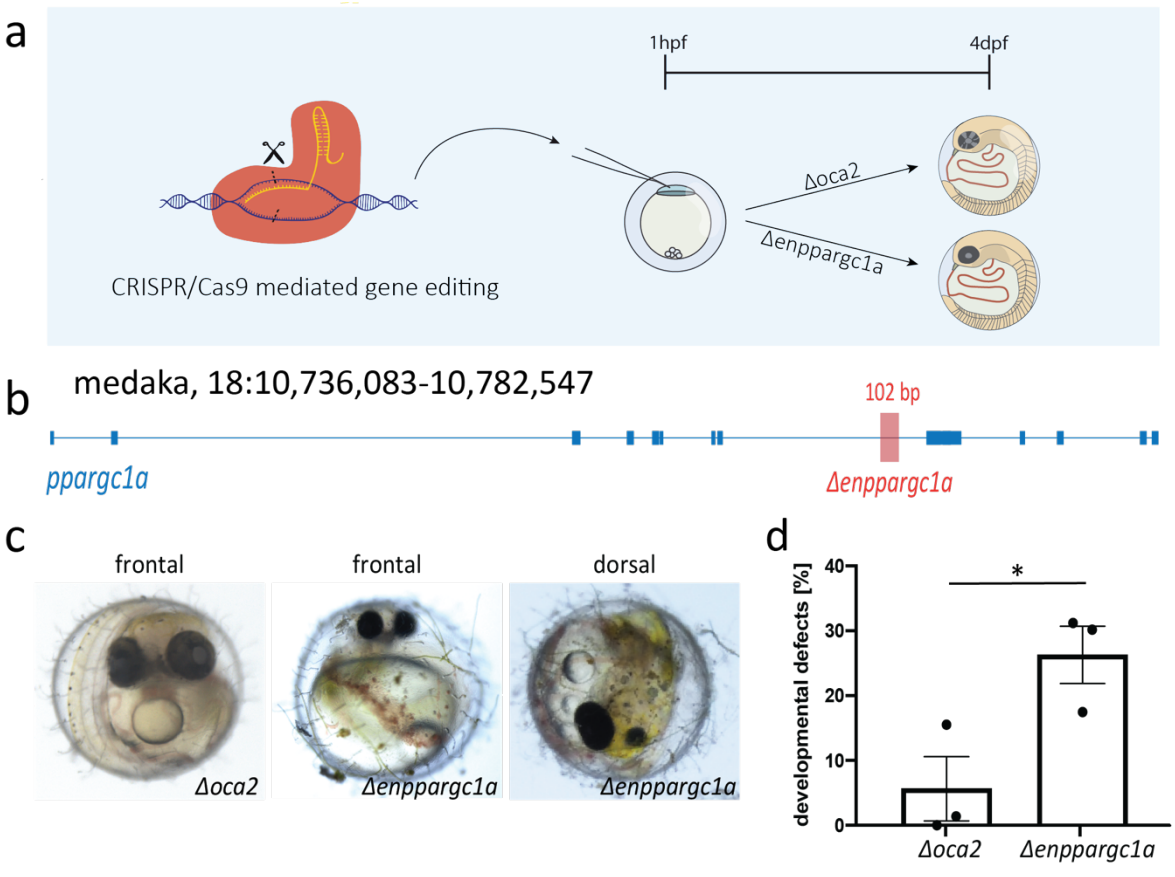


Fig. 5-2:Functional analysis of *enppargc1a in vivo*. a) Overview of the experimental set up. The endogenous enhancer of *pparc1a* was removed via CRISPR/Cas9. Targeting *oca2*, a gene responsible for the pigmentation in the eye, was used as a control. *b*) *ppargc1a* locus and endogenous *enppargc1a* in medaka. 102bp were targeted with 2 sgRNAs flanking the sequence. c) genetically modified medaka targeting *oca2* as a control (left, frontal) or *Δenppargc1a* embryos (middle, frontal and right, dorsal view). d) deletion of *enppargc1a* resulted in a significant increase (p=0.0360) in developmental defects, such as smaller bodies, deformed organs and underdeveloped embryos (26.25 %) compared to *Δoca2* control embryos (5.64%) (t-test, two-tailed).

Among all identified putative REs, only *enFLT1* and *enPPARGC1A* contained sequences that were highly conserved from human to fish. To further assess a potential function of the regulatory element *enPPARGC1A*, a short highly conserved sequence of 102 bp in *ppargc1a* (*enppargc1a*) was excised in medaka via CRISPR/Cas9 gene editing (Fig. 5-2a). As control injections, the gene *oca2* was targeted as described previously (Fig. 5-2b) and injected embryos were observed for heart and developmental defects (Fig. 5-2c). When injected with sgRNAs targeting *enppargc1a*, we observed a significant increase (p=0.0360) of embryos with developmental defects and growth malformations, such as smaller bodies, smaller eyes and cardiovascular deformities, in comparison to the control group (N=3 injections, embryos in total: *Δenppargc1a*: 252, *Δoca2*: 195) (Fig. 5-2d). Interestingly, the defects were not restricted to heart or the cardiovascular system but ranged from mild growth defects to severe malformations of body and organs, suggesting an essential role of this RE during embryonic development.

5.2.2 *Chromatin interaction reveals targets of enC5ORF35, enKCNH2 and enSMG6*

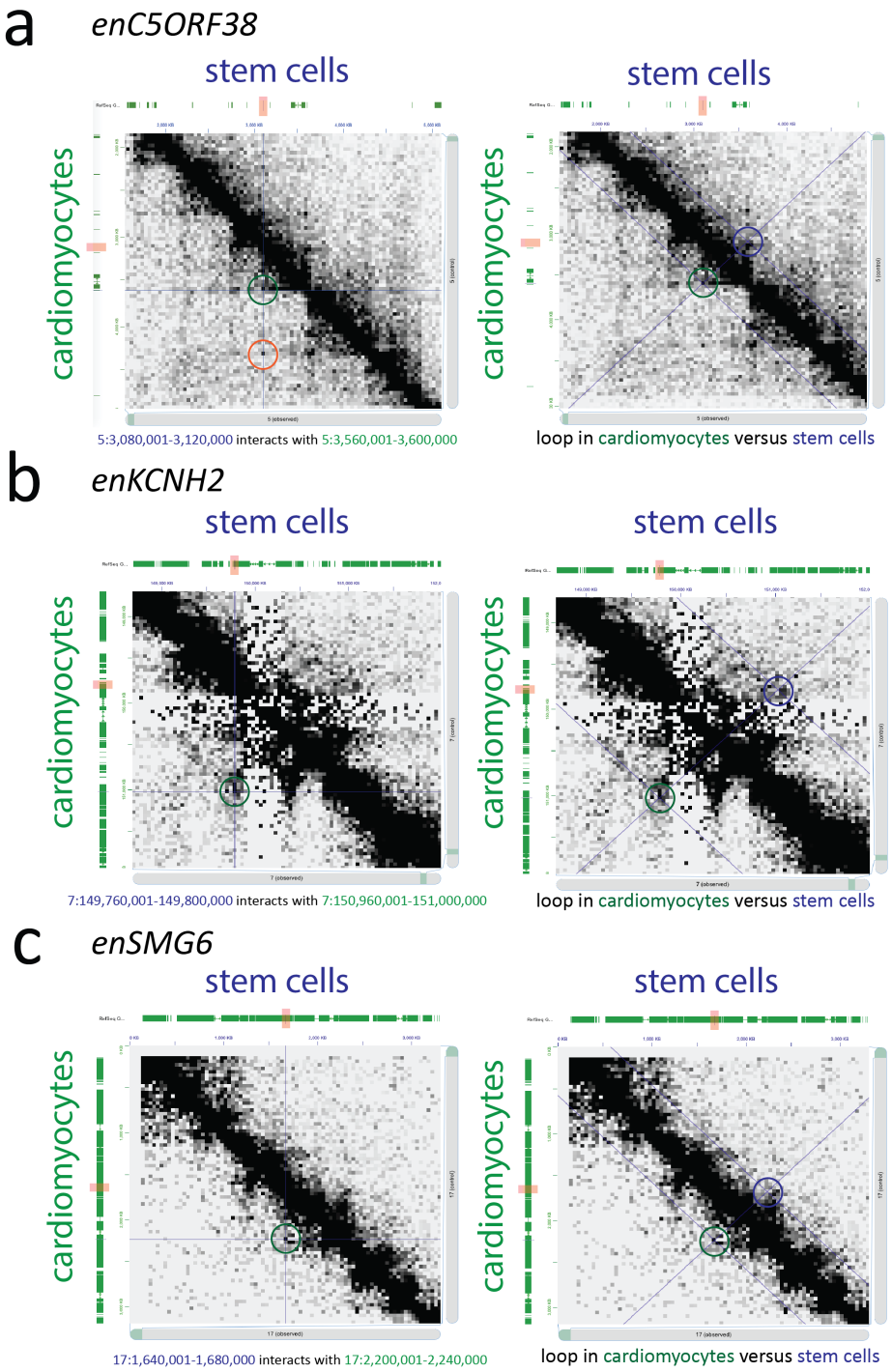


Fig. 5-3: Chromatin capture analysis of putative REs. a) 40kb resolution view on interaction map of *enC5ORF38*. In cardiomyocytes at d14, strong interaction of *enC5ORF38* with an intergenic region (red circle) and the promoter of *IRX1* (green circle) (left panel), compared to a weaker interaction in hESC (blue circle, middle panel). b) 40kb resolution view on interaction map of *enKCNH2*. In cardiomyocytes at d14, strong interaction of *enKCNH2* with the promoter of *ZNF467* and *SSPO* (green circle) (left panel), compared to a weaker interaction in hESC (blue circle, middle panel). c) 40kb resolution view on interaction map of *enSMG6*. In cardiomyocytes at d14, strong interaction of *enSMG6* with the possible target gene *RILP* (green circle) (left panel), compared to a weaker interaction in hESC (blue circle, middle panel). Darker colours indicate a higher interaction frequency (images and analysis in collaboration with Alessandro Bertero).

Bioinformatic analysis of genomic sequences in combination with datasets of histone modification and transcription factor ChIP-seq provide a powerful tool to predict non-coding REs in the genome. However, the association of putative enhancers to their potential target genes are often predictions based on proximity to genes and not based on actual interactions between the genomic regions. Chromosome conformation capture experiments aim to identify with which promoters or regions a specific enhancer or RE might interact. In order to determine candidates which were identified by our bioinformatic pipeline and their physical interactions in the genome, we used a publicly available dataset of HI-C in cardiomyocytes (Bertero *et al.*, 2019). In this study changes in the chromosome architecture were determined by comparing interactions and looping events at day 0 of cardiac differentiation (stem cells) with cardiomyocytes at day 14. We aimed to determine the interaction partners of *enCAV1*, *enC5ORF38*, *enKCNH2*, *enPPARGC1A* and *enSMG6*. In these HI-C heatmaps, darker colours indicate a higher interaction frequency (Lajoie *et al.*, 2015) and in turn indicating possible looping events between REs and promoters. Using this dataset, we identified putative looping events for the sequences *enC5ORF38*, *enKCNH2* and *enSMG6* (Fig. 5-3). For *enC5ORF38*, two loops were detected (Fig. 5-3 a). One interaction located in an intergenic region (red circle) and another one interacting with the promoter of *IRX1* (Fig. 5-3a, right panel), a gene known to be involved in cardiac development and specification (Guo *et al.*, 2017) (Fig. 5-3a left panel, green circle). Interactions with the gene *C5ORF38*, which was the predicted target gene by GREAT, were not detected. Interestingly, the loop between the putative enhancer and the promoter of *IRX1* was stronger in cardiomyocytes (green circle) compared to stem cells (red circle), suggesting a cardiac specificity of the RE (Fig. 5-3a middle panel). Regarding *enKCNH2*, several interactions were detected. However, the most promising looping event (Fig. 5-3b left panel, green circle) which was specific to cardiomyocytes (Fig. 5-3b middle panel), was identified to occur between the enhancer and the promoter regions of the genes *SSPO* and *ZNF467* (Fig. 5-3b, right panel). The putative enhancer of *SMG*6, *enSMG6*, interacts with a region containing the gene *RILP*. Here again, this looping event was specific to cardiomyocytes at day 14 of the differentiation assay and the interaction was not detected in stem cells (Fig. 5-3c).

5.2.3 *Putative cardiac specific CREs regulate gene expression in human cardiomyocytes*

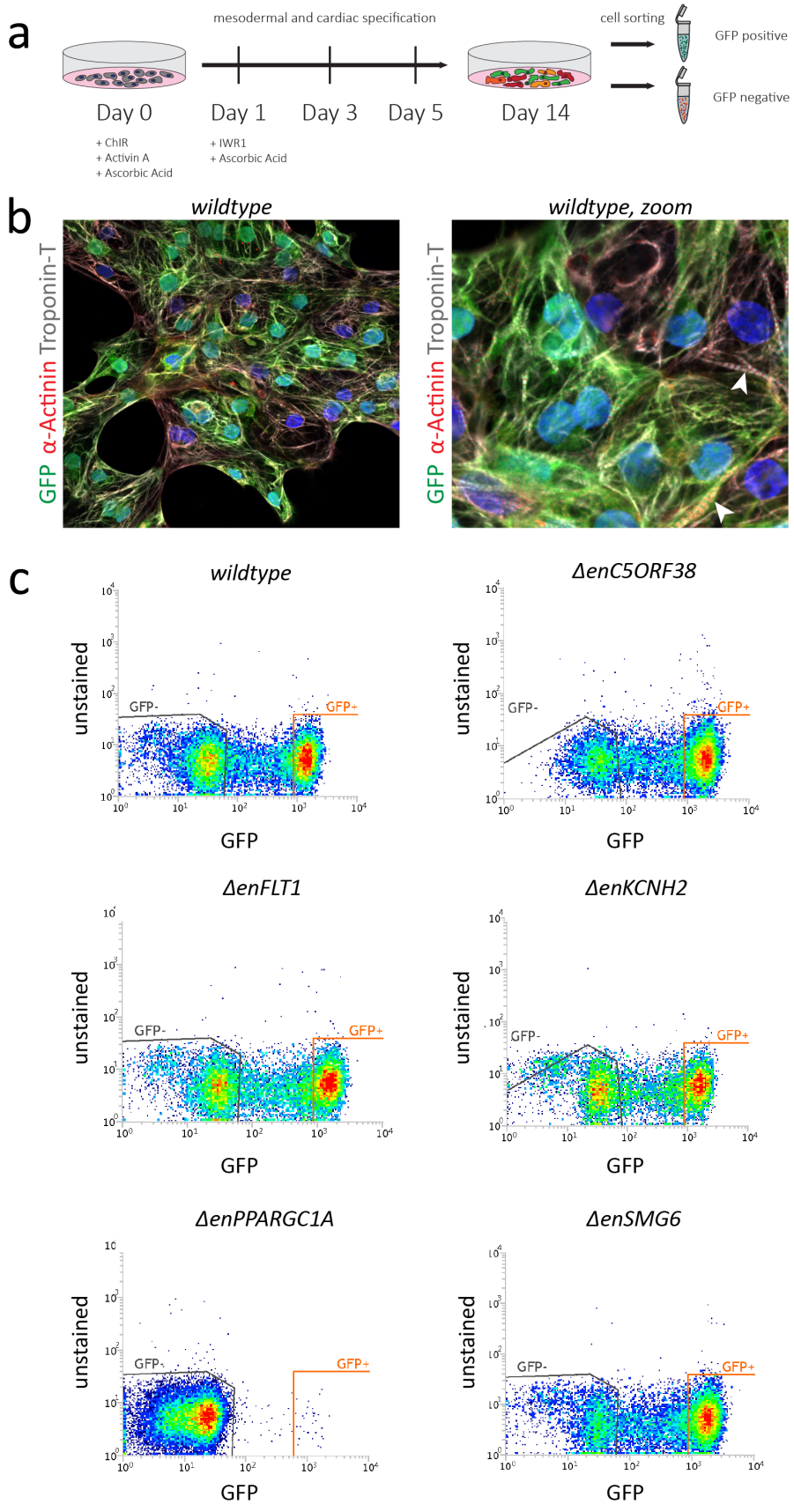


Fig. 5-4: Enhancers for *C5orf38*, *FLT1*, *KCNH2*, *PPARGC1A* and *SMG6* are not essential for cardiac differentiation. a) hESCs were differentiated into cardiomyocytes over the course of 14 days and sorted into cardiomyocyte and non-cardiomyocyte populations. b) Confocal imaging of *wildtype* (*NKX2-5*$^{eGFP/wt}$) cardiomyoyctes at day 14. Cells were positive for GFP as well as the cardiac specific markers alpha actinin and troponin-t in sarcomeres. Sarcomere structures are highlighted with arrowheads. c) Flow cytometry of RE knockout mutant cardiomyoyctes. GFP-positive and GFP-negative populations were collected for each line (if possible).

Enhancer knock-out lines and control lines were differentiated into mesodermal cells, cardiac progenitors and subsequently cardiomyocytes over the course of 14 days. All enhancer mutant lines were generated using the cell line *NKX2-5*$^{eGFP/wt}$ (Elliott *et al.*, 2011) and therefore GFP expression was activated once a heart muscle cell fate was established. The heterogeneous culture containing cardiomyocytes, endothelial cells and cardiac fibroblasts were then sorted into two populations of cardiomyocytes and non-cardiomyocytes for later processing and RNA extraction (Fig. 5-4a) based on GFP expression. Differentiation assays of all cell lines resulted in more than 60% phenotypically normal cardiomyocytes expressing specific cardiac markers such as Troponin T and α-Actinin (Fig. 5-4b). Flow cytometry analysis of cardiac differentiation assays at d14 and sorting for GFP revealed that all enhancer knockout cell lines yielded cardiomyocyte at day 14 of the differentiation assay (more than 3% of total) (Fig. 5-4c). However, additional flow cytometry experiments have revealed a GFP negative population of cardiomyocytes which were positive for cardiac-specific markers such as SIRPA and VCAM1. In addition, differentiation of *ΔenPPARGC1A* resulted in cells that started to beat around 10 days post induction (data not shown), despite FACS data showing a lack of GFP positive cells (0.03% of total).

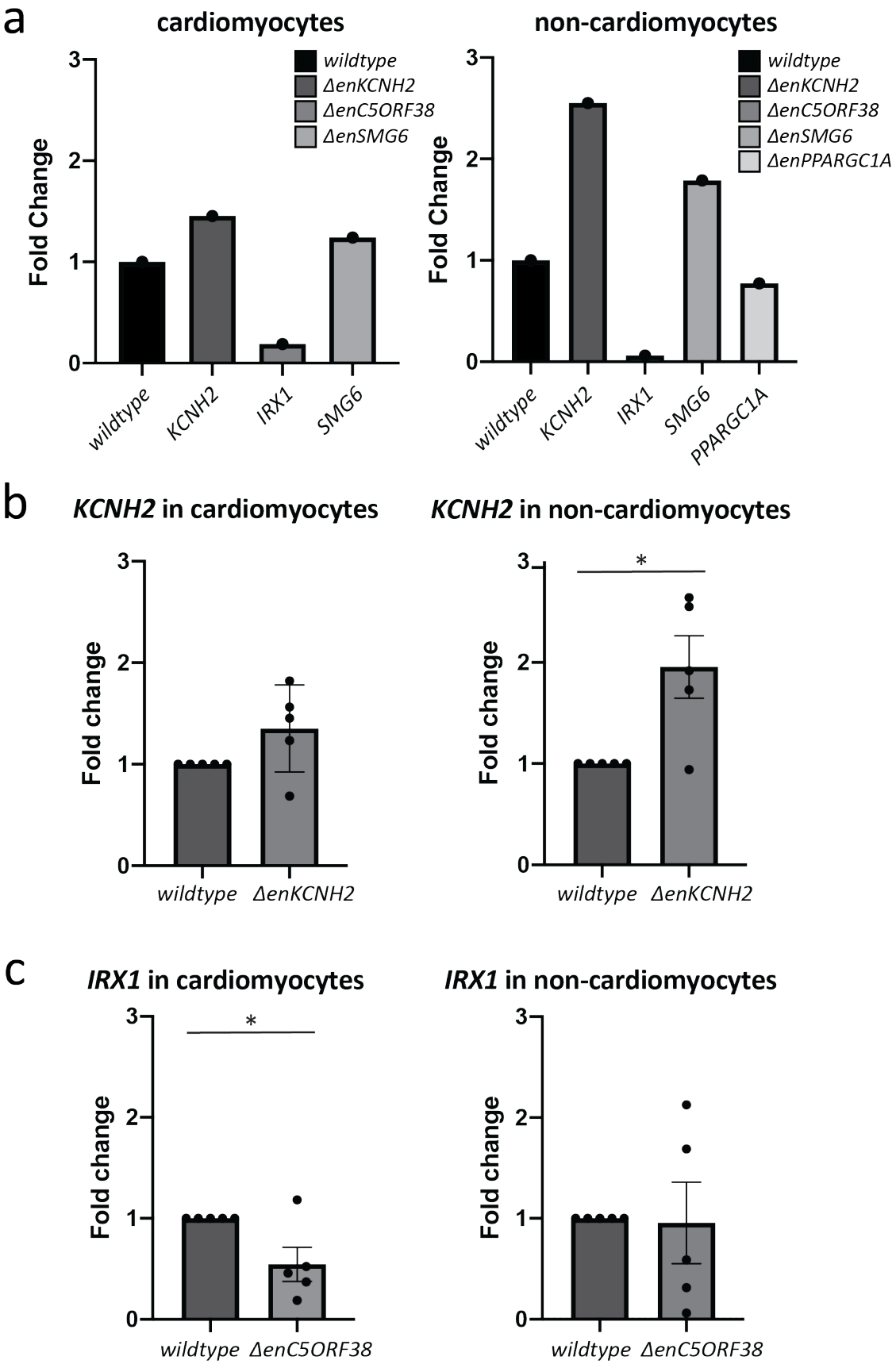


Fig. 5-5: Expression analysis of RE knockout in cardiomyocytes. a) preliminary results (n=1) of various target gene expression levels in *wildtype*, *ΔenKCNH2*, *ΔenC5orf38*, *ΔenSMG6*, *ΔenPAPRGC1A* in cardiomyocytes and non-cardiomyocytes. b) Expression levels of *KCNH2* in *ΔenKCNH2* cardiomyocytes and non-cardiomycoyctes compared to *wildtype* (n=5). c) Expression

levels of *IRX1* in *ΔenC5ORF38* cardiomyocytes and non-cardiomycoyctes compared to *wildtype* (n=5). Statistical significance indicated with asterisks (t-test, two tailed).

Since REs are required for the correct spatio-temporal expression of their target gene, we assessed whether the excision of the putative enhancer elements has an effect on the gene expression of their predicted target gene in cardiomyocytes. We performed gene expression assays of each predicted target gene for each enhancer knockout line, sorted on the basis of GFP expression into cardiomyocytes or non-cardiomyocyte populations. We initially investigated the expression levels of *KCNH2*, *IRX1*, *SMG6* and *PPARGC1A* in *ΔKCNH2*, *ΔenC5ORF38*, *ΔenSMG6* and *ΔenPPARGC1A* respectively and compared them to wildtype expression levels (Fig. 5-5a). In a preliminary screen (n=1), we observed altered expression levels between mutant cell lines and control in both cardiomyocytes and non-cardiomyocytes. No differences in expression levels of *PPARGC1A* in cardiomyocytes in *ΔenPPARGC1A* could be determined due to the lack of GFP positive cells. Subsequently, we focussed on the important cardiac functional genes *KCNH2* and *IRX1*. For *KCNH2* we could detect a significant increase of RNA levels in *ΔenKCNH2* (p=0.0147; N=5) in the non-cardiomyocyte population compared to the wildtype control. However, no significant difference of *KCNH2* expression levels were observed in the cardiomyocyte population (Fig. 5-5b). Interestingly, in regard to *IRX1* expression, we observed a significant decrease of *IRX1* levels in *ΔenC5ORF38* in cardiomyocytes (p=0.0275) compared to the wildtype cell line (Fig. 5-5c). Due to the fact that there was no significant difference in non-cardiomyocyte population of *IRX1* RNA levels this data hints towards a cardiomyocyte specific role for *enC5ORF38*.

5.3 Discussion

Recent studies investigating the role of enhancers during development and disease have highlighted the importance of non-coding regulatory elements. Here we identified the putative enhancers *enC5ORF38*, *enCAV1*, *enKCNH2*, *enPPARGC1A* and *enSMG6*. Due to its proximity to the closest gene, *enC5ORF38* has been associated to *C5ORF38*. As its function is largely unknown, *C5ORF38* could be a potential new candidate as an important gene during heart development and thus *enC5ORF38* might be an important regulator. Since this enhancer is

bound by transcription factors like NKX2-5 and the GATA factors it is likely to have a role in cardiovascular development. However, bioinformatic analysis of chromatin confirmation capture (Hi-C) in cardiomyocytes and human pluripotent stem cells, identified a cardiac specific looping event at the promoter of *IRX1*, suggesting a physical interaction with *IRX1* rather than *C5ORF38*. Indeed, when *IRX1* gene expression was assessed in *ΔenC5ORF38*, deletion of the enhancer resulted in lower mRNA levels in cardiomyocytes compared to wildtype cells. This highly indicates that *enC5ORF38* is directly or indirectly regulating *IRX1* gene expression in heart muscle cells. The primate specific *enC5ORF38* is located in an intergenic area between several *iroquois* homeobox genes. As *iroquois* genes play an important part in the formation of the heart (Kim *et al.*, 2012) *enC5ORF38* could potentially regulate multiple heart specific targets, again suggesting an important role of the regulatory element. In order to understand the phenotypical changes of the enhancer mutant and to investigate the lack of *IRX1* expression on cardiomyocytes, we generated a gene mutant of *IRX1* by deleting the first exon (*ΔexIRX1*) of the gene via CRIPSR/Cas9 gene editing. Both *ΔenC5ORF38* and *ΔexIRX1* hESCs can be successfully differentiated into cardiomyocytes and future analysis will focus on comparing cell physiology and morphology compared to wildtype cells. Preliminary results assessing the electrophysiological properties of the cells are promising, however further experiments are required to define the *ΔenC5ORF38* and *ΔexIRX1* phenotypes. To determine the molecular network behind *C5ORF38* and *IRX1,* bulk-RNA sequencing will be performed to identify potential interaction partners and networks that are regulated by *ΔenC5ORF38.*

Next, we identified a potential cardiac specific RE, *enCAV1,* located in intron 4 of *CAV1,* which encodes a structural protein (CAV1) in the plasma membrane of cells (Boscher and Nabi, 2012). This protein has been implemented as an important contributor of blood vessel growth, regulating the permeability of blood vessels, angiogenesis and mechanotransduction (Bernatchez, 2020). However, we could not detect any specific looping events for this region in cardiomyocytes. We generated enhancer knockout cell lines by excising the region in hESCs and iPSCs. Although we have not performed any experiments on these cell lines due to prioritising other candidates, this region provides a promising putative regulatory element. In addition, a SNP (rs3807989) associated to electrocardiographic traits, atrial fibrillation and QRS duration (Welter *et al.*, 2014) directly falling into that region are supportive of the hypothesis that this RE plays a role in maintaining integrity of the cardiac conduction system. Future work will focus on the functional and molecular profiling of *enCAV1* cardiomyocytes.

We identified a putative RE, which we named *enKCNH2*, located in intron 2 of the gene *KCNH2*. *KCNH2* or *hERG* is a gene encoding for a well-studied voltage-gated potassium channel that is required for the correct repolarization in the human heart and, specifically, cardiomyocytes (Guo *et al.*, 2019a). Disruption of this gene can lead to abnormal electrophysiological properties and can subsequently lead to arrhythmia and sudden cardiac death (SCD) (Jenewein *et al.*, 2018; Sanguinetti and Tristani-Firouzi, 2006). Since this gene is transcribed into different isoforms, the gene regulation of this locus has to be tightly controlled and regulatory perturbations may underlie disease. When assessing *enKCNH2* targets on a structural level, we identified potential loops targeting regions controlling the genes *SSPO* and *ZNF467*. However, since *enKCNH2* is located in an intronic region, we assumed that the most likely target is *KCNH2* itself. In order to test this, we excised *enKCNH2* via CRISPR/Cas9 in hESC. *ΔenKCNH2* cells differentiated into cardiomyocytes and started beating at 7 to 9 days post induction. Interestingly, removal of the RE lead to a significant increase of *KCNH2* expression in non-cardiomyocytes while there was no change of *KCNH2* mRNA levels in cardiomyocytes. An increase of transcription upon the deletion of *ΔenKCNH2* strongly suggests the RE functions as a repressor, in particular in non-cardiomyocyte cell types. A recent study investigating the regulatory network of *KCNH2* and its transcripts, identified the same RE, termed it *CRE5* and found that the element repressed the activity of a cardiac reporter (ZED vector) in zebrafish (Bessa *et al.*, 2009), supporting our initial finding. However, it is surprising that there was no significant effect on gene expression in cardiomyocytes, which suggests the cellular specificity of the repressor is restricted to non-cardiomyocytes. Since the non-cardiomyocyte population consisted of smooth muscle cells and endothelial cells, it needs to be determined whether the activity is specific to one of these two cell types.

Interestingly we did see a decrease in *KCNH2* expression in *ΔenKCNH2* hESCs compared to wildtype stem cells (data not shown) suggesting a change of regulatory function over time and in different cell types. However, further experiments need to be conducted to fully understand its tissue specificity and function of *enKCNH2* during heart development and disease. Since *KCNH2* has an important role in the electrophysiology of these cells, assessing these properties via electrophysiological profiling, would be informative.

The putative RE, *enPPARGC1A,* was identified in PPARGC1A's 9th intron. Besides the cardiac specific binding of NKX2-5, the sequence also contains several TFBS like Oct1, Sox5, PBX1, MEIS1A, MYB possibly indicating an important role for embryonic development (Rosenbloom *et al.*, 2015). A sequence of about 630 bp containing these TFBS is ultra-conserved from human to fish, suggesting a functional role during development. PPARGC1A has been shown to be essential in the regulation of mitochondrial biogenesis and the determination of muscle fibres (Zhang *et al.*, 2019a). Specifically, in the heart, it is important in providing the energy needed to regulate cardiac metabolism. As a coactivator it induces downstream genes that in turn drive the metabolism and energy levels in the heart (Di *et al.*, 2018). However, *enPPARGC1A* may have a more global role rather than a cardiac specific function. its Indeed, when we knocked out the highly conserved region in medaka via targeted gene editing, we saw a significant increase of developmental defects and malformations highlighting a much broader function that was not only restricted to the heart. Besides its role in mitochondrial biogenesis, a previous study in zebrafish has linked *ppargc1a* to ciliated cell development and reported similar phenotypes (Chambers *et al.*, 2020). However, if the phenotypes that were observed in our enhancer deletion model are linked to reduced levels of *ppargc1a* mRNA remains to be elucidated. Interestingly, when *ΔenPPARGC1A* hESCs were differentiated into cardiomyocytes, we did not see GFP positive cells suggesting differentiation along the cardiomyocyte lineage is compromised. However, FACS analysis of cardiac specific markers such as SIRPA and VCAM clearly indicated a double positive population of cardiomyocytes (data not shown). A possible explanation for this could be that the disruption of *enPPARGC1A* might have an effect on *NKX2-5* expression which in turn could lead to a decrease of GFP positive cells. However, since NKX2-5 is crucial for cardiomyocyte differentiation, a perturbation in this system would lead to hESC not being able to differentiate and follow a cardiac fate. The fact that we are able to generate cardiomyocytes using *ΔenPPARGC1A* cells, suggests that there might have been a silencing event repressing reporter gene function. Due technical limitation arising from the inability to obtain *ΔenPPARGC1A* GFP positive cells we were unable to assess *PPARGC1A* expression levels in hESC derived cardiomyocytes. Future work will focus on increasing replicate experiments to address the putative link between *PPARGC1A* levels and expression of *NKX2-5* during hESC cardiac differentiation. In this way, we may identify a novel link between enhancer regulated mitochondrial levels and cardiac development.

SMG6 functions as a mediator of embryonic stem cell differentiation through nonsense-mediated decay and has been linked to coronary artery disease (van der Harst and Verweij, 2018). It specifically functions as an endonuclease that cleaves mRNA containing nonsense mutations. This initiates the degradation process of the RNA as a protection mechanism (Huntzinger *et al.*, 2008).Given *SMG6* functional role, the identified regulatory element, *enSGM6*, might be crucial to facilitate differentiation from a naive to a more specialised cardiac lineage. Also located in an intronic region, removal of *enSMG6*, did not lead to any significant changes in morphology of the RE knockout cells when differentiated into cardiomyocytes. In addition, we detected a looping event of the enhancer with the promoter region of *RILP* in chromatin confirmation capture in cardiomyocytes at day 14. However, preliminary gene expression data of *SMG6* levels cardiomyocytes and non-cardiomyocytes showed that *enSMG6* was dispensable for gene regulation in early human cardiogenesis. Other new potential target genes of the *enSMG6* RE such as *RILP* were not examined and will be the focus of future work.

In regard to Hi-C data and possible connections for all putative REs, although very promising, one should be careful when analysing potential looping events. Sufficient sequencing depth is required to reliably determine interactions between promoters and enhancers. Very weak or unstable connections might not be able to be identified. Furthermore, since "self-ligation noise" has to be considered, local looping events in a range of less than 2kb are filtered out of the analysis (Pal *et al.*, 2019). This limitation makes it difficult to identify looping events of REs located in intronic regions or in close proximity to their target gene.

Taken together, we identified the putative regulatory elements *enC5ORF38*, *enCAV1*, *enKCNH2*, *enPPARGC1A* and *enSMG6* and prioritised the most promising candidates to assess their interactions in the genome as well as their effect on gene expression and development *in vivo* and *in vitro*. We show that *enC5ORF38* acts as an enhancer for *IRX1 in vitro* while *enKCNH2* may function in a repressive manner to regulate *KCNH2* expression. While we did not study the function of *enPPARGC1A in vitro*, we provide evidence that the disruption of a highly conserved sequence of the enhancer harbouring many TFBS, results in defects during medaka development. Future work should include the functional analysis of *enCAV1* and *enSMG6* and a more detailed assessment of the other candidates to shed some light on the precise spatio-temporal regulation of their target genes and biological processes. In conclusion, these findings

highlight the importance of regulatory elements and their role during development and differentiation as well as their potential contribution to the disease class referred to as enhanceropathies.

6 Chapter 6: Final Discussion

6.1 Identifying putative *cis*-regulatory elements

In this study, we aimed to identify and functionally assess putative cardiac-specific *cis*-REs with a combination of computational and experimental approaches. Further, we sought to address the role of the identified *cis*-REs during heart development in order to lay the groundwork for studying how variants in these REs may contribute to congenital cardiovascular disease. Identifying and defining these regulatory DNA regions is a challenge due to the lack of annotation and the volume of non-coding DNA in the genome (UK10K Consortium *et al.*, 2015). In order to define these *cis*-REs, we used a combination of publicly available and our own unpublished sequencing datasets from multiple species to find putative regulatory elements. We bioinformatically defined 19 primate-specific sequences as putative cardiovascular enhancer elements. To assess for evolutionary conservation of enhancer function, we used transgenic medaka to demonstrate that 3 enhancers drove transgene expression in the developing cardiovascular system.

We hypothesise, that the level of conservation was the major influence on function in the medaka as sequences with higher conservation are potentially more likely to be recognized by the endogenous transcription machinery. A functionally conserved RE does not have to show complete sequence conservation (Wang *et al.*, 2007; Wong *et al.*, 2020; Yuan *et al.*, 2018), but also, conservation of specific TFs does not necessarily mean they will be active in the tissue of interest nor does it mean that they will even function as an enhancer or show their state of activity (Villar *et al.*, 2015). Nevertheless, conservation has been used as a reliable method to predict enhancers in previous studies (Dickel *et al.*, 2018; Letelier *et al.*, 2018). However, only a fraction of candidates might be detected if the search is limited to or relying on conservation during the process of identifying enhancers. When we adjusted our bioinformatic pipeline to sort for ultra-conserved sequences from human to fish, we obtained a smaller set of candidates and all of those REs were implicated in the formation of the heart, highlighting the importance of conservation. Thus, the new set of putative cardiac-specific *cis*-regulatory elements may be informative in future *in vivo* experiments. To improve enhancer prediction and identification, machine learning and other computational algorithms provide powerful tools to predict

putative enhancers, especially in combination with experimental approaches (Barr *et al.*, 2019; Belokopytova *et al.*, 2020). For example, tools to predict TF binding (Dang *et al.*, 2018; Ettwiller *et al.*, 2007; Park *et al.*, 2020) and chromatin looping (Salameh *et al.*, 2020) can be implemented to identify both REs and the corresponding gene targets more effectively.

For our analysis, histone modification and transcription factor binding provided essential genomic landmarks to identify candidate REs. To identify active tissue-specific REs using histone modifications datasets, it is crucial to look for overlaps in different marks (*i.e.* H3K4me1 & H3K27ac) which increases the likelihood of finding enhancers (Catarino and Stark, 2018). However, there are no standard procedures about which histone marks should be used. Also, they are not perfect predictors, since it has not been shown that these modifications are directly involved with transcription (Catarino and Stark, 2018). Furthermore, not all REs harbour characteristic marks, which can make the interpretation of various combinations difficult. To improve enhancer predictions in the future, histone mark datasets can be integrated with predictions for open chromatin such as ATAC-seq (Cusanovich *et al.*, 2018; Galang *et al.*, 2020; Sim *et al.*, 2021). Nevertheless, using chromatin modifications proved to be a powerful filter for identifying the two different sets of putative cardiac-specific *cis*-regulatory elements involved in human cardiovascular development presented in this thesis.

Taken together, we provided two different sets of putative cardiac-specific *cis*-regulatory elements involved in human cardiovascular development.

6.2 Enhancer redundancy and robustness

In this study, we identified and functionally tested a small subset of putative cardiac-specific regulatory elements. Three candidates, *enFLT1*, *enKCNH2*, and *enC5ORF38*, were investigated in more detail *in vivo* and *in vitro*. Functional analysis of *enFLT1* revealed an essential role in the formation and development of blood vessels *in vivo* and correct gene expression in human endothelial cells derived from pluripotent stem cells. However, there was only a statistically non-significant, subtle effect on the behaviour of these cells during tube formation *in vitro*. Given the large number of regulatory elements implemented in the transcriptional regulation

of *FLT1* (Simons *et al.*, 2016; Thomas *et al.*, 2010), *enFLT1* is highly likely to be a shadow enhancer displaying redundant spatio-temporal activity (Waymack *et al.*, 2020).

Our results support the notion that shadow enhancers provide phenotypic robustness to a genetic network and act as a fail-safe mechanism against disruptions that might occur during transcriptional regulation. As many other studies have shown the importance of shadow enhancers as buffers during conditions of stress (El-Sherif and Levine, 2016; Frankel *et al.*, 2010), understanding the function of *enFLT1* will shed light on how *FLT1* is regulated during formation of the vasculature. *FLT1* is essential for the correct development of blood vessels in mammalian systems (Fong *et al.*, 1995). Knockout of *FLT1* in mice is embryonic lethal, highlighting the need for a strict safeguard mechanism to maintain correct *FLT1* expression. It has been shown that developmentally regulated genes have a higher number of redundant REs compared to housekeeping genes to maintain gene expression (Gorkin *et al.*, 2020). Due to the subtle effect on the phenotype when shadow enhancers are removed from the genome (Waymack *et al.*, 2020), it may be difficult to define a precise role for an enhancer in disease pathogenesis. However, there is striking evidence that redundant REs can contribute to the formation of enhanceropathies (Rickels and Shilatifard, 2018; Smith and Shilatifard, 2014). Indeed, we showed that disruption of *enFLT1* only slightly decreased mRNA levels of *FLT1* but had stronger effects on transcriptional levels and isoforms of other genes involved in blood vessel development. This suggests, that perturbations of *enFLT1* decrease the overall genetic robustness, which could make an organism more susceptible to environmental stress, manifesting as cardiovascular disease, whereas under ideal conditions, no phenotype is observed. Similar patterns of such decrease of robustness have been reported in other genes and organisms (El-Sherif and Levine, 2016; Frankel *et al.*, 2010). Since single deletions of redundant REs usually only cause strong phenotypes in an already compromised background (Antosova *et al.*, 2016; Sagai *et al.*, 2017), it remains to be investigated how the knockout of *enFLT1* would affect blood vessel growth and *FLT1* gene expression a disease setting. Furthermore, it has been shown that double knockouts of shadow enhancers can recapitulate the phenotype of a gene mutant (Dickel *et al.*, 2018; Hay *et al.*, 2016). Thus, it would be interesting to see what effect additional disruptions of the complex array of REs controlling the expression of *FLT1* has on blood vessel formation.

In addition to *enFLT*1, we have also identified *enKCNH2*, a putative repressor element regulating the gene encoding for the potassium channel *KCNH2*, which is an important regulator of the excitation-contraction cycle of cardiomyocytes (Itzhaki *et al.*, 2011). As van den Boogaard *et al.* have shown, the gene regulatory network of KCNH2 is tightly controlled with multiple enhancers maintaining stable expression (van den Boogaard *et al.*, 2019). Consistent with their findings of the same RE (CRE5), we observed an increase of target gene expression when the sequence was deleted, suggesting it to be a repressive element. Here again, the identified repressor seems to be a part of a redundant gene expression network which, when perturbed, has only mild effects on gene expression and phenotype. These findings support the notion that genes essential for development, in particular cardiovascular development, are controlled by a network of shadow enhancers or regulators to buffer expression. Interestingly, *enC5ORF38* did not seem to follow the same pattern as the other candidates. The fact that this regulatory element is located in a gene desert, raises the question whether it is part of a bigger network regulating one specific gene or if it is a single enhancer acting on multiple targets. Chromatin interaction revealed that *enC5ORF38* has multiple looping points within the vicinity of cardiac specific genes, suggesting a role during heart development. Indeed, the primate specific enhancer directly interacts and regulates *IRX1* expression. It remains to be elucidated whether the other *iroquois* related genes in its proximity are also under its control, in different cell types or developmental stages.

In conclusion, we explored cardiac specific *cis*-regulatory elements and whether these sequences are regulatory components that buffer gene expression to ensure the correct gene expression profiles are maintained during development. While most enhancers display a degree of redundancy in the transcriptional regulation they control, shadow enhancers are far from being dispensable. Their essential role as a fail-safe mechanism during development highlights the importance of tightly controlled gene regulation to subsequently prevent the formation of disease. How partially redundant REs exactly coordinate and modulate gene expression to provide genetic robustness, remains a major question in the field. In this context, studies such as those presented in this PhD thesis serve as important foundations to precisely annotate and describe regulatory elements for future studies.

6.3 The relevance of enhancer in human disease

We postulated that *enFLT1* is very likely to provide robustness to *FLT1* transcriptional regulation and, as part of a network, buffer against loss-of function induced by non-coding disruptions. Interestingly, non-redundant regulatory elements in humans are more likely to cause a disease when mutated compared to their redundant counterparts, which is also portrayed in the lack of genetic variants in *cis*-REs that were associated to disease (Wang and Goldstein, 2020). This bias may explain the difficulty in identifying REs based solely on location of SNPs associated to specific diseases. Thus, coupling genetic studies with the functional characterisation of enhancers presented in this thesis is likely to be more informative regarding the disease pathogenesis arising from perturbed regulatory elements.

It has been shown that, the more redundant REs a gene has, the more likely it is to be highly implicated in pathogenicity (Li *et al.*, 2016). Also, high sequence conservation levels could imply an essential function, which when disrupted leads to disease. However, even ultra-conserved enhancers, like *enFLT1*, are not necessarily disease causative when disrupted. Often, deletions of these REs only cause subtle phenotypes as well (Ahituv *et al.*, 2007; Nolte *et al.*, 2014). The disruption of *enFLT1*, as well as *enPPARGC1A*, only mildly altered cardiovascular cell differentiation and function in human cells *in vitro,* however, the deletion in the medaka genome resulted in a higher formation of blood clots and blood vessel disruptions and whole-body malformations respectively. This raises the question if the gene regulation of these redundant enhancers is fundamentally different in fish than humans or if the effects observed are solely due to the differences *in vivo* versus *in vitro* assays used in this study. Shadow enhancers have been shown to be important in the transcriptional regulation in many different organisms including plants (Poulsen and Chua, 1988), flies (Waymack *et al.*, 2020), fish (Letelier *et al.*, 2018) and mice (Nakada *et al.*, 2004), although, the composition of REs might differ across different species. Therefore, we can safely assume that gene regulation follows the same pattern across evolution. Genome duplication events have created a huge diversity between the same genes throughout the animal kingdom, which would explain different outcomes of enhancer knockouts in different model organisms.

6.4 Future perspectives

In conclusion, the data presented in this study highlights the importance of regulatory elements and their role in development and lays the groundwork to discover their contribution to the formation of disease, in particular in the cardiovascular system. The lack of information regarding enhanceropathies of the heart is striking but together with this project, more and more studies are aiming to shed light on heart disease formation caused by perturbations in *cis*-regulatory elements (Villar *et al.*, 2020).

New technologies in the fields of enhancer biology and transcriptional regulation have provided significant advances toward understanding the mechanisms behind gene regulation. Modern computational approaches have improved the way enhancers and their role in gene regulation are predicted (Belokopytova *et al.*, 2020; Salameh *et al.*, 2020). Technologies to identify enhancers using chromatin accessibility histone marks and transcription factor binding in a high-throughput manner have revolutionised the field (van Arensbergen *et al.*, 2019) and gene editing tools have enabled studies to link regulatory elements in development to the formation of disease (Arvanitis *et al.*, 2020; Galang *et al.*, 2020). However, there is still more work needed to fully understand the origin and function of regulatory elements in human development. The availability of cost-effective large-scale whole genome sequencing in combination with high throughput functional assays of enhancers and their variants will help to elucidate the mechanism behind REs and cardiovascular development and subsequently pathogenicity in humans (van Arensbergen *et al.*, 2019; Karczewski *et al.*, 2020; Man *et al.*, 2019).

7 References

Adler, A., Novelli, V., Amin, A.S., Abiusi, E., Care, M., Nannenberg, E.A., Feilotter, H., Amenta, S., Mazza, D., Bikker, H., et al. (2020). An International, Multicentered, Evidence-Based Reappraisal of Genes Reported to Cause Congenital Long QT Syndrome. Circulation *141*, 418–428.

Agoglia, R.M., and Fraser, H.B. (2016). Disentangling Sources of Selection on Exonic Transcriptional Enhancers. Mol. Biol. Evol. *33*, 585–590.

Agopian, A.J., Goldmuntz, E., Hakonarson, H., Sewda, A., Taylor, D., and Mitchell, L.E. (2017). Genome-Wide Association Studies and Meta-Analyses for Congenital Heart Defects. Circ. Cardiovasc. Genet. *10*, e001449.

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L.A., and Rubin, E.M. (2007). Deletion of ultraconserved elements yields viable mice. PLoS Biol. *5*, 1906–1911.

Akhtar-Zaidi, B., Cowper-Sal{middle dot}lari, R., Corradin, O., Saiakhova, A., Bartels, C.F., Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M.F., et al. (2012). Epigenomic Enhancer Profiling Defines a Signature of Colon Cancer. Science (80-.). *336*, 736–739.

Anderson, D.J., Kaplan, D.I., Bell, K.M., Koutsis, K., Haynes, J.M., Mills, R.J., Phelan, D.G., Qian, E.L., Leitoguinho, A.R., Arasaratnam, D., et al. (2018). NKX2-5 regulates human cardiomyogenesis via a HEY2 dependent transcriptional network. Nat. Commun. *9*, 1–13.

Anderson, K.M., Anderson, D.M., McAnally, J.R., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016). Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. Nature *539*, 433–436.

Andersson, R., Refsing Andersen, P., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T., and Sandelin, A. (2014). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. Nat. Commun. *5*, 5336.

Antosova, B., Smolikova, J., Klimova, L., Lachova, J., Bendova, M., Kozmikova, I., Machon, O., and Kozmik, Z. (2016). The Gene Regulatory Network of Lens Induction Is Wired through Meis-Dependent Shadow Enhancers of Pax6. PLoS Genet. *12*.

van Arensbergen, J., Pagie, L., FitzPatrick, V.D., de Haas, M., Baltissen, M.P., Comoglio, F., van der Weide, R.H., Teunissen, H., Võsa, U., Franke, L., et al. (2019). High-throughput identification of human SNPs affecting regulatory element activity. Nat. Genet. *51*, 1160–1169.

Arnaoutova, I., and Kleinman, H.K. (2010). In vitro angiogenesis: Endothelial cell tube formation on gelled basement membrane extract. Nat. Protoc. *5*, 628–635.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, L.M., Rath, M., and Stark, A. (2013). Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. Science (80-.). *339*, 1074–1077.

Arnold, C.D., Gerlach, D., Spies, D., Matts, J.A., Sytnikova, Y.A., Pagani, M., Lau, N.C., and Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. Nat. Genet. *46*, 685–692.

Arnold, C.D., Zabidi, M.A., Pagani, M., Rath, M., Schernhuber, K., Kazmar, T., and Stark, A. (2017). Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. Nat. Biotechnol. *35*, 136–144.

Arnosti, D.N., Barolo, S., Levine, M., and Small, S. (1996). The eve stripe 2 enhancer employs

multiple modes of transcriptional synergy. Development *122*.

Arvanitis, M., Tampakakis, E., Zhang, Y., Wang, W., Auton, A., Agee, M., Aslibekyan, S., Bell, R.K., Bryc, K., Clark, S.K., et al. (2020). Genome-wide association and multi-omic analyses reveal ACTN2 as a gene linked to heart failure. Nat. Commun. *11*, 1122.

Bafunno, V., Firinu, D., D'Apolito, M., Cordisco, G., Loffredo, S., Leccese, A., Bova, M., Barca, M.P., Santacroce, R., Cicardi, M., et al. (2018). Mutation of the angiopoietin-1 gene (ANGPT1) associates with a new type of hereditary angioedema. J. Allergy Clin. Immunol. *141*, 1009–1017.

Bahr, C., von Paleske, L., Uslu, V. V., Remeseiro, S., Takayama, N., Ng, S.W., Murison, A., Langenfeld, K., Petretich, M., Scognamiglio, R., et al. (2018). A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. Nature *553*, 515–520.

Bakkers, J. (2011). Zebrafish as a model to study cardiac development and human cardiac disease. Cardiovasc. Res. *91*, 279–288.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S., and Natoli, G. (2014). Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. Mol. Cell *54*, 844–857.

Barr, K., Reinitz, J., and Radulescu, O. (2019). An in silico analysis of robust but fragile gene regulation links enhancer length to robustness. PLoS Comput. Biol. *15*.

Belokopytova, P.S., Nuriddinov, M.A., Mozheiko, E.A., Fishman, D., and Fishman, V. (2020). Quantitative prediction of enhancer–promoter interactions. Genome Res. *30*, 72–84.

van Berkum, N.L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L.A., Dekker, J., and Lander, E.S. (2010). Hi-C: A method to study the three-dimensional architecture of genomes. J. Vis. Exp.

Bernatchez, P. (2020). Endothelial caveolin and its scaffolding domain in cancer. Cancer Metastasis Rev. *39*, 471–483.

Bertero, A., Fields, P.A., Ramani, V., Bonora, G., Yardimci, G.G., Reinecke, H., Pabon, L., Noble, W.S., Shendure, J., and Murry, C.E. (2019). Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. Nat. Commun. *10*, 1–19.

Bessa, J., Tena, J.J., de la Calle-Mustienes, E., Fernández-Miñán, A., Naranjo, S., Fernández, A., Montoliu, L., Akalin, A., Lenhard, B., Casares, F., et al. (2009). Zebrafish enhancer detection (ZED) vector: A new tool to facilitate transgenesis and the functional analysis of *cis* -regulatory regions in zebrafish. Dev. Dyn. *238*, 2409–2417.

Birnbaum, R.Y., Clowney, E.J., Agamy, O., Kim, M.J., Zhao, J., Yamanaka, T., Pappalardo, Z., Clarke, S.L., Wenger, A.M., Nguyen, L., et al. (2012). Coding exons function as tissue-specific enhancers of nearby genes. Genome Res. *22*, 1059–1068.

Blackman, R.K., Sanicola, M., Raftery, L.A., Gillevet, T., and Gelbart, W.M. (1991). An extensive 3' cis-regulatory region directs the imaginal disk expression of decapentaplegic, a member of the TGF-beta family in Drosophila. Development *111*, 657–666.

Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2010). ChIP-Seq identification of weakly conserved heart enhancers. Nat. Genet. *42*, 806–810.

Blue, G.M., Kirk, E.P., Sholler, G.F., Harvey, R.P., and Winlaw, D.S. (2012). Congenital heart

disease: current knowledge about causes and inheritance. Med. J. Aust. *197*, 155–159.

Boffelli, D., Nobrega, M.A., and Rubin, E.M. (2004). Comparative genomics at the vertebrate extremes. Nat. Rev. Genet. *5*, 456–465.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. *44*, 148–156.

van den Boogaard, M., Smemo, S., Burnicka-Turek, O., Arnolds, D.E., van de Werken, H.J.G., Klous, P., McKean, D., Muehlschlegel, J.D., Moosmann, J., Toka, O., et al. (2014). A common genetic variant within SCN10A modulates cardiac SCN5A expression. J. Clin. Invest. *124*, 1844–1852.

van den Boogaard, M., van Weerd, J.H., Bawazeer, A.C., Hooijkaas, I.B., van de Werken, H.J.G., Tessadori, F., de Laat, W., Barnett, P., Bakkers, J., and Christoffels, V.M. (2019). Identification and Characterization of a Transcribed Distal Enhancer Involved in Cardiac Kcnh2 Regulation. Cell Rep. *28*, 2704-2714.e5.

Boscher, C., and Nabi, I.R. (2012). CAVEOLIN-1: Role in cell signaling. Adv. Exp. Med. Biol. *729*, 29–50.

Bothma, J.P., Garcia, H.G., Ng, S., Perry, M.W., Gregor, T., and Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. Elife *4*.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.-L., Ruan, Y., Wei, C.-L., Ng, H.H., et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. *18*, 1752–1762.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322.

Brady, P.D., Van Houdt, J., Callewaert, B., Deprest, J., Devriendt, K., and Vermeesch, J.R. (2014). Exome sequencing identifies ZFPM2 as a cause of familial isolated congenital diaphragmatic hernia and possibly cardiovascular malformations. Eur. J. Med. Genet. *57*, 247–252.

Bruneau, B.G. (2013). Signaling and transcriptional networks in heart development and regeneration. Cold Spring Harb. Perspect. Biol. *5*, a008292.

Buckingham, M., Meilhac, S., and Zaffran, S. (2005). Building the mammalian heart from two sources of myocardial cells. Nat. Rev. Genet. *6*, 826–835.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Bush, W.S., and Moore, J.H. (2012). Chapter 11: Genome-wide association studies. PLoS Comput. Biol. *8*, e1002822.

Bushman, F.D. (2003). Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. Cell *115*, 135–138.

Bussmann, J., Bos, F.L., Urasaki, A., Kawakami, K., Duckers, H.J., and Schulte-Merker, S. (2010). Arteries provide essential guidance cues for lymphatic endothelial cells in the zebrafish trunk. Development *137*, 2653–2657.

Bywater, M.J., Burkhart, D.L., Straube, J., Sabò, A., Pendino, V., Hudson, J.E., Quaife-Ryan, G.A., Porrello, E.R., Rae, J., Parton, R.G., et al. (2020). Reactivation of Myc transcription in the mouse heart unlocks its proliferative capacity. Nat. Commun. *11*.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.

Carroll, S.B. (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell *134*, 25–36.

Catarino, R.R., and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. Genes Dev. *32*, 202–223.

Catarino, R.R., Neumayr, C., and Stark, A. (2017). Promoting transcription over long distances. Nat. Genet. *49*, 972–973.

Chahal, G., Tyagi, S., and Ramialison, M. (2019). Navigating the non-coding genome in heart development and Congenital Heart Disease. Differentiation *107*, 11–23.

Chambers, J.M., Addiego, A., Flores-Mireles, A.L., and Wingert, R.A. (2020). Ppargc1a Controls Ciliated Cell Development by Regulating Prostaglandin Biosynthesis. Cell Rep. *33*.

Chang, S.W., Mislankar, M., Misra, C., Huang, N., Dajusta, D.G., Harrison, S.M., Mcbride, K.L., Baker, L.A., and Garg, V. (2013). Genetic abnormalities in FOXP1 are associated with congenital heart defects. Hum. Mutat. *34*, 1226–1230.

Charité, J., McFadden, D.G., Merlo, G., Levi, G., Clouthier, D.E., Yanagisawa, M., Richardson, J.A., and Olson, E.N. (2001). Role of Dlx6 in regulation of an endothelin-1-dependent, dHAND branchial arch enhancer. Genes Dev. *15*, 3039–3049.

Chen, T.T., Luque, A., Lee, S., Anderson, S.M., Segura, T., and Iruela-Arispe, M.L. (2010). Anchorage of VEGF to the extracellular matrix conveys differential signaling responses to endothelial cells. J. Cell Biol. *188*, 595–609.

Chen, W., Dai, D., Wang, C., Huang, T., Zhai, L., and Deng, Z. (2013). Genetic dissection of the polyoxin building block-carbamoylpolyoxamic acid biosynthesis revealing the "pathway redundancy" in metabolic networks. Microb. Cell Fact. *12*, 121.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science *351*, 1083–1087.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Corradin, O., and Scacheri, P.C. (2014). Enhancer variants: evaluating functions in common disease. Genome Med. *6*, 85.

Cretekos, C.J., Wang, Y., Green, E.D., Martin, J.F., Rasweiler, J.J., Behringer, R.R., IV, and Behringer, R.R. (2008). Regulatory divergence modifies limb length between mammals. Genes Dev. *22*, 141–151.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Crocker, J., Abe, N., Rinaldi, L., McGregor, A.P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., et al. (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. Cell *160*, 191–203.

Cunningham, T.J., Lancman, J.J., Berenguer, M., Dong, P.D.S., and Duester, G. (2018). Genomic Knockout of Two Presumed Forelimb Tbx5 Enhancers Reveals They Are Nonessential for Limb Development. Cell Rep. *23*, 3146–3151.

Cusanovich, D.A., Hill, A.J., Aghamirzaie, D., Daza, R.M., Pliner, H.A., Berletch, J.B., Filippova, G.N., Huang, X., Christiansen, L., DeWitt, W.S., et al. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. Cell *174*, 1309-1324.e18.

d'Apolito, M., Santacroce, R., Colia, A.L., Cordisco, G., Maffione, A.B., and Margaglione, M. (2019). Angiopoietin-1 haploinsufficiency affects the endothelial barrier and causes hereditary angioedema. Clin. Exp. Allergy *49*, 626–635.

Dang, L.T., Tondl, M., Chiu, M.H.H., Revote, J., Paten, B., Tano, V., Tokolyi, A., Besse, F., Quaife-Ryan, G., Cumming, H., et al. (2018). Trawler Web: An online de novo motif discovery tool for next-generation sequencing datasets. BMC Genomics *19*.

Das, S., Goldstone, A.B., Wang, H., Farry, J., D'Amato, G., Paulsen, M.J., Eskandari, A., Hironaka, C.E., Phansalkar, R., Sharma, B., et al. (2019). A Unique Collateral Artery Development Program Promotes Neonatal Heart Regeneration. Cell *176*, 1128-1142.e18.

Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.-B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., et al. (2012). DNaseI sensitivity QTLs are a major determinant of human expression variation. Nature *482*, 390–394.

Dekker, J., Belmont, A.S., Guttman, M., Leshyk, V.O., Lis, J.T., Lomvardas, S., Mirny, L.A., O'Shea, C.C., Park, P.J., Ren, B., et al. (2017). The 4D nucleome project. Nature *549*, 219–226.

Desjardins, C., and Naya, F. (2016). The Function of the MEF2 Family of Transcription Factors in Cardiac Development, Cardiogenomics, and Direct Reprogramming. J. Cardiovasc. Dev. Dis. *3*, 26.

Di, W., Lv, J., Jiang, S., Lu, C., Yang, Z., Ma, Z., Hu, W., Yang, Y., and Xu, B. (2018). PGC-1: The energetic regulator in cardiac metabolism. Curr. Issues Mol. Biol. *28*, 29–46.

Dib, S., Denarier, E., Dionne, N., Beaudoin, M., Friedman, H.H., and Peterson, A.C. (2011). Regulatory modules function in a non-autonomous manner to control transcription of the mbp gene. Nucleic Acids Res. *39*, 2548–2558.

Dickel, D.E., Barozzi, I., Zhu, Y., Fukuda-Yuzawa, Y., Osterwalder, M., Mannion, B.J., May, D., Spurrell, C.H., Plajzer-Frick, I., Pickle, C.S., et al. (2016). Genome-wide compendium and functional assessment of in vivo heart enhancers. Nat. Commun. *7*, 12923.

Dickel, D.E., Ypsilanti, A.R., Pla, R., Zhu, Y., Barozzi, I., Mannion, B.J., Khin, Y.S., Fukuda-Yuzawa, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Ultraconserved Enhancers Are Required for Normal Development. Cell *172*, 491-499.e15.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N. V., Scruggs, B.S., Nady, N., Still, C.D., Garcia,

B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Mol. Cell *66*, 568-576.e4.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Dunipace, L., Ozdemir, A., and Stathopoulos, A. (2011). Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. Development *138*, 4075–4084.

Duque, T., and Sinha, S. (2015). What does it take to evolve an enhancer? A simulation-based study of factors influencing the emergence of combinatorial regulation. Genome Biol. Evol. *7*, 1415–1431.

Durbin, R.M., Altshuler, D.L., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Collins, F.S., De La Vega, F.M., Donnelly, P., et al. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.

Edmunds, J.W., and Mahadevan, L.C. (2004). MAP kinases as structural adaptors and enzymatic activators in transcription complexes. J. Cell Sci. *117*, 3715–3723.

Edwards, J.J., and Gelb, B.D. (2016). Genetics of congenital heart disease. Curr. Opin. Cardiol. *31*, 235–241.

El-Sherif, E., and Levine, M. (2016). Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the Drosophila Embryo. Curr. Biol. *26*, 1164–1169.

Elliott, D.A., Braam, S.R., Koutsis, K., Ng, E.S., Jenny, R., Lagerqvist, E.L., Biben, C., Hatzistavrou, T., Hirst, C.E., Yu, Q.C., et al. (2011). NKX2-5 eGFP/w hESCs for isolation of human cardiac progenitors and cardiomyocytes. Nat. Methods *8*, 1037–1043.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Erceg, J., Saunders, T.E., Girardot, C., Devos, D.P., Hufnagel, L., and Furlong, E.E.M. (2014). Subtle Changes in Motif Positioning Cause Tissue-Specific Effects on Robustness of an Enhancer's Activity. PLoS Genet. *10*, e1004060.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Ettwiller, L., Paten, B., Ramialison, M., Birney, E., and Wittbrodt, J. (2007). Trawler: de novo regulatory motif discovery pipeline for chromatin immunoprecipitation. Nat. Methods *4*, 563–565.

Fabre, P.J., Benke, A., Joye, E., Nguyen Huynh, T.H., Manley, S., and Duboule, D. (2015). Nanoscale spatial organization of the HoxD gene cluster in distinct transcriptional states. Proc. Natl. Acad. Sci. U. S. A. *112*, 13964–13969.

Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., and Levine, M.S. (2015). Suboptimization of developmental enhancers. Science *350*, 325–328.

Fearnley, C.J., Llewelyn Roderick, H., and Bootman, M.D. (2011). Calcium signaling in cardiac myocytes. Cold Spring Harb. Perspect. Biol. *3*.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. Nat.

Rev. Genet. *9*, 397–405.

Fong, G.H., Rossant, J., Gertsenstein, M., and Breitman, M.L. (1995). Role of the Flt-1 receptor tyrosine kinase in regulating the assembly of vascular endothelium. Nature *376*, 66–70.

Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature *466*, 490–493.

Frankel, N., Erezyilmaz, D.F., McGregor, A.P., Wang, S., Payre, F., and Stern, D.L. (2011). Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. Nature *474*, 598–603.

French, J.D., and Edwards, S.L. (2020). The Role of Noncoding Variants in Heritable Disease. Trends Genet. *0*.

Fujita, M., Isogai, S., and Kudo, A. (2006). Vascular anatomy of the developing medaka, Oryzias latipes: a complementary fish model for cardiovascular research on vertebrates. Dev. Dyn. *235*, 734–746.

Furutani-Seiki, M., and Wittbrodt, J. (2004). Medaka and zebrafish, an evolutionary twin study. Mech. Dev. *121*, 629–637.

Gaborit, N., Sakuma, R., Wylie, J.N., Kim, K.H., Zhang, S.S., Hui, C.C., and Bruneau, B.G. (2012). Cooperative and antagonistic roles for Irx3 and Irx5 in cardiac morphogenesis and postnatal physiology. Dev. *139*, 4007–4019.

Galang, G., Mandla, R., Ruan, H., Jung, C., Sinha, T., Stone, N.R., Wu, R.S., Mannion, B.J., Allu, P.K.R., Chang, K., et al. (2020). ATAC-Seq Reveals an Isl1 Enhancer That Regulates Sinoatrial Node Development and Function. Circ. Res. *127*, 1502–1518.

Garry, D.J., and Olson, E.N. (2006). A Common Progenitor at the Heart of Development. Cell *127*, 1101–1104.

Ghim, J., Moon, J.S., Lee, C.S., Lee, J., Song, P., Lee, A., Jang, J.H., Kim, D., Yoon, J.H., Koh, Y.J., et al. (2014). Endothelial deletion of phospholipase d2 reduces hypoxic response and pathological angiogenesis. Arterioscler. Thromb. Vasc. Biol. *34*, 1697–1703.

Ghosh, A.K. (2020). p300 in cardiac development and accelerated cardiac aging. Aging Dis. *11*, 916–926.

Gierten, J., Pylatiuk, C., Hammouda, O.T., Schock, C., Stegmaier, J., Wittbrodt, J., Gehrig, J., and Loosli, F. (2020). Automated high-throughput heartbeat quantification in medaka and zebrafish embryos under physiological conditions. Sci. Rep. *10*.

Gifford, C.A., Ranade, S.S., Samarakoon, R., Salunga, H.T., Yvanka De Soysa, T., Huang, Y., Zhou, P., Elfenbein, A., Wyman, S.K., Bui, Y.K., et al. (2019). Oligogenic inheritance of a human heart disease involving a genetic modifier. Science (80-.). *364*, 865–870.

De Gobbi, M., Viprakasit, V., Hughes, J.R., Fisher, C., Buckle, V.J., Ayyub, H., Gibbons, R.J., Vernimmen, D., Yoshinaga, Y., de Jong, P., et al. (2006). A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. Science *312*, 1215–1217.

Gomez-Velazquez, M., Badia-Careaga, C., Lechuga-Vieco, A.V., Nieto-Arellano, R., Tena, J.J., Rollan, I., Alvarez, A., Torroja, C., Caceres, E.F., Roy, A., et al. (2017). CTCF counter-regulates cardiomyocyte development and maturation programs in the embryonic heart. PLoS Genet. *13*.

Goode, D.K., Callaway, H.A., Cerda, G.A., Lewis, K.E., and Elgar, G. (2011). Minor change, major difference: divergent functions of highly conserved cis-regulatory elements subsequent to

whole genome duplication events. Development *138*, 879–884.

Gorkin, D.U., Barozzi, I., Zhao, Y., Zhang, Y., Huang, H., Lee, A.Y., Li, B., Chiou, J., Wildberg, A., Ding, B., et al. (2020). An atlas of dynamic chromatin landscapes in mouse fetal development. Nature *583*, 744–751.

Guo, C., Wang, Q., Wang, Y., Yang, L., Luo, H., Cao, X.F., An, L., Qiu, Y., Du, M., Ma, X., et al. (2017). Exome sequencing reveals novel IRXI mutation in congenital heart disease. Mol. Med. Rep. *15*, 3193–3197.

Guo, F., Sun, Y., Wang, X., Wang, H., Wang, J., Gong, T., Chen, X., Zhang, P., Su, L., Fu, G., et al. (2019a). Patient-Specific and Gene-Corrected Induced Pluripotent Stem Cell-Derived Cardiomyocytes Elucidate Single-Cell Phenotype of Short QT Syndrome. Circ. Res. *124*, 66–78.

Guo, G.L., Sun, L.Q., Sun, M.H., and Xu, H.M. (2019b). LncRNA SLC8A1-AS1 protects against myocardial damage through activation of cGMP-PKG signaling pathway by inhibiting SLC8A1 in mice models of myocardial infarction. J. Cell. Physiol. *234*, 9019–9032.

Gurvitz, M., Burns, K.M., Brindis, R., Broberg, C.S., Daniels, C.J., Fuller, S.M.P.N., Honein, M.A., Khairy, P., Kuehl, K.S., Landzberg, M.J., et al. (2016). Emerging Research Directions in Adult Congenital Heart Disease: A Report From an NHLBI/ACHA Working Group. J. Am. Coll. Cardiol. *67*, 1956–1964.

Gutierrez-Triana, J.A., Herget, U., Castillo-Ramirez, L.A., Lutz, M., Yeh, C.-M., De Marco, R.J., and Ryu, S. (2015). Manipulation of Interrenal Cell Function in Developing Zebrafish Using Genetically Targeted Ablation and an Optogenetic Tool. Endocrinology *156*, 3394–3401.

Haines, J.L., Hauser, M.A., Schmidt, S., Scott, W.K., Olson, L.M., Gallins, P., Spencer, K.L., Kwan, S.Y., Noureddine, M., Gilbert, J.R., et al. (2005). Complement Factor H Variant Increases the Risk of Age-Related Macular Degeneration. Science (80-.). *308*, 419–421.

van der Harst, P., and Verweij, N. (2018). Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery DiseaseNovelty and Significance. Circ. Res. *122*, 433–443.

Hay, D., Hughes, J.R., Babbs, C., Davies, J.O.J., Graham, B.J., Hanssen, L., Kassouf, M.T., Marieke Oudelaar, A.M., Sharpe, J.A., Suciu, M.C., et al. (2016). Genetic dissection of the α-globin super-enhancer in vivo. Nat. Genet. *48*, 895–903.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone Modifications at Human Enhancers Reflect Global Cell Type-Specific Gene Expression. Nature *459*, 108.

Hilgers, L., and Schwarzer, J. (2019). The untapped potential of medaka and its wild relatives. Elife *8*.

Hille, S., Dierck, F., Kühl, C., Sosna, J., Adam-Klages, S., Adam, D., Lüllmann-Rauch, R., Frey, N., and Kuhn, C. (2016). Dyrk1a regulates the cardiomyocyte cell cycle via D-cyclin-dependent Rb/E2f-signalling. Cardiovasc. Res. *110*, 381–394.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. U. S. A. *106*, 9362–9367.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hong, J.-W., Hendrix, D.A., and Levine, M.S. (2008). Shadow Enhancers as a Source of

Evolutionary Novelty. Science (80-.). *321*, 1314–1314.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

Huntzinger, E., Kashima, I., Fauser, M., Saulière, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. RNA *14*, 2609–2617.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. *31*, 227–229.

Iampietro, C., Gummalla, M., Mutero, A., Karch, F., and Maeda, R.K. (2010). Initiator elements function to determine the activity state of BX-C enhancers. PLoS Genet. *6*, e1001260.

Itzhaki, I., Maizels, L., Huber, I., Zwi-Dantsis, L., Caspi, O., Winterstern, A., Feldman, O., Gepstein, A., Arbel, G., Hammerman, H., et al. (2011). Modelling the long QT syndrome with induced pluripotent stem cells. Nature *471*, 225–230.

Iwamatsu, T. (2004). Stages of normal development in the medaka Oryzias latipes. Mech. Dev. *121*, 605–618.

Jacques, P.-É., Jeyakani, J., and Bourque, G. (2013). The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet. *9*, e1003504.

Jenewein, T., Neumann, T., Erkapic, D., Kuniss, M., Verhoff, M.A., Thiel, G., and Kauferstein, S. (2018). Influence of genetic modifiers on sudden cardiac death cases. Int. J. Legal Med. *132*, 379–385.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science (80-.). *337*, 816–821.

Juven-Gershon, T., Cheng, S., and Kadonaga, J.T. (2006). Rational design of a super core promoter that enhances gene expression. Nat. Methods *3*, 917–922.

Karczewski, K.J., Dudley, J.T., Kukurba, K.R., Chen, R., Butte, A.J., Montgomery, S.B., and Snyder, M. (2013). Systematic functional regulatory assessment of disease-associated variants. Proc. Natl. Acad. Sci. *110*, 9607–9612.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.

Kawakami, K., and Shima, A. (1999). Identification of the Tol2 transposase of the medaka fish Oryzias latipes that catalyzes excision of a nonautonomous Tol2 element in zebrafish Danio rerio. Gene *240*, 239–244.

Kawakami, K., Largaespada, D.A., and Ivics, Z. (2017). Transposons As Tools for Functional Genomics in Vertebrate Models. Trends Genet. *33*, 784–801.

Kehat, I., Kenyagin-Karsenti, D., Snir, M., Segev, H., Amit, M., Gepstein, A., Livne, E., Binah, O., Itskovitz-Eldor, J., and Gepstein, L. (2001). Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. J. Clin. Invest. *108*, 407–414.

Kehat, I., Gepstein, A., Spira, A., Itskovitz-Eldor, J., and Gepstein, L. (2002). High-resolution electrophysiological assessment of human embryonic stem cell-derived cardiomyocytes: A

novel in vitro model for the study of conduction. Circ. Res. *91*, 659–661.

Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M., and Tsui, L.C. (1989). Identification of the cystic fibrosis gene: genetic analysis. Science *245*, 1073–1080.

Kikas, T., Inno, R., Ratnik, K., Rull, K., and Laan, M. (2020). C-allele of rs4769613 Near FLT1 Represents a High-Confidence Placental Risk Factor for Preeclampsia. Hypertens. (Dallas, Tex. 1979) *76*, 884–891.

Kim, K.H., Rosen, A., Bruneau, B.G., Hui, C.C., and Backx, P.H. (2012). Iroquois homeodomain transcription factors in heart development and function. Circ. Res. *110*, 1513–1524.

Kleinjan, D.A., and Lettice, L.A. (2008). Long Range Gene Control and Genetic Disease. In Advances in Genetics, pp. 339–388.

Kleinjan, D.A., Bancewicz, R.M., Gautier, P., Dahm, R., Schonthaler, H.B., Damante, G., Seawright, A., Hever, A.M., Yeyati, P.L., van Heyningen, V., et al. (2008). Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence. PLoS Genet. *4*, e29.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat. Struct. Mol. Biol. *18*, 956–963.

Koefoed, K., Veland, I.R., Pedersen, L.B., Larsen, L.A., and Christensen, S.T. (2014). Cilia and coordination of signaling networks during heart development. Organogenesis *10*, 108–125.

Korzh, V., Kondrychyn, I., and Winata, C. (2018). The Zebrafish as a New Model System for Experimental Biology. Cytol. Genet. *52*, 406–415.

Kothary, R., Clapoff, S., Darling, S., Perry, M.D., Moran, L.A., and Rossant, J. (1989). Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. Development *105*, 707–714.

Krishnan, A., Samtani, R., Dhanantwari, P., Lee, E., Yamada, S., Shiota, K., Donofrio, M.T., Leatherbury, L., and Lo, C.W. (2014). A detailed comparison of mouse and human cardiac development. Pediatr. Res. *76*, 500–507.

Krueger, J., Liu, D., Scholz, K., Zimmer, A., Shi, Y., Klein, C., Siekmann, A., Schulte-Merker, S., Cudmore, M., Ahmed, A., et al. (2011). Flt1 acts as a negative regulator of tip cell formation and branching morphogenesis in the zebrafish embryo. Development *138*, 2111–2120.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature *512*, 91–95.

Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J., Tissières, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A., et al. (2016). Progressive Loss of Function in a Limb Enhancer during Snake Evolution. Cell *167*, 633-642.e11.

Kvon, E.Z., Zhu, Y., Kelman, G., Novak, C.S., Plajzer-Frick, I., Kato, M., Garvin, T.H., Pham, Q., Harrington, A.N., Hunter, R.D., et al. (2020). Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. Cell.

Kwasnieski, J.C., Mogno, I., Myers, C.A., Corbo, J.C., and Cohen, B.A. (2012). Complex effects of

nucleotide variants in a mammalian cis-regulatory element. Proc. Natl. Acad. Sci. U. S. A. *109*, 19498–19503.

Lai, S., Kuenne, C., Kuan, J., Lai, H., Tsedeke, A.T., Guenther, S., Looso, M., Nauheim, B., Sequencing, D., and Nauheim, B. (2017). Reciprocal analyses in zebrafish and medaka reveal that harnessing the immune response promotes cardiac regeneration. 1–20.

Lai, S.L., Marín-Juez, R., and Stainier, D.Y.R. (2019). Immune responses in cardiac repair and regeneration: a comparative point of view. Cell. Mol. Life Sci. *76*, 1365–1380.

Lajoie, B.R., Dekker, J., and Kaplan, N. (2015). The Hitchhiker's guide to Hi-C analysis: practical guidelines. Methods *72*, 65–75.

Lan, X., and Pritchard, J.K. (2016). Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. Science *352*, 1009–1013.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Leddin, M., Perrod, C., Hoogenkamp, M., Ghani, S., Assi, S., Heinz, S., Wilson, N.K., Follows, G., Schönheit, J., Vockentanz, L., et al. (2011). Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. Blood *117*, 2827–2838.

Lee, C.S., Ghim, J., Song, P., Suh, P.-G., and Ryu, S.H. (2016). Loss of phospholipase D2 impairs VEGF-induced angiogenesis. BMB Rep. *49*, 191–196.

Letelier, J., De La Calle-Mustienes, E., Pieretti, J., Naranjo, S., Maeso, I., Nakamura, T., Pascual-Anaya, J., Shubin, N.H., Schneider, I., Martinez-Morales, J.R., et al. (2018). A conserved Shh cis-regulatory module highlights a common developmental origin of unpaired and paired fins. Nat. Genet. *50*, 504–509.

Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.

Levine, M. (2010). Transcriptional Enhancers in Animal Development and Evolution. Curr. Biol. *20*, R754–R763.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: Recent insights and future perspectives. Nat. Rev. Genet. *17*, 207–223.

Li, Y., Rivera, C.M., Ishii, H., Jin, F., Selvaraj, S., Lee, A.Y., Dixon, J.R., and Ren, B. (2014). CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. PLoS One *9*, e114485.

Liao, Y., Smyth, G.K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. *47*.

Lin, C.-Y., Chiang, C.-Y., and Tsai, H.-J. (2016). Zebrafish and Medaka: new model organisms for modern biomedical research. J. Biomed. Sci. *23*.

Lischik, C.Q., Adelmann, L., and Wittbrodt, J. (2019). Enhanced in vivo-imaging in medaka by optimized anaesthesia, fluorescent protein selection and removal of pigmentation. PLoS One *14*, e0212956.

Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C.L., Lindberg, E.L., Kanda, M., Polanski, K., Heinig, M., Lee, M., et al. (2020). Cells of the adult human heart. Nature *588*, 466–472.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170–1187.

Ludwig, M.Z., Bergman, C., Patel, N.H., and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. Nature *403*, 564–567.

Lynch, V.J., Leclerc, R.D., May, G., and Wagner, G.P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. *43*, 1154–1159.

Madonna, R., and De Caterina, R. (2009). VEGF receptor switching in heart development and disease. Cardiovasc. Res. *84*, 4–6.

Maeda, R.K., and Karch, F. (2011). Gene expression in time and space: additive vs hierarchical organization of cis-regulatory regions. Curr. Opin. Genet. Dev. *21*, 187–193.

Man, J.C.K., Mohan, R.A., Boogaard, M. van den, Hilvering, C.R.E., Jenkins, C., Wakker, V., Bianchi, V., Laat, W. de, Barnett, P., Boukens, B.J., et al. (2019). Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. Nat. Commun. *10*.

Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., et al. (2014). Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science *346*, 1373–1377.

Marinić, M., Aktas, T., Ruf, S., and Spitz, F. (2013). An Integrated Holo-Enhancer Unit Defines Tissue and Gene Specificity of the Fgf8 Regulatory Landscape. Dev. Cell *24*, 530–542.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

May, D., Blow, M.J., Kaplan, T., McCulley, D.J., Jensen, B.C., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., et al. (2012). Large-scale discovery of enhancers from human heart tissue. Nat. Genet. *44*, 89–93.

McCulley, D.J., and Black, B.L. (2012). Transcription Factor Pathways and Congenital Heart Disease. In Current Topics in Developmental Biology, (Academic Press Inc.), pp. 253–277.

McDermott, M.I., Wang, Y., Wakelam, M.J.O., and Bankaitis, V.A. (2020). Mammalian phospholipase D: Function, and therapeutics. Prog. Lipid Res. *78*.

McFadden, D.G., Charité, J., Richardson, J.A., Srivastava, D., Firulli, A.B., and Olson, E.N. (2000). A GATA-dependent right ventricular enhancer controls dHAND transcription in the developing heart. Development *127*, 5331–5341.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Melgar, M.F., Collins, F.S., and Sethupathy, P. (2011). Discovery of active enhancers through bidirectional expression of short transcripts. Genome Biol. *12*, R113.

Mendis, S., Puska, P., Norrving, B., World Health Organization., World Heart Federation., and World Stroke Organization. (2011). Global atlas on cardiovascular disease prevention and control (World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization).

Menendez, D., Krysiak, O., Inga, A., Krysiak, B., Resnick, M.A., and Schönfelder, G. (2006). A SNP

in the flt-1 promoter integrates the VEGF system into the p53 transcriptional network. Proc. Natl. Acad. Sci. U. S. A. *103*, 1406–1411.

Mieczkowski, J., Cook, A., Bowman, S.K., Mueller, B., Alver, B.H., Kundu, S., Deaton, A.M., Urban, J.A., Larschan, E., Park, P.J., et al. (2016). MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nat. Commun. *7*, 11485.

Mihaly, J., Barges, S., Sipos, L., Maeda, R., Cléard, F., Hogga, I., Bender, W., Gyurkovics, H., and Karch, F. (2006). Dissecting the regulatory landscape of the Abd-B gene of the bithorax complex. Development *133*, 2983–2993.

Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I.E., Males, M., Viales, R.R., and Furlong, E.E.M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev. *32*, 42–57.

Miller, J.A., and Widom, J. (2003). Collaborative competition mechanism for gene activation in vivo. Mol. Cell. Biol. *23*, 1623–1632.

Mills, R.J., Titmarsh, D.M., Koenig, X., Parker, B.L., Ryall, J.G., Quaife-Ryan, G.A., Voges, H.K., Hodson, M.P., Ferguson, C., Drowley, L., et al. (2017). Functional screening in human cardiac organoids reveals a metabolic mechanism for cardiomyocyte cell cycle arrest. Proc. Natl. Acad. Sci. U. S. A. *114*, E8372–E8381.

Minhas, R., Paterek, A., Łapiński, M., Bazała, M., Korzh, V., and Winata, C.L. (2019). A novel conserved enhancer at zebrafish zic3 and zic6 loci drives neural expression. Dev. Dyn. *248*, 837–849.

Montefiori, L.E., Sobreira, D.R., Sakabe, N.J., Aneas, I., Joslin, A.C., Hansen, G.T., Bozek, G., Moskowitz, I.P., McNally, E.M., and Nóbrega, M.A. (2018). A promoter interaction map for cardiovascular disease genetics. Elife *7*, 1–35.

Muerdter, F., Boryń, Ł.M., Woodfin, A.R., Neumayr, C., Rath, M., Zabidi, M.A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R.R., et al. (2018). Resolving systematic errors in widely used enhancer activity assays in human cells. Nat. Methods *15*, 141–149.

Mummery, C., Ward, D., Van Den Brink, C.E., Bird, S.D., Doevendans, P.A., Opthof, T., Brutel De La Riviere, A., Tertoolen, L., Van Der Heyden, M., and Pera, M. (2002). Cardiomyocyte differentiation of mouse and human embryonic stem cells. J. Anat. *200*, 233–242.

Mummery, C.L., Zhang, J., Ng, E.S., Elliott, D.A., Elefanty, A.G., and Kamp, T.J. (2012). Differentiation of human embryonic stem cells and induced pluripotent stem cells to cardiomyocytes: A methods overview. Circ. Res. *111*, 344–358.

Murtha, M., Tokcaer-Keskin, Z., Tang, Z., Strino, F., Chen, X., Wang, Y., Xi, X., Basilico, C., Brown, S., Bonneau, R., et al. (2014). FIREWACh: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. Nat. Methods *11*, 559–565.

Nakada, Y., Parab, P., Simmons, A., Omer-Abdalla, A., and Johnson, J.E. (2004). Separable enhancer sequences regulate the expression of the neural bHLH transcription factor neurogenin 1. Dev. Biol. *271*, 479–487.

Nesmith, J.E., Chappell, J.C., Cluceru, J.G., and Bautch, V.L. (2017). Blood vessel anastomosis is spatially regulated by Flt1 during angiogenesis. Development *144*.

Neumann, H., Hancock, S.M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D., and Chin, J.W. (2009). A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Mol. Cell *36*,

153–163.

Nguyen, C.T., Lu, Q., Wang, Y., and Chen, J.N. (2008). Zebrafish as a model for cardiovascular development and disease. Drug Discov. Today Dis. Model. *5*, 135–140.

Nikpay, M., Soubeyrand, S., Tahmasbi, R., and McPherson, R. (2020). Multiomics Screening Identifies Molecular Biomarkers Causally Associated with the Risk of Coronary Artery Disease. Circ. Genomic Precis. Med. *13*, 649–659.

Nistala, R., and Savin, V. (2017). Diabetes, hypertension, and chronic kidney disease progression: Role of DPP4. Am. J. Physiol. - Ren. Physiol. *312*, F661–F670.

Nolte, M.J., Wang, Y., Deng, J.M., Swinton, P.G., Wei, C., Guindani, M., Schwartz, R.J., and Behringer, R.R. (2014). Functional analysis of limb transcriptional enhancers in the mouse. Evol. Dev. *16*, 207–223.

Nowak-Sliwinska, P., Alitalo, K., Allen, E., Anisimov, A., Aplin, A.C., Auerbach, R., Augustin, H.G., Bates, D.O., van Beijnum, J.R., Bender, R.H.F., et al. (2018). Consensus guidelines for the use and interpretation of angiogenesis assays. Angiogenesis *21*, 425–532.

Nowicka, M., and Robinson, M.D. (2016). DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. F1000Research *5*.

Olson, E.N. (2006). Gene regulatory networks in the evolution and development of the heart. Science *313*, 1922–1927.

Osterwalder, M., Barozzi, I., Tissiéres, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature *554*, 239–243.

van Ouwerkerk, A.F., Bosada, F.M., van Duijvenboden, K., Hill, M.C., Montefiori, L.E., Scholman, K.T., Liu, J., de Vries, A.A.F., Boukens, B.J., Ellinor, P.T., et al. (2019). Identification of atrial fibrillation associated genes and functional non-coding variants. Nat. Commun. *10*.

Owen, L.A., Morrison, M.A., Ahn, J., Woo, S.J., Sato, H., Robinson, R., Morgan, D.J., Zacharaki, F., Simeonova, M., Uehara, H., et al. (2014). FLT1 genetic variation predisposes to neovascular AMD in ethnically diverse populations and alters systemic FLT1 expression. Investig. Ophthalmol. Vis. Sci. *55*, 3543–3554.

Pal, K., Forcato, M., and Ferrari, F. (2019). Hi-C analysis: from data generation to integration. Biophys. Rev. *11*, 67–78.

Park, S., Koh, Y., Jeon, H., Kim, H., Yeo, Y., and Kang, J. (2020). Enhancing the interpretability of transcription factor binding site prediction using attention mechanism. Sci. Rep. *10*.

Pasquali, L., Gaulton, K.J., Rodríguez-Seguí, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, İ., Tena, J.J., Morán, I., Gómez-Marín, C., van de Bunt, M., et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat. Genet. *46*, 136–143.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nat. Methods *14*, 417–419.

Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.-I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. *30*, 265–270.

Peden, J.F., Hopewell, J.C., Saleheen, D., Chambers, J.C., Hager, J., Soranzo, N., Collins, R., Danesh, J., Elliott, P., Farrall, M., et al. (2011). A genome-wide association study in Europeans

and South Asians identifies five new loci for coronary artery disease. Nat. Genet. *43*, 339–346.

Pengelly, A.R., Copur, O., Jackle, H., Herzig, A., and Muller, J. (2013). A Histone Mutant Reproduces the Phenotype Caused by Loss of Histone-Modifying Factor Polycomb. Science (80-.). *339*, 698–699.

Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. Nat. Rev. Genet. *14*, 288–295.

Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc. Natl. Acad. Sci. U. S. A. *108*, 13570–13575.

Pinto, A.R., Ilinykh, A., Ivey, M.J., Kuwabara, J.T., D'antoni, M.L., Debuque, R., Chandran, A., Wang, L., Arora, K., Rosenthal, N.A., et al. (2016). Revisiting cardiac cellular composition. Circ. Res. *118*, 400–409.

Pomerantz, M.M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M.P., Doddapaneni, H., Beckwith, C.A., Chan, J.A., Hills, A., Davis, M., et al. (2009). The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat. Genet. *41*, 882–884.

Poon, K.L., and Brand, T. (2013). The zebrafish model system in cardiovascular research: A tiny fish with mighty prospects. Glob. Cardiol. Sci. Pract. *2013*, 4.

Poulsen, C., and Chua, N.H. (1988). Dissection of 5′ upstream sequences for selective expression of the Nicotiana plumbaginifolia rbcS-8B gene. MGG Mol. Gen. Genet. *214*, 16–23.

Pradeepa, M.M., Grimes, G.R., Kumar, Y., Olley, G., Taylor, G.C.A., Schneider, R., and Bickmore, W.A. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. Nat. Genet. *48*, 681–686.

Prazak, L., Fujioka, M., and Gergen, J.P. (2010). Non-additive interactions involving two distinct elements mediate sloppy-paired regulation by pair-rule transcription factors. Dev. Biol. *344*, 1048–1059.

Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. Cell *163*, 68–83.

Pusztaszeri, M.P., Seelentag, W., and Bosman, F.T. (2006). Immunohistochemical expression of endothelial markers CD31, CD34, von Willebrand factor, and Fli-1 in normal human tissues. J. Histochem. Cytochem. *54*, 385–395.

Quaife-Ryan, G.A., Sim, C.B., Ziemann, M., Kaspi, A., Rafehi, H., Ramialison, M., El-Osta, A., Hudson, J.E., and Porrello, E.R. (2017). Multicellular transcriptional analysis of mammalian heart regeneration. Circulation *136*, 1123–1139.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rebeiz, M., Jikomes, N., Kassner, V.A., and Carroll, S.B. (2011). Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. Proc. Natl. Acad. Sci. U. S. A. *108*, 10036–10043.

Richards, A.A., and Garg, V. (2010). Genetics of congenital heart disease. Curr. Cardiol. Rev. *6*, 91–97.

Rickels, R., and Shilatifard, A. (2018). Enhancer Logic and Mechanics in Development and Disease. Trends Cell Biol. *28*, 608–630.

Rickels, R., Herz, H.-M., Sze, C.C., Cao, K., Morgan, M.A., Collings, C.K., Gause, M., Takahashi, Y.-H., Wang, L., Rendleman, E.J., et al. (2017). Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nat. Genet. *49*, 1647–1653.

Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*, 896–902.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–329.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Roe, J.-S., Mercan, F., Rivera, K., Pappin, D.J., and Vakoc, C.R. (2015). BET Bromodomain Inhibition Suppresses the Function of Hematopoietic Transcription Factors in Acute Myeloid Leukemia. Mol. Cell *58*, 1028–1039.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. *43*, D670–D681.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development *132*, 797–803.

Sagai, T., Amano, T., Maeno, A., Kiyonari, H., Seo, H., Cho, S.W., and Shiroishi, T. (2017). SHH signaling directed by two oral epithelium-specific enhancers controls tooth and oral development. Sci. Rep. *7*.

Sakabe, N.J., Savic, D., and Nobrega, M.A. (2012). Transcriptional enhancers in development and disease. Genome Biol. *13*, 238.

Salameh, T.J., Wang, X., Song, F., Zhang, B., Wright, S.M., Khunsriraksakul, C., Ruan, Y., and Yue, F. (2020). A supervised learning framework for chromatin loop detection in genome-wide contact maps. Nat. Commun. *11*, 1–12.

Sanguinetti, M.C., and Tristani-Firouzi, M. (2006). hERG potassium channels and cardiac arrhythmia. Nature *440*, 463–469.

Santenard, A., Ziegler-Birling, C., Koch, M., Tora, L., Bannister, A.J., and Torres-Padilla, M.-E. (2010). Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. Nat. Cell Biol. *12*, 853–862.

Sarro, R., Kocher, A.A., Emera, D., Uebbing, S., Dutrow, E. V., Weatherbee, S.D., Nottoli, T., and Noonan, J.P. (2018). Disrupting the three-dimensional regulatory topology of the Pitx1 locus results in overtly normal development. Dev. *145*.

Schaffner, W. (2015). Enhancers, enhancers – from their discovery to today's universe of transcription enhancers. Biol. Chem. *396*, 311–327.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer,

P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225–1228.

Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. Mol. Cell *58*, 1101–1112.

Shi, X., Richard, J., Zirbes, K.M., Gong, W., Lin, G., Kyba, M., Thomson, J.A., Koyano-Nakagawa, N., and Garry, D.J. (2014). Cooperative interaction of Etv2 and Gata2 regulates the development of endothelial and hematopoietic lineages. Dev. Biol. *389*, 208–218.

Shibuya, M. (2006). Vascular endothelial growth factor receptor-1 (VEGFR-1/Flt-1): a dual regulator for angiogenesis. Angiogenesis *9*, 225–230.

Shibuya, M., Yamaguchi, S., Yamane, A., Ikeda, T., Tojo, A., Matsushime, H., and Sato, M. (1990). Nucleotide sequence and expression of a novel human receptor-type tyrosine kinase gene (flt) closely relatd to the fms family. Oncogene *5*, 519–524.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272–286.

Sim, C.B., Phipson, B., Ziemann, M., Rafehi, H., Mills, R.J., Watt, K.I., Abu-Bonsrah, K.D., Kalathur, R.K.R., Voges, H.K., Dinh, D.T., et al. (2021). Sex-Specific Control of Human Heart Maturation by the Progesterone Receptor. Circulation CIRCULATIONAHA.120.051921.

Simons, M., Gordon, E., and Claesson-Welsh, L. (2016). Mechanisms and regulation of endothelial VEGF receptor signalling. Nat. Rev. Mol. Cell Biol. *17*, 611–625.

Slatkin, M. (2008). Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. *9*, 477–485.

Smemo, S., Campos, L.C., Moskowitz, I.P., Krieger, J.E., Pereira, A.C., and Nobrega, M.A. (2012). Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. Hum. Mol. Genet. *21*, 3255–3263.

Smith, E., and Shilatifard, A. (2014). Enhancer biology and enhanceropathies. Nat. Struct. Mol. Biol. *21*, 210–219.

Smith, R.P., Taher, L., Patwardhan, R.P., Kim, M.J., Inoue, F., Shendure, J., Ovcharenko, I., and Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. Nat. Genet. *45*, 1021–1028.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Srivastava, D. (2006). Making or Breaking the Heart: From Lineage Determination to Morphogenesis. Cell *126*, 1037–1048.

Srivastava, D., and Kwon, C. (2010). Signaling pathways involved in cardiogenesis. In Handbook of Cell Signaling, 2/E, (Elsevier Inc.), pp. 2601–2609.

St Johnston, R.D., Hoffmann, F.M., Blackman, R.K., Segal, D., Grimaila, R., Padgett, R.W., Irick, H.A., and Gelbart, W.M. (1990). Molecular organization of the decapentaplegic gene in Drosophila melanogaster. Genes Dev. *4*, 1114–1127.

Stefater, J.A., Lewkowich, I., Rao, S., Mariggi, G., Carpenter, A.C., Burr, A.R., Fan, J., Ajima, R., Molkentin, J.D., Williams, B.O., et al. (2011). Regulation of angiogenesis by a non-canonical Wnt - Flt1 pathway in myeloid cells. Nature *474*, 511–516.

Stemmer, M., Thumberger, T., del Sol Keyer, M., Wittbrodt, J., and Mateo, J.L. (2015). CCTop:

An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS One *10*, e0124633.

Stergachis, A.B., Haugen, E., Shafer, A., Fu, W., Vernot, B., Reynolds, A., Raubitschek, A., Ziegler, S., LeProust, E.M., Akey, J.M., et al. (2013). Exonic transcription factor binding directs codon choice and affects protein evolution. Science *342*, 1367–1372.

Stine, Z.E., McGaughey, D.M., Bessling, S.L., Li, S., and McCallion, A.S. (2011). Steroid hormone modulation of RET through two estrogen responsive enhancers in breast cancer. Hum. Mol. Genet. *20*, 3746–3756.

Stolper, J., Ambrosio, E.M., Danciu, D.P., Buono, L., Elliott, D.A., Naruse, K., Martínez-Morales, J.R., Marciniak-Czochra, A., and Centanin, L. (2019). Stem cell topography splits growth and homeostatic functions in the fish gill. Elife *8*.

Su, M., Han, D., Boyd-Kirkup, J., Yu, X., and Han, J.-D.J. (2014). Evolution of Alu Elements toward Enhancers. Cell Rep. *7*, 376–385.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell *131*, 861–872.

Tam, P.P., Parameswaran, M., Kinder, S.J., and Weinberger, R.P. (1997). The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation. Development *124*, 1631–1642.

Taylor, J.S., and Raes, J. (2004). Duplication and Divergence: The Evolution of New Genes and Old Ideas. Annu. Rev. Genet. *38*, 615–643.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. Cell *83*, 1091–1100.

Thåström, A., Lowary, P.., Widlund, H.., Cao, H., Kubista, M., and Widom, J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. J. Mol. Biol. *288*, 213–229.

Thermes, V., Grabher, C., Ristoratore, F., Bourrat, F., Choulika, A., Wittbrodt, J., and Joly, J.S. (2002). I-SceI meganuclease mediates highly efficient transgenesis in fish. Mech. Dev. *118*, 91–98.

Thomas, C.P., Raikwar, N.S., Kelley, E.A., and Liu, K.Z. (2010). Alternate processing of Flt1 transcripts is directed by conserved cis-elements within an intronic region of FLT1 that reciprocally regulates splicing and polyadenylation. Nucleic Acids Res. *38*, 5130–5140.

Tillo, D., Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Field, Y., Lieb, J.D., Widom, J., Segal, E., and Hughes, T.R. (2010). High Nucleosome Occupancy Is Encoded at Human Regulatory Sequences. PLoS One *5*, e9129.

Ting, C.N., Rosenberg, M.P., Snow, C.M., Samuelson, L.C., and Meisler, M.H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. Genes Dev. *6*, 1457–1465.

Tuan, D., Kong, S., and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. Proc. Natl. Acad. Sci. U. S. A. *89*, 11219–11223.

Al Turki, S., Manickaraj, A.K., Mercer, C.L., Gerety, S.S., Hitz, M.P., Lindsay, S., D'Alessandro, L.C.A., Swaminathan, G.J., Bentham, J., Arndt, A.K., et al. (2014). Rare variants in NR2F2 cause congenital heart defects in humans. Am. J. Hum. Genet. *94*, 574–585.

UK10K Consortium, T.U., Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S.,

Perry, J.R.B., Xu, C., Futema, M., et al. (2015). The UK10K project identifies rare variants in health and disease. Nature *526*, 82–90.

Vierstra, J., Rynes, E., Sandstrom, R., Zhang, M., Canfield, T., Hansen, R.S., Stehling-Sun, S., Sabo, P.J., Byron, R., Humbert, R., et al. (2014). Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science *346*, 1007–1012.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell *160*, 554–566.

Villar, D., Frost, S., Deloukas, P., and Tinker, A. (2020). The contribution of non-coding regulatory elements to cardiovascular disease. Open Biol. *10*, 200088.

Virani, S.S., Alonso, A., Benjamin, E.J., Bittencourt, M.S., Callaway, C.W., Carson, A.P., Chamberlain, A.M., Chang, A.R., Cheng, S., Delling, F.N., et al. (2020). Heart disease and stroke statistics—2020 update: A report from the American Heart Association. Circulation *141*, E139–E596.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser - A database of tissue-specific human enhancers. Nucleic Acids Res. *35*, 88–92.

Visel, A., Akiyama, J.A., Shoukry, M., Afzal, V., Rubin, E.M., and Pennacchio, L.A. (2009). Functional autonomy of distant-acting human enhancers. Genomics *93*, 509–513.

Wang, X., and Goldstein, D.B. (2020). Enhancer Domains Predict Gene Pathogenicity and Inform Gene Discovery in Complex Disease. Am. J. Hum. Genet. *106*, 215–233.

Wang, Y., and Wang, J.-G. (2018). Genome-Wide Association Studies of Hypertension and Several Other Cardiovascular Diseases. Pulse *6*, 169–186.

Wang, Q.F., Prabhakar, S., Chanan, S., Cheng, J.F., Rubin, E.M., and Boffelli, D. (2007). Detection of weakly conserved ancestral mammalian regulatory sequences by primate comparisons. Genome Biol. *8*.

Wang, S.-P., Tang, Z., Chen, C.-W., Shimada, M., Koche, R.P., Wang, L.-H., Nakadai, T., Chramiec, A., Krivtsov, A. V., Armstrong, S.A., et al. (2017). A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. Mol. Cell *67*, 308-321.e6.

Ward, L.D., and Kellis, M. (2012). Evidence of Abundant Purifying Selection in Humans for Recently Acquired Regulatory Functions. Science (80-.). *337*, 1675–1678.

Warren, J.S., Tracy, C.M., Miller, M.R., Makaju, A., Szulik, M.W., Oka, S. ichi, Yuzyuk, T.N., Cox, J.E., Kumar, A., Lozier, B.K., et al. (2018). Histone methyltransferase Smyd1 regulates mitochondrial energetics in the heart. Proc. Natl. Acad. Sci. U. S. A. *115*, E7871–E7880.

Waymack, R., Fletcher, A., Enciso, G., and Wunderlich, Z. (2020). Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic. Elife *9*, 1–57.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42*, D1001–D1006.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wild, R., Klems, A., Takamiya, M., Hayashi, Y., Strähle, U., Ando, K., Mochizuki, N., van Impel, A.,

Schulte-Merker, S., Krueger, J., et al. (2017). Neuronal sFlt1 and Vegfaa determine venous sprouting and spinal cord vascularization. Nat. Commun. *8*, 13991.

Williamson, I., Lettice, L.A., Hill, R.E., and Bickmore, W.A. (2016). Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. Development *143*, 2994–3001.

Wirtwein, M., Melander, O., Sjogren, M., Hoffmann, M., Narkiewicz, K., Gruchala, M., and Sobiczewski, W. (2016). The Relationship Between Gene Polymorphisms and Dipping Profile in Patients With Coronary Heart Disease. Am. J. Hypertens. *29*, 1094–1102.

Wittbrodt, J., Shima, A., and Schartl, M. (2002). Medaka - A model organism from the Far East. Nat. Rev. Genet. *3*, 53–64.

Wittkopp, P.J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. *13*, 59–69.

Wong, E.S., Zheng, D., Tan, S.Z., Bower, N.I., Garside, V., Vanwalleghem, G., Gaiti, F., Scott, E., Hogan, B.M., Kikuchi, K., et al. (2020). Deep conservation of the enhancer regulatory code in animals. Science (80-.). *370*.

Xin, M., Olson, E.N., and Bassel-Duby, R. (2013). Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. Nat. Rev. Mol. Cell Biol. *14*, 529–541.

Xing, K., and He, X. (2015). Reassessing the “Duon” Hypothesis of Protein Evolution. Mol. Biol. Evol. *32*, 1056–1062.

Xu, X., Li, Y., Huang, Y., Ye, H., Han, L., Ji, H., Chen, X., Wu, N., Jiang, D., Xu, L., et al. (2018). Impact of gender and age on the association of the BUD13-ZNF259 rs964184 polymorphism with coronary heart disease. Anatol. J. Cardiol. *19*, 42–49.

Yang, F., Jin, C., Jiang, Y.J., Li, J., Di, Y., and Fu, D.L. (2011). Potential role of soluble VEGFR-1 in antiangiogenesis therapy for cancer. Expert Rev. Anticancer Ther. *11*, 541–549.

Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2019). Ensembl 2020. Nucleic Acids Res. gkz966.

Young, R.S., Kumar, Y., Bickmore, W.A., and Taylor, M.S. (2017). Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. Genome Biol. *18*, 242.

Yuan, X., and Braun, T. (2017). Multimodal Regulation of Cardiac Myocyte Proliferation. Circ. Res. *121*, 293–309.

Yuan, X., Song, M., Devine, P., Bruneau, B.G., Scott, I.C., and Wilson, M.D. (2018). Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. Nat. Commun. *9*.

Yuh, C.H., and Davidson, E.H. (1996). Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. Development *122*, 1069–1082.

Yuh, C.H., Bolouri, H., and Davidson, E.H. (1998). Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. Science *279*, 1896–1902.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zhang, G.M., Guo, Y.X., Deng, M.T., Wan, Y.J., Deng, K.P., Xiao, S.H., Meng, F.X., Wang, F., and Lei, Z.H. (2019a). Effect of PPARGC1A on the development and metabolism of early rabbit embryos in vitro. Mol. Reprod. Dev. *86*, 1758–1770.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C.,

Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Y., Li, T., Preissl, S., Amaral, M.L., Grinstein, J.D., Farah, E.N., Destici, E., Qiu, Y., Hu, R., Lee, A.Y., et al. (2019b). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat. Genet. *51*, 1380–1388.

Zhou, H.Y., Katsman, Y., Dhaliwal, N.K., Davidson, S., Macpherson, N.N., Sakthidevi, M., Collura, F., and Mitchell, J.A. (2014). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. Genes Dev. *28*, 2699–2711.

(2018). MonashBioinformaticsPlatform/RNAsik-pipe: JOSS ready.

(2019). drpowell/degust 4.1.1.

8 Appendices

8.2 Supplementary data

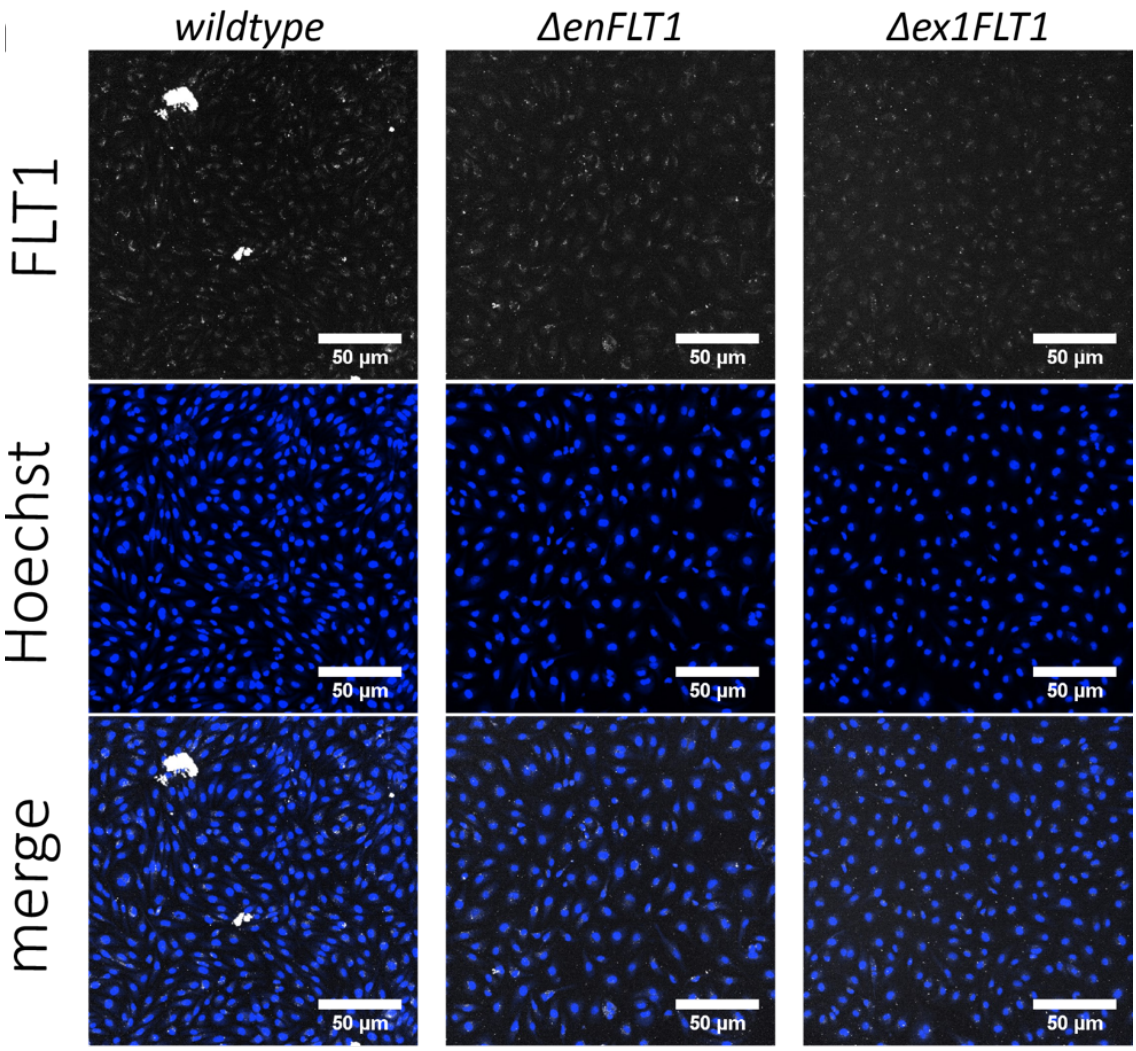
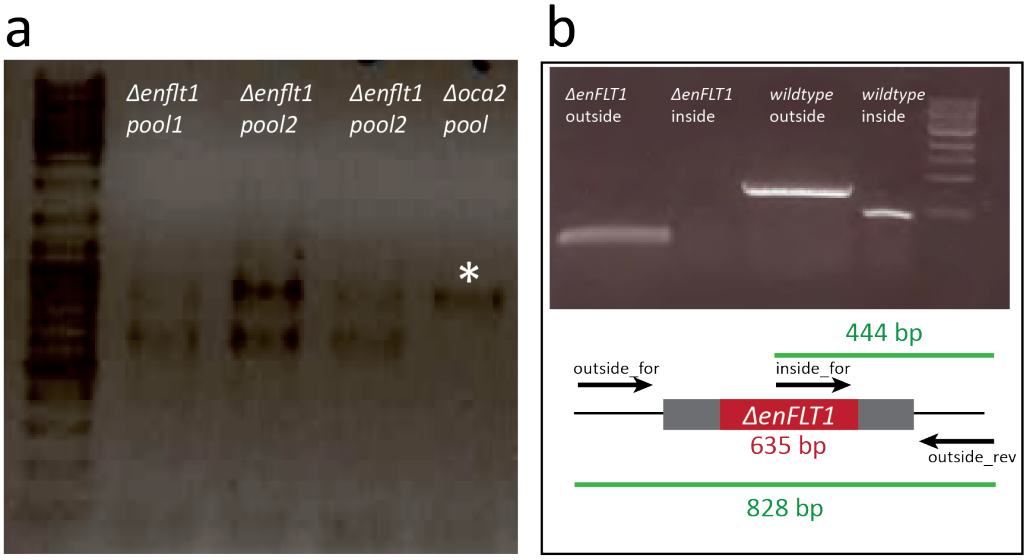


Fig. 8-1: Expression of FLT1 does not seem to be perturbed in *ΔenFLT1* and *Δex1FLT* pluripotent stem cell derived endothelial cells. Immunohistochemistry of FLT1 in wildtype, *ΔenFLT1* and *Δex1FLT* endothelial cells at d14 derived from sorted CD34/CD31 double positive endothelial cells. Nuclei were labelled with Hoechst. Stainings were inconsistent and not considered for analysis.



c

hg19_dna range=chr13:28982398-28983225 5'pad=0 3'pad=0 strand=+ repeatMasking=none

GTTCACTTGGGGCTGTTGACCTCTGACACCCCACACCTCCCCCATCCCAC ACCCATGACTCTGAGGGCAGTTACTAAGGGCACAAGCCCTAGTATGGAAA CTATAGGCCAGACTTATTGCCAGCTCTTCCACATAAAAAGGACAGATTCT GTGGCTATTTTCTAATTTAGGGGCTGTTATTTTTCCCTCTTTGTAATTGG GGCTATAAAAATAGAATGTCATTTAGAAGCCTTACTATATAACAGCACTT GAAAAGAGATTTAGGGGGAGGGGAATGTGCCTAAGCCTTAAGAATACATC AAGCTCTTAAGGAAGTAGATTGTGCCATGTTTGAAGATGCAAAGATAATC ACTGTGCTGGGAACTGTCTTGAGGGGGGAGGAAGCCGAACAACAAGAGGA AAGCAGCGTTCGCTCCTGATTTCGCACAGGAAGTCCACAGCACCTCCTGC CAAAGCGAGATAAGAAAGCTCCCATGCACTATTCTGGACAAGGCCACTGT TATCAGCCTCAGCTCGGCGCTGTCCTCTCTTTGCTCCAAGCCTCCGTCCC TTTCCTAATGAGCTATATGACTCCTACCTCCATTTGCATACAGCATGCCA AGTACAATGGCAGGGAGGTATTTATAACCCCTGAAAGGAAAAAAAAAATT ATGGGCTAAGAGTGTACTCCATGTCATCTCTGCTGATTGCTCTGGACCTG AAACAACTTAATTTCGGACTAGAGATTGATATAAATAAATTAAAAGAAAG CAGCCAATTATGTTTCCAACACATGGAGCTAGGACCTCATGTGTGATATG TCAGCCTTCTACATTCCAAGTGGGAGCC

Fig. 8-2: Genotyping and sequencing of *Δenflt1* medaka and *ΔenFLT1* human embryonic stem cells. a) Screening PCR of gene edited embryos injected with CRISPR/Cas9 targeting *enflt1* or *oca2*. Asterisk indicate the wildtype band at 879bp. Excision of enhancer led to a band with a size of 634 bp. Samples contained pooled gDNA from several injected embryos. b) Screening PCR targeting the deleted RE in hESCs. 3 Primer PCR approach was used to amplify gene edited locus. Band in *ΔenFLT1* outside, was cut out and sent for sequencing. c) Sequence of *enFLT1* in *ΔenFLT1*. The region highlighted in red was excised by CRISPR/Cas9 gene editing to generate *ΔenFLT1*.

Table 8-1: Uniquely differentially expressed genes in *ΔenFLT1*

#	Gene-ID	Gene name
1	ENSG00000236956	*NF1P8*
2	ENSG00000053501	*USE1*
3	ENSG00000155816	*FMN2*
4	ENSG00000092051	*JPH4*
5	ENSG00000178538	*CA8*
6	ENSG00000036448	*MYOM2*

7	ENSG00000104856	*RELB*
8	ENSG00000152467	*ZSCAN1*
9	ENSG00000287872	*AC026956.2*
10	ENSG00000112149	*CD83*
11	ENSG00000167637	*ZNF283*
12	ENSG00000115008	*IL1A*
13	ENSG00000140873	*ADAMTS18*
14	ENSG00000183166	*CALN1*
15	ENSG00000197857	*ZNF44*
16	ENSG00000287774	*AC025614.2*
17	ENSG00000223749	*MIR503HG*
18	ENSG00000231431	*FAR2P4*
19	ENSG00000104332	*SFRP1*
20	ENSG00000171295	*ZNF440*
21	ENSG00000203950	*RTL8A*
22	ENSG00000230450	*NEK2P4*
23	ENSG00000229109	*AL137847.1*
24	ENSG00000106688	*SLC1A1*
25	ENSG00000130513	*GDF15*
26	ENSG00000256073	*URB1-AS1*
27	ENSG00000184058	*TBX1*
28	ENSG00000154188	*ANGPT1*
29	ENSG00000198208	*RPS6KL1*
30	ENSG00000197044	*ZNF441*
31	ENSG00000170373	*CST1*
32	ENSG00000132170	*PPARG*

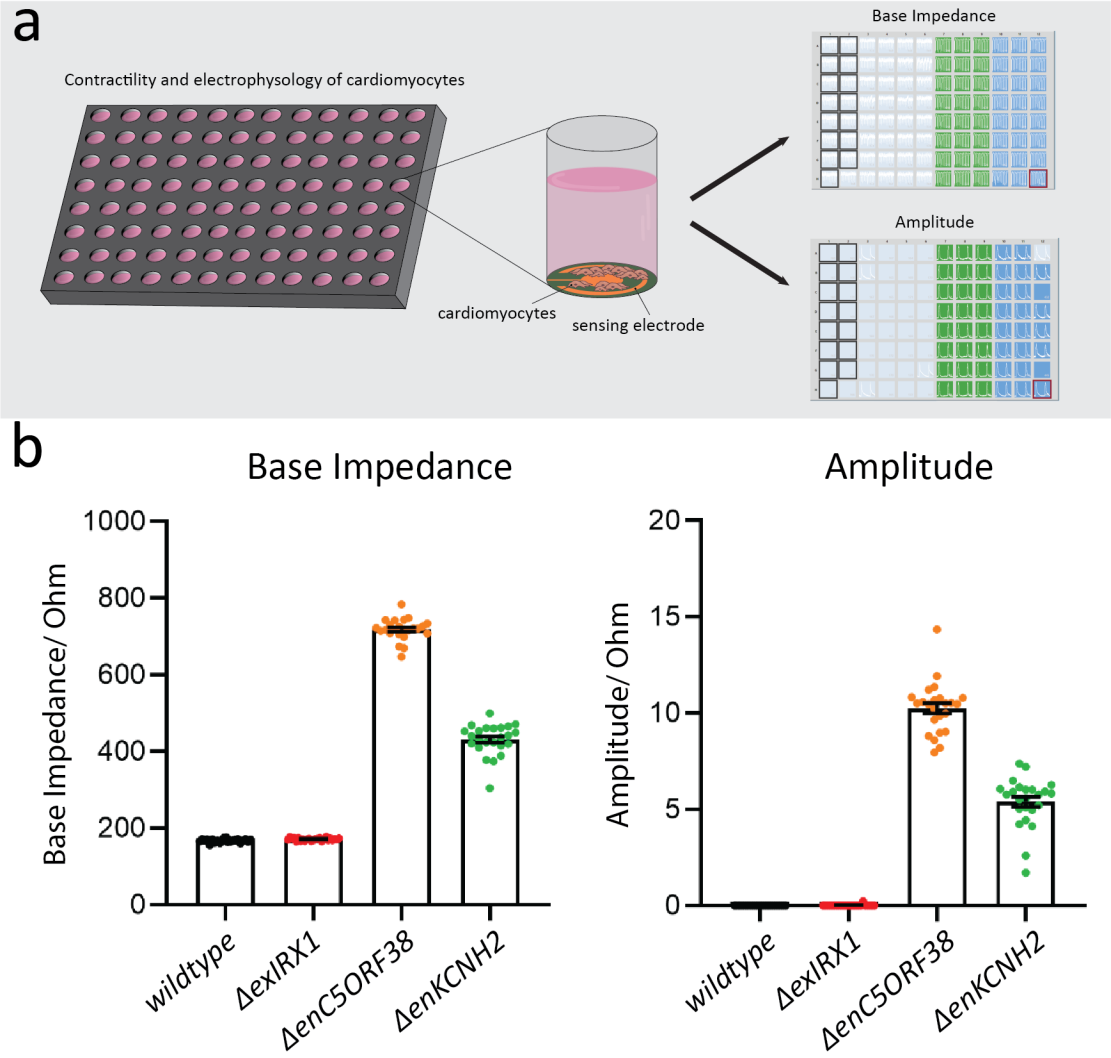


Fig. 8-3: Electrophysiological analysis of RE knockout cardiomyocytes. a) Cells were plated onto 96-well plates (CardioExcyte 96) and the base impedance and amplitude were measured. b) Preliminary results (n=1) of electrophysiological properties (impedance and amplitude) in *wildtype*, *ΔexIRX1*, *ΔenC5ORF38* and *ΔenKCNH2* cells. Measurements of control groups (*wildtype* and *ΔexIRX1*) failed.

8.3 Research output

8.3.1 Understanding cis-regulatory elements in cardiovascular development

The work of this study investigates the role of the human *FLT1* regulatory element in cardiovascular development using pluripotent stem cell derived endothelial cells and medaka as an *in vivo* model organism. In this study, I developed an *in vitro* and *in vivo* enhancer knockout model and a high throughput vessel assay. Furthermore, for this project and for the first time,

we established the Japanese rice fish as a model organism in Australia. The data presented in this research article encompasses an extensive part of my PhD project and is described in detail in Chapter 3 and 4 of this thesis. It is currently under revision at eLife and is available as a preprint at bioRxiv.

The human FLT1 regulatory element directs vascular expression and modulates angiogenesis pathways *in vitro* and *in vivo*

Julian Stolper, Holly K. Voges, Michael See, Neda Rahmani Mehdiabadi, Gulrez Chahal, Mark Drvodelic, Michael Eichenlaub, Tanya Labonne, Benjamin G. Schultz, Alejandro Hidalgo, Lazaro Centanin, Jochen Wittbrodt, Enzo R. Porrello, David A. Elliott, Mirana Ramialison

bioRxiv 2021.03.03.433738; doi: https://doi.org/10.1101/2021.03.03.433738

8.3.2 Understanding cardiac biology in the fish model system

This review article summarizes recent work in the field of spatial transcriptomics in cardiac research in zebrafish. We have reviewed state of the art publications and summarized the current standards. In light of this PhD thesis, this article was written to highlight and understand the cardiac development and biology in the fish model system. I contributed to this work as a co-first author.

Towards spatio-temporally resolved developmental cardiac gene regulatory networks in zebrafish

Jeannette C. Hallab*, Hieu T. Nim*, Julian Stolper*, Gulrez Chahal*, Lisa Waylen*, Francesca Bolk*, David A Elliott, Enzo R. Porrello, Mirana Ramialison

*co-first authors

Under review in Briefings in Functional Genomics

8.3.3 Understanding cardiac transcription factors and their role in CHD

This work was published as a book chapter summarizing human transcription factor networks of the heart and their role of binding regulatory elements to initiate gene expression. I

contributed as a co-author to this article as part of understanding the intricate network of master regulators of heart development, such as NKX2-5 and their role in the contribution of congenital heart diseases.

Human Cardiac Transcription Factor Networks

Akriti Varshney*, Gulrez Chahal*, Luana Santos*, Julian Stolper*, Jeannette C. Hallab*, Hieu T. Nim*, Martin Nikolov, Adrian Yip, Mirana Ramialison

Systems Medicine, Volume 2, Pages 429-453 (2021). 10.1016/B978-0-12-801238-3.11597-1

*co-first authors

8.3.4 Understanding and developing skillsets for the medaka as a model stystem

These research articles were published as part of a collaboration in order to understand stem cells *in vivo* during organogenesis and organ homeostasis in the medaka model system. For the study I published as a first author in eLife, I imaged and analysed more than 9000 functional units of the gill and discovered four different stem cell populations contributing to postembryonic organ growth. This work also laid the groundwork to introduce medaka as a new model organism in Australia and to develop the skillset to utilize these fish in my main PhD project.

Identifying stem cell numbers and functional heterogeneities during post-embryonic organ growth

Diana-Patricia Danciu, Julian Stolper, Lázaro Centanin, Anna Marciniak-Czochra,

bioRxiv 2020.04.08.032110; doi: https://doi.org/10.1101/2020.04.08.032110

Stem cell topography splits growth and homeostatic functions in the fish gill

Julian Stolper, Elizabeth Mayela Ambrosió, Diana Patricia Danciu, Lorena Buono, David A. Elliott, Kiyoshi Naruse, Juan R. Martínez-Morales, Anna Marciniak-Czochra, Lázaro Centanin.

Elife. (2019) PMID: 31090541

Neural stem cells induce the formation of their physical niche during organogenesis.

Ali Seleit, Isabel Krämer, Bea F. Riebesehl, Elizabeth Mayela Ambrosió, Julian S. Stolper, Colin Q. Lischik, Nicolas Dross, Lázaro Centanin.

Elife. (2017) PMID: 28950935

8.3.5 Understanding enhancer grammar and TFBS composition

This research article was published to highlight the importance of TrawlerWeb, a TF motif discovery tool for next-generation sequencing. I contributed to test the efficacy of Trawler across other software and participation in this project helped me to understand motif composition of regulatory elements and transcription factor binding.

TrawlerWeb: an online de novo motif discovery tool for next-generation sequencing datasets.

Louis T. Dang, Markus Tondl, Man Ho H. Chiu, Jerico Revote, Benedict Paten, Vincent Tano, Alex Tokolyi, Florence Besse, Greg Quaife-Ryan, Helen Cumming, Mark J. Drvodelic, Michael P. Eichenlaub, Jeannette C. Hallab, Julian S. Stolper, Fernando J. Rossello, Marie A. Bogoyevitch, David A. Jans, Hieu T. Nim, Enzo R. Porrello, James E. Hudson, Mirana Ramialison.

BMC Genomics. (2018) PMID: 29621972

8.4 COVID-19 related impacts

For the assessors of this thesis, COVID-19 has had a significant impact on this project. Due to working in a hospital, laboratory work was stopped for 2 months in my final year of my PhD and working hours were severely impacted for the remainder of 2020. Strict lockdown policies in Melbourne have led to only limited access to facilities and institutes. Several planned *in vivo* (*e.g.* crossing of fish lines) and *in vitro* experiments were not able to be finalised. Nevertheless, we were able to write and submit a manuscript detailing a large proportion of the work conducted over the course of my PhD.