

Dissecting the upstream regulation of the GATA genes during heart development

A Thesis submitted for the degree of Doctor of Philosophy
Hilary Term 2019

Department of Physiology, Anatomy and Genetics
University of Oxford

Caroline Wennington (née Telfer)

St Hugh's College

Supervised by

Dr Filipa Simões, Professor Paul Riley and Professor Roger Patient

Abstract

Dissecting the upstream regulation of the GATA genes during heart development. Caroline Wennington (née Telfer), St Hugh's College, DPhil Cardiovascular Regenerative Medicine, HT 2019.

The cardiac GATA transcription factors (TFs) play vital roles in the developing myocardium and potentially during regeneration. While their function as cardiac effectors has been well studied, the precise molecular mechanisms driving their activity have yet to be parsed. Here, I have investigated the upstream regulation of the cardiac GATA factors during cardiogenesis in zebrafish by interrogating their cis-regulatory genomic landscapes. Using cardiomyocyte-specific ATAC-Seq datasets, I have identified putative enhancers around the *gata4*, *5* and *6* loci that are accessible during cardiac development and regeneration and tested their activity *in vivo* using a novel, highly efficient reporter system in zebrafish.

I observed heart-specific reporter activity driven by the majority of the newly identified putative gata5 and gata6 enhancers, within particular subpopulations of these TFs domains of expression. The reported activity of Tg(Gata5-E5-mCherry) was of particular interest. I characterised Gata5-E5-mCherry positive populations using a combination of microscopic and transcriptomic analyses. I then tested the requirement of overrepresented TF binding site motifs for the activity of Gata5-E5 by assessing changes in enhancer-driven reporter expression after binding site perturbation. Gata5-E5 was active in a population of extracardiac cells adjacent to the inflow tract that expressed epithelial markers, hypothesised to act as support cells for second heart field progenitor cells. Gata5-E5 was also active in a population of CMs in the inflow tract region that expressed the hallmarks of cardiac conduction system cells. Binding site studies revealed a potential upstream regulation by Gata5, Nkx2.5/Nkx2.7 and Tbx2a/2b/5a to drive the activity of gata5 in these cells.

Gata5-E5 was also active in CMs throughout the atrium and ventricle and binding site studies suggested an upstream role for Gata5, Meis1b, Nkx2.5/Nkx2.7 and Tbx20 in regulating Gata5-E5 activity in this subpopulation of CMs.

Future experiments will involve ChIP-Seq to test the direct interactions suggested by binding site perturbation studies, and Capture-C to investigate topologically associating domains, ultimately revealing the functional interaction between these enhancers and their regulated GATA genes. Taken together, my work has begun to uncover the exact upstream spatiotemporal regulation of the cardiac GATA factors in various subpopulations of the heart during early cardiogenesis at high resolution. Comprehension of these complex networks is critical to understand the molecular basis of GATA-related congenital heart defects.

For Mum and Dad,

my heroes

“I’ve learned that I still have a lot to learn..”

-Maya Angelou

Acknowledgements

I am in the extremely fortunate position that this is a very long list, so bear with me!

First and foremost, Filipa. Words can't do justice to how much I appreciate everything she has done for me. She has been the most wonderful mentor, both inside and outside of the lab and I truly admire her passion, drive and loyalty. None of this work would have been possible without her outstanding support and encouragement and I will carry everything she has taught me for years to come. I am so grateful to Tatjana, for taking a chance on me. I am regularly in awe of her brilliant mind and 'can-do' attitude to novel technologies, and I so admire the way she runs her lab with passion, determination and a giant heart. I would like to thank Paul, for his amazing insight and ability to always see the 'big picture'. Even at the toughest points of this process, I don't think I once left a meeting with him without feeling excited about my work. Thanks go to Roger for giving me the opportunity to work for him and for creating an inspiring work environment. A huge thank you to the British Heart Foundation for funding my DPhil programme.

To Patient lab members, in particular: Michael, for his unshakeable patience and kindness, and for demonstrating what you can achieve with hard work and dedication; Jana for teaching me everything she knows about zebrafish and for being an imaging guru and helping me so much with confocal microscopy; Arif, for his constant support and encouragement; Aldo for his vast knowledge, for always sharing new papers, and for his company when in the lab during antisocial hours; Lorna for her contagious enthusiasm and supportive nature; Lei for demonstrating the value of hard work. Thanks also go to Patient lab members past: Tomek, Flo, Rui, Nicola, Monika, Rosella, Tessa and Phil. I would also like to thank Marella for her advice and support.

I would like to thank the Riley lab for their insightful comments during lab meetings, and for the brilliant socials. In particular I would like to thank Dee for always being so helpful, Katie and Tertia for always checking in and making sure I am doing ok and for their amazing organisational skills. Thank you to my fellow students on this programme, Ellie and Tonia, for sharing in the highs and the lows of DPhil life. Thanks go to Sarah Noujaim for her words of wisdom, advice and encouragement.

Huge thanks go to the Sauka-Spengler lab: to Martyna for being encouraging and inspiring in equal measure; Vanessa for her endless knowledge of molecular biology, which she was always willing to share, her admirable work ethic and her kindness; Ruth for her unwavering patience and sequencing wisdom; Irving for all of his help with my single-cell sequencing experiments and for always making time to answer even the most inane of questions; Ivan for his amazing insight into all things developmental biology, and for his compassionate nature; Sarah for always being so helpful, and for making all of our shared tasks fun and Mara for her encouragement and kindness.

Thanks go to the aquarium team, Alison, Andy, Greg, Kelly, Adam and Matt, for looking after my fish so well and for creating such a fun and supportive working environment; especially to Greg, for always reminding me to eat lunch! I would further like to acknowledge the staff at the various facilities I used for their invaluable work, Chris at the WIMM imaging facility, Kevin and Sally at the WIMM FACS facility and Neil at the WIMM single cell facility.

Endless thanks to Lizzie, who has been there through literally every stage of my life, always both supporting and inspiring me, and for picking up the pieces more times than I would care to admit! I will always be grateful to my primary school self for her choice of friends: for Rozz, Shez, Hats, Oli and Lou (and now Corn, Andy and Matt too!). I am so lucky to have grown up in the knowledge that I will always have my support team of strong, intelligent and always hilarious women. We will always be each other's biggest cheerleaders and that makes this world a lot less scary! A huge thank you to Nina and Mark (and now Peter!), my first friends in Oxford, for being an absolute pillar of strength throughout this whole process. Thank you for always cheering me up when I needed it, and for so many fun nights that I will always treasure. Thank you to my Sheffield girls, Jo, Kate, Hannah, Sara, Debbie and Freya, for always being supportive and brilliant despite having to navigate the ever-demanding Junior Doctor years. Thank you to my Future Science Group crew for always making me laugh. To Niamh, for being a shining light and for her endless optimism and support and to Katy and Sven (and now Rudi), for their amazing encouragement and support.

Massive thanks go to my family. To my mum and dad, for their relentless (if sometimes misguided) belief in my abilities; knowing that they will always be there for me has been the most amazing safety net for everything I do in life. For fully supporting me in everything I do and for always being interested, and for raising me with love and humour in equal measure. Also, to my extended family, in the UK and Australia, for always believing in me. Thank you to my new family: Steve, Laura, Oliver, William, Alison and Tiago, for accepting me as one of your own and for always cheering me on.

Last, but by no means least, my Matthew, for believing in me and for never failing to make me laugh. For always calling me out when I'm being ridiculous and supporting me when I'm not. I could not have done this without his unwavering support. Of course, not forgetting my four-legged therapists, Alpha and Orla, for always greeting me with excitement when I got home from the lab, for sitting with me while I wrote every word in this thesis, and for generally making life better.

Table of Contents

TABLE OF FIGURES **14**

TABLE OF TABLES **17**

LIST OF ABBREVIATIONS **18**

1) INTRODUCTION **20**

1.1 The zebrafish as a model organism to study the developing and regenerating heart *20*

1.2 Heart development in the zebrafish *20*

1.2.1. The first and second heart field *25*

1.2.2 Trabeculation *27*

1.2.3. Cardiac conduction system *28*

1.2.4 Epicardium *29*

1.2.5 Neural crest *29*

1.3 Heart regeneration in the zebrafish *30*

1.4 The cardiac GATA TFs *35*

1.5 The cis-regulatory landscape *41*

1.5.1. Transcriptional regulation *41*

1.5.2. The eukaryotic transcriptional process *41*

1.5.3. Transcriptional regulatory elements *43*

1.5.4. The chromatin landscape *49*

1.6 Enhancer identification *49*

1.6.1. TF binding *50*

1.6.2 Comparative genomic features *50*

1.6.3. Epigenetic features *52*

1.6.3 eRNA *56*

1.6 Transgenesis to test putative enhancers in vivo *57*

1.7 THESIS AIMS 59

2) MATERIALS AND METHODS **60**

2.1 Zebrafish husbandry 60
2.2 Zebrafish lines 60
2.3 In Situ Hybridisation 61
2.3.1 In vitro transcription of probes for whole-mount in situ hybridisation 61
2.3.2 Embryo fixation and preparation for in situ hybridisation 61
2.3.3 Whole mount in situ hybridisation 62
2.4 Ac/Ds-mediated enhancer screening 63
2.4.1 Genomic DNA extraction 64
2.4.2 Polymerase chain reaction (PCR) 64
2.4.3 Gel purification 67
2.4.4 Vector restriction digest 67
2.4.5 In-Fusion cloning 67
2.4.6 Transcription of Ac RNA 68
2.4.7 Zebrafish embryo micro-injection 68
2.5 Zebrafish cryoinjury 69
2.5.1 Cryoinjury and heart extraction 69
2.5.2 Paraffin embedding and sectioning 69
2.6 Immunocytochemistry (ICC) 70
2.6.1 Embryo whole-mount ICC 70
2.6.2 ICC on adult heart sections 70
2.7 Hybridisation chain reaction 71
2.8 CRISPR-Cas9 editing 72
2.8.1 Synthesis of sgRNA 72
2.8.2 Cas9 mRNA synthesis 74
2.8.3 CRISPR-Cas9 injections 75
2.8.4 High resolution melt analysis (HRMA) 75
2.9 Microscopy and digital image analysis 76
2.9.1 Microscopy 76
2.9.2 Image analysis 77

2.10 10X SINGLE CELL RNA-SEQ 77

2.10.1 Sequencing *77*

2.10.2 Analysis *78*

2.10.3 GO term analysis *80*

2.11 BINDING SITE ANALYSIS 80

2.11.1 JASPAR analysis *80*

2.11.2 Binding site perturbation constructs *80*

2.12 BLINDING, QUANTIFICATION AND STATISTICS 81

3) IDENTIFICATION OF PUTATIVE ENHANCERS OF THE CARDIAC GATA GENES AND *IN VIVO* TESTING 82

3.1 INTRODUCTION 82

3.2 RESULTS 83

3.2.1 Putative enhancers for the cardiac GATA genes can be predicted using ATAC-Seq and ChIP-seq datasets. *83*

3.2.2. Putative enhancers for gata4 showed no activity in vivo *86*

3.2.3. Putative enhancers for gata5 showed heart-specific activity in vivo *88*

3.2.4. Putative enhancers for gata6 showed heart-specific activity in vivo *90*

3.2.5. Gata5-E5 labels a subset of cells in and adjacent to the inflow tract of the developing heart. *91*

3.2.6 Gata5-E6 labels a subset of cells in and adjacent to the inflow tract of the developing heart at early stages that becomes restricted to the ventricle and AVC by 5dpf *93*

3.2.7 Gata5-E8 activity in the F1 generation did not reflect the heart-specific patterns observed in the transient F0 analysis *95*

3.2.8 Gata6-E12 labels a population of CMs that becomes restricted to the developing ventricle. *97*

3.3 SUMMARY 98

4) CHARACTERISATION OF THE NOVEL GATA ENHANCERS GATA5-E5 AND GATA6-E12 99

4.1 INTRODUCTION 99

4.2 RESULTS 101

4.2.1 Tg(Gata5-E5 mCherry) was expressed in a subset of gata5-expressing cells 101

4.2.2 Tg(Gata6-E12-mCherry) reported expression throughout the early heart tube, becoming restricted to the developing ventricle by 3dpf 103

4.2.3 Hybridisation chain reaction showed that Gata5-E5 was active in the same cells as endogenous gata5 was expressed 105

4.2.4 Hybridisation chain reaction revealed that Gata6-E12 was active in the same cells as endogenous gata6 was expressed 106

4.2.5 Gata5-E5 was also active in the adult heart and its activity increased around the injury border zone following cryoinjury 107

4.2.6 Gata6-E12 was active in the adult heart, in the same cell population as it was active in the embryo 110

4.2.7 Gata5-E5 KO reduced gata5 expression and enhancer activity in the inflow tract-adjacent cell population 111

4.2.8 Excision of Gata6-E12 led to a small decrease in enhancer-driven reporter expression that was rescued by 5dpf 116

4.2.9 Gata5-E5-cherry expression pattern resembled that of islet1 119

4.2.10 Islet1 expression was reduced in the region adjacent to the inflow tract upon Gata5-E5-KO, suggesting a loss of SHF progenitor cells, and in the putative CCS 121

4.2.11 Analysis of the hearts of Gata5-E5-KO embryos at 5dpf showed a reduction in atrial area and atrial hypoplasia 123

4.3 SUMMARY 125

5) SINGLE CELL TRANSCRIPTOMICS IDENTIFIES THE SUBPOPULATIONS IN WHICH GATA5-E5 IS ACTIVE, AND PROVIDES CANDIDATE FACTORS FOR ITS UPSTREAM REGULATION 128

5.1 INTRODUCTION 128

5.2 RESULTS 130

5.2.1 Single cell transcriptome results 130

5.2.2 Initial dataset clustering 131

5.2.3 Reclustered results based on mCherry transcript expression 132

5.2.4 Cluster 0 was enriched in transcripts typical of epithelial cells *135*

5.2.5 Cluster 1 was enriched in transcripts for genes associated with protein synthesis *141*

5.2.6 Clusters 2 and 3 were enriched for transcripts inferring CM identity *141*

5.2.8 Clusters 4, 5 and 7 were enriched for neural markers *142*

5.2.7 Cluster 6 was enriched in transcripts encoding markers of the cardiac conduction system *142*

5.3 SUMMARY 147

6) BINDING SITE PERTURBATION REVEALS UPSTREAM REGULATION OF GATA5-E5 149

6.1 INTRODUCTION 149

6.2 RESULTS 151

6.2.1 Candidate TF Selection *151*

6.2.2 Gata5 was necessary for the activity of Gata5-E5 during cardiogenesis from 2dpf *154*

6.2.3 Meis-Gata5-E5 binding interactions were required in the ventricle and AVC populations where Gata5-E5 activity was reported *158*

6.2.4 NKX-Gata5-E5 binding interactions were required in the atrium and AVC populations *161*

6.2.5 TBX-Gata5-E5 binding interactions were required in the atrium and AVC populations *166*

6.2.6 PRRX/SHOX-Gata5-E5 binding interactions were not required for enhancer activity *170*

6.2.7 FOX binding interactions are not required for the activity of Gata5-E5 ..... *175*

6.3 SUMMARY 178

7) DISCUSSION AND FUTURE DIRECTIONS 180

7.1 INVESTIGATING THE UPSTREAM REGULATION OF THE CARDIAC GATA FACTORS BY INTERROGATING THE CIS-REGULATORY ENVIRONMENT 180

7.2 INVESTIGATING THE FUNCTION OF GATA5-E5 AND GATA6-E12 BY OBSERVING THE CONSEQUENCES OF THEIR GENOMIC DELETION 186

7.3 CHARACTERISING THE GATA5-E5-POSITIVE SUBPOPULATIONS OF CELLS OF THE DEVELOPING HEART BY SINGLE CELL TRANSCRIPTOMICS 189

Cluster 0 ..... *191*

Cluster 1 ..... *193*

Cluster 2 and 3 ..... *193*

Cluster 6 ..... *194*

Cluster 4, 5 and 7 ..... *195*

7.4 IDENTIFYING THE UPSTREAM REGULATION OF *GATA5* USING BINDING SITE ANALYSIS AND PERTURBATION 197

7.5 FINAL REMARKS 206

REFERENCES 209

SUPPLEMENTARY FIGURES 229

Table of Figures

Figure 1-1 Stages of heart development in the zebrafish 24

Figure 1-2 ATAC-Seq schematic..... 54

Figure 1-3 Histone marks at regulatory elements 56

Figure 2-1 Cellranger aggregate output from two replicate runs..... 79

Figure 3-1 Summary of ATAC-Seq and ChIP-Seq data collection..... 84

Figure 3-2 In situ hybridisation to show the expression of gata4, gata5 and gata6 at 48hpf 86

Figure 3-3 Workflow for screening putative enhancers 87

Figure 3-4 Selected putative enhancers for gata4 did not show activity in vivo. ... 88

Figure 3-5 Putative gata5 enhancers showed heart-specific activity in vivo 89

Figure 3-6 3 out of 4 tested putative gata6 enhancers showed heart-specific activity in vivo 91

Figure 3-7 In the stable transgenic line, Gata5-E5-driven reporter shows expression in a subpopulation of cells in the inflow region 92

Figure 3-8 In the stable transgenic line, Gata5-E6-driven reporter shows expression in a subpopulation of cells in the inflow region early on that becomes restricted to the ventricle and AVC by 5dpf 95
Figure 3-9 Gata5-E8 reporter expression in the stable transgenic line did not reflect what was observed in the F0 generation 96
Figure 3-10 In the stable transgenic line, Gata6-E12-driven reporter shows expression in a subpopulation of CMs that becomes restricted to the ventricle. .. 97
Figure 4-1 In the F2 generation of the stable transgenic line, Tg(Gata5-E5-mCherry), Gata5-E5 was active in specific subpopulations of the developing heart. 103
Figure 4-2 Tg(Gata6-E12-mCherry) expression became restricted to the ventricular CMs and bulbus arteriosus by 3dpf 105
Figure 4-3 HCR confirmed that Gata5-E5-mCherry was expressed in the same cells as endogenous gata5. 106
Figure 4-4 HCR showed that Gata6-E12-mCherry was expressed in the same cells as endogenous gata6 transcripts. 107
Figure 4-5 Gata5-E5 was active in the adult heart and is activated in the injury-proximal region following cryoinjury 109
Figure 4-6 Gata6-E12 was active in the ventricle of the adult heart 111
Figure 4-7 Excision of the endogenous Gata5-E5 enhancer led to a decrease in enhancer-driven reporter expression and a reduction in endogenous gata5 expression. 116
Figure 4-8 Gata6-E12 excision led to a transient decrease of enhancer-driven reporter expression 119

Figure 4-9 Gata5-E5 activity overlapped with Islet expression in the early cardiac conduction system cells. 121
Figure 4-10 Reduced levels of Islet protein in the region adjacent to the inflow tract and in the inflow tract after Gata5-E5 excision suggests a loss of putative posterior SHF progenitor cells and CCS cells in these embryos..... 123
Figure 4-11 Gata5-E5 excision caused a reduction in atrium area and atrial cell numbers by 5dpf..... 125
Figure 5 -1 Workflow of 10x Chromium single-cell sequencing 129
Figure 5 -2 Marker genes identified 12 distinct populations within the single cell dataset..... 132
Figure 5 -3 Marker genes identified 8 distinct populations within the single cell dataset..... 133
Figure 5-4 The single cell data set contained 8 distinct cell clusters. 135
Figure 5 -5 Transcripts encoding the cluster 0 markers prrx1a, jam2b, col1a1b and adamts15a were present in the extracardiac region adjacent to the inflow tract. 141
Figure 5 -6 Transcripts encoding the cluster 6 markers shox2, sema3fb, col14a1 were present in the inflow tract..... 147
Figure 6-1 Workflow for binding site perturbation studies 154
Figure 6 -2 gata5 and gata6 are expressed in clusters 0, 2, 3 and 6. 155
Figure 6 -3 Gata5 binding is necessary for the activity of Gata5-E5 from 2dpf... 157
Figure 6-4 meis1a and meis1b were expressed throughout the cell clusters studied..... 158
Figure 6-5 Meis binding was necessary for the activity of Gata5-E5 in the ventricle and AVC. 160
Figure 6-6 Nkx2.5 was strongly expressed in clusters 2, 3 and 6 164

Figure 6 -7 Nkx binding was necessary for the activity of Gata5-E5 in the atrium and AVC. 165
Figure 6 -8 tbx1, 5a, 2a, 2b, 3a and 20 transcripts were enriched in Gata5-E5-mCherry-expressing cells 167
Figure 6 -9 Tbx binding was necessary for the activity of Gata5-E5 in the atrium and AVC. 169
Figure 6 -10 shox2 and prrx1a expression colocalised with regions of Gata5-E5-mCherry expression in the CCS and adjacent to the inflow tract, respectively ... 172
Figure 6 -11 Shox/Prrx binding was not necessary for the activity of Gata5-E5 in any of the populations studied. 174
Figure 6 -12 foxp4 and foxp1a were expressed within the single cell dataset 176
Figure 6-13 Fox binding was not necessary for the activity of Gata5-E5 in any of the populations studied. 177
Figure 7 -1 Summary of subpopulations showing Gata5-E5 activity at 48hpf 196
Figure 7 -2 Summary of binding site perturbation studies 205

Table of Tables

Table 1 Details of whole-mount in situ hybridisation probes used in this thesis ... 61
Table 2 Details of genomic locations of putative enhancers 64
Table 3 Details of primers used to amplify putative enhancer regions 66
Table 4 Details of the oligos used for sgRNA synthesis 74
Table 5 Details of HRMA primers to test sgRNA cutting efficiency 76
Table 6 Predicted binding sites within Gata5-E5 152

List of abbreviations

A - atrium

ATAC - assay for transposase accessible chromatin

AVC- atrioventricular canal

AVN- atrioventricular node

BA - bulbus arteriosus

CCS- cardiac conduction system

CM - cardiomyocyte

dpf - days post fertilisation

dpi- days post injury

EMT - epithelial-to-mesenchymal transition

FACS - fluorescence assisted cell sorting

FHF- first heart field

FPKM - fragments per kilobase of transcript per million transcripts

GFP - green fluorescent protein

HCR - hybridisation chain reaction

hpf - hours post fertilisation

ICC- immunocytochemistry

PBS - phosphate buffered saline

PFA – paraformaldehyde

PIC- preinitiation complex

RA- retinoic acid

SAN- sinoatrial node

SHF- second heart field

SSC - sodium chloride sodium citrate

TF- transcription factor

Tg - transgenic

t-SNE - t-distributed stochastic neighbour embedding

V – ventricle

1) Introduction

1.1 The zebrafish as a model organism to study the developing and regenerating heart

The zebrafish is a small tropical freshwater teleost that is a popular and powerful vertebrate model organism in research, with applications in development, regeneration, genetics and toxicology. Zebrafish offer several advantages as a model organism: they can be housed in large numbers; they have a short generation time; they can produce hundreds of offspring at weekly intervals; adult fish can breed year-round in laboratory conditions; external fertilisation facilitates easy embryo manipulation; the genome has been fully sequenced; and there are a large number of tools available for fluorescent reporter transgenesis and both forward and reverse genetics (Bakkers, 2011).

The zebrafish is a particularly powerful model in which to study the heart as their embryos are transparent, allowing visualisation of organ development using simple light microscopy (Kimmel et al., 1995). One particular advantage is that zebrafish develop a functional heart within 24 hours post-fertilisation (hpf) but are able to survive without circulation for the first 4-5 days of embryonic development, allowing investigation of malformations that would be otherwise lethal in other vertebrate organisms. Despite the anatomical differences between zebrafish and human hearts, many of the important genes and pathways have been conserved (Moorman and Christoffels, 2003).

1.2 Heart development in the zebrafish

The heart is the first organ to form and become functional in the vertebrate embryo (summarised in Figure 1-1). The zebrafish heart comprises just two chambers; one

atrium and one ventricle, as well as the sinus venosus (inflow tract) and the bulbus arteriosus (outflow tract). The bulbus arteriosus is unique to the teleost lineage (Grimes et al., 2010). It is a chamber-like structure comprising smooth muscle, collagen and elastic fibres, making it highly distensible. It is believed to be an evolutionary adaption to its short distance from the gill apparatus and to maintain constant blood perfusion and compensate for high pressure in a manner similar to the mammalian aortic arch (Grimes and Kirby, 2009). The mature zebrafish heart has two valves; the atrioventricular (AV) valve and the bulboventricular (BV) valve.

Despite the morphological differences between the zebrafish heart and the mammalian 4-chambered heart, the developmental mechanisms are remarkably similar. The network of cardiac transcription factors required for specification and differentiation of cardiac precursors, including GATA, T-BOX, NKX and HAND factors is conserved (Stainier et al., 1993).

Cardiac progenitor cells that give rise to the heart are specified just before gastrulation at around 5hpf (Figure 1-1a). Fate-mapping studies have revealed that two populations of myocardial progenitor cells reside bilaterally in the lateral marginal zone: the ventricular progenitor cells and the atrial progenitor cells (Keegan et al., 2004). These pools of cells are maintained and restricted in size by retinoic acid signalling. Even at such an early stage, the atrial and ventricular progenitor cells are already segregated (Keegan et al., 2004): the atrial progenitors reside more ventrally in the lateral marginal zone than the ventricular progenitor cells (Keegan et al., 2005). By contrast to this organised distribution, atrial and ventricular endocardial progenitor cells are located across the lateral marginal zone, without apparent chamber-specific spatial organisation (Keegan et al., 2004).

Whilst maintaining their relative position, the cardiac progenitor cells, which express *gata4, gata5, gata6* and *hand2*, move dorsally to reside in the posterior section of the anterior lateral plate mesoderm (ALPM) by the 6-9 somite stage. The medially located ventricular cardiomyocytes (CMs) are the first to differentiate and begin to express *nkx2.5* and *myl7* at the 12 somite stage, followed slightly later by differentiation of the lateral atrial CMs at the 15 somite stage (Berdougo E, 2003). This region is referred to as the cardiac field and it is restricted posteriorly by pectoral fin precursors (Waxman et al., 2008) and anteriorly by the haematopoietic and cranial vascular precursors (Schoenebeck et al., 2007).

As they differentiate, the bilateral populations of myocardial precursors migrate inwards towards the embryonic midline (Figure 1-1b). This process is regulated both cell-autonomously and non-cell-autonomously by the overlying anterior endoderm, the underlying yolk syncytial layer, and by the endocardium (Stainier et al., 1996, Yelon, 2001). By 18hpf, these populations fuse and form a cone-shaped structure, with ventricular precursors at the dorsal apex, and endocardial precursors forming a central lumen (Figure 1-1c) (Stainier et al., 1993). Many zebrafish mutations that perturb the medial migration of myocardial precursors have been identified, including *faust* (*fau*), *miles apart* (*mil*), *hands off* (*han*), *bonnie and clyde* (*bon*), *one-eye pinhead* (*oep*), *casanova* (*cas*), *natter* (*nat*) and *two-of-heart* (*toh*) (Alexander et al., 1998). These mutations lead to a phenotype of cardia bifida, whereby two separate hearts form in lateral loci (Chen et al., 1996), and help us to understand the critical requirements for proper migration of myocardial precursors to the midline. *han, fau* and *oep* mutations all disrupt myocardial differentiation, suggesting that this process is critical for differentiation (Schier et al., 1997, Reiter, 1999, Yelon et al., 2000). The cardia bifida phenotype in mutants lacking the anterior endoderm

(*cas, bon, fau* and *oep*) suggested that this germ layer is critical for myocardial migration. This was corroborated by transplantation experiments demonstrating that transplanted wild type endoderm was sufficient to rescue myocardial migration in *cas* mutants (David and Rosa, 2001).

These cells then involute asymmetrically to form a linear tube by 24hpf (Figure 1-1d), which has been reported to act as a suction pump (Rohr et al., 2008, Forouhar et al., 2006). The heart tube then extends and undergoes a series of asymmetric morphological changes leading to rightwards looping of the heart from 36hpf and the displacement of the ventricle to the right of the atrium (Figure 1-1e-f). In order to prevent retrograde blood flow at this point, the necessity for cardiac valves arises. At 55hpf, squamous endocardial cells in the AVC differentiate to assume a cuboidal shape (Beis et al., 2005). At 60hpf, these cells migrate into an extracellular matrix named the cardiac jelly where they proliferate, forming endocardial cushions which subsequently differentiate into mature valves by 72hpf (Figure 1-1g) (Beis et al., 2005). Many signalling pathways are involved in coordinating the formation of the cardiac valves, including Notch, Bmp, TGF-β and Wnt (Yelon, 2001).

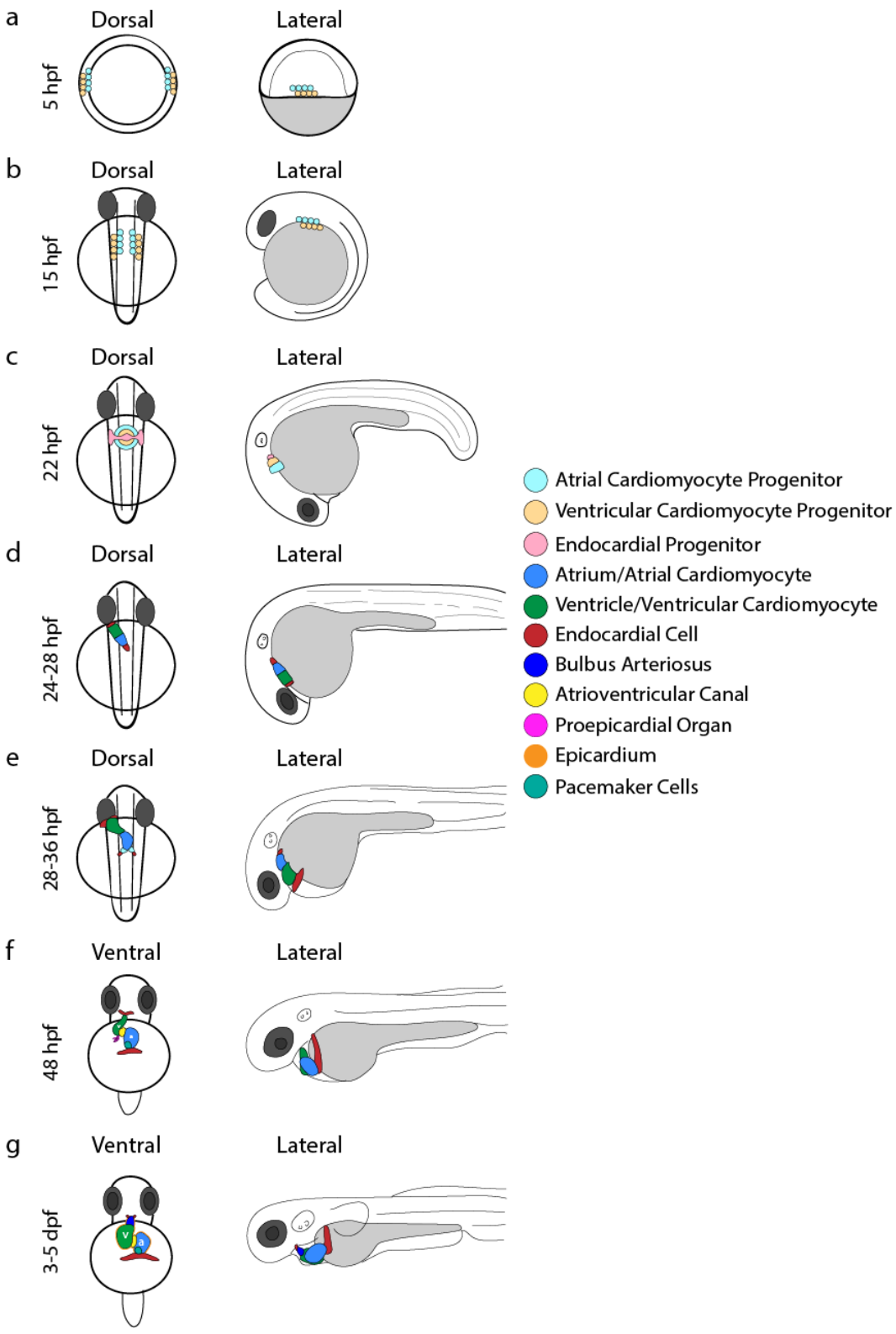


Figure 1-1 Stages of heart development in the zebrafish

a) At 5hpf, cardiac progenitor cells are located bilaterally in the lateral marginal zone. Atrial progenitor cells sit more ventrally than ventricular progenitor cells. b) Cardiac progenitors migrate bilaterally to the ALPM by 15 hpf. c) Cardiac progenitors and early endocardial cells fuse to form the cardiac disk by 22 hpf, which then starts to elongate

between 22-24 hpf. d) From 24-28hpf, linear heart tube elongation progresses and begins leftward migration. e) From 28-36 hpf, the heart tube begins looping and SHF cells begin to be added to the venous and arterial poles. At 36 hpf, the constriction at the position of the AVC is visible. f) By 48hpf, the two-chambered heart has formed. Extra-cardiac pro-epicardial cells are located near the AVC. The early pacemaker is specified at the inflow tract region of the developing heart. g) The bulbus arteriosus forms at the outflow tract. Epicardial cells begin to cover the myocardium. (Figure adapted from (Bakkers, 2011, Brown et al., 2016))

1.2.1. The first and second heart field

Mirroring what has been recorded in mammals (reviewed in (Günthel et al., 2018)), studies utilising transgenic fish expressing EGFP and dsRed under the control of the CM-specific *myl7* promoter demonstrated that the growth of the developing zebrafish heart consists of two distinct phases of CM differentiation (de Pater et al., 2009). This occurs through contribution from two spatiotemporally distinct progenitor populations (de Pater et al., 2009). Firstly, cardiac progenitor cells from the first heart field (FHF) form the initial heart tube, and then another population of progenitor cells, the second heart field (SHF) progenitors, differentiate and accrete to each pole of the developing heart, extending the atrium and ventricle and forming the inflow and outflow tracts. This has been confirmed by photoconversion assays using *Tg(myl7:kaede)* which demonstrate late-differentiating CMs being added to the venous and arterial poles (de Pater et al., 2009). Zebrafish SHF precursors are specified in the ALPM and express Gata4 and Nkx2.5. Recent studies have confirmed that the zebrafish SHF comprises a progenitor cell sheath, where cells are continuously added to and gradually differentiate at the poles of the developing heart (Felker et al., 2018). There is a large body of work investigating SHF contributions to the arterial pole of the developing zebrafish heart, confirming that mammalian signalling pathways are conserved. *Tbx1* plays an important role in SHF cell proliferation, as demonstrated by the hypoplastic outflow tract seen in zebrafish

tbx1 mutants (Hami et al., 2011). Morpholino studies have shown that *mef2cb* is necessary for the SHF-derived contribution to the outflow tract, mirroring the phenotype seen in *Mef2c* mutant mice (Lazic and Scott, 2011, Lin et al., 1997). De Pater et al. have demonstrated that the accretion of SHF-derived cells to the arterial pole requires Fgf signalling (specifically, *fgf8*) (de Pater et al., 2009). The hedgehog signalling pathway is essential for anterior SHF development as demonstrated by a small outflow tract in the *smoothened* mutant (Hami et al., 2011). TGF-β signalling has also been shown to be important in the anterior SHF, highlighted by the discovery that Ltbp3, a secreted protein regulating TGF-β ligand availability, is necessary for proper outflow tract development (Zhou et al., 2011). It has since been discovered that Nkx2.5 promotes the maintenance of *ltbp3* expression (Guner-Ataman et al., 2013).

In the mouse, the transcription factor Isl1 is expressed throughout the SHF and is required for both arterial and venous pole development, as demonstrated by defects in the development of both poles in *Isl1*-knockout mice (Cai et al., 2003). Isl1 is also expressed in zebrafish SHF progenitors, but, by contrast to what has been observed in mice, it is only required for differentiation at the venous pole of the heart. In addition to being immotile and displaying bradycardia, *isl1* mutant zebrafish embryos have a significantly reduced number of cells in the venous pole (de Pater et al., 2009), suggesting that CMs fail to differentiate in this region. Zebrafish have 4 Islet family members; isl1, isl2a, isl2b and isl1l, and all show high homology to each other and to the mouse Isl1. A recent study by *Witzel* et al. (Witzel et al., 2017) demonstrated the power of dissecting Islet-dependent gene regulatory networks in specific regions of the developing heart over time. Specifically, this study determined that *isl2b* is critical for arterial pole development from the SHF and it

controls key transcription factor expression including *mef2ca, mef2cb, hand2* and *tbx20* (Witzel et al., 2017).

By contrast to what is known about the anterior SHF that gives rise to the arterial pole of the heart, there is a paucity of knowledge about the cells of the posterior SHF, including their gene expression profiles, regulation and the precise timing of their accretion to the venous pole. The focus of all published work on the development of the inflow tract centres around the previously mentioned role of Isl1. Zebrafish studies have also highlighted a role for the LIM domain protein, Ajuba, which acts upstream of *isl1* and acts as a negative regulator to control the level of SHF differentiation (Witzel et al., 2017). In a recent study by Felker et al. (Felker et al., 2018), authors generated an EGFP reporter driven by the 3.2kb upstream region of *tbx1*. The *tbx1* reporter-expressing cells were present in a population adjacent to the inflow tract, and coexpressed the SHF marker Isl1. Lineage tracing showed that these cells contributed to the venous pole of the developing heart, confirming that the *tbx1* reporter is labelling a population of SHF progenitor cells.

1.2.2 Trabeculation

Following cardiac looping and chamber ballooning, at 60hpf ventricular CMs delaminate to form trabeculae; organised ridges lined by endocardial cells that act to increase the myocardial surface area for blood oxygenation (Tu and Chi, 2012). Many factors contribute to the process of trabeculation including signalling pathways such as Angiopoietin/Tie2, Semaphorin/Plexin and Neuregulin/ErbB2/ErbB4 (Peshkovsky et al., 2011). Haemodynamic flow is also critical for the proper formation of trabeculae (Peshkovsky et al., 2011). The outermost cortical muscle layer of the heart is formed on the surface of the ventricular myocardium as juveniles mature (Gupta and Poss, 2012).

1.2.3. Cardiac conduction system

The cardiac conduction system (CCS) is a vital system within the heart, responsible for coordinating and maintaining rhythmic contraction and propelling blood around the body. In the mammalian CCS, initial electrical impulses are generated in the pacemaker cells of the sinoatrial node (SAN). This is a specialised region of myocardium expressing a unique transcription factor profile when compared with working myocardium, including *shox2 (Blaschke Rüdiger et al., 2007)* and *isl1* (Tessadori et al., 2012). Impulses are then disseminated through specialised atrial CMs before being delayed at a region of slowly conducting CMs located in the atrioventricular canal, named the atrioventricular node (AVN). The electrical impulse then travels through the His-Purkinje system, comprising rapidly conducting ventricular CMs, allowing effective ejection of blood from the ventricles. A recent study by Hulsmans et al highlighted a role for tissue-resident macrophages in facilitating electrical conduction through the atrioventricular node via connexin 43-containing gap junctions (Hulsmans et al., 2017). Despite the previously described morphological differences between the mammalian 4-chambered heart and the zebrafish 2-chambered heart, studies using optogenetics, optical mapping and electrophysiologic analysis have revealed significant similarities in the conduction system (Sedmera et al., 2003, Arrenberg et al., 2010). From 24hpf, conduction is initiated at the venous pole and travels unidirectionally across the heart as synchronised peristaltic contractions. During cardiac looping, a conduction delay is created at the AV boundary, causing a transition from peristaltic contraction to sequential contraction of the atrium followed by the ventricle. Endocardial signalling is required for the development of AV conduction as demonstrated by the

endocardium-deficient *cloche* (*clo*) mutant which fails to develop AV conduction delay (Milan et al., 2006).

1.2.4 Epicardium

Throughout early linear heart tube and looping stages, two cellular layers make up the heart wall; the myocardium and the endocardium. A mesothelial layer called the epicardium is added to the already-looped heart and covers the outer surface of the myocardium. The epicardium derives from an extracardiac structure called the proepicardium, characterised by expression of *wt1, tbx18* and *tcf21*, which can be distinguished morphologically at 48hpf on the ventral surface of the looped heart (Serluca, 2008). By 72hpf, these cells translocate to the ventricular surface, where they collapse and proliferate, forming a continuous layer. The epicardial cell layer then begins to colonise the atrium by 5dpf (Peralta M, 2013, Simões and Riley, 2018).

1.2.5 Neural crest

Neural crest cells are an ectodermally derived population of cells that are specified in the dorsal region of the neural tube (Le Douarin, 1999). They delaminate and migrate throughout the developing embryo to contribute to a wide host of tissues, including melanocytes, neurons, thymus, craniofacial skeleton, gut smooth muscle and the developing heart (Le Douarin, 1999). The subset of neural crest cells that contribute to the heart are referred to as Cardiac Neural Crest cells (CNCC). In chick and mouse models, CNCC originate between the otic vesicle and the third somite and migrate via the pharyngeal arches 3, 4 and 6 where they then form the smooth muscle layer of the great vessels (Jiang et al., 2000). Some CNCC also contribute to the cardiac outflow tract cushion, dividing the aorta and pulmonary trunks (Kirby et al., 1983). Unlike in other models, some zebrafish CNCC directly

contribute to the myocardium (Li et al., 2003, Mongera et al., 2013). Lineage tracing experiments in zebrafish have revealed a broader region of specification, with CNCC originating between rhombomere 1 and the sixth somite (Sato and Yost, 2003). Similar lineage tracing experiments also revealed that CNCC contribute to the heart in two waves: one stream migrating through pharyngeal arches 1 and 2, which integrates into the myocardium of the early heart tube between 24 and 30hpf; and a second wave 50 hours later, migrating along the sixth pharyngeal arch artery to contribute to the bulbus arteriosus (Cavanaugh et al., 2015). Ablation of CNCC in zebrafish causes defects in myocardial maturation and prevents recruitment of SHF progenitors at the arterial pole of the developing heart (Cavanaugh et al., 2015).

1.3 Heart regeneration in the zebrafish

Despite major advances in knowledge and efficiency of treatments, cardiovascular disease remains the leading cause of mortality and morbidity worldwide (WHO). The adult mammalian heart has a limited capacity to regenerate after injury; instead, after myocardial infarction (MI), lost myocardium is replaced by collagen-rich, non-contractile scar tissue. This contributes to degeneration of heart function and often leads to heart failure, a dramatically debilitating condition for which the only current cure is heart transplant. There is, therefore, a significant clinical need to identify new therapeutic strategies for cardiac muscle regeneration after injury. One avenue that has been explored and has even reached clinical trials is transplantation of exogenous cells (N, 2014, Van Berlo, 2014). However, thus far, these methods have yielded minimal improvement in cardiac function. The most promising avenue appears to be endogenous sources of regeneration. One possibility is to induce surviving host CMs to re-enter the cell cycle, proliferate and form new myocardium

(Foglia and Poss, 2016). Another is the prospect of activating resident cardiovascular progenitor cells to produce *de novo* CMs (Foglia and Poss, 2016). If there was a subset of cells that had the potential to regenerate some, or all, lost myocardium it would be extremely therapeutically valuable.

To find such cell types, one approach to regenerative medicine is to identify examples of successful cardiac regeneration in nature, elucidate their underlying mechanisms and attempt to apply insights gained to augment regeneration in human hearts after injury through provision of appropriate stimuli. By contrast to humans, zebrafish display a remarkable regenerative capacity, being able to recover from severe cardiac injury by creating new heart muscle with minimal scarring (Poss et al., 2002).

The original model of heart regeneration used by researchers in zebrafish comprises a ventricular resection injury, in which up to 20% of the ventricular apex is surgically removed and lost tissue, including coronary vasculature, myocardium and endocardial cells, is subsequently rapidly replaced (Poss et al., 2002). However, the parallels between the resection model in zebrafish and myocardial infarction in mammals are limited. Ventricular resection causes a large blood clot and leaves only a negligible scar region with a different composition to that formed after myocardial infarction in mammals. This model also lacks the ischaemia-induced cell death and requirement for removal of cell debris (Juan Manuel González-Rosa, 2011). Because of these limitations, cryoinjury is now utilised as an alternative injury model that closely mimics the post-injury environment observed in a post-MI human patient. Using a copper filament previously cooled in liquid nitrogen, roughly 25% of the ventricle is damaged at the ventral apex. This model is therefore proposed to better resemble MI in mammals, as an injury area is formed,

with the destruction of all cardiac cell types mirroring the necrosis seen after mammalian infarction, ensuring formation of an extensive fibrotic scar (Juan Manuel González-Rosa, 2011).

Current dogma suggests that existing CMs are the main source of new muscle after injury, with the hypothesis that these cells dedifferentiate, divide and mature, in a way that recapitulates the developmental programme (Kikuchi and Poss, 2012). Dedifferentiation is the process by which a specialised cell returns to an earlier developmental stage. In this case, CM dedifferentiation is characterised by reduced sarcomeric myofibril organisation and re-expression of embryonic marker genes, including foetal myosin genes and transcription factors such as *gata4, hand2, nkx2.5, tbx5*, and *tbx20* (Kikuchi, 2014). Kikuchi et al. used a reporter line to show that the majority of the regenerating CMs activate the regulatory sequences of *gata4* and maintains this expression throughout regeneration (Kikuchi et al., 2010) .

This hypothesis has been further supported by the work of Jopling et al. (Jopling et al., 2010) who aimed to directly identify the source of newly formed CMs during regeneration using lineage tracing in the adult zebrafish. The authors used a tamoxifen-inducible Cre-Lox approach to genetically label mature CMs using the CM-specific *myl7* promoter. In embryos at 48hpf, myl7-expressing cells were labelled via tamoxifen treatment and then these GFP-labelled CMs were analysed months later during adult heart regeneration. By observing whether GFP expression was present in the regenerated tissue, Jopling et al. were able to understand whether the new CMs were derived from existing, previously labelled, CMs or a different source, which would be GFP-negative. At 7 days post-amputation (dpa), the missing tissue had been replaced by a collagen/fibrin clot and GFP was only detected in the remaining cardiac tissue. However, by 14 dpa, uniform expression

of GFP was observed in the regenerating CMs (Jopling et al., 2010). The authors therefore concluded that new CMs arise from pre-existing, previously GFP-labelled, CMs (Jopling et al., 2010). BrdU treatment confirmed that these GFP-positive cells had re-entered the cell cycle (Jopling et al., 2010). Regenerating hearts were also analysed using transmission electron microscopy in order to observe any cytoskeletal changes in the cells. In the uninjured hearts, CMs displayed an organised sarcomeric myofibrillar structure, however, after amputation, CMs displayed disordered actin and myosin. Jopling et al. concluded that pre-existing CMs undergo dedifferentiation in order to facilitate cell cycle re-entry and proliferation (Jopling et al., 2010). The neonatal mouse has been shown to possess the capacity for heart regeneration, but only for a 7-day window of time after birth (Porrello et al., 2011). Disassembly of sarcomeric structures has been noted in proliferating CMs of the embryonic mouse heart (Ahuja et al., 2004). Cases have also been reported where neonatal human patients have fully functionally recovered from MI (Haubner Bernhard et al., 2016). These studies in mammals (and observations in humans) suggest that mechanisms used for regenerating zebrafish hearts might also be conserved in mammalian hearts. Fundamental insights gained from the study of cardiac regeneration in zebrafish and neonatal mice might therefore be translational to cardiac repair in humans in the future. Therefore, it is a logical strategy to use genes important during heart development as a starting point for investigations into their role during cardiac regeneration.

A number of signalling pathways have also been shown to be important for cardiac regeneration in zebrafish. A correct balance in Notch signalling is required for CM proliferation and it is believed this is exerted through paracrine signalling from the endocardium and/or the epicardium (Zhao et al., 2014, Raya et al., 2003). BMP

signalling is required for CM dedifferentiation and proliferation (Wu et al., 2016). FGF and VEGF signalling both promote revascularisation after injury (Lepilina et al., 2006, Marín-Juez et al., 2016). Retinoic acid, IGF, Hh, Jak/Stat and TGFβ signalling are all required for CM proliferation (Kikuchi et al., 2011, Lepilina et al., 2006, Choi et al., 2013, Fang et al., 2013). Because of the degree of conservation between the mechanisms of heart development in zebrafish and mammals, deciphering the cellular and molecular mechanisms regulating cardiac regeneration in the zebrafish will inform the development of therapeutic strategies to treat human heart disease.

The most pressing question in the field is: if zebrafish have the capacity to regenerate their heart muscle, then why can't we? Many groups have proposed hypotheses to answer this question. A leading hypothesis is that the difference in regenerative capacity is largely due to the DNA content of their CMs. As they mature, mammalian CMs become polypoid, meaning that they contain greater than two complete sets of homologous chromosomes (either within a single nucleus or separated into multiple nuclei) (Davoli and Lange, 2011). Polyploidisation correlates with terminal differentiation in mammals and occurs when the diploid (containing two sets of homologous chromososmes) immature cells replicate their DNA without mitosis and/or cytokinesis (Orr-Weaver, 2015, Brodsky et al., 1994). By contrast, in organisms that are able to regenerate their hearts, e.g. zebrafish, salamanders and neonatal mice, the majority of the CMs are diploid (Oberpriller et al., 1988, Wills et al., 2008). In neonatal mice, the regenerative window coincides with the time in which the majority of the CMs are diploid (Porrello et al., 2011). Until recently, this was just a correlative observation, but a study by González-Rosa et al (González-Rosa et al., 2018) has now demonstrated compelling evidence that polyploidisation may contribute to the inability of mammals to regenerate their hearts. The authors

transiently inhibited cytokinesis in zebrafish to induce polyploidisation of the CMs to adopt the ploidy of mammalian adult CMs and demonstrated that this was sufficient to reduce the regenerative capacity of these fish (González-Rosa et al., 2018).

In most organisms, cardiac regenerative potential has been shown to correlate with a number of other factors and a positive associatin is correlated with: a low pressure circulatory system (Ostadal, 2012), an inability to regulate body temperature (Ostadal, 2012) low metabolic state (Lopaschuk and Jaswal, 2010), hypoxia (Puente et al., 2014), immature myofilament composition (Gomes et al., 2016) and an immature innate immune system (Buchmann, 2014, Lavine et al., 2014). To date, many of these are simply correlative observations that require further study.

1.4 The cardiac GATA TFs

The GATA family of transcription factors have been shown to be important for both development and regeneration. This family comprises six evolutionarily conserved zinc finger-containing proteins which recognise and bind to the consensus DNA sequence (A/T)GATA(A/G) (Gove et al., 1997). This sequence exists in the promoters of many cardiac muscle-specific genes. GATA factors exist throughout eukaryotes, where they play vital roles in development, specifically in cell-fate specification, differentiation regulation, cell proliferation control and cell movement (Gove et al., 1997). The GATA family is divided into two subfamilies defined by expression pattern and sequence homology. Gata-1, -2 and -3 are involved in haematopoiesis and ectodermal patterning, and Gata-4, -5 and -6 are expressed in cardiac tissue and endodermal derivatives (Peterkin et al., 2005).

Early research has shown that *Gata4* is critical for CM differentiation of the embryonal carcinoma cell line P19 (Grépin et al., 1995) (Grepin et al., 1997) . Knockdown of *Gata4* in embryonic stem cells using an antisense approach blocks cardiogenesis at an early stage (Grépin et al., 1995), whereas overexpression enhances cardiogenesis (Grepin et al., 1997). In line with this, injection of *gata-4, -5* and *-6* mRNA into fertilised *Xenopus* eggs led to expression of alpha-myosin heavy chain (αMHC) and cardiac actin (Jiang and Evans, 1996). Despite this early evidence, *Gata4*-null mice do form differentiated myocardium but show ventral midline fusion defects, resulting in cardia bifida, suggesting that *Gata4* is important for later events in heart development (Kuo CT1, 1997). A suggested reason for the minimal disruption seen in the *Gata*4-null heart is that increased levels of *Gata6* may be compensating for the loss of *Gata4* (Kuo CT1, 1997). *GATA4* has also been implicated in human congenital heart defects, with mutations being associated with atrial-septal defect. *gata4* knockdown in zebrafish embryos has no effect on induction or maturation of the myocardium. However, *gata5* and *gata6* morphant defects were made worse by combinatorial knockdown of *gata4*, supporting their functional redundancy (Peterkin et al., 2007a).

Gata5-null mice show little-to-no cardiac phenotype, whereas cardia bifida and loss of CMs and endocardial cells are seen in zebrafish depleted for *gata5* (*faust* mutants) (Reiter, 1999). The early regulator *nkx2.5* was also downregulated in these mutants and the expression of many terminal differentiation markers (e.g.*myl7)* was lost, meaning both early and late cardiac formation signalling events were disrupted (Reiter, 1999). Consistent with this observation, *gata5* overexpression induces the ectopic expression of myocardial genes including *nkx2.5* and has been shown to produce ectopic foci of beating myocardial tissue

(Reiter, 1999). *Gata5* has also been shown to be critical for endocardial differentiation of a mouse cell line derived from cardiac mesoderm (Nemer and Nemer, 2002). In those cells that had been induced towards an endocardial fate by retinoic acid, inhibition of *Gata5* blocked terminal differentiation before the endocardium was specified.

Gata6-null mice die before heart induction due to an extra-embryonic defect at E5.5–E7.5(Morrisey et al., 1998), meaning little is known about its role in cardiogenesis. *Gata6*$^{-/-}$ embryonic stem cell transplants are, however, able to contribute to heart tissue in chimeric mouse embryos. A likely explanation is that the result of loss of *Gata6* is non-cell-autonomous and can therefore be rescued by signals from surrounding wild-type cells (Peterkin et al., 2007b). It is believed that this signal is BMP; there is a striking similarity between the phenotypes of *Bmp4* inhibition and *gata6* depletion (Peterkin et al., 2003).

gata6-MO knockdown in *Xenopus* resulted in heartless embryos (Peterkin et al., 2003). In zebrafish, GATA6 MO knockdown causes variable cardiac phenotypes including cardia bifida and partially fused heart tube. Early cardiac markers, such as *nkx2.5*, are left initially intact but depleting with time. This suggests that *gata6* is required for maintenance of *nkx2.5,* and although not necessary for specification of myocardium, it is critical for initiation of subsequent CM differentiation programmes such as contractile machinery gene expression (Peterkin et al., 2003).

In *Xenopus*, *gata6* expression in heart precursors decreases with the onset of expression of cardiac machinery genes (Gove et al., 1997). Overexpression of *gata6* with exogenous mRNA beyond this time window delayed terminal CM differentiation (Gove et al., 1997). This overexpression was, however, temporary

and, after exogenous mRNA had degraded, the myocardium was thicker than in the control, suggesting that *GATA6* holds cardiac precursors in a proliferating progenitor state until downregulation leads to their maturation (Gove et al., 1997).

As well as being important for cardiac precursor specification, *gata4, gata5* and *gata6* have also been shown to have a crucial role in specification of haemangioblasts (Peterkin et al., 2009). Haemangioblasts are a transient population of blood/endothelial precursors that reside in the ALPM between the 5- and 10-somite stages and co-express genes that are later expressed in either blood or endothelium. They eventually give rise to myeloid cells, head endothelium and endocardium. The cardiac precursors reside more posteriorly to the haemangioblast population in the ALPM (Peterkin et al., 2009). Studies have shown that loss of the haemangioblast programmes causes the expansion of cardiac progenitors (Schoenebeck et al., 2007) and ectopic expression of the cardiac regulator *nkx2.*5 inhibited the haemangioblast programme (Simões et al., 2011), suggesting an antagonistic relationship between the two programmes. The co-dependence of anterior haemangioblasts and cardiac precursors on *gata4, gata5* and *gata6* places these genes at the apex of genetic regulatory hierarchy specifying these two ALPM derivatives.

In the adult heart, stimuli such as injury, ventricular wall stress, change in workload or neural-hormonal stimuli, can lead to an increase in ventricular and/or septal wall thickness known as cardiac hypertrophy (van Berlo et al., 2010). Hypertrophy is a temporary compensatory response developed to normalise heart function in times of peak stress; prolonged cardiac hypertrophy is a risk factor for heart failure (van Berlo et al., 2010). With pathologic stimuli, ventricular myocytes increase expression of skeletal α-actin, β-myosin heavy chain (βMHC), b-type natriuretic

peptide (BNP), and atrial natriuretic factor (ANF) as well as activating specific intracellular signalling cascades such as mitogen-activated protein kinase (MAPK), protein kinase C (PKC), and insulin-like growth factor-1 (IGF1) (van Berlo et al., 2010). These in turn further regulate intracellular transcriptional regulatory proteins modulating gene expression (e.g.MEF2 and NFAT) to enable hypertrophic growth of the heart (characterised by increased cell size without cell division, increased cell RNA and protein content and increased production and assembly of contractile proteins into sarcomeric units) (Glennon et al., 1995, van Berlo et al., 2010). Many regulatory sequences for genes required for hypertrophic growth contain GATA binding sites.

Liang et al. (Liang et al., 2001) showed that overexpression of *Gata4* or *Gata*6 in cultured mouse CMs was sufficient to induce a hypertrophic response. Van Berlo et al. (van Berlo et al., 2010) investigated the role of GATA6 in pathological cardiac hypertrophy in mouse models, showing that both GATA4 and GATA6 protein levels increase in pressure overload-induced hypertrophic hearts. Because *Gata6*-null mutants are not viable, *Gata6fl/fl* mice were crossed with βMHC-cre mice to produce a later deletion, bypassing embryonic lethality (van Berlo et al., 2010). 8-week-old Gata6fl/fl MHC-cre mice were exposed to 2 weeks of transverse aortic constriction stimulation, mimicking pathologic hypertrophy-inducing stimuli. The hypertrophic response was significantly reduced in Gata6fl/fl βMHC-cre mice compared with the control group. Gata6fl/fl βMHC-cre mice also showed greater sensitivity to heart failure, measured by reduced ventricular fractional shortening and increased mRNA for foetal genes associated with stress and heart failure (ANF, αMHC, and skeletal α-actin). Conversely, overexpression of *Gata6* was shown to enhance cardiac hypertrophy (van Berlo et al., 2010). At present, the phenomenon of hypertrophy

has not been studied extensively in zebrafish. One documented case is the research of Sun et al. (Sun et al., 2009), who observe cardiac hypertrophy alongside cardiac hyperplasia (increased number of CMs) in an anaemic zebrafish line (*tr265/tr265* mutant). In a recent study, González-Rosa et al. also noted that expression of a dominant-negative form of the protein Ect2 in zebrafish CMs induced polyploidisation, led to impaired cytokinesis and caused CMs to grow by hypertrophy (González-Rosa et al., 2018).

Although *gata5* and *gata6* have not been investigated in the context of heart repair, a significant body of work exists on the importance of *gata4.* Kikuchi et al. (Kikuchi et al., 2010) tracked expression of Gata4, using a transgenic *gata4:EGFP* reporter line and concluded that a subpopulation of CMs within the cortical layer switches on *gata4*, proliferates and contributes the majority, if not all of the regenerating muscle. Upon *gata4* inhibition using a CM-specific, inducible dominant-negative *gata4* construct (Kikuchi et al., 2010), regeneration after resection injury was impaired and persistent scarring was observed.

As a result of the plethora of clinical presentations, in addition to the abundance of studies in model organisms, we have learnt a huge amount about the factors and processes downstream of the cardiac GATA factors. However, we know relatively little about the upstream regulatory activity. A powerful approach to understanding the factors involved in the upstream regulation of a gene is to study the gene's cis-regulatory environment.

1.5 The cis-regulatory landscape

1.5.1. Transcriptional regulation

The robust and precise spatiotemporal expression of genes is paramount for the accurate accomplishment of biological processes such as development. Using a relatively limited set of protein-coding genes (~21,000 in the human genome (Pertea et al., 2018)), extremely complex processes are orchestrated, and this spatiotemporal specificity is achieved with the help of transcriptional regulatory elements (Heinz et al., 2015). The ability to identify and study such elements presents the possibility for comprehensive understanding of development in both normal and pathological conditions.

1.5.2. The eukaryotic transcriptional process

In eukaryotes, protein-coding genes are transcribed by RNA polymerase II and this process is divided into three phases: initiation, when the initiation complex is recruited to the promoter to begin RNA synthesis; elongation, when the RNA strand is being synthesised; and termination, when RNA polymerase II is released along with the transcribed mRNA (Nechaev and Adelman, 2011). The majority of protein-coding gene expression is thought to be controlled at the level of transcriptional initiation, although some evidence suggests that elongation might also be a limiting step.

Three groups of factors are involved in transcription via RNA polymerase II: general transcription factors, activators and coactivators. General transcription factors (TFs) include TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, and they bind sequentially to the core promoter to form a transcription preinitiation complex (PIC), bringing RNA polymerase II to the transcriptional start site (TSS). TFIID binding is the first step in PIC assembly (Maston et al., 2006). It is a protein complex that comprises a TATA-

box-binding protein and a set of associated factors. TFIIB is then recruited, followed by TFIIF-polymerase II complex and the binding of TFIIH. The Mediator is a conserved multisubunit protein complex that acts as a functional bridge between TFs, RNA polymerase II and the general transcription factors, promoting the assembly of the PIC on core promoters (Soutourina, 2017).

This PIC assembly on the core promoter is only sufficient to orchestrate low levels of basal transcription (Ptashne and Gann, 1997). This activity can be enhanced by activators; sequence-specific DNA-binding proteins whose recognition sequences typically lie upstream of the core promoter (Weake and Workman, 2010). Activators typically contain an activation domain, in addition to their DNA-binding domain, that is necessary to stimulate transcription (Ptashne and Gann, 1997). As well as their role in increasing PIC formation, activators can also function by promoting later steps in transcription such as elongation or reinitiation (Lee and Young, 2000). They can also recruit chromatin remodellers to provide a suitable chromatin landscape for PIC assembly (Lemon and Tjian, 2000).

Co-activators can be recruited to chromatin by a range of TFs and tend to be associated with the most active enhancers. They are required for the function of DNA-binding activators, but do not specifically bind by themselves (Malik and Roeder, 2010). Co-activator recruitment is more frequent at a given enhancer when more TFs are co-bound (Heinz et al., 2015). They do not directly bind to DNA but are instead recruited by DNA-bound activators. They are large proteins that comprise multiple interaction sites for TFs (Heinz et al., 2015). Co-activators typically function to integrate intracellular signals ((Heinz et al., 2015)) and modify and remodel the chromatin state of enhancers and include histone acetyltransferases, such as p300, histone methyltransferases, such as MLL3/4,

chromatin remodelling factors, such as Brg1, and factors that regulate crosstalk with transcription machinery, such as the Mediator complex (Malik and Roeder, 2010, Krasnov et al., 2016, Taatjes et al., 2004).

1.5.3. Transcriptional regulatory elements

To achieve maximal rates of transcription by RNA polymerase II, two main classes of *cis*-regulatory elements are required: proximal elements, which include the core and proximal promoter, and distal elements which include silencers, insulators and enhancers.

- Core promoter

The core promoter is the minimum sequence required for accurate transcriptional initiation by acting as a docking site for PIC assembly (Maston et al., 2006). They lie in the vicinity of the TSS and contain multiple core promoter elements, including the TATA box, the binding site for the TATA box-binding protein subunit of the general transcription factor TFIID. Although the TATA box was the first core promoter element to be described, metazoan core promoters can also comprise many other elements, including: Initiator element (Inr), Downstream Promoter Element (DPE), TFIIB-Recognition Element (BRE), and Motif Ten Element (MTE) (Juven-Gershon et al., 2008).

- Proximal promoter

The proximal promoter is the region immediately upstream of the core promoter (within 100-200bp of the transcription initiation site). They contain multiple binding sites for activators and may also contain tethering elements for recruiting distal enhancers (Maston et al., 2006, Calhoun et al., 2002).

- Silencers

Silencers are cis-regulatory elements that function to silence or repress the transcription of a target gene (Ogbourne and Antalis, 1998). They are binding sites for negative TFs, or transcriptional repressors (Privalsky, 2004). Repressors can also function via the recruitment of negative cofactors called corepressors. Although they can be positioned close to, or within the intron of their target gene, silencers can also be situated far from their cognate promoter and can act independent of orientation. Repressors can act by competing with activators for a common binding site, by protein-protein interaction to block the function of a nearby activator, by inhibiting the formation of the PIC, or by inhibiting access of activators or general transcription factors to the promoter via histone-modifier recruitment or chromatin stabilisation to create an unfavourable chromatin environment (Maston et al., 2006).

- Insulators

Insulators play a key role in establishing the active chromatin landscape and partitioning the genome into discrete functional regions. They are typically 0.5-3kb in length and act as boundary elements and prevent the interference of transcriptional activity from separate neighbouring genes. They achieve this by blocking unspecific enhancer-promoter interactions and also by forming a barrier to the spread of heterochromatin. Unlike enhancers and silencers, insulators function in a position-dependent manner. However, these features all share the feature of being orientation-independent. In Drosophila, many trans-activating factors have been shown to regulate insulator activity (West et al., 2002). By contrast, only one protein has been discovered that mediates this process in vertebrates: CCCTC-binding factor (CTCF) (Heger and Wiehe, 2014). CTCF activity can be modulated by interaction with cofactors, DNA methylation or by post-translational modification.

CTCF binding sites are particularly enriched in conserved non-coding regions of DNA (Xie et al., 2007), in particular in the boundaries of topological associated domains (Dixon et al., 2012). It is present at 55,000-65,000 sites in mammalian genomes. Although CTCF has traditionally been referred to as an insulator because of its ability to create boundaries between topologically associated domains, recent studies have also highlighted a role for CTCF in facilitating interactions between regulatory sequences and so the term 'architectural' protein may be more appropriate (Ong and Corces, 2014).

- Enhancers

Enhancers are the major regulators of developmental gene expression. They are cis-regulatory elements that can increase or activate the transcription of genes. Enhancers are short regions of DNA, typically 100bp-1kb in length, which function as a platform for the binding of many TFs through short, specific DNA sequence motifs (Spitz and Furlong, 2012). They integrate the regulatory cues from these TFs to direct precise spatial and temporal patterns of gene expression. General transcription factors and RNA polymerase II are recruited to enhancers, forming a hub for the assembly of the PIC. They can be found in intergenic regions, in introns or exons of the genes they regulate, or even in neighbouring genes. Their function is independent of orientation and distance from their cognate promoter (Long et al., 2016). They can act at great distances from their target gene, for example, the limb enhancer of *Shh* resides 1Mb away within the intron of another gene (Lettice, 2003). This presents a challenge when trying to identify enhancers and assign them to target promoters. It also allows a gene to be regulated by multiple enhancers at different times, enabling a combinatorial complexity of spatiotemporal gene regulation (Long et al., 2016). It is thought that cooperativity of multiple TFs in close

proximity can act to overcome the energetic barrier of nucleosome eviction, aiding enhancer activation (Spitz and Furlong, 2012). Cooperativity can be direct, whereby TFs physically associate, or indirect, whereby TFs collectively compete with the same nucleosome to expose the underlying DNA in the absence of direct protein-protein interaction (Miller and Widom, 2003).

TFs bind to a series of consensus motifs. The term ‘enhancer grammar’ refers to the number, order, type, orientation, spacing and binding affinity of these motifs and this is still very much an area of active research (Long et al., 2016). There are two main models that have been put forward to describe the organisation of enhancers.

The ‘enhanceosome’ model requires rigid motif organisation and spacing and suggests that alterations in individual binding sites can have drastic effects on enhancer output (Merika and Thanos, 2001, Thanos and Maniatis, 1995). A classic example of this model is the mammalian virus-inducible enhancer of interferon-β (IFNβ) (Thanos and Maniatis, 1995). IFNβ is normally silent but is strongly induced specifically upon viral infection due to an enhancer located upstream of the TSS. This enhancer comprises three regulatory domains, none of which function in isolation. If two or more copies of any of these regulatory domains are present, then they can act as a virus-inducible enhancer; however these modified enhancers display altered transcription levels and can respond to cues other than viral infection (Thanos and Maniatis, 1995, Merika and Thanos, 2001). The ‘billboard’ model maintains the requirement for TF binding sites, but the order and spacing of these binding sites is flexible within the enhancer i.e. the enhancer comprises multiple functional units that can independently regulate the expression of genes (Kulkarni and Arnosti, 2003). These enhancers may contain suboptimal binding sites that facilitate rapid motif turnover across species and may assist enhancer specificity.

An example of this can be demonstrated by the Otx-a enhancer in Ciona, which contains GATA and ETS motifs that differ from the consensus motifs (Farley et al., 2015). When the motifs are modified to match the consensus motifs, robust but ectopic enhancer activity patterns result as the enhancer appears to be responsive to lower levels of GATA or ETS. Therefore, a combination of multiple suboptimal enhancers may result in specificity of regulation while maintaining signal strength (Farley et al., 2015).

The billboard model, whereby TF binding events drive the activity of an enhancer irrespective of the order and orientation of the motifs, has generally been more broadly supported with a host of examples. However, some recent studies have suggested that the need for more rigid enhancer grammar is more prevalent than originally thought. Recent studies have challenged the previously held view that cooperative binding is represented by composite motifs (motifs comprising complete consensus motifs for both TFs). In particular, a study by Jolma et al (Jolma et al., 2015), authors demonstrated that the majority of TF-TF interactions have a novel consensus motif that is weakly bound by each TF in isolation, meaning that these motifs have been overlooked in previous analyses and the importance of defined motif spacing has been underestimated (Jolma et al., 2015). It is therefore likely that the architecture of most enhancers falls somewhere in between the rigid motif structure of the enhanceosome model and the flexible structure of the billboard model (Long et al., 2016). One interesting hypothesis is that enhanceosomes are necessary for more ‘switch-like’ activity for the control of genes that require the integration of strict regulatory inputs. In accord, many genes encoding TFs important for development are surrounded by clusters of highly conserved enhancers. Because of their enhanceosome characteristics (rigid structural

requirements) and developmental importance, mutations in these enhancers are often associated with disease in humans, e.g. preaxial polydactyly results from mutations in the highly conserved *Shh* enhancer (Lettice, 2003).

Enhancers can be brought into close proximity with their associated promoters via a looping mechanism, whereby the intervening DNA loops out. This interaction between enhancer and promoter is thought to be facilitated by cohesin, which encircles both strands of DNA to maintain proximity between the two (Kagey et al., 2010). The Mediator complex has also been shown to interact with cohesin in order to support enhancer-promoter contacts (Phillips-Cremins et al., 2013). Many enhancers act in a modular fashion, where each regulatory element regulates a subset of the global expression pattern of a gene. This has been elegantly demonstrated in studies of the α-globin locus in primary mouse erythroid cells, which is regulated by a cluster of enhancer elements (known as a 'super-enhancer'), with each element acting independently in an additive fashion to contribute to the total expression pattern of the gene (Hay et al., 2016). However, in some cases, the overall expression pattern of a gene differs from the sum of the individual activities of each of its enhancers. For example, several studies in both mice and Drosophila have shown that the spatiotemporal expression pattern of enhancer elements tested individually, differed from the expression patterns observed when multiple enhancers were tested together (Dunipace et al., 2011, Dib et al., 2011, Prazak et al., 2010). In most cases, the effect of multiple enhancers collaborating is to prevent ectopic expression, perhaps by recruitment of repressors which act to suppress the activity of enhancers in ectopic regions (Perry et al., 2011, Long et al., 2016).

1.5.4. The chromatin landscape

Enhancers are classically able to activate transcription from large distances, which raises the question of how enhancers establish a selective interaction with their cognate promoters. Chromatin conformation capture (3C)-based assays have helped to tackle this question (Dixon et al., 2012). These techniques are based on the digestion and re-ligation of chromatin in order to understand the 3D structure and proximity of sequences within the nucleus. Analysis in multiple species has revealed a complex higher order 3D organisation of entire chromosomes and that the genome is segmented into self-interacting portions, referred to as 'topologically associated domains' (TADs). This compartmentalisation of the genome creates local regulatory environments where enhancers can only communicate with promoters within their designated TAD, explaining the specificity of enhancer-promoter interactions. It is not clear how the 3D organisation of the genome is established, but it is known that cohesion and the Mediator complex are involved in the formation of these structures (Heinz et al., 2015). Interestingly, TAD boundaries are conserved across eukaryotes, and are typically largely invariant between cell types (Harmston et al., 2017). TADs have been shown to represent relatively robust structural features of chromatin and any structural rearrangements between eukaryotic genomes occur in between these boundaries (Vietri Rudan et al., 2015). The boundaries of TADs are enriched for the insulator protein CTCF, suggesting that they may play a role in establishing these structures (Dixon et al., 2012).

1.6 Enhancer identification

Due to the fact that enhancers can be located at great distances from their target promoters, historically they have been very difficult to identify. Traditionally,

regulatory regions were identified by 'promoter bashing', whereby serial deletions were made, and the function of the fragments tested in reporter assays. The rapid advances in sequencing technologies in recent years have meant a dramatic increase in the volume of available genomic data, greatly improving the potential for identifying enhancers on a genome-wide scale. Prediction of active enhancers can be approached based on the following known characteristics of enhancers: a) TF binding b) comparative genomic features c) epigenetic features.

1.6.1. TF binding

Because enhancers serve as a platform for the combinatorial binding of TFs, one strategy for enhancer prediction is to map known TF binding sites and scan for motif clustering. TF-enhancer interactions can then be tested by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) (Yáñez-Cuna et al., 2013). Multiple databases have been generated based on experimental data characterising TF binding sites and by using these data, one can generate a statistical match between the sequence and the site matrix in order to predict binding. This, however, generates a number of false positive sites for TF binding due to chance. It is also likely that some binding sites may not have been fully characterised, meaning some TF binding sites could be missed. Therefore, while TF binding site monitoring approaches may provide an indication of cis-regulatory element location, they must be used in conjunction with another of the methods described below.

1.6.2 Comparative genomic features

Now that high throughput sequencing is commonplace, the genomes of many species have been sequenced in full. Using a technique known as phylogenetic footprinting (Zhang and Gerstein, 2003), these genomic sequences can be

compared against each other. Sequences that are found to be conserved across large evolutionary distances, known as conserved non-coding elements (CNEs) are then treated as candidates for being functionally important, with the assumption that important TF binding sites will be conserved throughout evolution (Woolfe et al., 2005). Interestingly, CNEs have been shown to cluster around developmental regulatory genes which are part of conserved gene-regulatory networks across vertebrates, consistent with strong selective pressure (Woolfe et al., 2005). There are many studies demonstrating the significance of conserved enhancers, for example the discovery of a *Sox9* mandibular enhancer that when mutated, causes Pierre Robin sequence (cleft palate) (Benko et al., 2009). These comparative studies assume that a high level of sequence conservation represents functionality, however there are also studies where functional conservation is observed without sequence similarity (Fisher et al., 2006). In addition, there are numerous examples of highly conserved genomic sequences for which no enhancer function could be demonstrated in transgenic assays.

Other caveats that must be considered with this method are that not all TF binding sites are conserved across species; degeneracy of binding sites means that perfect sequence conservation is not necessarily required for TF binding. Also, because of the redundancy of many enhancers, specific regulatory elements may not be conserved across species, despite the conservation of gene-expression patterns. The level of conservation can also depend on the tissue type. Of note, relatively few heart enhancers have been identified using this approach: in some of the largest datasets mining sequence conservation for embryonic enhancers, less than 2% of the tested sequences were found to be heart enhancers, compared with 16% for forebrain, 14% for midbrain and 5% for limb enhancers (Visel et al., 2008,

Pennacchio et al., 2006). A study by Blow et al. concluded that heart enhancers are three times less well conserved than other tissue-specific enhancers in vertebrates (Blow et al., 2010). It is also known that enhancers evolve much more rapidly than promoter regions, meaning that relatively few are deeply conserved. This feature was demonstrated elegantly in a study by Villar et al. in which the speed of enhancer and promoter evolution was calculated using liver tissue from 20 different mammals (Villar et al., 2015). Active enhancers were found to have a mean lifetime three times shorter than active promoters. This strongly suggests that the strategy of comparative sequence conservation for the prediction of enhancers should be used in conjunction with another method outlined below.

1.6.3. Epigenetic features

-Accessible DNA

Active enhancers can be characterised by regions of accessible DNA, devoid of nucleosomes. This allows for the binding of TFs and chromatin remodelling complexes, which can act as a good indicator for the presence of a regulatory element (Buenrostro et al., 2013).

Open chromatin structure leaves the regions vulnerable to enzymatic cleavage by DNAse I or micrococcal nuclease (MNase). When coupled to deep sequencing (DNAse-seq (Song and Crawford, 2010) and MNase-seq (Zaret, 2005)), open chromatin regions can be mapped at the genome-wide level, exposing multiple potential regulatory regions. Nucleosome-free regions can also be identified using Formaldehyde-Assisted Isolation of Regulatory Elements followed by deep sequencing (FAIRE-Seq), which depletes closed DNA after chemical crosslinking with formaldehyde (Giresi et al., 2007).

Although reliable and informative, these bulk sequencing approaches bypass cellular heterogeneity. These methods often require input material of tens to hundreds of millions of cells, making them technically challenging for developmental biology, especially when studying rare populations.

ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing) is a method that utilises a hyperactive transposase, Tn5, that can cut regions of increased accessibility and ligate adapters for high-throughput sequencing (Buenrostro et al., 2015) (summarised in Figure 1-2). Genome-wide mapping of insertion ends can then be used to assay the regulatory landscape of chromatin, identifying regions of increased accessibility. Standard ATAC-seq was already revolutionary in that it could be performed on samples of only hundreds of cells, but recent technological advances have meant that chromatin can now be profiled by ATAC-seq even at the single-cell level (Chen et al., 2018). Such technological breakthrough allows the study of regulatory elements in novel and rare cell populations.

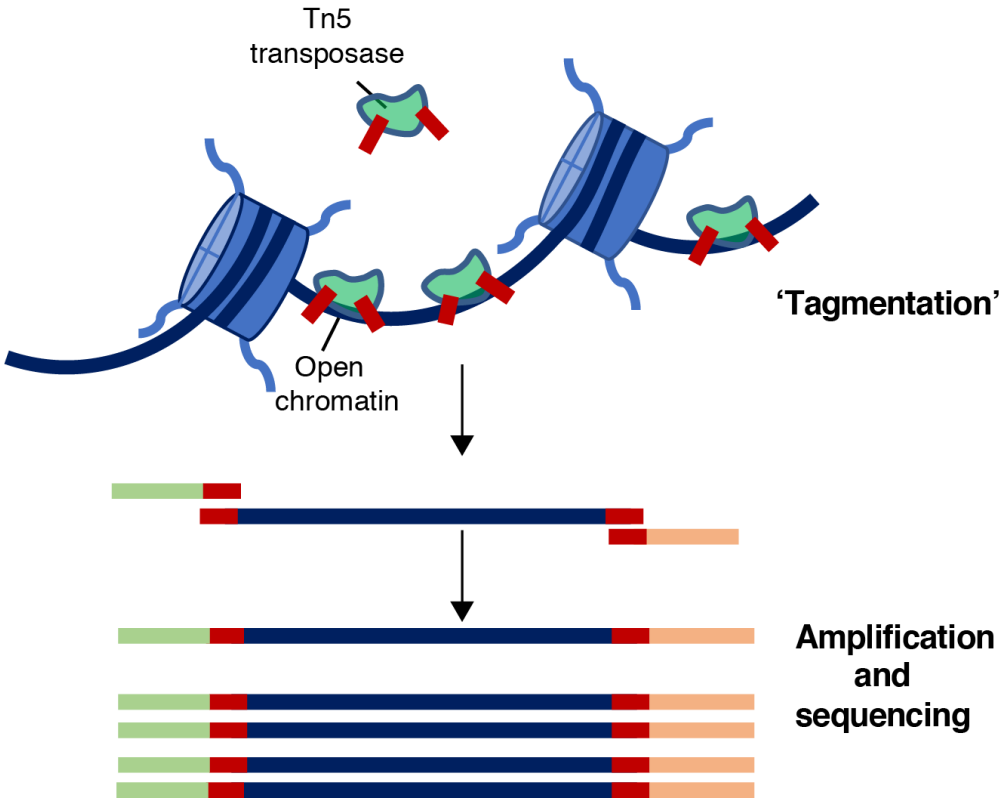


Figure 1-2 ATAC-Seq schematic

The transposase Tn5, loaded with sequencing adaptors, simultaneously cuts and ligates in regions of open chromatin and generates sequencing library fragments that can be PCR-amplified.

-Histone marks

One mechanism of epigenetic regulation of gene expression is mediated by post-translational modifications of histone proteins, leading to modification of chromatin structure. These modifications include acetylation, methylation and phosphorylation of the core histones, H2A, H2B, H3 and H4 (summarised in Figure 1-3) (Shlyueva et al., 2014b).

Nucleosomes in the vicinity of active enhancers typically show the presence of certain histone modifications, including histone H3 lysine 4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac) at their amino termini. Whether the

presence of these modifications is a cause or a consequence of enhancer function or TF occupancy is yet to be revealed (Spitz and Furlong, 2012). Knowledge of typical histone modifications can be exploited in the genome-wide search for active regulatory elements using ChIP-seq. H3K27ac is particularly highly correlated with enhancer activity and expression of the nearest gene (Creyghton et al., 2010, Rada-Iglesias et al., 2011). Transient positioning of H3K79me3 and RNA polymerase II at enhancers has also been shown to correlate with active enhancers and to give an accurate read-out of the timing of enhancer activity during development (Bonn et al., 2012, Santos-Rosa et al., 2002). By contrast, H3K9me3 correlates with heterochromatic regions that are transcriptionally silent, and silent or repressed enhancers are typically occupied by nucleosomes containing the H3K27me3 mark, linked to the Polycomb repressive complex (Ringrose and Paro, 2004). Genome-wide analysis of histone modifications has also revealed the occurrence of bivalent chromatin at poised enhancers, carrying histone modifications associated with both active (H3K4me1) and repressed (H3K27me3) states (Rada-Iglesias et al., 2011). Poised enhancers have the potential to become active and their putative targets typically function later in development (Creyghton et al., 2010).

The histone acetyltransferase p300 is a transcriptional coactivator shown to be associated with active enhancers. P300 ChIP-Seq has been used to predict enhancers in many species, including mice and humans and 52-82% of sites bound by p300 function as enhancers (Visel et al., 2009).

These methods are restricted by the requirement for large numbers of cells (on the order of hundreds of thousands of cells) and the lack of appropriate antibodies, particularly in zebrafish.

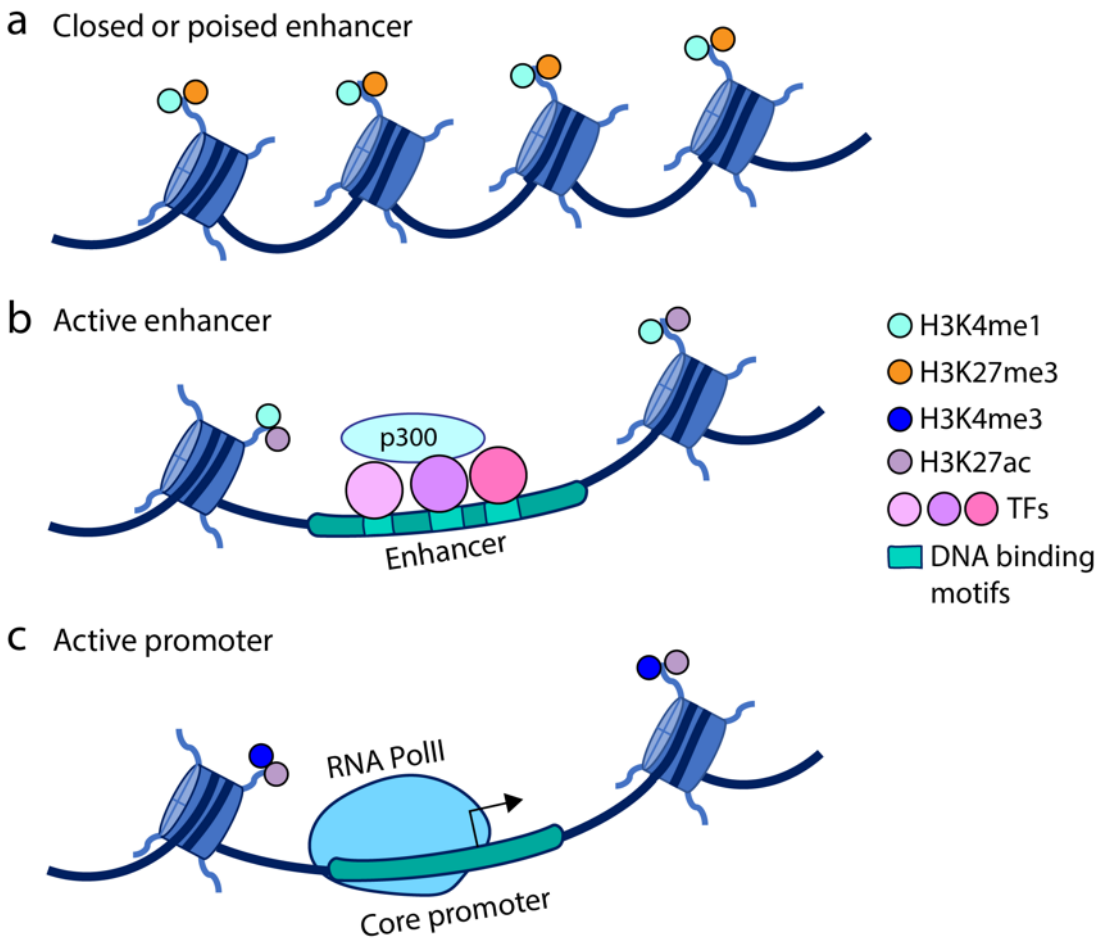


Figure 1-3 Histone marks at regulatory elements

Closed/poised enhancer regions have densely positioned nucleosomes which restrict access for TFs, RNA polymerase II and other factors. They can be marked by the active H3K4me1 and repressive Polycomb protein-associated H3K27me3 marks.

Histones flanking active enhancers often bear H3K27ac and H3K4me1 marks. The histone acetyltransferase p300 is also associated with active enhancers. Nucleosomes with H3K27ac and H3K4me3 flank active promoters.

(Figure adapted from (Shlyueva et al., 2014b))

1.6.3 eRNA

More recently, it was discovered that enhancers can also recruit RNA polymerase II and drive transcription of the surrounding DNA, producing enhancer RNAs (eRNAs) (Kim et al., 2010b, Arner et al., 2015). The resulting transcripts are not polyadenylated or spliced, and are quickly degraded by the exosome (Heinz et al., 2015). The recruitment of Pol II to enhancers and signalling-dependent eRNA

expression has been linked with changes in nearby gene expression, inferring a functional link between eRNAs and gene expression (Heinz et al., 2015). Enhancers that are transcribed tend to have higher binding ofco-activators, enrichment of active histone marks such as H3K27ac, and greater chromatin accessibility than non-transcribing enhancers (Li et al., 2016). The biological function of eRNAs is still a topic for debate. Some suggest that transcription initiation at enhancers could aid chromatin decompaction, assisting recruitment of TFs to the enhancer. Regardless of the function, the presence of eRNAs can be exploited for the identification of active enhancers. Techniques measuring nascent transcription, such as precision run-on sequencing (PRO-seq) and native elongating transcript sequencing (NET-seq) are able to detect and measure these non-polyadenylated, relatively unstable RNA molecules (Core et al., 2008).

1.6 Transgenesis to test putative enhancers *in vivo*

Although the above techniques are extremely informative for predicting putative enhancers, active enhancers must still be validated and characterised functionally *in vivo*. ENCODE and modENCODE are massive consortia using data from ChIP-based methods and DNAse-seq experiments to create databases of putative regulatory elements in *Drosophila* (Roy et al., 2010)*, C.elegans* (Gerstein et al., 2010)*,* mouse (Yue et al., 2014) and human (Dunham et al., 2012). When surveyed, it was found that only about 26% of these predicted enhancers showed activity *in vivo* (Kwasnieski et al., 2014), highlighting the importance of functional validation in these studies. Transgenesis is an effective method to assess putative regulatory element activity *in vivo*. Active enhancers can be tested in developing embryos by placing the putative enhancer DNA sequence upstream of a minimal promoter and

a reporter gene in a vector containing transposon recognition sequences. Zebrafish are particularly amenable to this technique as their transparency in early development allows fluorescent proteins driven by active enhancers to be directly visualised in developing embryos. Tol2-mediated genomic transposition has historically been the method of choice for generating transgenic reporter lines in zebrafish (Ishibashi et al., 2013). This method, however, is not well suited to transient assays as a high proportion of the F0 generation show highly mosaic expression, as well as non-specific background expression (Chong-Morrison et al., 2018). In recent years, members of the Sauka-Spengler lab have worked to re-purpose the AcDs system from maize in order to test and characterise putative enhancers in a transient manner. Chong-Morrison et al. have found vast improvements when compared with the traditional Tol2-based approach (Chong-Morrison et al., 2018). The Ac/Ds system was first identified in maize in the 1940s by Barbara McClintock (McClintock, 1950) and the components were fully characterised in the 1980s (N. Fedoroff, 1983). This study revealed that transposition can occur in the presence of an Ac (Activator) transposase and two Ds (Dissociation) elements. This approach was first adapted for use in zebrafish in order to facilitate the integration of a reporter construct by Emelyanov et al. (Emelyanov et al., 2006) in 2006, but the authors overlooked the potential of using this technique in transient, in the F0 generation. Experiments within the Sauka-Spengler lab have validated the use of the Ac/Ds system for both germline transmission, in the creation of transgenic lines, and for somatic integration-based experiments (Chong-Morrison et al., 2018).

1.7 Thesis Aims

Working hypothesis:

An in-depth analysis of the putative enhancers regulating the cardiac GATA genes will provide novel information on the precise temporal and spatial control of these genes in the various subpopulations present in the heart during early cardiogenesis.

Aims:

1) To predict putative enhancers around the cardiac GATA gene loci that are accessible during development and regeneration in zebrafish CMs.

2) To characterise the activity of putative enhancers for the cardiac GATA genes *in vivo* using a novel, highly efficient, reporter system.

3) To verify the putative enhancer activity detected in mosaic transient transgenesis, which show potentially interesting expression patterns by the establishment of stable zebrafish transgenic lines.

4) To use these stable transgenic lines to investigate the potential consequences of endogenous enhancer deletion on GATA expression and heart development.

5) To profile the subpopulations derived from a selected enhancer-driven reporter line using single-cell RNA sequencing in order to understand the TF signatures of each population, thereby informing which factors may be responsible for the upstream regulation of the chosen GATA factor in different subpopulations of the developing heart.

6) To analyse the spatial localisation of the subpopulations identified through single-cell transcriptomics using visualisation of subpopulation-specific marker genes.

7) To use the newly-generated knowledge of the transcriptome, as well as binding site prediction to identify potential TF-enhancer binding interactions necessary for the enhancer-driven activity of the chosen GATA factor in different subpopulations of cells in the heart.

8) To perturb the predicted binding sites required for the TF-enhancer binding interactions *in vivo*, in order to understand which are functional and at what time point during development, and therefore contributing to dissection of the precise regulatory activity of the GATA gene during cardiogenesis.

2) Materials and Methods

2.1 Zebrafish husbandry

Zebrafish were maintained in a circulating water system at 28.5°C, pH7.1-7.2 and a light/dark cycle of 14 hours of light and 10 hours of darkness. Fish were fed 3 times daily on a mixed diet of high protein pellets, flake food and enriched brine shrimp. All zebrafish work was approved by the Research Ethics Committee of the University of Oxford.

2.2 Zebrafish lines

The wild type (WT) zebrafish lines used in this study originated from Nigel Holder's lab at King's College London. The transgenic reporter lines *Tg(myl7:GFP)* (Burns et al., 2005) and *Tg(kdrl:GFP)* (Jin et al., 2005) have been previously described. Transgenic lines were maintained as heterozygous carriers which were then incrossed or outcrossed where necessary. I generated *Tg(Gata5-E5-mCherry), Tg(Gata5-E6-EGFP), Tg(Gata5-E8-mCherry), Tg(Gata6-E12-mCherry)* for the purpose of this study.

2.3 In Situ Hybridisation

2.3.1 In vitro transcription of probes for whole-mount in situ hybridisation

Probe	Restriction Site	RNA Polymerase	Reference	Source
gata4	EcoRI	T7	(Reiter, 1999)	D. Stainier
gata5	EcoRI	T7	(Reiter, 1999)	A. Rodaway
gata6	BamHI	T7	(Peterkin et al., 2007a)	A. Gibson

Table 1 Details of whole-mount in situ hybridisation probes used in this thesis

Restriction enzymes (detailed in Table 1) were used to generate linear templates for the synthesis of antisense RNA probes. 1µg linearised template plasmid was incubated for 3-12 hours at 37°C with 10x concentrated NTP Labeling Mixture (DIG), 10x concentrated Transcription Buffer, 10U RNase inhibitor (RNAseOut, Invitrogen) and 20U T7 RNA Polymerase (Roche). Residual template DNA was digested with DNaseI (Roche) for 15 min at 37°C. The reaction was stopped with 1µl 0.5M EDTA and the volume was made up to 100 µl with dH2O. Probes were precipitated with 1/3 volume of 10M NH_4Acetate and 2.5 volumes of 100% EtOH on ice for a minimum of 10 min. Probes were then pelleted by centrifuging for 10 min at 4°C, washed with 70% EtOH and air-dried. Probes were resuspended in RNAse-free H_20 and stored at -80°C until use.

2.3.2 Embryo fixation and preparation for in situ hybridisation

Embryos were collected following natural spawning and incubated at 28.5°C until the required time points. Embryos were euthanised in 5g/L ethyl-3-amino benzoate methanesulfonate (MS222) and fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) overnight at 4°C. Embryos were then washed three times in PBStw (1xPBS; 0.1% Tween 20) and then taken through the following wash steps

for 5 mins each: 25%, 50% and 75% methanol in PBSTw, then 100% methanol, before storage at -20°C for at least 24 hours.

2.3.3 Whole mount in situ hybridisation

Fixed embryos stored at -20°C in 100% methanol were rehydrated through a series of 5-minute washes of 75%, 50% and 25% methanol in PBSTw and then washed three times for 5 minutes in 100% PBSTw. 48hpf embryos were treated with 10µg/ml proteinase K (Sigma) for 12 minutes at room temperature. Proteinaise K activity was then stopped by washing 2x5 minutes in 0.1M glycine and embryos were re-fixed for 20 minutes at room temperature in 4% PFA and then washed 5x 5 minutes with PBSTw. Embryos were then transferred to 50% Hybe$^{+/+}$ (50% Formamide, 5x SSC, 9.2 mM Citric Acid, 0.1% Tween, 0.05% tRNA, 0.005% Heparin) for 5 minutes and then pre-hybridised in 100% Hybe $^{+/+}$ for at least 1 hour at 65°C. The 100% Hybe$^{+/+}$ was then replaced with 100% Hybe$^{+/+}$ containing a 1:200 dilution of the appropriate RNA probe and incubated overnight at 65°C.

Probes were washed off using pre-heated Hybe$^{-/-}$ (50% Formamide, 5x SSC, 9.2 mM Citric Acid, 0.1% Tween) then embryos were taken through a series of 10 min washes at 65°C in 75%, 50% and 25% Hybe-/- in 2x SSC (20x stock: 3 M NaCl, 0. 3 M Na3C6H5O7, pH 7. 0), followed by a wash in 2x SSC, all at 65°C. Embryos were then washed four times for 15 minutes at 65°C in 0.2x SSC were followed by a series of 25%, 50%,75% and 100% MABTw (0.1 M maleic Acid, 0.15 M NaCl, 0. 1% Tween, pH 7.5) in 0.2x SSC before blocking in MAB-Block (0.1 M maleic Acid, 0.15 M NaCl, pH 7.5, 2% Blocking Reagent) for 1 hour. Anti-Digoxigenin (Roche) antibody was diluted in MAB-Block (1:5000) and incubated overnight at 4°C. Embryos were then washed 8x at RT in MABtw.

Embryos were equilibrated by washing 3x for 5 min in freshly prepared AP buffer (100 mM Tris-HCl pH 9. 5, 100 mM NaCl, 50 mM MgCl2, 0.1% Tween). Signals were developed using BM Purple^{TM} (Roche) (diluted 1:1 in AP buffer)

Staining was developed in the dark at RT until sufficient staining had developed (this step could be sped up by developing at 32°C or slowed down by incubating at 4°C). To stop the staining process, embryos were washed in PBStw and fixed in PFA for at least 20 min. Embryos were then washed 3x in PBS for 5 min and then transferred to 80% glycerol in PBSTw. Pictures were taken on a dissecting microscope in 80% glycerol.

2.4 Ac/Ds-mediated enhancer screening

4 candidate putative enhancer sequences were predicted for each of the cardiac GATA genes due to their high accessibility and bound p300 based on previously generated ATAC-Seq and p300 ChIP-Seq datasets. Table 2 shows the genomic locations of the chosen putative enhancers. In order to validate the candidate putative enhancers *in vivo,* a novel reporter system was used.

Putative Enhancer	Genomic location
Gata4-E1	chr20:53,146,956-53,148,207
Gata4-E2	chr20:53,153,012-53,154,121
Gata4-E3	chr20:53,302,183-53,303,132
Gata4-E4	chr20:53,309,939-53,311,160
Gata5-E5	chr23:7,434,815-7,435,541
Gata5-E6	chr23:7,485,872-7,486,654
Gata5-E7	chr23:7,488,539-7,489,412

Gata5-E8	chr23:7,494,250-7,495,151
Gata6-E9	chr2:4,402,879-4,403,742
Gata6-E10	chr2:4,391,468-4,392,431
Gata6-E11	chr2:4,436,111-4,437,209
Gata6-E12	chr2:4,416,079-4,417,144

Table 2 Details of genomic locations of putative enhancers

2.4.1 Genomic DNA extraction

Genomic DNA was extracted from a clutch of 5dpf WT embryos using the DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer's instructions.

2.4.2 Polymerase chain reaction (PCR)

Putative enhancers were amplified from genomic DNA using the primers detailed in Table 3.

Primers	Sequence	Fragment size (bp)
gata4en1_FW	*tcgagtttacgtaccgctag* GAATAAACTCAGTTAGTAAATGAC	1252
gata4en1_RV	*TATCGCCGCAAGCTTgctag* ATTGTAGTGCTTTCTGTTATTTGCC	
gata4en2_FW	*tcgagtttacgtaccgctag* ATCTTTGTCCATATTTGACATATTC	1110
gata4en2_RV	*TATCGCCGCAAGCTTgctag* AGTACAGATGAAGTCTAAATTATTCGCCC	
gata4en3_FW	*tcgagtttacgtaccgctag* TCCAGCATATGTTTTACACAGCGG	950
gata4en3_RV	*TATCGCCGCAAGCTTgctag* AGAGCAGTTTGGAGTTAGTTTGTGG	

gata4en4_FW	t*cgagtttacgtaccgctag* TTCTCACATAATAAAAAAAAAC	1222
gata4en4_RV	*TATCGCCGCAAGCTTgctag* CCGCCGGAGGGAATCCTTTTCC	
gata5en5_FW	*tcgagtttacgtaccgctag* TTAAATCACTGTAAAATACGCTTATAC	727
gata5en5_RV	*TATCGCCGCAAGCTTgctag* CATGAATGATAAAAATAAATCACAG	
gata5en6_FW	*tcgagtttacgtaccgctag* ATTTTAATAAATATATCTGTTTAAT	783
gata5en6_RV	*TATCGCCGCAAGCTTgctag* CCCATTATTTAACTTTCTGTGAAGC	
gata5en7_FW	*tcgagtttacgtaccgctag* AGGGTGGCGTAGGACCGGCGGGGC	874
gata5en7_RV	*TATCGCCGCAAGCTTgctag* ATAGAAGAAAAAACTACATTTTAG	
gata5en8_FW	*tcgagtttacgtaccgctag* GTGGGGTATGGGAAGGATAGGTTAG	852
gata5en8_RV	*TATCGCCGCAAGCTTgctag* TCAGCCAGAAAAGATTTAGGTGG	
gata6en9_FW	t*cgagtttacgtaccgctag* TTAAAGAATAAATAGTGTTGACTCG	864
gata6en9_RV	*TATCGCCGCAAGCTTgctag* TAGTGTGCTCTAAAATTGGC	
gata6en10_FW	*tcgagtttacgtaccgctag* TTGAGTGTGTAAGTGCATAAGTAC	964
gata6en10_RV	*TATCGCCGCAAGCTTgctag* GAATCAATGAATAAATGTGTAAATTC	
gata6en11_FW	*tcgagtttacgtaccgctag* ATAAGTTTATGTAGTATAACAG	1099
gata6en11_RV	*TATCGCCGCAAGCTTgctag* TTGTGGCGAATTTAAATAATCAAACG	

gata6en12_FW	*tcgagtttacgtaccgctag* AGATCGTACAAATTCATAGTAATC	1066
gata6en12_RV	*TATCGCCGCAAGCTTgctag* GTGTACTGTTATTATGCTCCCAGC	

Table 3 Details of primers used to amplify putative enhancer regions

The sequence in italics corresponds to the sequences adjacent to the NheI cut site on the recipient vector allowing the required fragments to be integrated using In-Fusion™ HD Cloning (638909, Takara Bio).

PCR was conducted using Phusion High-Fidelity PCR Master Mix (NEB), in a 2-step reaction under the following conditions:

95°C	5 min	x1
95°C	30 sec	x5
55°C	30 sec	
72°C	1 min/kb	
95°C	30 sec	x25
63°C	30 sec	
72°C	1 min/kb	
95°C	5 min	x1
10°C	10 min	x1

2.4.3 Gel purification

PCR products were run on a 0.6% agarose gel electrophoresis and bands of expected product size were cut out under UV light. DNA bands were purified using the Wizard® SV Gel and PCR Clean-Up System according to the manufacturer's instructions. DNA was quantified using a NanoDrop (ND)-1000 (Thermofischer).

2.4.4 Vector restriction digest

The Ac/Ds vector contained a NheI cutting site upstream of an E1b minimal promoter-mCherry reporter sequence and flanked by Ds recognition sites of the Ac transposase (Emelyanov et al., 2006). DNA was digested at 37 °C for 3-12 hours with Nhe1 restriction enzyme (NEB) in the provided buffer. Cut and uncut fragments were analysed on an agarose gel and cut fragments were gel purified as above.

2.4.5 In-Fusion cloning

The amplified putative enhancer sequences (which now also have the homologous sequences adjacent to the NheI cutting site on the recipient vector on each end of the sequence) were inserted into the digested Ac/Ds vector using In-Fusion™

HD Cloning (638909, Takara Bio) according to manufacturer's instructions. When the enhancer DNA sequence was <1kb, transformation was performed into chemically competent Stellar cells. DNA was incubated with Stellar cells on ice for 30 mins before being subjected to heat shock at 42 °C for 45 sec and allowed to recover for 1 hour at 37 °C in 200 µl SOC, shaking at 225 rpm, before plating onto LB agar containing 50µg/ml ampicillin. For the enhancer sequences >1kb, the In-Fusion material was ethanol precipitated with 10M NH4Ac at a final concentration of 0.3M and 100% ethanol at -80°C for 20 mins. Samples were then spun at 12,000g for 30mins at 4°C and the resultant DNA pellets were washed in 70% ethanol, air

dried and resuspended in RNAse free water. DNA samples were then incubated with competent DH10 β cells on ice for 5 min. The mixture was then loaded into electroporation cuvettes and electroporated at 1660V and allowed to recover for 1 hour at 37 °C in 200 µl SOC, shaking at 225 rpm, before plating onto LB agar containing 50µg/ml ampicillin. Liquid cultures were set up from the resulting colonies and plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (QIAGEN), according to the manufacturer's instructions. Samples were quantified using a NanoDrop and correct cloning was confirmed by Sanger sequencing using the in-house sequencing facility.

2.4.6 Transcription of Ac RNA

Ac transposase was transcribed from the pAC-SP6 vector (Emelyanov et al., 2006). The vector was linearised with BamHI as described above, and Ac was transcribed using the mMESSAGE mMACHINE™ SP6 Transcription Kit (AM1340, ThermoFisher Scientific) according to manufacturer's instructions. Synthesised RNA was precipitated with 10M NH4Ac at a final concentration of 0.3M with 100% ethanol at -80°C for 20 mins. Samples were then spun at 12,000g for 30mins at 4°C and the resultant RNA pellets were washed in 70% ethanol, air dried and resuspended in RNAse free water. A sample was run on an RNAse free 1% agarose/TAE gel to confirm that the RNA had been transcribed and the samples were quantified using Qubit Fluorometric Quantitation (ThermoFisher).

2.4.7 Zebrafish embryo micro-injection

Breeding pairs were set up in individual tanks, separated by a divider, about 12-18 hours before injection. In the morning, dividers were removed, and eggs were collected from natural spawning and aligned against a glass slide in a petri dish. Embryos were injected at the 1-cell stage with 30pg of DNA plasmid and 8pg of *in*

vitro transcribed Ac transposase in 1nl. When screening enhancers, 50-100 embryos were injected per enhancer and then observed throughout development.

2.5 Zebrafish cryoinjury

2.5.1 Cryoinjury and heart extraction

All procedures/protocols were carried out in accordance with British Home Office regulations. Cryoinjury was performed as described in (González-Rosa and Mercader, 2012). Adult zebrafish were immersed in 160 mg/l MS222 (Sigma) and placed ventral side up in a firm sponge under a dissecting microscope. The ventral scales covering the heart were removed using forceps and then a small incision was made through the body wall and the pericardium to the heart using microdissection scissors. The ventricle of the heart was exposed by very gently pressing the abdomen of the fish. The exposed heart was dried using sterile tissue. A cryoprobe with a copper filament was cooled by dipping into liquid nitrogen and placed onto the ventricular surface until thawing was observed. Sham surgery, which was used as a control, involved all steps routinely done before cryoinjury, but the ventricular apex was briefly touched with a metal probe at room temperature. After surgery, fish were placed in a tank of fresh water and water was pipetted gently across the gills until the fish began to swim again.

At 3 days post-injury, fish were euthanised in 4g/l MS222 and hearts were extracted, washed in Ringer solution, and fixed overnight at 4°C in 4% PFA in PBS.

2.5.2 Paraffin embedding and sectioning

Hearts were washed 2x 20 minutes with PBSTw and then washed for 30 minutes in 50% then 70% ethanol in PBSTw followed by 6x30 minute washes in 100% ethanol.

Hearts were then washed in 100% Xylene for 2x 30 minutes, then in 50% Xylene: paraffin for 45 minutes and then washed into 100% paraffin wax and embedded in paraffin blocks. Sections were cut using a microtome (Leica) at a thickness of 10 µm.

2.6 Immunocytochemistry (ICC)

2.6.1 Embryo whole-mount ICC

Embryos were collected following natural spawning and incubated at 28.5°C until the required time points. Embryos were euthanised in 5g/L ethyl-3-amino benzoate methanesulfonate (MS222) and fixed in 4% PFA in PBS for 1 hour at room temperature. Embryos were then washed 3 times for 5 minutes in PBSTw (0.5%) and blocked using 5% goat serum for 1 hour. Primary antibodies used were: chicken α-GFP (ab13970, Abcam), rabbit α-mCherry (GTX128508, GeneTex), mouse α-mCherry (632543, Clontech), mouse α-Isl1/2 (Developmental Studies Hybridoma Bank, Iowa City, IA, USA, clone 39.4D5), and they were incubated with embryos overnight at 4°C in PBSTw with 2% goat serum at 1:500 dilution. The following day, embryos were washed 6x 15 minutes in PBSTw at room temperature. The following secondary antibodies were used and incubated overnight in PBS-T, 2% goat serum at a dilution of 1:1000: goat α-chicken-488nm (ThermoFisher), goat α-rabbit-555nm (ThermoFisher), goat α-rabbit-633nm (ThermoFisher), goat α-mouse-633nm (ThermoFisher) and Hoechst reagent 33258. Embryos were washed 3 times in PBSTw and mounted on Superfrost slides using VectaShield (H-1400, Vector Laboratories Inc) mounting medium, or in 1% low-melting agarose .

2.6.2 ICC on adult heart sections

Slides were de-waxed in Xylene for 2x 7 minutes. Slides were then washed 2x 5 minutes in 100% ethanol, followed by 1-minute washes in 96%, 90%, 80% and 70%

ethanol in PBSTw then 5 minutes in PBTw. Sections were then pressure cooked for 4 minutes in Antigen unmasking solution (H-3300, Vector Laboratories Inc), cooled and then blocked using TNB (NEL702001KT, Perkin Elmer) followed by primary antibody incubation overnight in TNB (1:200 dilution) at room temperature in a humidified slide chamber. Sections were washed 3x 5 minutes with PBSTw and then incubated with secondary antibody in TNB (1:200) and Hoechst (1:1000) for 2 hours at room temperature. Sections were washed 3x 5mins with PBSTw and then mounted in VectaShield (H-1400, Vector Laboratories Inc) mounting medium.

2.7 Hybridisation chain reaction

Hybridisation chain reaction was performed using a modified protocol from (Choi et al., 2018). Embryos were collected from natural spawning and, at the required time points, were euthanised in 5g/L MS222 and fixed in 4% paraformaldehyde overnight at 4°C. Embryos were subsequently washed into methanol and stored at -20°C (at least overnight). On day 1 of the protocol, embryos were rehydrated into PBS-Triton (0.1%), permeabilised with 30ug/ml of proteinase K for 12minutes at room temperature, post-fixed in 4% PFA for 20 minutes at room temperature, washed 5x 5 minutes with PBSTw then pre-hybridised with probe-hybridisation buffer (30% formamide, 5x sodium chloride sodium citrate (SSC), 9mM citric acid (pH6.0), 0.1% Tween20, 50 µg/mL heparin, 1x Denhardt's solution, 10% dextran sulphate) for 30 minutes at 37°C. Samples were then incubated with 2nM of probes against the target mRNAs in hybridisation buffer overnight at 37°C. Probes against *gata5* (amplifier B4), *gata6* (amplifier B4), *myl7* (amplifier B5), *mCherry* (amplifier B1), *nkx2.5* (amplifier B2), *prrx1a* (amplifier B2), *shox2* (amplifier B2), *jam2b* (amplifier B4), *adamts15* (amplifier B5), *col1a1b* (amplifier B5), *cxcl12a* (amplifier B3), *sema3fb* (amplifier B5), *isl1* (amplifier B5), and *col4a1* (amplifier B4) were ordered

directly from Molecular Instruments. Excess probes were removed by washing embryos 4x 15 minutes with 30% probe wash buffer (30% formamide, 5x SSC, 9mM Citric acid (pH 6.0), 0.1% Tween20, 50µg/ml heparin) and then 3x 5 minutes with 5x SSCT (5x SSC, 0.1% Triton) at room temperature. Subsequently, embryos were pre-amplified in amplification buffer (5x SSC, 0.1% Tween20, 10% dextran sulphate) for 30 minutes at room temperature. Fluorescently labelled hairpins were heated to 95˚C and allowed to cool down in the hot block for 3-4 hours. Hairpins were diluted to 30nM in amplification buffer and embryos were incubated with the hairpin mix overnight in the dark at room temperature. The next day, embryos were incubated with Hoechst reagent (1:1000, 5x SSC-T) for 30 minutes at room temperature and excess hairpins were washed off with 5x SSC-T (2x 5 minutes, 2x 30 minutes, 1x 5 minutes). Embryos were mounted on Superfrost slides using VectaShield mounting medium.

2.8 CRISPR-Cas9 editing

2.8.1 Synthesis of sgRNA

sgRNAs were designed using the online tool CCtop (Stemmer et al., 2015) followed by a cloning free synthesis protocol ((Varshney, 2015)) using the primers below. Guide RNA sequences are listed in Table 4.

Forward oligo-

GAAATTAATACGACTCACTATAGGG xxxxxxxxxxxxxxxxxxx *GTTTTA GAGCTAGAAATAGC*

(Bold= T7 promoter sequence, xxxxxx represents the 18-20 nucleotide guide RNA sequence, italics= overlap region with tracrRNA)

Reverse oligo (same for all guides)-

AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATT

TAAaCTTGCTATgCTGtttCCAGCAtagctctTAAAC

Name	Sequence
Enh5 5'4	**GAAATTAATACGACTCACTATAGGG**CTGTAAACTCGCTAACATTT**GTTTTA GAGCTAGAAATAGC**
Enh5 5'5	**GAAATTAATACGACTCACTATAGGG**ATGTAACATATTGTGAAAT**GTTTTA GAGCTAGAAATAGC**
Enh5 5'6	**GAAATTAATACGACTCACTATAGGG** ACAATATGTTACATCCATAA GTTTTA GAGCTAGAAATAGC
Enh5 3'1	**GAAATTAATACGACTCACTATAGGG**ACATTCATTTAAATTAGCTG**GTTTTA GAGCTAGAAATAGC**
Enh5 3'2	**GAAATTAATACGACTCACTATAGGG**TCATTCATGAAAATAATGTT**GTTTTA GAGCTAGAAATAGC**
Enh5 3'3	**GAAATTAATACGACTCACTATAGGG**CATGAAAATAATGTTAGGTT**GTTTTA GAGCTAGAAATAGC**
Enh5 3'4	**GAAATTAATACGACTCACTATAGGG**ATATTGCTGCATGAAGATGT**GTTTTA GAGCTAGAAATAGC**
Enh12 5'1	**GAAATTAATACGACTCACTATAGGG**TCATCCCAAGTTATTAAAAT**GTTTTA GAGCTAGAAATAGC**
Enh12 5'4	**GAAATTAATACGACTCACTATAGGG** TTGCGTTAGGCCATCGTAAG GTTTTA GAGCTAGAAATAGC
Enh12 5'5	**GAAATTAATACGACTCACTATAGGG** TTTACGTGAAACTGTCCGGC GTTTTA GAGCTAGAAATAGC
Enh12 3'2	**GAAATTAATACGACTCACTATAGGG**TCTAGTATGCATACTAACAT**GTTTTA GAGCTAGAAATAGC**
Enh12 3'4	**GAAATTAATACGACTCACTATAGGG** TCAACATTACTTAAGTATGT GTTTTA GAGCTAGAAATAGC
Enh12 3'5	**GAAATTAATACGACTCACTATAGGG** AAACCTATAGACTCAGTACC GTTTTA GAGCTAGAAATAGC

Table 4 Details of the oligos used for sgRNA synthesis

PCR was performed using the Phusion High-Fidelity PCR Master Mix (NEB) using the following programme:

95°C	5 min	x1
95°C	30 sec	x35
60°C	30 sec	
72°C	30 sec	
72°C	10 min	x1
10°C	∞	

Annealing and extension were confirmed on a 2% agarose gel using 5ul of the PCR product. The PCR product was cleaned up with the PCR Purification Kit (e.g. QIAGEN). RNA was *in vitro* transcribed using the NEB High Scribe T7 Quick High Yield Kit. A 30 µl reaction mix containing 500 ng of template DNA, 10 µl of NTP buffer mix and 2 µl T7 RNA Polymerase was incubated at 37°C for 4 hours. Template DNA was removed using 2 µl of DNAseI for an additional 15 minutes. RNA was purified with the MEGA Clear kit (Ambion) and eluted in 50 µl. RNA quality was confirmed on an agarose gel and quantified using Qubit Fluorometric Quantitation (ThermoFisher).

2.8.2 Cas9 mRNA synthesis

Cas9 mRNA was transcribed from the pT3TS-Cas9 vector (Emelyanov et al., 2006). The vector was linearised with XbaI as described above, and Cas9 was transcribed

using the mMESSAGE mMACHINE™ T3 Transcription Kit (AM1348, ThermoFisher Scientific) according to manufacturer's instructions. Synthesised RNA was precipitated with 10M NH4Ac at a final concentration of 0.3M with 100% ethanol at -80°C for 20 mins. Samples were then spun at 12,000g for 30mins at 4°C and the resultant RNA pellets were washed in 70% ethanol, air dried and resuspended in RNAse free water. A sample was run on an RNAse free 1% agarose/TAE gel to confirm that the RNA had been transcribed and the samples were quantified using Qubit Fluorometric Quantitation (ThermoFisher).

2.8.3 CRISPR-Cas9 injections

Embryos were injected at the one-cell stage with 250pg Cas9 mRNA, 100pg of each sgRNA in 1nl. For comparison, as a control, a clutch of embryos were also injected with only Cas9 mRNA (no sgRNA).

2.8.4 High resolution melt analysis (HRMA)

At 24hpf, injected embryos (including control Cas9-only injected embryos) were individually collected and genomic DNA was extracted using the HotSHOT method (Truett et al., 2000). Briefly, collected embryos were immersed in 75µl alkaline lysis buffer (25 mM NaOH, 0.2 mM disodium EDTA, pH 12.0) and incubated at 95°C for 1 hr. Samples were cooled to 4°C before adding 75µl neutralising buffer (40 mM Tris-HCl, pH 5.0).

Primers were designed using the Primer3 online tool to generate a ~150bp product spanning each targeted cut site, shown in Table 5. A PCR reaction mix was set up containing 5µl of 2x Hotshot Diamond PCR Master Mix (Client Lifescience, #HS002-TS), 1µl of 10x LC Green plus Dye (BioFire Diagnostics, #BCHM-ASY-0005), 0.5µl of 10mM forward and reverse primer, 0.5µl of the DNA template ad ddH20 up to

10µl per reaction. The reaction was completed using a C1000 TouchTM Bio-Rad thermal cycler. Cycling conditions: 95°C for 5min, 49X (95°C for 20s, 60°C for 20s, 72°C for 30s), 72°C for 2min, 95°C for 30s, 25°C for 30s and 95°C for 30s. HRMA was performed by increasing the temperature from 65°C to 95°C in 0.2°C increments (10s/step). Raw data was plotted as fluorescence vs. temperature melt curves by the BioRad Precision Melt Analysis software. The fluorescence signal intensity was normalised so that the y-axis represented Relative Fluorescence Units (RFU) raging from the baseline noise value of 0 to the maximum fluorescence signal of 1.0. Melting curves were then temperature shifted and difference plots generated for comparison of the sgRNA/Cas9-injected embryos melt profiles with the Cas9-only injected control embryos, allowing assessment of successful sgRNAs.

HRMA primers	Sequence
5' enh5 Fwd	TGGCACTTGTGTGATTTTGG
5' enh5 Rev	AAGAGGTGTCACTGGGGTTG
3' enh12 Fwd	TGCTGAGGCTGCTTATGTCA
3' enh12 Rev	GTCTATAGGTTTTTAGCACCAACAT

Table 5 Details of HRMA primers to test sgRNA cutting efficiency

2.9 Microscopy and digital image analysis

2.9.1 Microscopy

Embryos were either stained and mounted onto glass slides using VectaShield mounting medium or euthanised in 5g/L MS222 and mounted into plastic dishes using 1% low melting agarose. Imaging was performed on an upright LSM780 confocal microscope (Zeiss). Embryos mounted in agarose were imaged using a

20x dipping objective and embryos mounted under glass coverslips were imaged with a 40x water objective.

Brightfield microscopy was performed for embryos subjected to whole-mount in situ hybridisation. Images were captured with a Nikon DS-Ri2 camera attached to a Nikon SMZ1000 dissecting microscope using the NIS-Elements BR 4.50.00 software.

2.9.2 Image analysis

All confocal images were processed using the Zen software (Zeiss) or Fiji-ImageJ. Fiji-ImageJ software was used to calculate both fluorescent intensities and atrial area. In order to count fluorescent cells, images were 3D-rendered using Imaris (Bitplane) and then nuclei were counted.

2.10 10x Single Cell RNA-Seq

2.10.1 Sequencing

Single cells were obtained from FAC-sorting and collected 7.5k cells in the first replicate and 7k cells in the second replicate. Libraries were prepared at the WIMM Core Single Cell Facility. Prior to sequencing, libraries were quantified using Qubit Fluorometric Quantitation (ThermoFisher) and Kapa to obtain a 4nM final library pool. Libraries were denatured, neutralised and diluted down to a final concentration of 1.8 pM, 1% of PhiX added and then loaded onto the NextSeq 150 cycle High Output 550/500 kit (Illumina). Sequencing was carried out in Standalone mode on an Illumina NextSeq sequencing machine using paired ends, 26 cycles for Read 1, 98 cycles for Read 2, 8 cycles for Index 1 and 0 cycles for Index 2 (according to manufacturer's instructions).

2.10.2 Analysis

Bioinformatic analysis was performed by Dr Irving Ling. Single cell RNA-seq reads were demultiplexed using Cellranger (v2.2.0) and mapped to the zebrafish *Danio rerio* v10 transcriptome generated using the Cellranger “mkref” function. The first replicate run was spiked with cells from chicken and therefore citrine and mCherry transcripts were included into the zv10 and galGal5 FASTA and GTF files to allow correct assignment of cells to the appropriate organism. We combined two separate 10X sequencing runs from two replicates and aggregated the data using the “aggr” function.

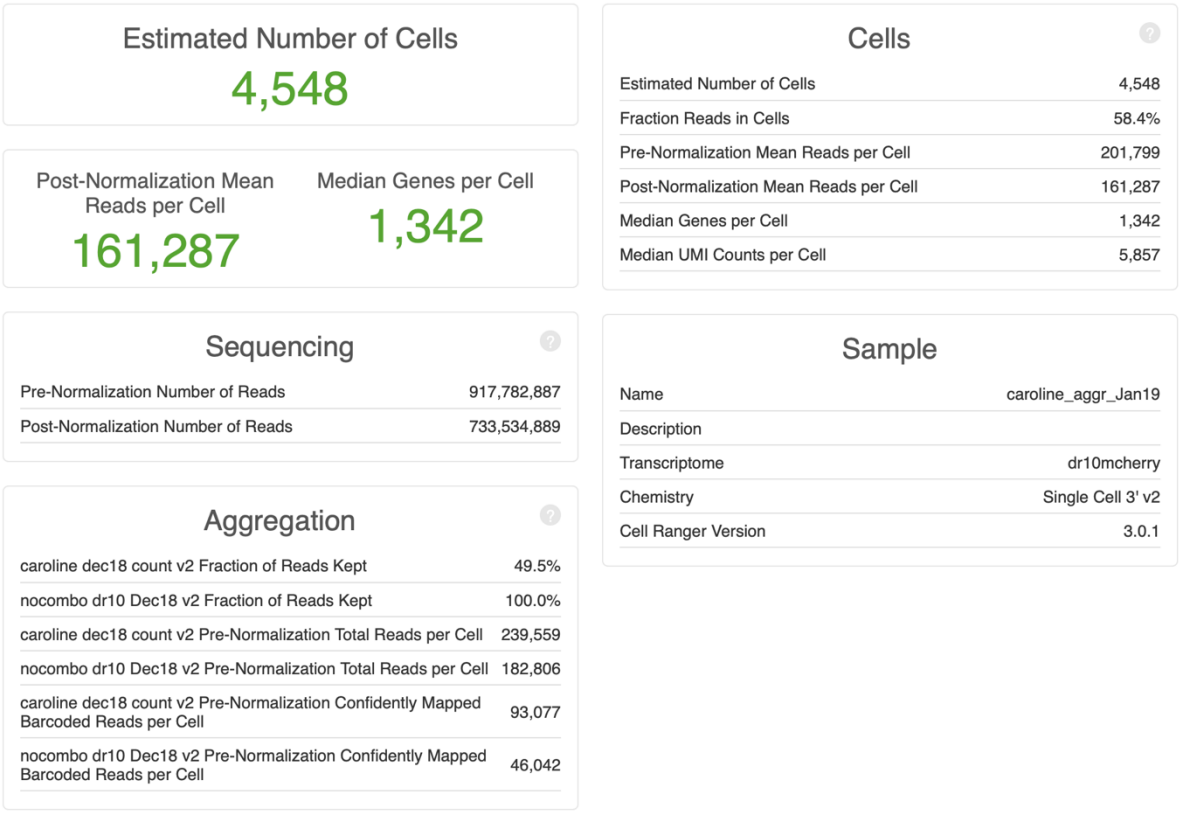


Figure 2-1 Cellranger aggregate output from two replicate runs.

A total of 4,548 cells were recovered from the runs with median genes per cell of 1,342. Note that the percentage of aggregation in the first replicate was 49.5% due to the spike in of chicken neural crest cells and therefore, offered a lower fraction of reads in cells. Median UMI counts per cell was 5,857.

Seurat package (v2.3.4) in R was used for further downstream analysis. After removing mitochondrial genes and filtering cells with low number of unique genes detected in each cell or aberrantly high gene count (which could indicate low quality cells, empty droplets, doublets or multiplets), we recovered 3976 cells (87.4% of total number of cells sequenced). Linear dimensional reduction was carried out and statistically significant principal components determined 12 cluster groups. A resolution of 0.5 was used to generate cluster definition before generating the tSNE plot. In each cluster, we recovered:

Cluster	0	1	2	3	4	5	6	7	8	9	10	11
Cells	761	583	549	486	359	340	335	213	143	93	66	48

We then subclustered 6, 7, 9, 10 and 11 (as clusters having the mCherry transcript) and use this for further downstream analysis. Running the pipeline again, we obtained 8 clusters using a resolution of 0.8. In each cluster, we recovered:

Cluster	0	1	2	3	4	5	6	7
Cells	169	135	123	118	85	66	42	17

2.10.3 GO term analysis

Enrichment for GO terms for biological processes were calculated using PANTHER version 14.1.

2.11 Binding site analysis

2.11.1 JASPAR analysis

Using the JASPAR 2014 database, the core regions of the selected putative enhancers were analysed for predicted transcription factor binding sites. The core was defined as the region at the centre of the ATAC peak and the following sequences were analysed using a minimum score of 87%:

2.11.2 Binding site perturbation constructs

For each TF family with predicted binding, mutations were designed in all putative binding motifs, according to the position weight matrix for each motif by substituting

the nucleotides with the highest affinity for the TF. Mutated sequences were re-analysed through JASPAR to confirm that the binding motif for the TF family was erased and no new TF binding sites were created.

Mutated sequences were ordered as double-stranded synthetic DNA (gBlocks Gene Fragments, Integrated DNA Technologies). Sequences were designed with In-Fusion arms on each end so that they could be cloned into the Gata5-E5-mCherry Ac/Ds vector by In-Fusion cloning after the core of the enhancer had been excised using the restriction enzymes NheI and SnaBI. Constructs were then injected as in 2.4.7 and imaged as in 2.9.1. At 2dpf, 3dpf and 5dpf, the numbers of positive cells in each heart were counted. The wild type, unmutated sequence was cloned into the construct to be used as a control to compare the expression patterns of the mutated enhancer-driven reporters against.

2.12 Blinding, quantification and statistics

The CRISPR knockout (section 2.8) and binding site perturbation (section 2.11.2) experiments were not blinded. However, on reflection, this would have been the proper way to conduct these experiments. In the CRISPR experiments, changes in fluorescence intensity after enhancer knockout were measured using fluorescence intensity analysis of confocal images on Fiji-ImageJ software. Significance was calculated using an unpaired Student's t-test. *= $p<0.0332$, **= $p<0.0021$, ***= $p<0.0002$, ****= $p<0.0001$. The effect of binding site perturbation was quantified by counting the number of fluorescent cells after injection of the respective mutated enhancer-driven reporter. In order to count fluorescent cells, confocal images were 3D-rendered using Imaris (Bitplane) and then nuclei were counted manually. Significance was calculated using an unpaired Student's t-test. *= $p<0.0332$, **= $p<0.0021$, ***= $p<0.0002$, ****= $p<0.0001$

3) Identification of putative enhancers of the cardiac GATA genes and *in vivo* testing

3.1 Introduction

Heart development is a complex process, controlled by highly evolutionarily conserved transcriptional networks. Differentiation of the multiple cell types that contribute to the heart requires the establishment of highly regulated spatio-temporal patterns of gene expression (Stainier, 2001). It is an emerging paradigm in regenerative medicine that developmental gene programmes may be reactivated upon injury to facilitate tissue regeneration. Comprehension of the complex networks involved in cardiogenesis is critical if we are to understand the molecular basis of congenital heart malformations and unlock the potential of reactivating developmental programmes following myocardial infarction.

Mutations in the cardiac GATA factors have been identified in human patients as the origin of multiple CHDs. Missense or nonsense mutations in *gata4* typically lead to a plethora of septal defects (Garg et al., 2003, Moskowitz et al., 2011). *Gata5* disruption can cause disorders such as tetralogy of Fallot (Wei et al., 2012), atrial fibrillation (Gu et al., 2012) and septal defects (D. Wei, 2013). *gata6* mutations have been associated with a number of congenital cardiac abnormalities, including ventricular septal defect, truncus arteriosus (Lin et al., 2010), tetralogy of Fallot and atrial septal defect (Li et al., 2014).

As a result of the numerous clinical presentations of GATA-derived abnormalities, and the extensive study of the gene family in model organisms, we have a detailed understanding of the downstream effectors of the cardiac GATA factors. There is,

however, a lack of understanding of the precise mechanisms acting upstream of the GATA factors during heart development.

To address this issue, I have investigated the upstream regulation of the GATA factors during cardiogenesis by interrogating their *cis*-regulatory landscapes. Specifically, by studying putative GATA enhancers. As outlined in chapter 1, active enhancers are typically devoid of nucleosomes, but not all open regions of chromatin are active enhancers. Putative enhancers must therefore be tested *in vivo*, e.g. by placing the candidate DNA sequence upstream of a minimal promoter and a reporter gene (Shlyueva et al., 2014a). In this chapter, I will describe my strategy for accurately predicting putative enhancer sequences based on next-generation sequencing datasets generated in my laboratory.

An in-depth analysis of the putative enhancers regulating the cardiac GATA genes will provide novel information on the precise temporal and spatial control of GATA gene expression and their critical upstream regulators.

3.2 Results

3.2.1 Putative enhancers for the cardiac GATA genes can be predicted using ATAC-Seq and ChIP-seq datasets.

Active enhancers are present within regions of open chromatin (Tsompana and Buck, 2014). As outlined in chapter 1, ATAC-Seq is a reliable method for identifying regions of open chromatin (Buenrostro et al., 2015). ATAC-Seq datasets from sorted CMs from 5 dpf embryos and adult hearts collected 3 days post-cryoinjury (dpi), 5dpi and appropriate control shams, generated by Dr Filipa Simões, formed the basis for this study (summarised in Figure 3-1).

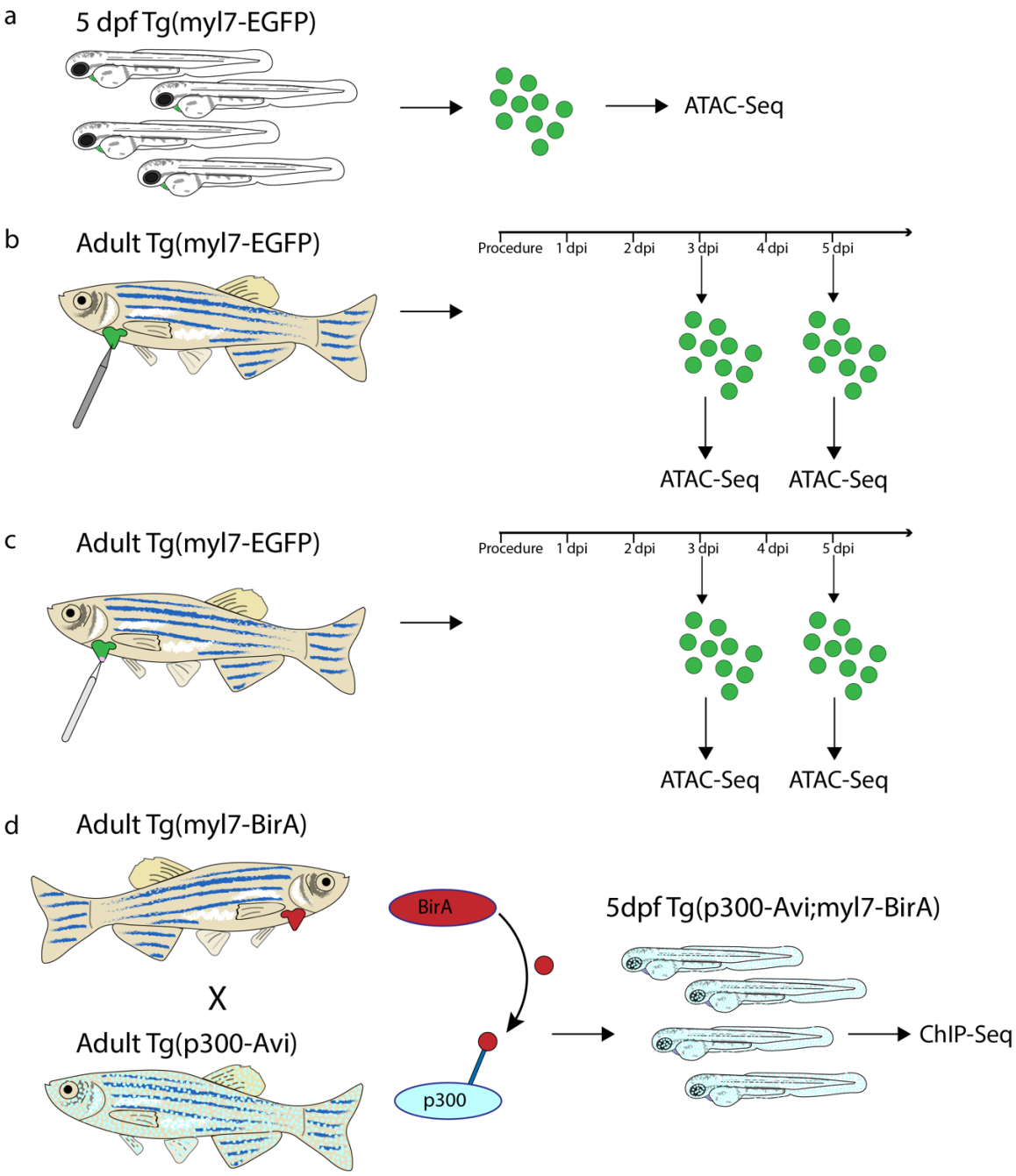


Figure 3-1 Summary of ATAC-Seq and ChIP-Seq data collection

a) ATAC-Seq was performed on CMs sorted from 5dpf embryos. b) ATAC-Seq was performed on CMs sorted from sham-operated hearts at 3dpi and 5dpi. c) ATAC-Seq was performed on CMs sorted from cryo-injured hearts at 3dpi and 5dpi. d) *In vivo* biotagging approach to conduct p300 ChIP-Seq: BirA was driven by the *myl7* promoter in *Tg(myl7-BirA)* and Avi-tagged p300 was expressed ubiquitously in all cells. p300 was thereby biotinylated in CMs, allowing purification of p300 using streptavidin beads and analysis of the regions bound by p300.

p300 is a conserved histone acetyltransferase that is known to associate with active enhancers and, as such, has previously been used to identify active enhancers in many other tissues (Shlyueva et al., 2014a). Dr Upeka Senanayake performed CM-specific p300 ChIP-Seq at 5dpf. Because of the lack of ChIP-grade antibodies available for zebrafish, a novel biotagging approach was used, as detailed in (Trinh et al., 2017). p300 was tagged with a biotin acceptor peptide (Avi tag) and expressed in every cell while biotin ligase was driven by the CM-specific *myl7* promoter. The Avi tag fused to p300 becomes biotinylated, allowing streptavidin-based affinity purification of p300 and its associated bound DNA specifically in CMs (outlined in Figure 3-1d).

Using both the available ATAC-Seq and p300 ChIP datasets in combination, I was able to identify the open chromatin regions around the cardiac GATA TFs loci that were also bound by p300. I then selected four putative enhancers for each of the cardiac GATA TFs, *gata4, gata5* and *gata6,* which are all expressed in the developing heart (Figure 3-2). Each selected region was accessible in both embryonic CMs and those in the injured adult heart. It is well known that enhancers can act over long distances, sometimes hundreds of kilobases, and often bypassing other genes, but in the absence of chromatin conformation capture data to establish the TADs of the GATA genes, I decided to focus on putative enhancers proximal to each of the cardiac GATA loci, using the 'nearest neighbour' approach (Hafez et al., 2017).

Although there are many ways to predict putative enhancers (as outlined in chapter 1), there is, to date, no definitive enhancer marker that reliably predicts whether an enhancer is active in a given cell type at a given time, therefore putative enhancers must still always be tested *in vivo*.

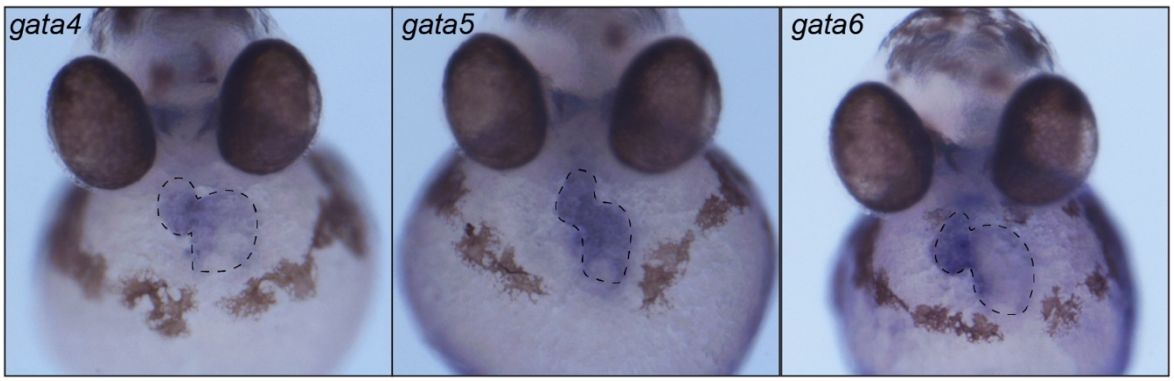


Figure 3-2 In situ hybridisation to show the expression of gata4, gata5 and gata6 at 48hpf

3.2.2. Putative enhancers for *gata4* showed no activity *in vivo*

In order to test the activity of the putative enhancers *in vivo*, each enhancer sequence was cloned upstream of a minimal promoter and a mCherry expression cassette in a vector for the novel and highly efficient Ac/Ds system (Chong-Morrison et al., 2018) using In-Fusion cloning. The Ac/Ds system allows analysis in the F0 generation, displaying little-to-no background ectopic expression. The E1b minimal promoter of adenovirus was used as it has previously been shown to be far superior to other alternatives such as EF-1α (Li et al., 2010). I then assessed enhancer-driven reporter expression during the first 5 days of development (workflow outlined in Figure 3-3).

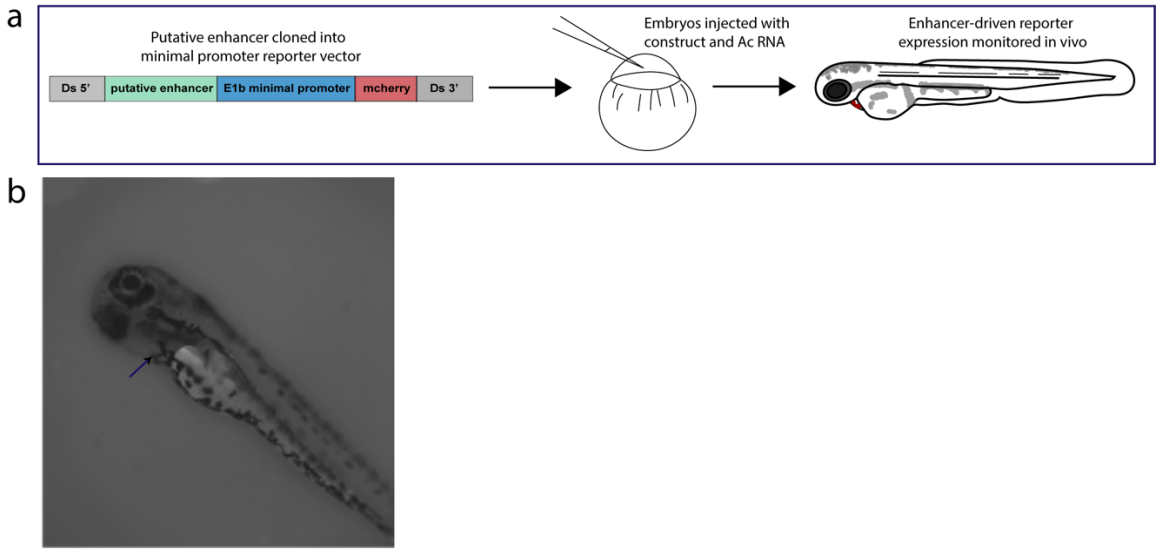


Figure 3-3 Workflow for screening putative enhancers

a) Workflow to assess the *in vivo* activity of selected putative enhancers. Putative enhancer sequences were cloned upstream of the E1b minimal promoter and an mCherry reporter in a cassette for the Ac/Ds transposon system. The construct was then injected into wild type embryos at the one-cell stage and embryos were monitored for enhancer-driven fluorescence. b) Embryos injected with an 'empty' vector, containing no putative enhancer sequence showed no fluorescence in the embryo. (Black arrow shows heart region)

I selected 4 putative enhancer regions for *gata4*; they varied in size from 950-1252bp, and two were located within the introns of *gata4* and the remaining two were 139kb and 147kb upstream of *gata4* (Figure 3-4).

No embryos injected with any of the *gata4* putative enhancer constructs showed any fluorescence (data not shown), suggesting that the putative enhancers selected for *gata4* are not active in the zebrafish embryo *in vivo* at the developmental times analysed. This highlights the importance and necessity of *in vivo* testing in addition to in silico prediction tools. Of note, the ATAC peaks of the chosen *gata4* putative enhancers are significantly smaller than those chosen for *gata5* and *gata6*.

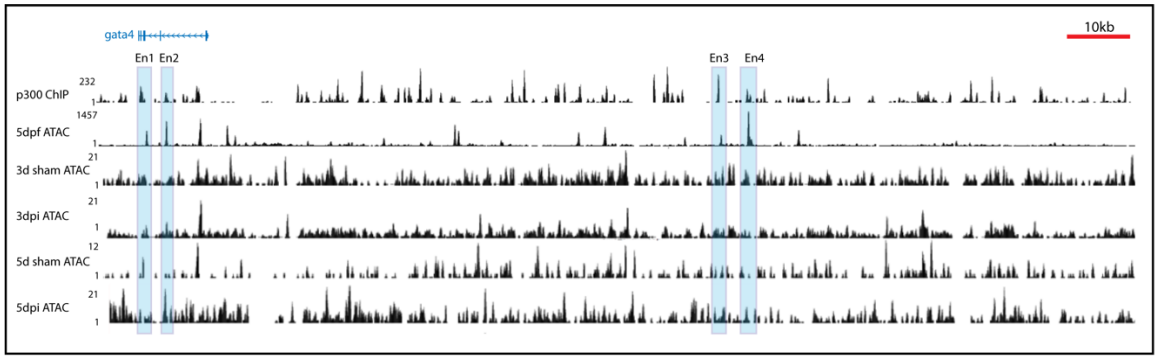


Figure 3-4 Selected putative enhancers for gata4 did not show activity in vivo.

ATAC-Seq datasets at the *gata4* locus from sorted Myl7-positive CMs at 5dpf and 3 and 5days post-cryoinjury and p300 ChIP datasets from sorted CMs at 5dpf. Blue bars indicate selected putative enhancers. No fluorescence was observed upon injection of any of the *gata4* putative enhancers tested (data not shown).

3.2.3. Putative enhancers for *gata5* showed heart-specific activity *in vivo*

By contrast to those tested around the *gata4* locus, all four putative enhancers identified and assigned to *gata5* were active *in vivo* during embryonic development and showed heart-specific enhancer-driven reporter expression at 3dpf (Figure 3-5). This heart-specific expression pattern was observed in all positive embryos (40-60 studied per enhancer). Interestingly, all four putative enhancers had larger ATAC peaks, both in the embryonic datasets and the adult. The enhancers ranged in size from 727bp to 874bp. Gata5-E5 lies 10kb upstream of the *gata5* promoter and Gata5-E6-8 lie 40-48.5kb downstream. I also injected some embryos with an 'empty vector' (i.e. lacking any putative enhancer sequence) as a control for signal specificity and these embryos exhibited no mCherry fluorescence (Figure 3-3). This confirms that any fluorescence seen is specific to the activity of the enhancer being tested.

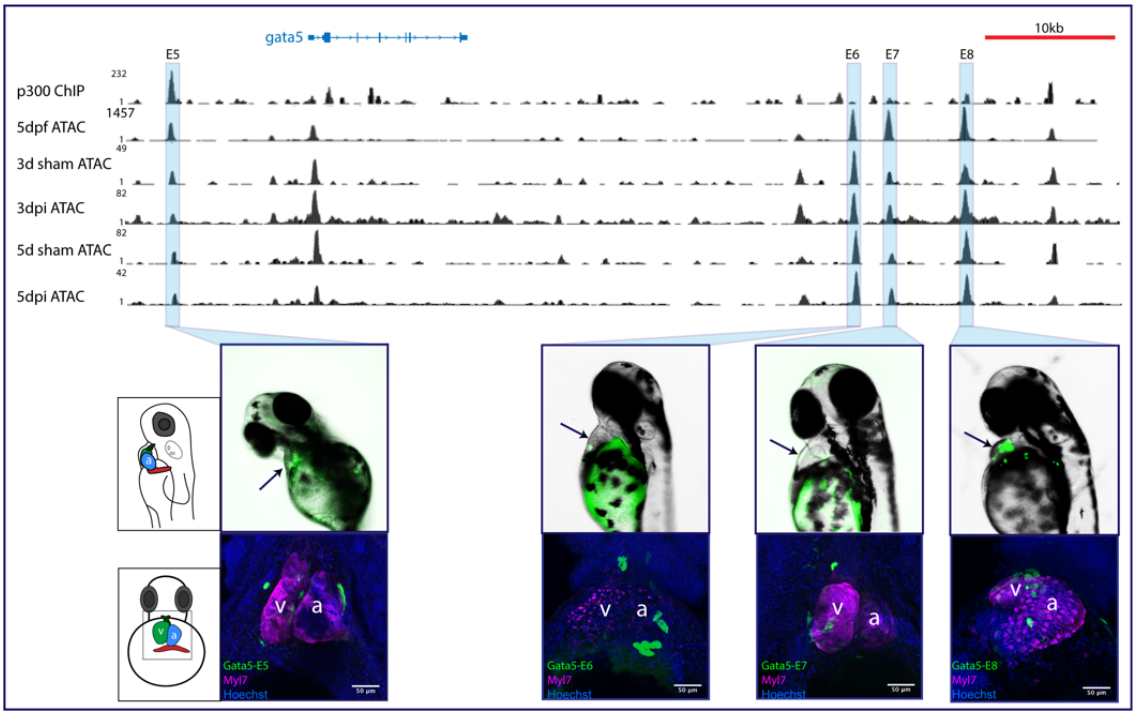


Figure 3-5 Putative gata5 enhancers showed heart-specific activity in vivo

ATAC-Seq datasets at the *gata5* locus from sorted Myl7-positive CMs at 5dpf and 3 and 5days post-cryoinjury and p300 ChIP data sets from sorted CMs at 5dpf. Blue bars indicate selected putative enhancers. Microscopic analysis of F0 embryos at 3dpf revealed heart-specific activity of each of the putative enhancers for *gata5.* In the lower panels, black arrows indicate activity of the enhancer-driven reporter. Confocal images show a ventral view of the enhancer-driven reporter expression in the heart at 3dpf. a=atrium, v=ventricle.

Of note, although I have generated transgenic enhancer constructs expressing the mCherry reporter, I have pseudocoloured this in green in all confocal images in order to make the expression clearer and easier to visualise in the images.

Even in transient, in the F0 generation, it was clear that these *gata5* enhancers were active within subpopulations of the developing heart and these subpopulations did not appear to completely overlap. However, *gata5* is expressed widely throughout the developing heart (Figure 3-2). Identifying the specific cells in which each enhancer is active will provide novel information regarding the precise upstream regulation of *gata5* in individual cell populations during cardiogenesis.

3.2.4. Putative enhancers for *gata6* showed heart-specific activity *in vivo*

The *gata6* putative enhancer regions varied in size, from 852bp to 1099bp and were all located 25-69kb downstream of the *gata6* gene (Figure 3-6). All but one of the *gata6* putative enhancers showed heart-specific activity *in vivo*. No fluorescence was observed during the first 5 days of development when the construct containing Gata6-E10 was injected. This was surprising given that the ATAC peak at this locus was the largest of all four, which has previously been correlated with activity *in vivo*. As is the case for *gata5*, *gata6* is expressed throughout the developing heart (Figure 3-2), however Gata6-E9, E11 and E12 appear to be active in subpopulations of the *gata6* expression domain. In order to tease apart the regulation of *gata6* in detail, it will be important to identify these subpopulations that revealed specific enhancer activity.

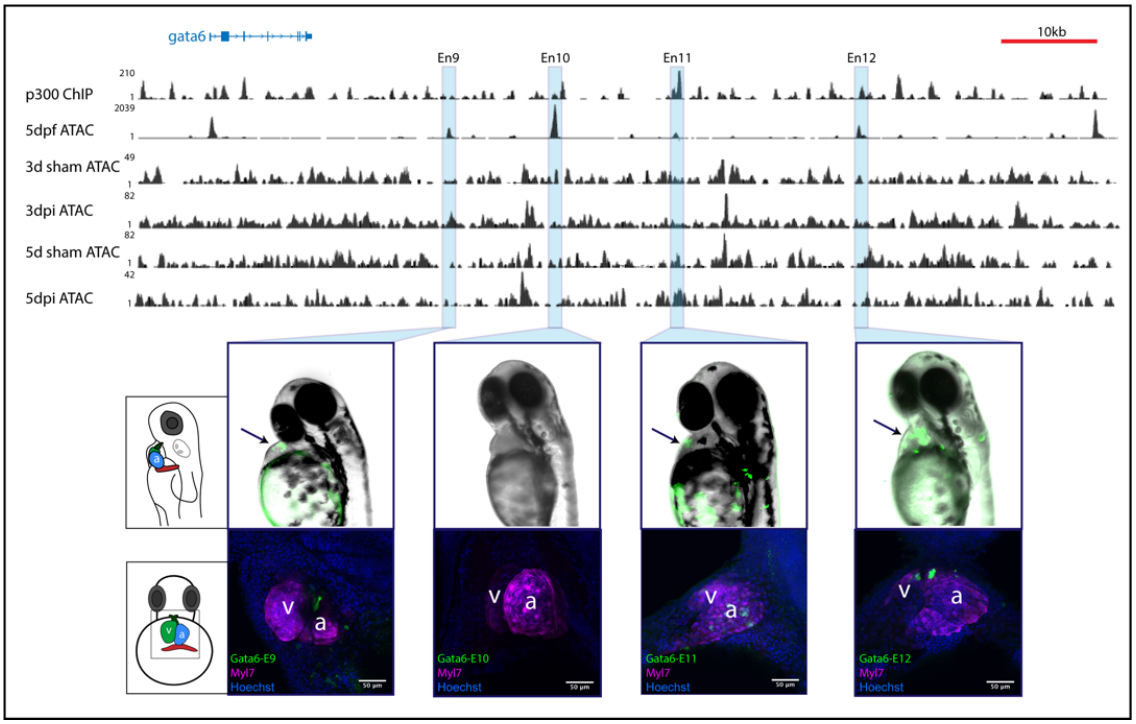


Figure 3-6 3 out of 4 tested putative gata6 enhancers showed heart-specific activity in vivo

ATAC-Seq datasets at the *gata6* locus from sorted Myl7-positive CMs at 5dpf and 3 and 5days post-cryoinjury and p300 ChIP data sets from sorted CMs at 5dpf. Blue bars indicate selected putative enhancers. Microscopic analysis of F0 embryos at 3dpf revealed heart-specific activity of each of the putative enhancers for *gata6*, with the exception of Gata6-E12, which showed no fluorescence. In the lower panels, black arrows indicate activity of the enhancer-driven reporter. Confocal images show a ventral view of the enhancer-driven reporter expression in the heart at 3dpf. a=atrium, v=ventricle.

3.2.5. Gata5-E5 labels a subset of cells in and adjacent to the inflow tract of the developing heart.

In order to verify the heart-specific activity shown by the putative GATA enhancers in the transient transgenic embryos, I sought to establish stable transgenic lines for those enhancers active in the heart in the F0 generation. After reaching sexual maturity, F0 embryos that were positive for the reporter transgene were outcrossed and their offspring screened for germline transmission of the enhancer-driven reporter activity. Once I identified a founder, I was then able to cross this fish with other available transgenic zebrafish expressing fluorophores reporting on different regions of the heart, specifically, the CM population (driven by the *myl7* promoter) and the endothelial/endocardial population (driven by the *kdrl* promoter). Figure 3-7a shows that the Gata5-E5 enhancer-driven reporter was heart-specific and was active in a population of CMs in the inflow region of the embryonic heart of 5dpf F1 embryos. When crossed with a *myl7*-reporter line and imaged at 48hpf, Gata5-E5-driven reporter expression overlapped with that of Myl7 in a subpopulation of CMs, particularly concentrated around the inflow region of the heart. Gata5-E5 activity did not overlap with *kdrl*-reporter expression (Figure 3-7b), suggesting that the enhancer was not active in the developing endocardium. Ideally, the offspring of multiple founders would be analysed, to prove that there is no variation due to the

mosaic integration of the transgene. However, all of the embryos analysed from the founder showed a consistent and reproducible enhancer-driven reporter expression pattern. I further explore the biological relevance behind this enhancer in chapter 4.

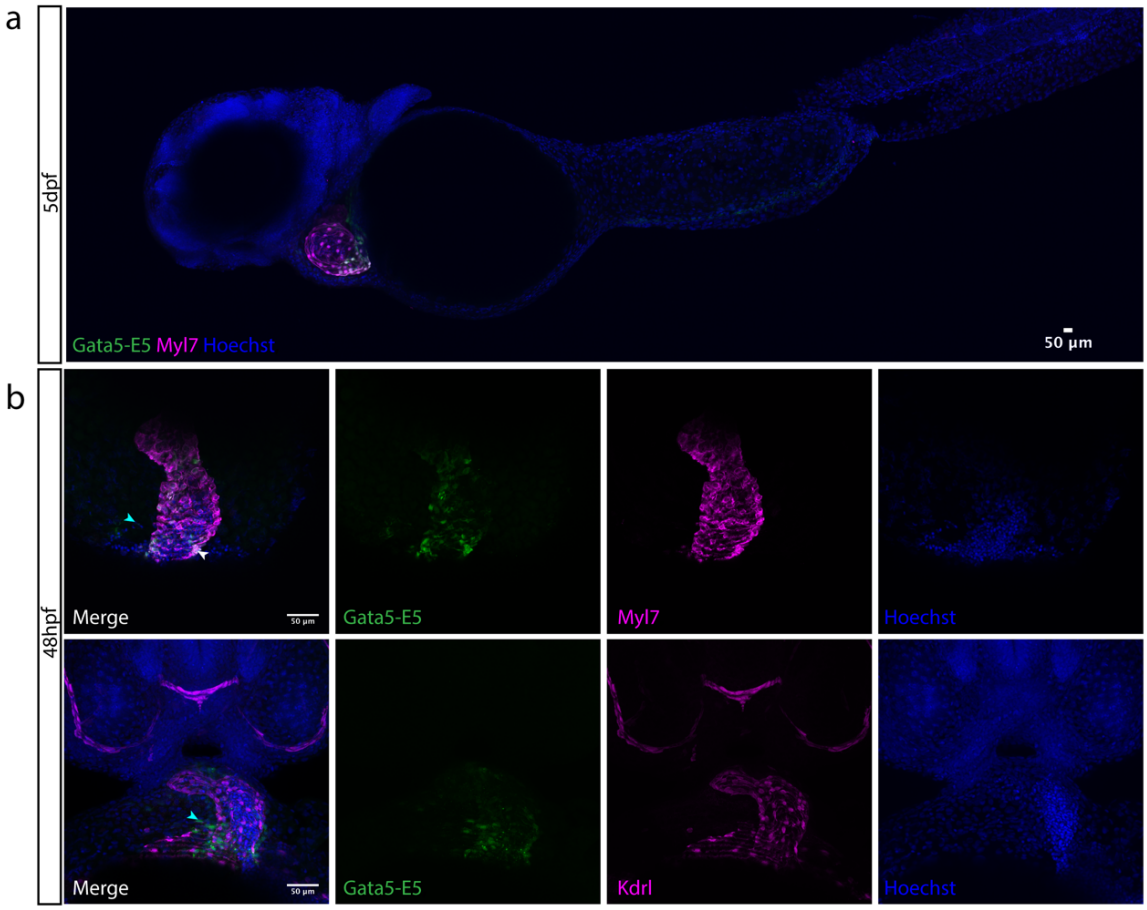


Figure 3-7 In the stable transgenic line, Gata5-E5-driven reporter shows expression in a subpopulation of cells in the inflow region

a) Microscopic analysis of Gata5-E5-mCherry (pseudocoloured green) expression in the Myl7-GFP (pseudocoloured magenta) line at 5 dpf. Expression of mCherry overlapped with Myl7-GFP in a subpopulation of cells within the inflow tract of the developing heart. (17 embryos analysed) b and c) Microscopic analysis at 48hpf in Tg(*Gata5-E5-mCherry;myl7-GFP*) (10 embryos analysed) (b) and *Tg(Gata5-E5-mCherry;kdrl-GFP)* (10 embryos analysed) (c) Enhancer-driven reporter expression showed overlap with Myl7 expression in the inflow tract, (depicted by white arrowhead), but not with the endothelial marker, Kdrl, suggesting that the enhancer is active in CMs at the venous pole of the heart, but not in the endocardium. Gata5-E5 was also active in a population of Myl7-, Kdrl-negative cells adjacent to the inflow tract (depicted by cyan arrowhead).

3.2.6 Gata5-E6 labels a subset of cells in and adjacent to the inflow tract of the developing heart at early stages that becomes restricted to the ventricle and AVC by 5dpf

In the F1 generation, Gata5-E6-driven reporter expression overlapped with the expression of Myl7-positive CMs in the developing heart tube at 24hpf (Figure 3-8a). There was also a population of Myl-negative cells surrounding the heart tube that showed enhancer activity. At 48hpf (Figure 3-8b) and 3dpf (Figure 3-8c), the Gata5-E6-driven reporter expression pattern greatly resembled that of Gata5-E5 as it was active within a population of CMs in the inflow tract region of the atrium, as well as a small population of Myl7-negative cells just adjacent to the inflow tract. Interestingly, Gata5-E6 appeared to become active in a different population of cells from 4dpf (Figure 3-8d), as reporter expression was seen in the bulbus arteriosus and ventricle at 4dpf. As the enhancer-driven reporter expression did not overlap with Myl7 reporter expression in the ventricle, and reporter-positive cells appeared to be inside the ventricle, it is likely that the Gata5-E6-driven reporter was expressed in the endocardium at this time point. By 5dpf (Figure 3-8e), the expression of the Gata5-E6-driven reporter was restricted to the AVC and what is predicted to be the ventricular endocardium.

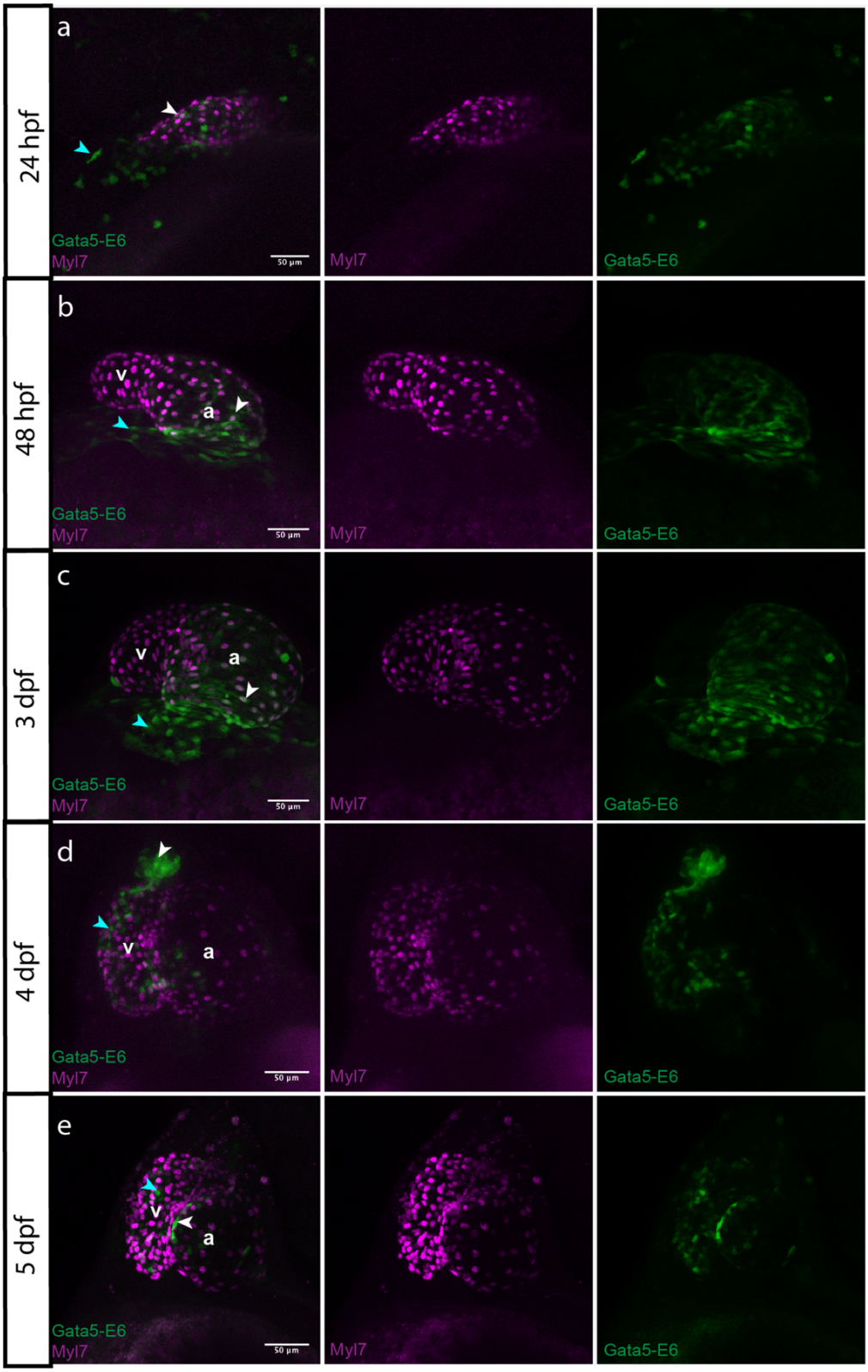

24 hpf
48 hpf
3 dpf
4 dpf
5 dpf
a
b
c
d
e
v
a
Gata5-E6
Myl7
50 µm
Myl7
Gata5-E6

Figure 3-8 In the stable transgenic line, Gata5-E6-driven reporter shows expression in a subpopulation of cells in the inflow region early on that becomes restricted to the ventricle and AVC by 5dpf

a) Microscopic analysis of Gata5-E6-mCherry (pseudocoloured green) expression in the Myl7-GFP (pseudocoloured magenta) background at 24hpf showed colocalisation of expression in some cells of the early heart tube (white arrowhead) and some Myl7-negative cells in the surrounding region (cyan arrowhead) (10 embryos analysed). b and c) At 48hpf (9 embryos analysed) and 3dpf (10 embryos analysed), Gata5-E6-driven reporter was seen in a population of Myl7-positive cells in the inflow tract (white arrowhead) and a population of Myl7-negative cells adjacent to the inflow tract (cyan arrowhead). d) by 4dpf, expression of the enhancer-driven reporter could be seen in the bulbus arteriosus (white arrowhead) and in a Myl7-negative population inside the ventricle, believed to be the endocardium (cyan arrowhead) (8 embryos analysed). At 5dpf, enhancer-driven reporter expression was seen in the AVC and the ventricular endocardium (10 embryos analysed).

3.2.7 Gata5-E8 activity in the F1 generation did not reflect the heart-specific patterns observed in the transient F0 analysis

In the Gata5-E8 F1 embryos, expression of the enhancer-driven reporter did not overlap with the Myl7-positive CMs (Figure 3-9a), but its expression mirrored that of Kdrl (Figure 3-9b), appearing in the endocardium as well as in the blood vessels throughout the embryo. This was a puzzling finding as a) no endocardial or blood vessel expression of enhancer-driven reporter was seen in the F0 injected embryos and b) Gata5 is not expressed in the blood vessels. It is of course possible that Gata5-E8 is not an enhancer regulating *gata5* expression and therefore is active in a different cell type to the native gene, but this still begs the question of why this additional reporter activity in blood vessels was not seen in the F0 embryos. This led me to speculate that the enhancer construct may have integrated into a region proximal to another regulatory element/active gene, which ‘hijacked’ its regulation causing ectopic expression of the fluorescent reporter.

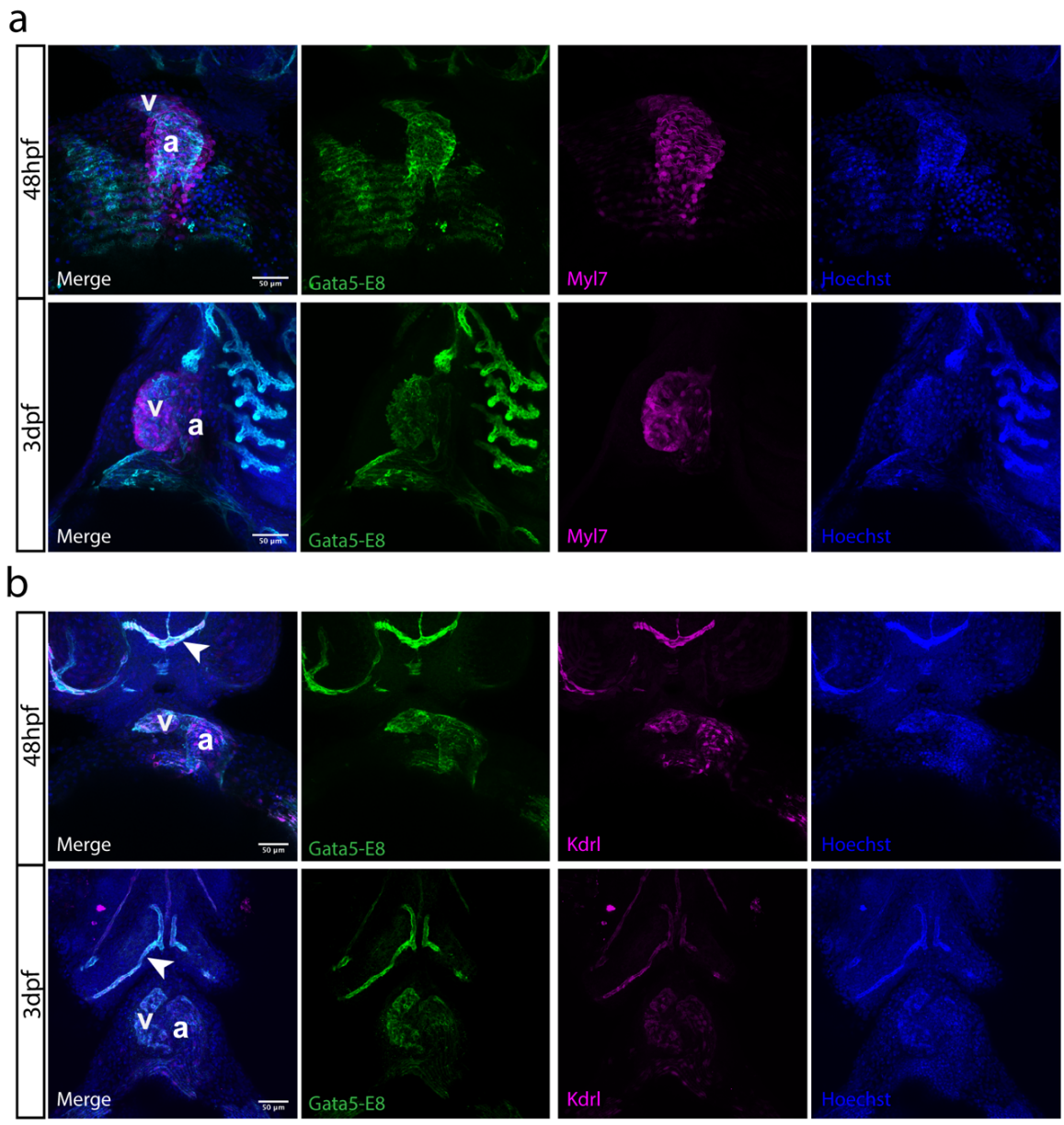


Figure 3-9 Gata5-E8 reporter expression in the stable transgenic line did not reflect what was observed in the F0 generation

Microscopic analysis of Gata5-E8-mCherry (pseudocoloured green) expression in the Myl7-GFP (pseudocoloured magenta) line at 48hpf (12 embryos analysed) and 3dpf (10 embryos analysed) showed no overlap, indicating that the reporter was not active in the developing myocardium. b) At 48hpf (10 embryos analysed) and 3dpf (10 embryos analysed), Tg(Gata5-E8-mCherry;kdrl-GFP) embryos showed almost perfect overlap between the enhancer-driven reporter and Kdrl in the endocardium and the head vessels (white arrowhead). v= ventricle, a= atrium.

3.2.8 Gata6-E12 labels a population of CMs that becomes restricted to the developing ventricle.

In all embryos studied, the Gata6-E12-driven reporter was expressed in a subpopulation of CMs, as shown by the overlap with the Tg(*myl7-GFP)* reporter in Figure 3-10. Unlike with Gata5-E5, I identified 4 separate founders for this enhancer and the offspring of all showed a consistent expression pattern, demonstrating the reproducibility of this approach. Interestingly, although early on at 48hpf, Gata6-E12 was active in cells of both chambers of the heart, this activity became restricted to the ventricle by 5dpf, despite the fact that *gata6* is expressed throughout the whole heart (Figure 3-2). I further explore the biological relevance behind this chamber-specific enhancer in chapter 4.

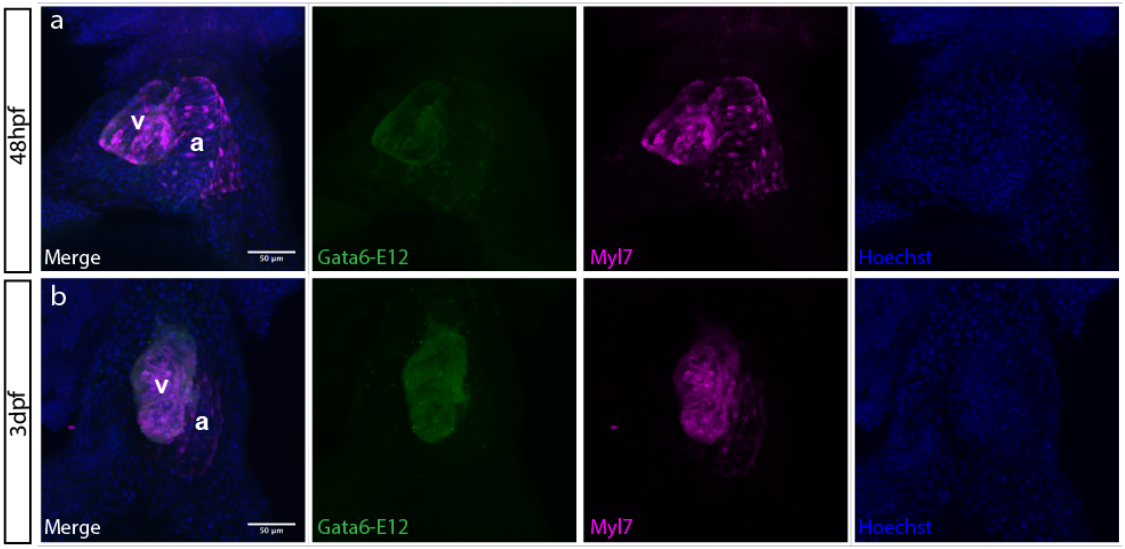


Figure 3-10 In the stable transgenic line, Gata6-E12-driven reporter shows expression in a subpopulation of CMs that becomes restricted to the ventricle.

a) Microscopic analysis of *Tg(Gata6-E12-mCherry;myl7-GFP)* embryos at 48hpf (13 embryos analysed) showed overlap between the expression driven by the two reporters in the ventricle as well as some cells of the atrium. The expression of Gata6-E12-driven reporter became restricted to the ventricle by 3dpf (11 embryos analysed) (b). v= ventricle, a= atrium.

3.3 Summary

Using ATAC-Seq and p300 ChIP datasets to inform on the open chromatin regions' positions, I have identified 4 putative enhancer regions around each of the cardiac GATA loci in zebrafish. In order to confirm whether these potential regulatory elements were active *in vivo*, I cloned the sequences upstream of a minimal promoter in reporter vectors for the highly efficient Ac/Ds system and injected these into embryos at the one cell stage. I was then able to track the activity of each enhancer throughout early cardiogenesis in the F0 generation.

None of the putative enhancers selected for *gata4* showed activity *in vivo*. Future experiments would include revisiting the sequencing datasets to select further putative *gata4* enhancers to test.

All putative enhancers tested for *gata5* showed activity *in vivo*. Despite the fact that *gata5* is expressed throughout the developing heart, it was already clear in the transient F0 embryos that the putative *gata5* enhancers were active in subpopulations representing only part of the cardiac *gata5* expression domain. A common issue with studying enhancer-driven reporter expression in a transient manner is the high degree of mosaicism observed in the injected F0 embryos. To combat this problem and to fully characterise the spatiotemporal activity of the putative enhancers I identified in this study, I generated stable lines expressing the enhancer-driven reporters. I observed that Gata5-E5 was active in a population of CMs in the inflow region of the heart. Gata5-E6 was active in a heterogeneous population of cells; in the inflow tract CMs and some extracardiac cells adjacent to the inflow tract early on and then in the ventricular endocardium, AVC and bulbus arteriosus at later stages. The enhancer-driven reporter activity for Gata5-E8 curiously mimicked the expression pattern of the endothelial cell marker Kdrl,

suggesting that the reporter construct may have integrated into a region where it became regulated by another regulatory element, or in a region where it could be 'hijacked' by other TFs. This highlights an important point that it would be prudent to test the integration sites of all enhancer-driven reporters in these lines. Integration sites could be performed using splinkerette PCR (Uren et al., 2009).

Although *gata6* is expressed throughout the developing heart, the Gata6-E12 enhancer was only expressed in a subpopulation of the *gata6* expression domain. Early in development, at 48hpf, Gata6-E12-mCherry was expressed in CMs throughout the atrium and ventricle, but by 3dpf its expression had become restricted to the ventricle. In the course of this project, I did not manage to make a stable transgenic line for the remaining *gata6* putative enhancers, but in future work it would be of interest to characterise the remaining enhancers and determine whether together, they are active in regions that make up the complete *gata6* domain of expression.

I have then selected one putative enhancer for *gata5* (Gata5-E5) and one for *gata6* (Gata6-E12) to study in further detail, as described in chapter 4.

4) Characterisation of the novel GATA enhancers Gata5-E5 and Gata6-E12

4.1 Introduction

Initial imaging of Gata5-E5 showed that this enhancer was active in a subpopulation of the *gata5*-expressing CMs in the inflow tract of the heart and an extracardiac population of cells that lay adjacent to the inflow tract during early development.

This provided a tool for an in-depth analysis of *gata5* regulation within specific heterogeneous subpopulations of the developing heart.

Despite the fact that *gata6* is highly expressed throughout the developing heart (Figure 3-2), Gata6-E12 activity was shown to become restricted to the ventricle by 3dpf. This is of particular interest as it can provide a tool for specifically probing the regulation of *gata6* in ventricular CMs and may highlight previously unknown differences between the regulatory dynamics between the developing atrium and ventricle. As yet, nothing is known about *gata6*-specific regulatory activity in such a particular subpopulation of the heart.

The fact that these two novel enhancer-reporter lines showed such specific expression patterns makes them extremely useful tools for investigating the upstream regulation of *gata5* and *gata6* throughout development in such specific subpopulations. Understanding of both the spatiotemporal dynamics of the activity of Gata5-E5 and Gata6-E12 and the TF-enhancer interactions that contribute to their activity has the potential to reveal the upstream regulation of the GATA factors in detail.

In this chapter, I characterise the enhancer-driven expression pattern of Gata5-E5-mCherry and Gata6-E12-mCherry transgenic lines in the developing and adult heart, to further understand which and how the specific cell populations in which *gata5* and *gata6,* respectively, are regulated. I also investigate the phenotypic consequences of deleting each of these two enhancers, in order to further understand their biological function and importance in the developing embryo.

4.2 Results

4.2.1 Tg*(Gata5-E5 mCherry)* was expressed in a subset of *gata5*-expressing cells

The CM-specific ATAC and ChIP-Seq datasets used as the basis for the selection of the *gata5* putative enhancers were generated from 5dpf embryos. As such, it was to be expected that, if Gata5-E5 were to be active *in vivo*, enhancer-driven reporter expression should be observed in CMs at the same time-point of development. However, since this is a novel putative enhancer, I decided to carefully analyse its activity across various stages of cardiogenesis. In all embryos studied, Gata5-E5 was active by the 6 somite stage in the ALPM, a time and place where CMs are being specified in the early embryo (Stainier et al., 1993) (Figure 4-1a). Although we know that *gata5* is of critical importance during these early stages of cardiogenesis, little is known about the upstream regulation of these very first steps leading to CM specification. By 24hpf, Gata5-E5-driven reporter expression could be seen throughout the early heart tube (Figure 4-1b). It then became restricted to specific subpopulations of cells by 48hpf (Figure 4-1c). One population, which was Myl7-negative, was lying just adjacent to the inflow tract and another one, which was Myl7-positive, was located within the inflow region. Cells within the AVC and some ventricular CMs also reported activity for Gata5-E5. The enhancer-driven mCherry reporter expression became restricted to the inflow tract, AVC and some ventricular CMs from 3-5dpf (Figure 4-1d-f).

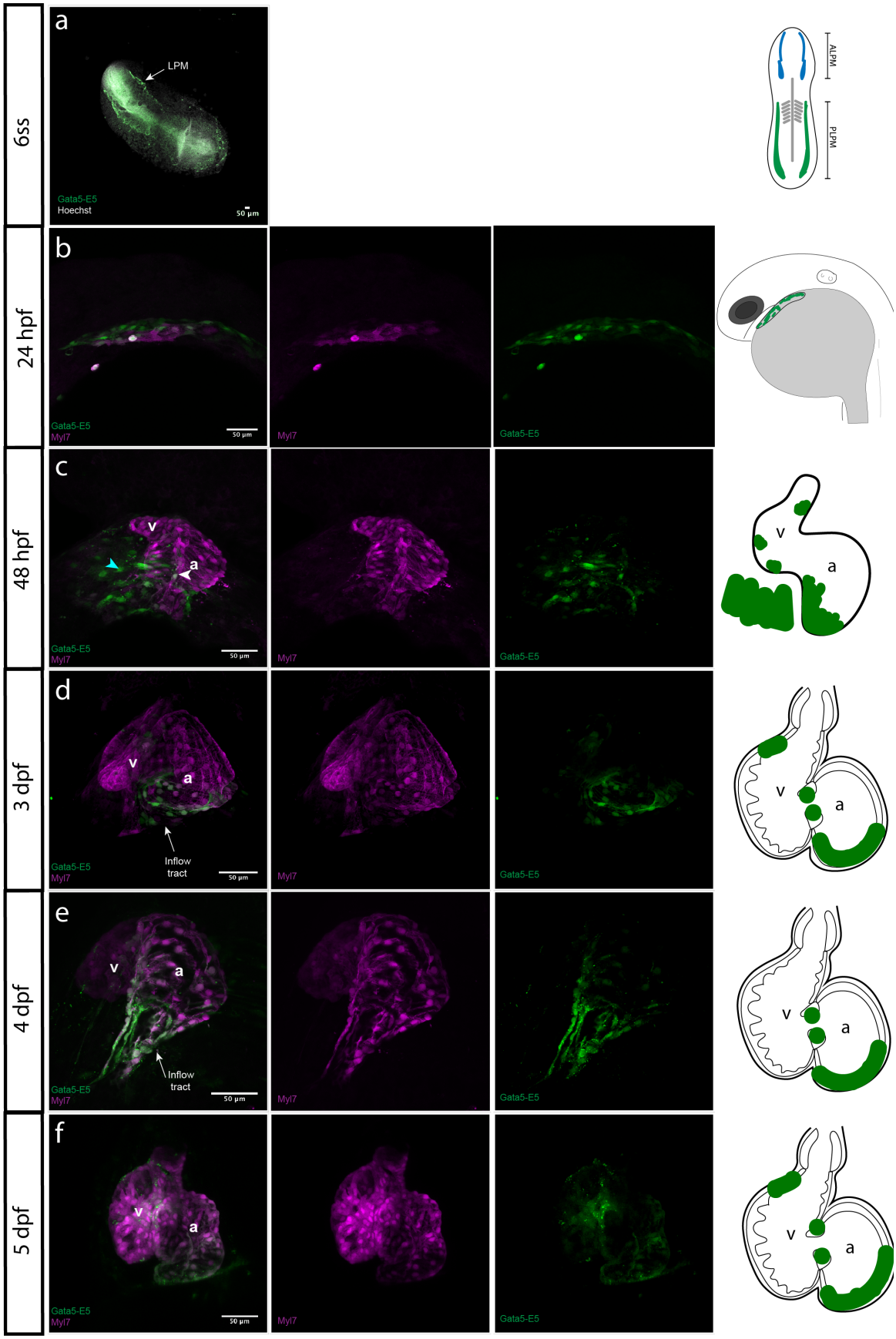

6ss
24 hpf
48 hpf
3 dpf
4 dpf
5 dpf
a
LPM
Gata5-E5
Hoechst
50 µm
ALPM
PLPM
b
Gata5-E5
Myl7
50 µm
Myl7
Gata5-E5
c
v
a
Gata5-E5
Myl7
50 µm
Myl7
Gata5-E5
v
a
d
v
a
Inflow
tract
Gata5-E5
Myl7
50 µm
Myl7
Gata5-E5
v
a
e
v
a
Inflow
tract
Gata5-E5
Myl7
50 µm
Myl7
Gata5-E5
v
a
f
v
a
Gata5-E5
Myl7
50 µm
Myl7
Gata5-E5
v
a

Figure 4-1 In the F2 generation of the stable transgenic line, Tg(Gata5-E5-mCherry), Gata5-E5 was active in specific subpopulations of the developing heart.

a) Microscopic analysis of a 6ss flat-mount *Tg(Gata5-E5-mCherry;myl7-GFP)* embryo showed expression of the Gata5-E5-driven reporter in cells of the lateral plate mesoderm. Schematic represents a dorsal view of a flat-mounted 6ss embryo, with the ALPM and PLPM labelled (12 embryos analysed) b) At 24hpf, Gata5-E5-mCherry was expressed in Myl7+ CMs throughout the early heart tube. Schematic shows a side view of a 24hpf embryo, with the Gata5-E5-mCherry cells represented in green. (15 embryos analysed) c) At 48hpf, Gata5-E5-mCherry was expressed in Myl7+ CMs in the inflow tract (white arrowhead) as well as a population of Myl7- cells just adjacent to the inflow tract (cyan arrowhead). The reporter was also expressed within the AVC and in certain cells within the ventricle. Schematic shows a ventral view of a 48hpf heart, with the Gata5-E5-mCherry cells represented in green.(100 embryos analysed) c-f) From 3-5dpf, Gata5-E5-mCherry was expressed in Myl7+ CMs in the inflow tract, the AVC and in certain cells within the ventricle (12, 15 and 20 embryos analysed, respectively). Schematics show ventral views of the heart, with the Gata5-E5-mCherry cells represented in green. a=atrium, v=ventricle.

4.2.2 Tg*(Gata6-E12-mCherry)* reported expression throughout the early heart tube, becoming restricted to the developing ventricle by 3dpf

To observe the activity of Gata6-E12 in CMs *in vivo*, Tg*(Gata6-E12-mCherry)* F1 transgenic zebrafish were crossed with the Tg(*myl7-GFP)* reporter line and mCherry:GFP double positive embryos were imaged each day for the first 5 days of development. Despite *gata6* being expressed throughout the developing heart, Gata6-E12 was only active in a subpopulation of the *gata6* expression domain. At 24hpf Gata6-E12-driven reporter expression showed a ‘salt and pepper’ distribution throughout the CMs of the early heart tube (Figure 4-2a). At 48hpf, Gata6-E12 was active in CMs of both the atrium and ventricle (Figure 4-2b). This expression became restricted to the ventricular CMs by 3dpf.

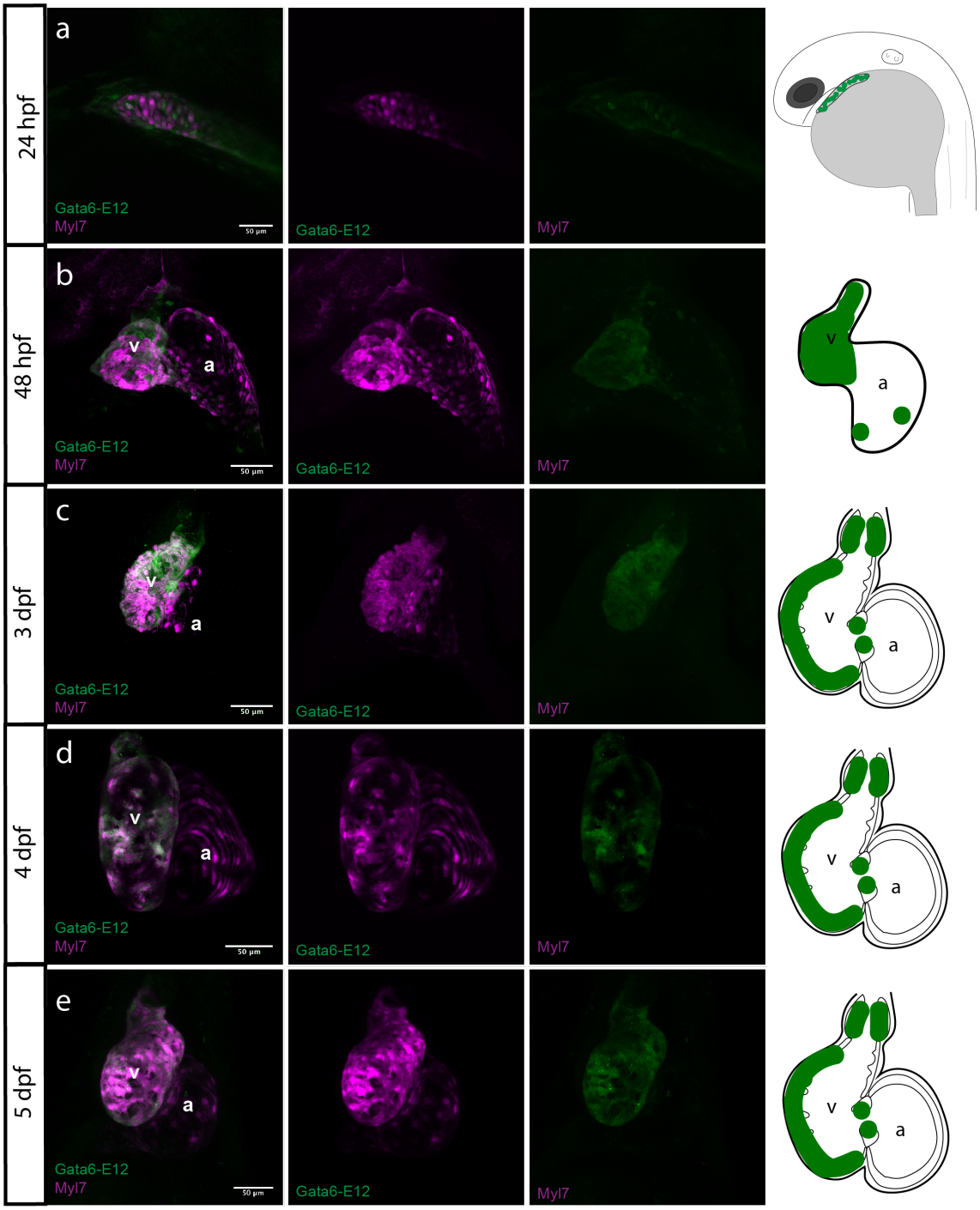

24 hpf
48 hpf
3 dpf
4 dpf
5 dpf
a
b
c
d
e
Gata6-E12
Myl7
50 µm
v
a

Figure 4-2 Tg(Gata6-E12-mCherry) expression became restricted to the ventricular CMs and bulbus arteriosus by 3dpf

a) Microscopic analysis of *Tg(Gata6-E12-mCherry;myl7-GFP)* embryos at 24hpf showed expression of the Gata6-E12-driven reporter in CMs throughout the early heart tube. Schematic shows a side view of a 24hpf embryo, with the Gata6-E12-mCherry cells represented in green (12 embryos analysed) b) At 48hpf, Gata6-E12-mCherry was expressed in myl7+ CMs predominantly in the ventricle but also in some atrial CMs. Schematic shows a ventral view of a 48hpf heart, with the Gata6-E12-mCherry cells represented in green. (10 embryos analysed) c-e) From 3dpf, Gata6-E12 expression was restricted to the ventricle and bulbus arteriosus. (15, 17 and 12 embryos analysed, respectively). Schematics show ventral views of the heart, with the Gata6-E12-mCherry cells represented in green. a=atrium, v=ventricle.

4.2.3 Hybridisation chain reaction showed that Gata5-E5 was active in the same cells as endogenous *gata5* was expressed

As it is known that proximity of an enhancer to a given promoter does not necessarily confer a link, simply being its 'nearest neighbour' does not guarantee that Gata5-E5 is regulating the activity of *gata5*. It was therefore important to confirm whether Gata5-E5 was active within the same cells as *gata5* endogenous expression. For this I used hybridisation chain reaction (HCR), an *in situ* staining technique that enables the simultaneous analysis of multiple gene expression patterns (Choi et al., 2018). This technique showed colocalisation of Gata5-E5-driven *mCherry* and *gata5* transcripts (Figure 4-3). One technical limitation of this method is that, as yet, we have no successful mode of labelling the cell membranes, therefore colocalisation within a cell was judged by proximity to the nucleus, marked by Hoechst staining. Although coexpression of Gata5-E5-driven *mCherry* and endogenous *gata5* transcripts does not definitively confirm that Gata5-E5 is a *gata5* enhancer, it strongly supports this idea.

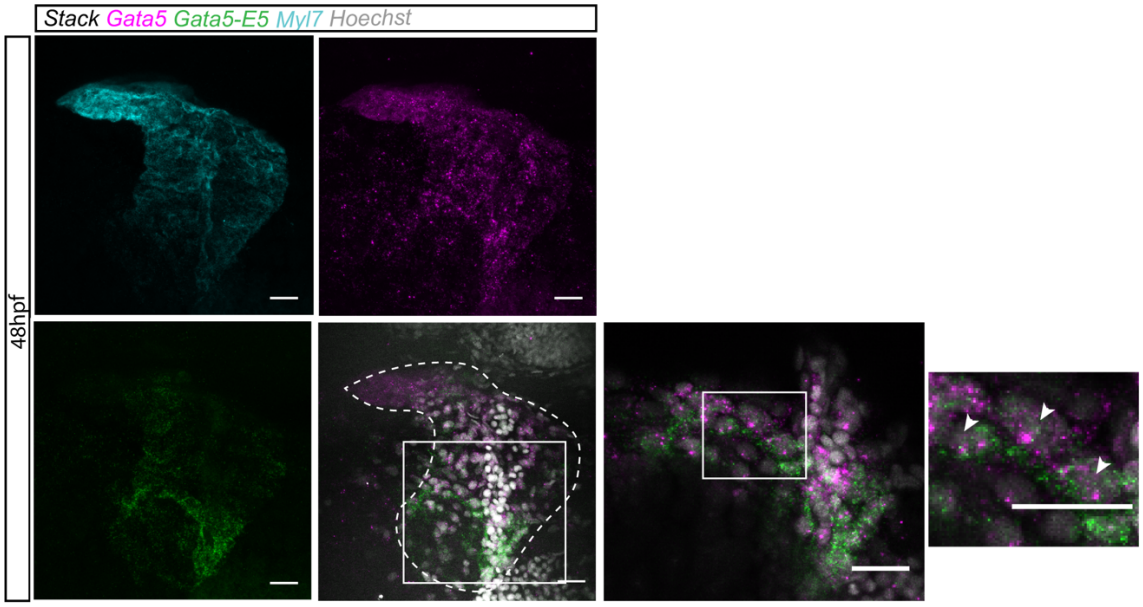


Figure 4-3 HCR confirmed that Gata5-E5-mCherry was expressed in the same cells as endogenous gata5.

Microscopic analysis of HCR in Tg*(Gata5-E5-mCherry)* embryos at 48hpf. Individual optical sections showed colocalisation of *mCherry* and *gata5* transcripts in the inflow tract of the developing heart (white arrowheads), based on proximity to a common nucleus (nuclei in grey). White boxes indicate the region zoomed into in subsequent images. Scale bars= 50µm. (15 embryos analysed)

4.2.4 Hybridisation chain reaction revealed that Gata6-E12 was active in the same cells as endogenous *gata6* was expressed

HCR was used to investigate whether Gata6-E12-*mCherry* reporter transcripts and *gata6* transcripts colocalise in the developing heart. Indeed, enhancer-driven reporter transcripts were present in the same cells as *gata6* transcripts at 48hpf and 5dpf (Figure 4-4), which strongly supports the hypothesis that Gata6-E12 is a *gata6* enhancer. This analysis also highlighted the fact that *gata6* was expressed throughout the developing heart but Gata6-E12 was only active in a subpopulation of cells within the *gata6* expression domain.

Therefore, *Tg(Gata6-E12-mCherry)* provides an extremely useful tool to dissect the differential regulation of such an important gene during cardiogenesis, focusing in a specific subpopulation in the ventricle.

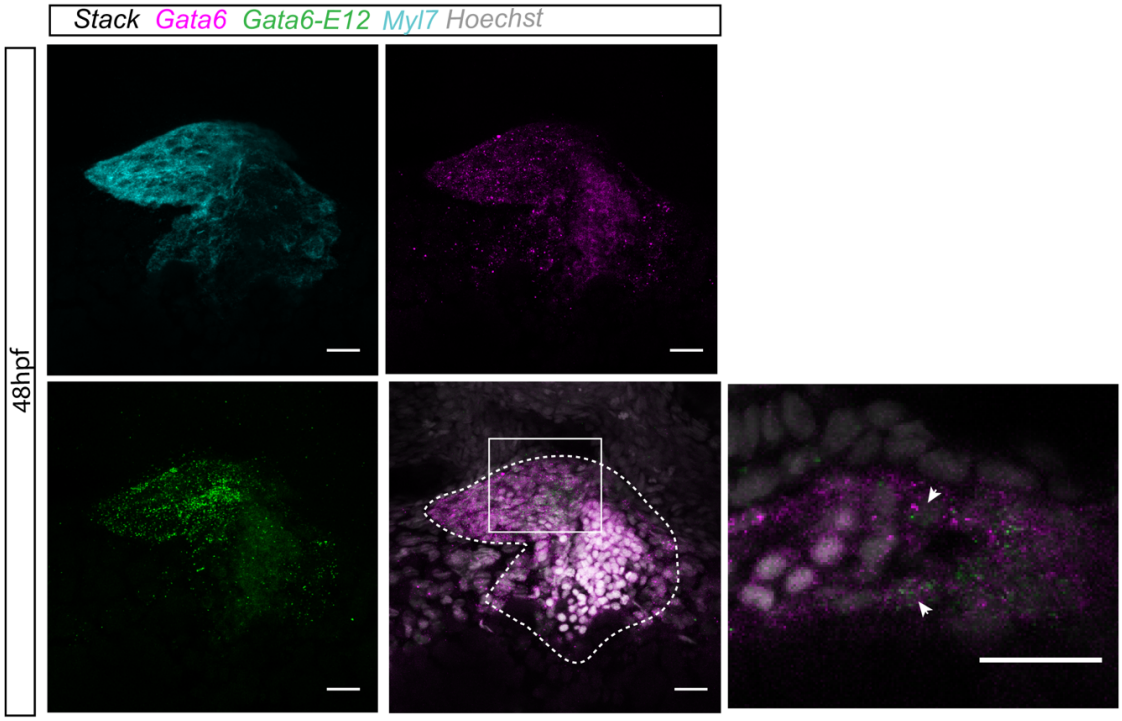


Figure 4-4 HCR showed that Gata6-E12-mCherry was expressed in the same cells as endogenous gata6 transcripts.

Microscopic analysis of HCR in Tg(*Gata6-E12-mCherry*) embryos at 48hpf. Individual optical sections showed colocalisation of *mCherry* and *gata6* transcripts in the ventricle of the developing heart (white arrowheads), based on proximity to a common nucleus (nuclei in grey). White boxes indicate the region zoomed into in the subsequent image. Scale bars= 50μm. (15 embryos analysed)

4.2.5 Gata5-E5 was also active in the adult heart and its activity increased around the injury border zone following cryoinjury

It has been postulated that adult heart regeneration in part recapitulates the mechanisms acting during cardiogenesis (Alexander and Bruneau, 2010). It was therefore of interest to know whether Gata5-E5 may be activated in the adult heart after injury. As Gata5-E5 was selected on the basis that there was an ATAC-Seq peak at this region in the injured adult CMs as well as embryonic, I predicted that

the enhancer would be active *in vivo* in the adult heart. In order to characterise the response of the enhancer to cardiac injury, I performed a cryoinjury to mimic a cardiac infarct (Chablais et al., 2011), as described in Chapter 1, in Tg*(Gata5-E5-mCherry;myl7-GFP)* zebrafish. I then collected hearts at 3 days post-injury (3dpi), the same timepoint used to collect ATAC-Seq data, and then performed antibody staining against mCherry to compare the expression of the enhancer-driven reporter in cryoinjured and control sham-operated hearts.

Sham-operated hearts, where the open hearts have been touched with a cryoprobe at room temperature (not dipped into liquid nitrogen), and therefore not injured, showed activity of Gata5-E5 in the atrial epicardium, in small regions of the endocardium and in a small number of cells within the ventricular myocardium (Figure 4-5). Interestingly, the expression pattern of the enhancer-driven reporter in the adult heart was markedly different to that observed the embryo, where, of all the cardiac cell layers, it was only active in the developing myocardium. In the cryoinjured heart, Gata5-E5-driven reporter expression appeared to be more widely spread throughout the endocardium and in the cells of the ventricular myocardium in the injury-proximal region (Figure 4-5). This area has been shown to be the site of proliferating CMs (Kikuchi et al., 2011) (Jopling et al., 2010), suggesting that Gata5-E5 might play a role in the dedifferentiating CMs that are the main, if not sole, contributors to the regenerating heart. It would be important to use a proliferation assay to confirm whether the Gata5-E5 enhancer is indeed active in the proliferating CMs.

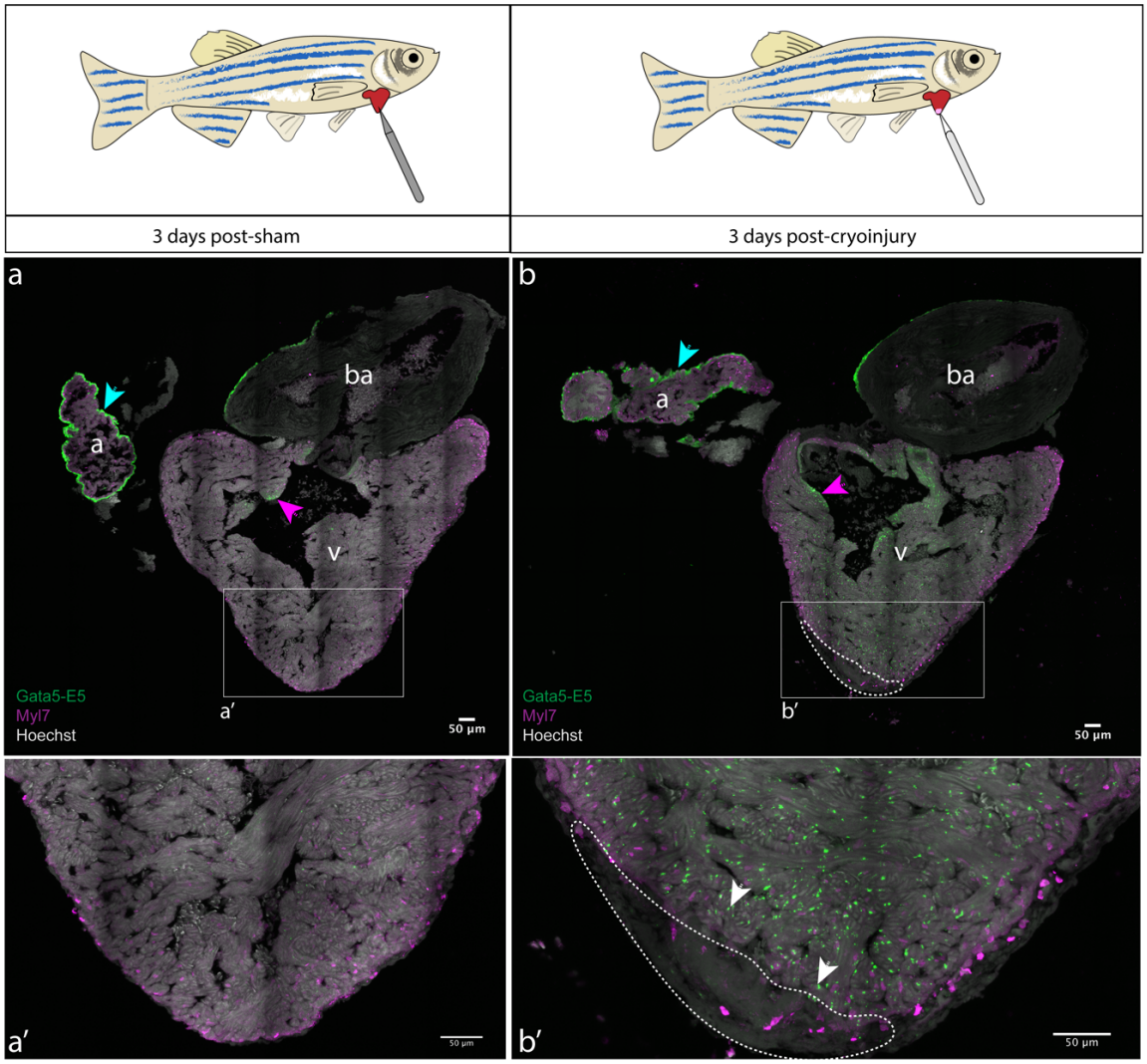


Figure 4-5 Gata5-E5 was active in the adult heart and is activated in the injury-proximal region following cryoinjury

a) In sham-operated hearts, Gata5-E5-driven reporter was expressed in the atrial epicardium (cyan arrowhead), and in some cells in the endocardium (magenta arrowhead). (a') shows the ventricular apex in higher magnification. b) After cryoinjury, expression of the Gata5-E5-driven reporter appeared to increase in the endocardium (magenta arrowhead) and in the CMs of the injury border zone (white arrowheads) and was maintained in the atrial epicardium (cyan arrowhead). a' shows the ventricular apex in higher magnification. a= atrium, v=ventricle, ba= bulbus arteriosus. (n=3)

4.2.6 Gata6-E12 was active in the adult heart, in the same cell population as it was active in the embryo

Unlike what was observed with the Gata5-E5 enhancer, Gata6-E12 maintained its activity in the same population of cells in the adult as in the embryo. The enhancer-driven reporter was expressed strongly and uniformly throughout the CMs of the ventricular myocardium (Figure 4-6). Upon cryoinjury, Gata6-E12-driven reporter expression increased in the injury border zone at 3dpi, indicating increased enhancer activity in proliferating CMs.

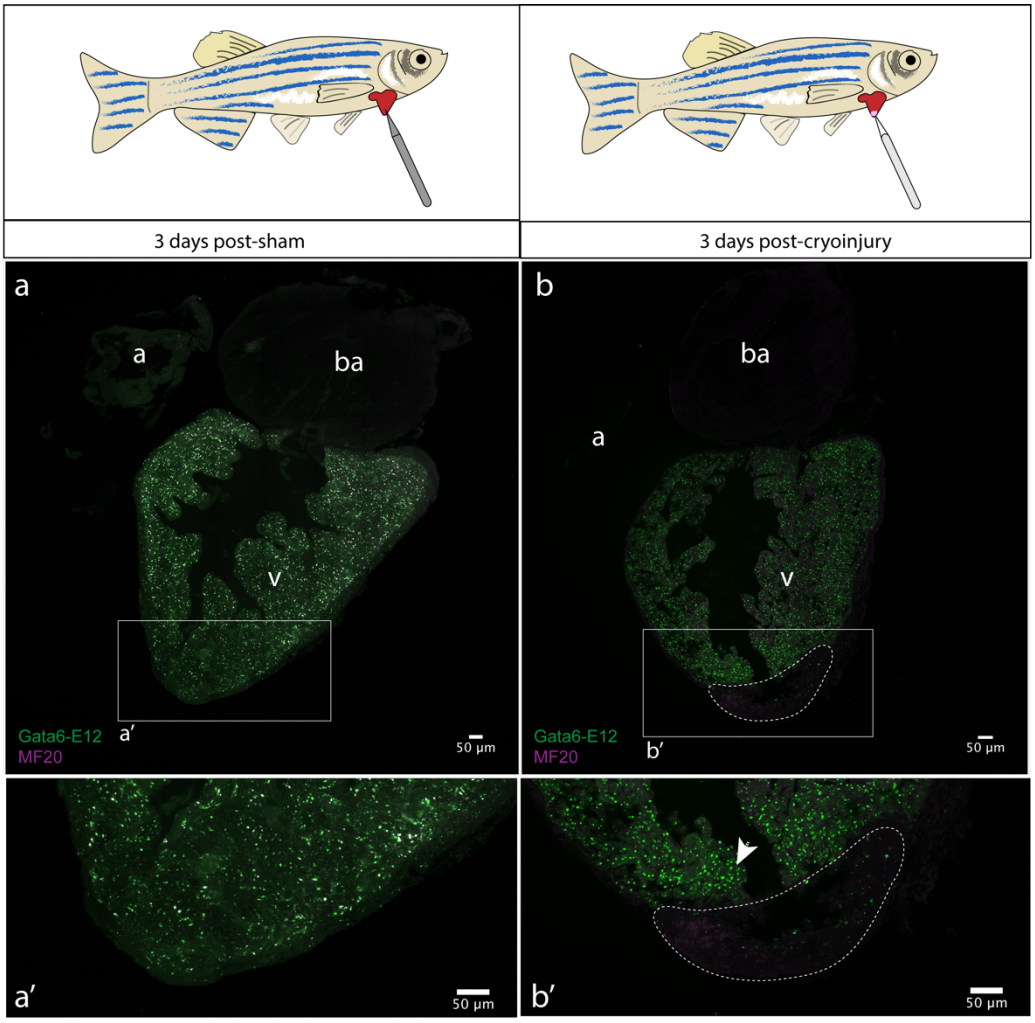


Figure 4-6 Gata6-E12 was active in the ventricle of the adult heart

a) Like in the embryo, Gata6-E12 was active throughout the CMs of the ventricle. a' shows the ventricular apex in higher magnification. b) After cryoinjury, Gata6-E12-driven reporter expression increased at the injury border zone (white arrowhead). (b') shows the ventricular apex in higher magnification. MF20 is a CM marker and the signal using the antibody was weak. This should be repeated using the Tg(*myl7-GFP*) line instead. a= atrium, v=ventricle, ba= bulbus arteriosus. (n=3)

4.2.7 Gata5-E5 KO reduced *gata5* expression and enhancer activity in the inflow tract-adjacent cell population

In order to understand the function of the Gata5-E5 putative enhancer *in vivo*, it was important to first investigate the consequences of its deletion. To address this, I used CRISPR-Cas9 technology. Since its widespread publication in 2012 (Jinek et al., 2012), the system based around the prokaryotic clustered, regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated proteins (Cas) has revolutionised genome editing in many model systems. This technology has been adapted from a prevalent bacterial defence mechanism that provides acquired resistance to invading pathogens. The type II CRISPR/Cas system involves the incorporation of short foreign DNA sequences into CRISPR loci in the bacterial genome. These are then transcribed into short CRISPR RNAs (crRNAs), which, when annealed to trans-activating crRNA (tracRNA), direct the Cas9 endonuclease protein to introduce a sequence-specific double-stranded break in foreign nucleic acids. This technology has been exploited and streamlined for use in model organisms, and now requires a CRISPR complex comprising just two components;

a single synthetic guide RNA (sgRNA) composed of a fusion of a target-specific 20bp crRNA and a tracRNA that can be delivered as an *in vitro* transcript or a DNA plasmid, and a Cas9 endonuclease, that can be delivered as an *in vitro* transcript, a DNA plasmid or as a recombinant protein (Shah et al., 2015). Once inside the cell, the CRISPR complex searches the genome for a protospacer adjacent motif (PAM) with the sequence 5'-NGG, where it unzips the DNA, enabling the binding of the sgRNA to its target region. The Cas9 endonuclease then cleaves a double-stranded break that is typically repaired by non-homologous end joining.

I aimed to exploit this system in order to efficiently excise the target enhancer. It has previously been shown that, due to the high efficiency of somatic mutation, it is possible to already detect specific phenotypes in injected F0 embryos (Shah et al., 2015). In order to allow detailed downstream analysis, sgRNAs were designed to target sites just outside, and therefore not present in, the putative enhancer region that was cloned into the enhancer-reporter construct. This approach led to excision of the endogenous enhancer, while maintaining the enhancer-driven reporter activity, allowing me to continue to track the specific target populations where the enhancer is active. sgRNAs were designed to target the 3' and 5' ends of the enhancer and were tested by high resolution melt analysis (HRMA). Selected sgRNAs were then injected in combination with Cas9 mRNA into embryos at the one-cell stage. Because of the importance of *gata5* activity during early cardiogenesis, its regulation is likely very robust. As a mechanism to protect developmental gene expression patterns from external influences, functional redundancy often exists within these transcriptional networks (Hong et al., 2008). Most developmental genes are regulated by multiple enhancers, and often each enhancer controls a specific spatiotemporal feature of the expression of a given

gene. Some genes will also have multiple enhancers with overlapping functions, driving transcription in a similar spatiotemporal pattern; a primary enhancer close to the TSS and a secondary, more distant enhancer, known as a 'shadow enhancer' (Hong et al., 2008, Frankel et al., 2010). According to the definition of enhancer redundancy, in these circumstances, the loss of one enhancer module has no discernible phenotypic effect. An example of this is seen in the mouse, where targeted deletion of an enhancer of the *Prx1* gene that has been shown to be sufficient to drive limb expression showed no developmental limb phenotype (Cretekos et al., 2008). It is therefore highly possible that there is a degree of enhancer redundancy within the GATA family, raising the possibility that no phenotype would be observed upon enhancer excision. However, this would not necessarily mean that the Gata5-E5 enhancer does not have an important regulatory role in *gata5* regulation during cardiogenesis.

To observe the effects of Gata5-E5 knockout, I analysed sgRNAs+Cas9-injected Tg(*Gata5-E5-mCherry;myl7-GFP*) embryos by confocal microscopy at 48hpf and 5dpf, along with Cas9-only injected controls. At 48hpf, mCherry intensity was significantly reduced in the Gata5-E5 knockout group, as compared with the Cas9-only group (Figure 4-7). Specifically, the knockout embryos showed a decreased level of enhancer-driven reporter expression in the region adjacent to the inflow tract. Because the sgRNAs were designed to target Cas9 activity outside of the enhancer sequence cloned into the reporter construct, I knew that this decrease in mCherry expression was not a direct result of the disruption of the enhancer cloned in the reporter construct, but of the endogenous one. I therefore deduced that either the cells that would usually express the Gata5-E5-driven reporter were absent, or, alternatively, that the activity of the enhancer cloned within the reporter construct

had been reduced due to alteration of the available TFs that regulate its activity. In situ hybridisation for *gata5* transcripts in Cas9-only controls and Gata5-E5 KO embryos revealed a decrease in *gata5* in the atrial region of the mutant embryos; the same region where I saw a loss of Gata5-E5-driven mCherry fluorescence in these embryos (Figure 4-7b). Together, these data suggest that the excision of endogenous Gata5-E5 causes a decrease in *gata5* endogenous expression in the cells surrounding the inflow tract. In addition, Gata5-E5 activity in these cells seems to rely on the expression of *gata5* in an autoregulatory loop manner, i.e., Gata5 binds to its own enhancer and is necessary for its activity. Alternatively, the loss of fluorescence observed in the KO embryos may be a result of a loss of *gata5*-expressing cells in the region proximal to the inflow tract following Gata5-E5 excision. Together, these data add another layer of evidence that Gata5-E5 is indeed an enhancer for *gata5.*

At 5dpf, the mCherry fluorescence intensity was also significantly lower in Gata5-E5 KO embryos, when compared with Cas9-only control embryos. It will be important to distinguish whether the cells adjacent to the inflow tract are losing the fluorescence of the enhancer-driven reporter but are still present, or whether this population of cells is absent altogether in the Gata5-E5-KO embryos. It will be important to establish a stable Gata5-E5 knockout line in order to validate these findings and demonstrate their reproducibility due to the high level of variability in the transient studies.

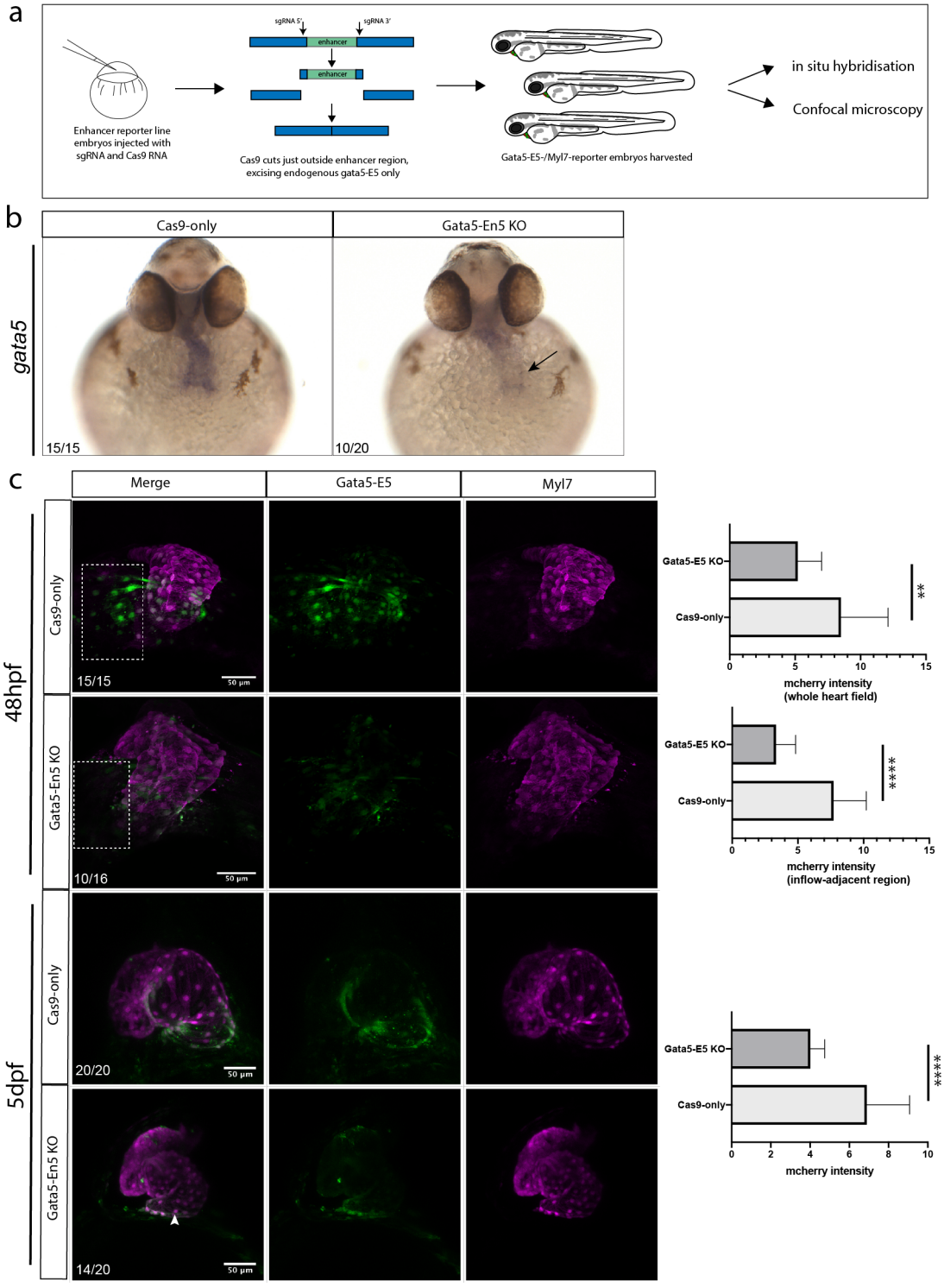

a
sgRNA 5'
sgRNA 3'
enhancer
enhancer
Enhancer reporter line embryos injected with sgRNA and Cas9 RNA
Cas9 cuts just outside enhancer region, excising endogenous gata5-E5 only
Gata5-E5-/Myl7-reporter embryos harvested
in situ hybridisation
Confocal microscopy
b
gata5
Cas9-only
Gata5-En5 KO
15/15
10/20
c
Merge
Gata5-E5
Myl7
48hpf
5dpf
Cas9-only
Gata5-En5 KO
Cas9-only
Gata5-En5 KO
15/15
10/16
20/20
14/20
50 µm
Gata5-E5 KO
Cas9-only
**
mcherry intensity (whole heart field)
Gata5-E5 KO
Cas9-only
****
mcherry intensity (inflow-adjacent region)
Gata5-E5 KO
Cas9-only
****
mcherry intensity

Figure 4-7 Excision of the endogenous Gata5-E5 enhancer led to a decrease in enhancer-driven reporter expression and a reduction in endogenous gata5 expression.

a) Workflow for CRISPR-Cas9 excision of the endogenous Gata5-E5 (but not the sequence cloned into the reporter construct. b) in situ hybridisation showed a decrease in endogenous *gata5* expression in the region surrounding the atrium in Gata5-E5-KO embryos (black arrow). c) At 48hpf, Tg(*Gata5-E5-mCherry; myl7-GFP*) embryos in which the Gata5-E5 enhancer had been excised showed significantly less fluorescence than the Cas9-only control. Significance was calculated using an unpaired Student's t-test. ** = p<0.0021. Confocal images showed that this decrease in fluorescence occurred within the Myl7-negative population adjacent to the inflow tract, marked with a white box. At 5dpf, significantly less fluorescence was seen within the atrium (white arrowhead) and AVC of Gata5-E5-KO embryos. Significance was calculated using an unpaired Student's t-test. **** = p<0.0001. Representative embryos, anterior to the top. Scale bar 50μm.

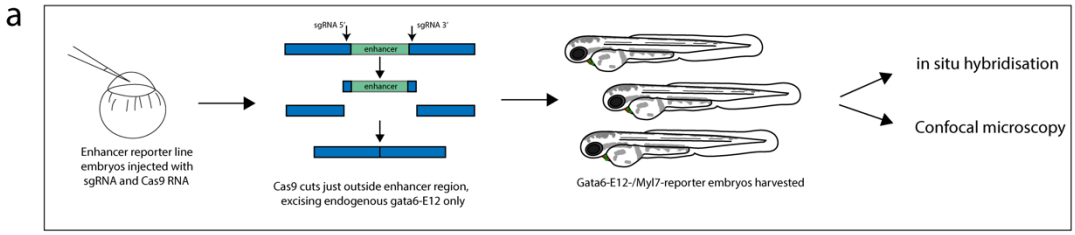
4.2.8 Excision of Gata6-E12 led to a small decrease in enhancer-driven reporter expression that was rescued by 5dpf

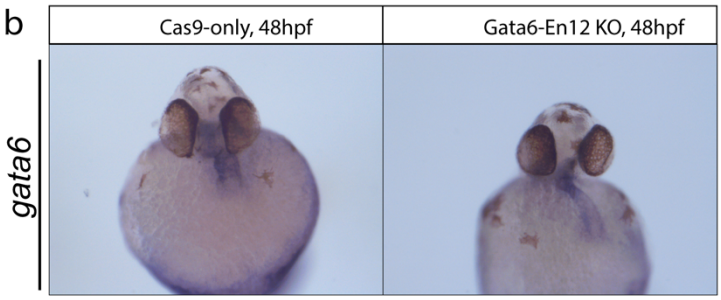
I used the same method to investigate the consequences of the deletion of the Gata6-E12 putative enhancer. Again, CRISPR-Cas9 was employed to excise the endogenous enhancer, with sgRNAs/Cas9 targeting regions just outside the sequence cloned into the enhancer reporter. Selected sgRNAs were injected in combination with Cas9 mRNA into Tg(*Gata6-E12-mCherry;myl7-GFP)* embryos at the one-cell stage. Embryos were analysed by confocal microscopy and in situ hybridisation at 48hpf and 5dpf to investigate whether any changes in reporter activity and/or endogenous *gata6* expression could be observed, when compared to Cas9-only controls.

Figure 4-8 shows that a small but statistically significant difference in fluorescence intensity was seen in Gata6-E12-KO embryos at 48hpf. This suggests that, as predicted for Gata5-E5, an autoregulatory loop may exist in which Gata6 regulates its own expression via the Gata6-E12 enhancer. Therefore, knockdown of Gata6-E12 would reduce the levels of available Gata6, thereby reducing enhancer-driven

reporter expression in these cells. In situ hybridisation does not reveal any obvious differences between the expression of *gata6* in Cas9-only control embryos and Gata6-E12-KO embryos. However, this method is by no means high resolution or quantitative and, as suggested by the small decrease in enhancer-driven reporter expression, the changes in *gata6* levels may be very subtle and therefore undetectable by this technique.

By 5dpf, the fluorescence intensity of the enhancer-driven reporter was comparable to that of the Cas9-only control. This suggests that any phenotype generated by the knockout of Gata6-E12 was recovered to control levels by 5dpf. As redundancy is very common within the GATA family, I speculate that compensation might be occurring to rescue the mild phenotype observed at 48hpf, either by other enhancers within the *gata6* domain, or by the other cardiac GATA factors.


a
sgRNA 5'
sgRNA 3'
enhancer
enhancer
Enhancer reporter line embryos injected with sgRNA and Cas9 RNA
Cas9 cuts just outside enhancer region, excising endogenous gata6-E12 only
Gata6-E12-/Myl7-reporter embryos harvested
in situ hybridisation
Confocal microscopy


b
Cas9-only, 48hpf
Gata6-En12 KO, 48hpf
gata6

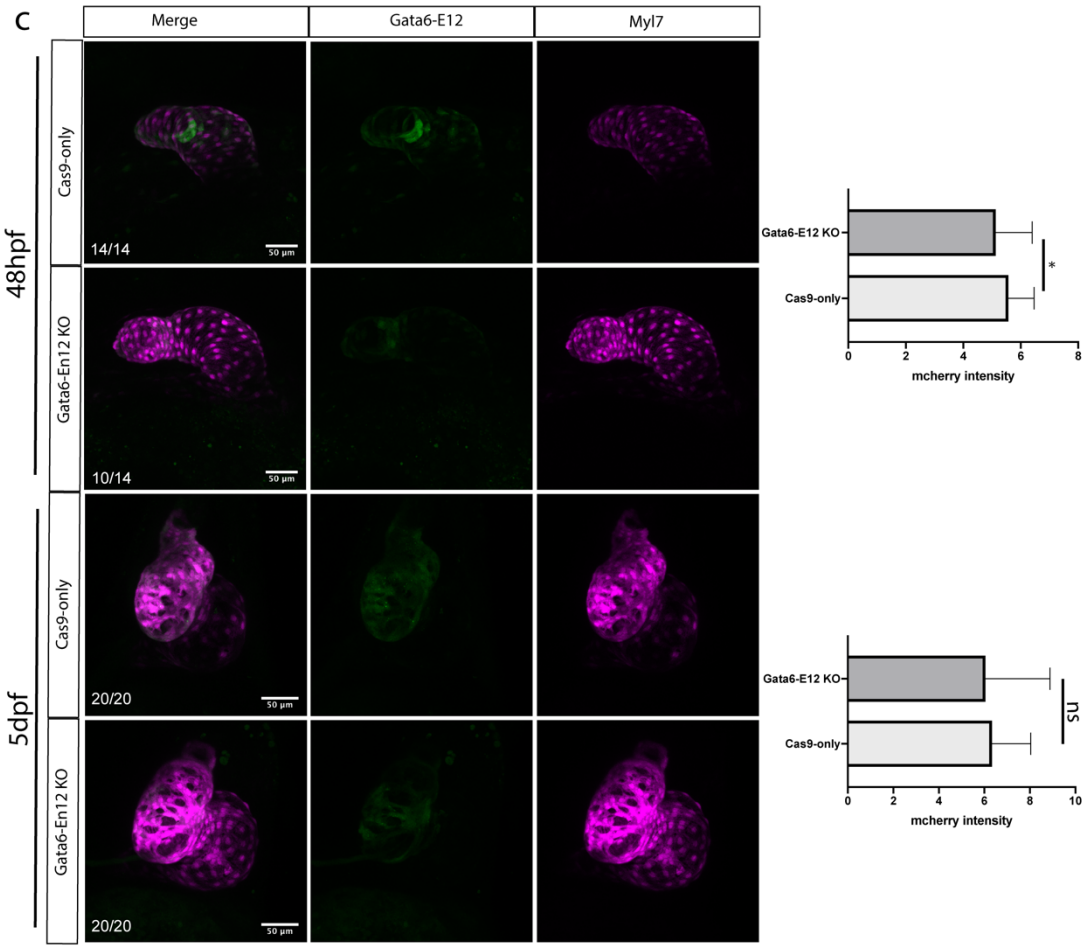

c
Merge
Gata6-E12
Myl7
48hpf
5dpf
Cas9-only
Gata6-En12 KO
Cas9-only
Gata6-En12 KO
14/14
10/14
20/20
20/20
50 µm
Gata6-E12 KO
Cas9-only
*
0
2
4
6
8
mcherry intensity
Gata6-E12 KO
Cas9-only
ns
0
2
4
6
8
10
mcherry intensity

Figure 4-8 Gata6-E12 excision led to a transient decrease of enhancer-driven reporter expression

a) Workflow for CRISPR-Cas9 excision of the endogenous Gata6-E12 (but not the sequence cloned into the reporter construct). b) in situ hybridisation showed no obvious difference in endogenous *gata6* expression in the heart in Gata6-E12-KO embryos, when compared to control siblings. c) At 48hpf, Tg(*Gata6-E12-mCherry; myl7-GFP*) embryos in which the Gata6-E12 enhancer had been excised showed significantly less fluorescence than the Cas9-only control. Significance was calculated using an unpaired Student's t-test. *= $p<0.0332$. At 5dpf, fluorescence was comparable between Cas9-only controls and Gata6-E12-KO embryos. Significance was calculated using an unpaired Student's t-test. Representative embryos, anterior to the top. Scale bar 50μm.

4.2.9 Gata5-E5-cherry expression pattern resembled that of *islet1*

After turning to the literature, I noticed that the expression pattern of the Gata5-E5-driven reporter bore a striking resemblance to the published expression of the transcription factor, *islet1(isl1)* (Tessadori et al., 2012).

Isl1 has long been accepted as a marker for SHF progenitors in other models, such as the mouse, where it coordinates and regulates the transcriptional cascade leading to SHF development and cardiac differentiation (Cai et al., 2003). As described in Chapter 1, SHF progenitors differentiate after the formation of the FHF-derived heart tube, where they form the outflow and inflow tracts of the developing heart. Study of the SHF is particularly important as the SHF-derived heart structures are perturbed in many common congenital heart defects (Gittenberger-de Groot et al., 2013). After some initial dispute over the existence of a SHF in zebrafish, it has been confirmed that zebrafish possess a population of SHF-like late-differentiating cardiac progenitors that contribute to both the atrial and ventricular poles of the heart tube (de Pater et al., 2009). It has been confirmed that *isl1* is also expressed in zebrafish SHF progenitors (Cai et al., 2003) but, by contrast to what has been observed in the mouse model, where *isl1* is required for both SHF fields, in the

zebrafish it is only required for CM differentiation at the venous pole (inflow tract) (Witzel et al., 2017).

Isl1 gets downregulated as the mesodermal progenitors differentiate into mature CMs, with the exception of CMs of the sinoatrial node (SAN) in both the embryonic and adult heart (Sun et al., 2007). As outlined in chapter 1, the CCS comprises several important components; the SAN consists of specialised pacemaker cells that generate spontaneous electrical impulses to trigger contraction of the atrium, the AVN then delays the impulse, maintaining a relaxed ventricle as it fills with blood before ventricular contraction is initiated (Christoffels Vincent et al., 2010). An optogenetic study has localised the functional pacemaker at the venous pole of the embryonic heart (Arrenberg et al., 2010). *Isl1* labels pacemaker cells between the sinus venosus and the atrium, an expression pattern that resembles that of Gata5-E5-driven mCherry.

Because of the similarity between the expression patterns of *isl1* and Gata5-E5 mCherry, I investigated whether the two overlapped using ICC at 48hpf in Tg(*Gata5-E5*-mCherry;*myl7*-GFP) embryos. ICC using an anti-pan-islet antibody (from the Developmental Studies Hybridoma Bank) and an anti-mCherry antibody demonstrated overlap between the two in the Myl7-positive CMs of the inflow tract (Figure 4-9). This led me to hypothesise that Gata5-E5 was active in a subset of cells composing the early CCS in both the CMs of the inflow tract and the AVC. In each embryo, one or two cells in the Myl7-negative region adjacent to the inflow tract expressed both Gata5-E5-mCherry and islet, but the majority only expressed one of the markers (Figure 4-9). The putative SHF cells expressing the islet marker were in very close proximity to the Gata5-E5-mCherry-expressing cells, suggesting that the two cell types may communicate or somehow be linked.

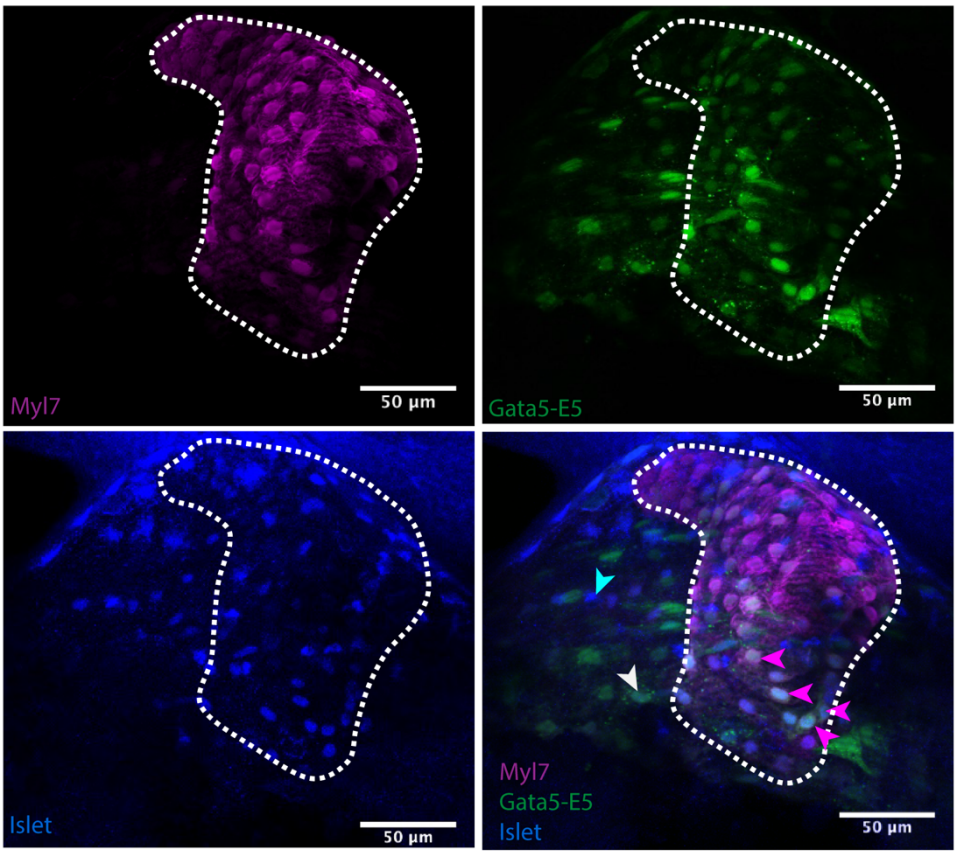


Figure 4-9 Gata5-E5 activity overlapped with Islet expression in the early cardiac conduction system cells.

Microscopic analysis of Tg*(Gata5-E5-mCherry;myl7-GFP)* embryos stained with an anti-pan-islet antibody at 48hpf. A population of Myl7-positive, Gata5-E5-mCherry-positive, islet-positive cells (magenta arrowheads) at the inflow tract represent the putative early CCS. Another population of Myl7-negative, Gata5-E5-mCherry-positive, Islet-negative cells lay just adjacent to the inflow tract (white arrowhead), in close proximity to the Islet-positive putative SHF progenitor cells (cyan arrowhead). Representative embryos, anterior to the top. Scale bar 50µm.

4.2.10 Islet1 expression was reduced in the region adjacent to the inflow tract upon Gata5-E5-KO, suggesting a loss of SHF progenitor cells, and in the putative CCS

Because the Myl7-negative, Gata5-E5-mCherry-positive cells adjacent to the inflow tract were present in the same region as the Islet-positive putative SHF cells, I

investigated the expression of the Islet marker after Gata5-E5 KO. ICC in 48hpf Gata5-E5-KO embryos showed significantly fewer Islet-positive cells in the region adjacent to the inflow tract, when compared to Cas9-only controls (Figure 4-10). This suggests that the Gata5-E5-KO embryos may either have lost expression of *isl1* in this region or, in fact lost a subset of their posterior SHF progenitor cells upon knockout of the endogenous Gata5-E5 enhancer. Considering the little overlap of expression between Gata5-E5-driven mCherry-expressing cells and Islet-expressing cells (Figure 4-9), the loss of SHF cells might be a favoured hypothesis. Of note, nuclear staining could be used in order to investigate whether indeed SHF cells are absent in the Gata5-E5-KO embryos, or if, alternatively, these cells are still present but have downregulated islet expression. I also investigated whether the Myl7-positive, Gata5-E5-mCherry-positive, Islet-positive population, hypothesised to be the early CCS cells, were affected by Gata5-E5 knockout. ICC in 48hpf Gata5-E5-KO embryos showed fewer Islet/Myl7/Gata5-E5-mCherry-positive cells in the inflow tract, when compared to Cas9-only controls, suggesting that fewer CCS cells were specified in the Gata5-E5-KO embryos (Figure 4-10).

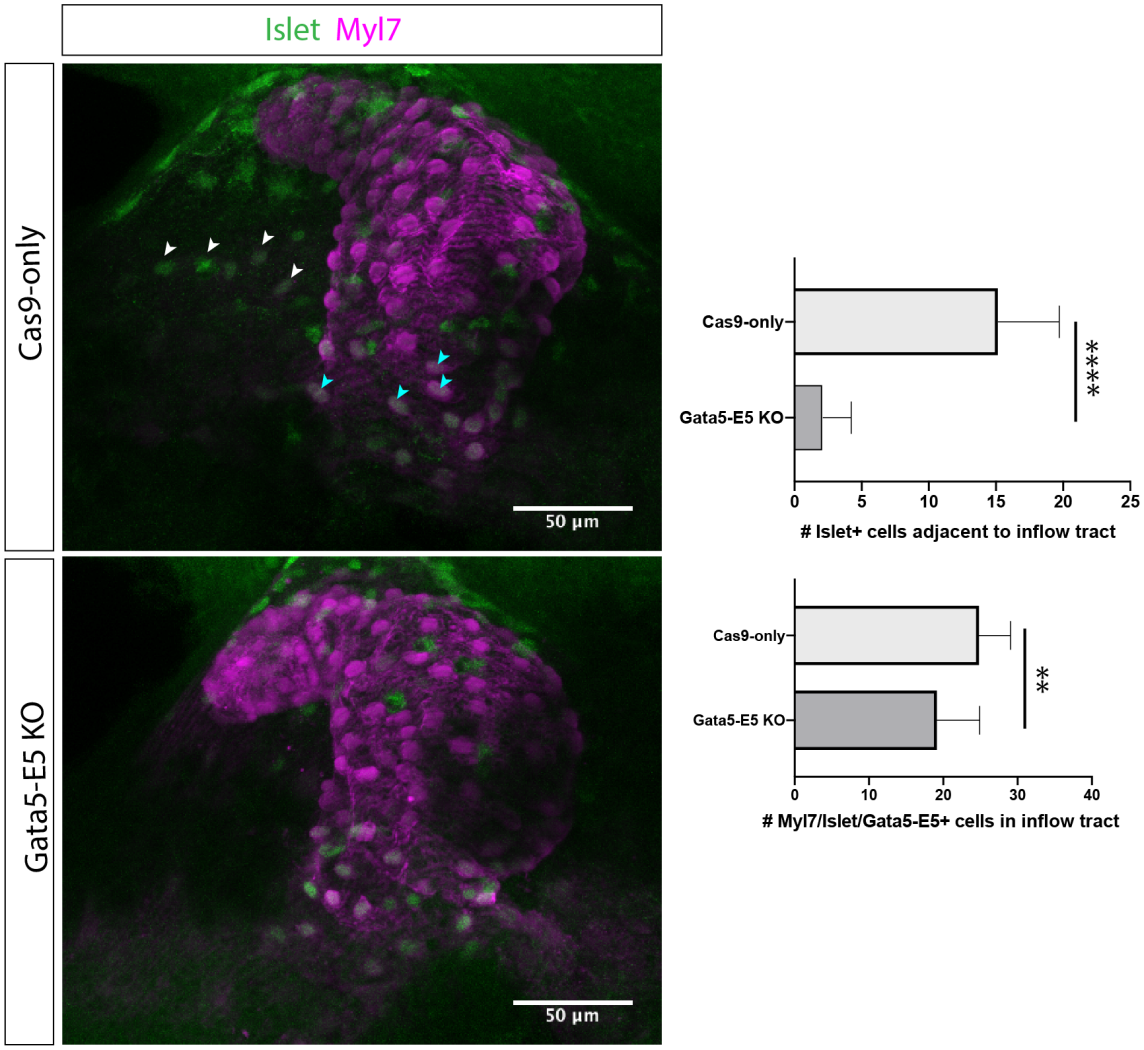


Figure 4-10 Reduced levels of Islet protein in the region adjacent to the inflow tract and in the inflow tract after Gata5-E5 excision suggests a loss of putative posterior SHF progenitor cells and CCS cells in these embryos.

Gata5-E5-KO embryos showed a reduction in Islet-positive cells (green) in the region surrounding the inflow tract in the Gata5-E5-KO embryos (white arrowheads) when compared to control siblings (n=16 in each group). The number of Islet-positive cells in each condition was quantified and significance was calculated using an unpaired Student's t-test. **** = $p<0.0001$. The number of Islet/Myl7/Gata5-E5-mCherry-positive cells (cyan arrowhead) in each condition was counted and significance was calculated using an unpaired Student's t-test. **= $p<0.0021$ Representative embryos, anterior to the top. Scale bar 50µm.

4.2.11 Analysis of the hearts of Gata5-E5-KO embryos at 5dpf showed a reduction in atrial area and atrial hypoplasia

The reduction of Isl1-positive cells around the venous pole of Gata5-E5-KO embryos at 48hpf strongly suggested that a population of posterior SHF progenitors

had been lost or perturbed. It was therefore important to establish whether this defect could have an effect on downstream processes during later stages of cardiogenesis, such as the accretion of cells to the developing atrium. Inspection of confocal images of Gata5-E5-KO embryos at 5dpf showed a potential decrease in the size of the atrium region, as compared with the Cas9-only controls (a representative sample is shown in Figure 4-11a). This observation was confirmed when the area of the atrium was calculated, and the difference between KO and sibling embryos shown to be statistically significant (Figure 4-11b). To ascertain whether this reduction was a result of hypoplasia or simply a reduction in volume without loss of cells, I counted the nuclei of CMs in both conditions. There were significantly fewer cells in the atria of Gata5-E5-KO embryos than in Cas9-only control embryos (Figure 4-11c). To see if this difference was particularly accentuated in the inflow tract, I counted cells within this region, defined as the quarter of the atrium proximal to the sinus venosus. Significantly fewer CMs were present within the inflow tract of Gata5-E5-KO nuclei than in Cas9-only controls, suggesting a defect in the accretion of SHF-derived CMs in these KO embryos.

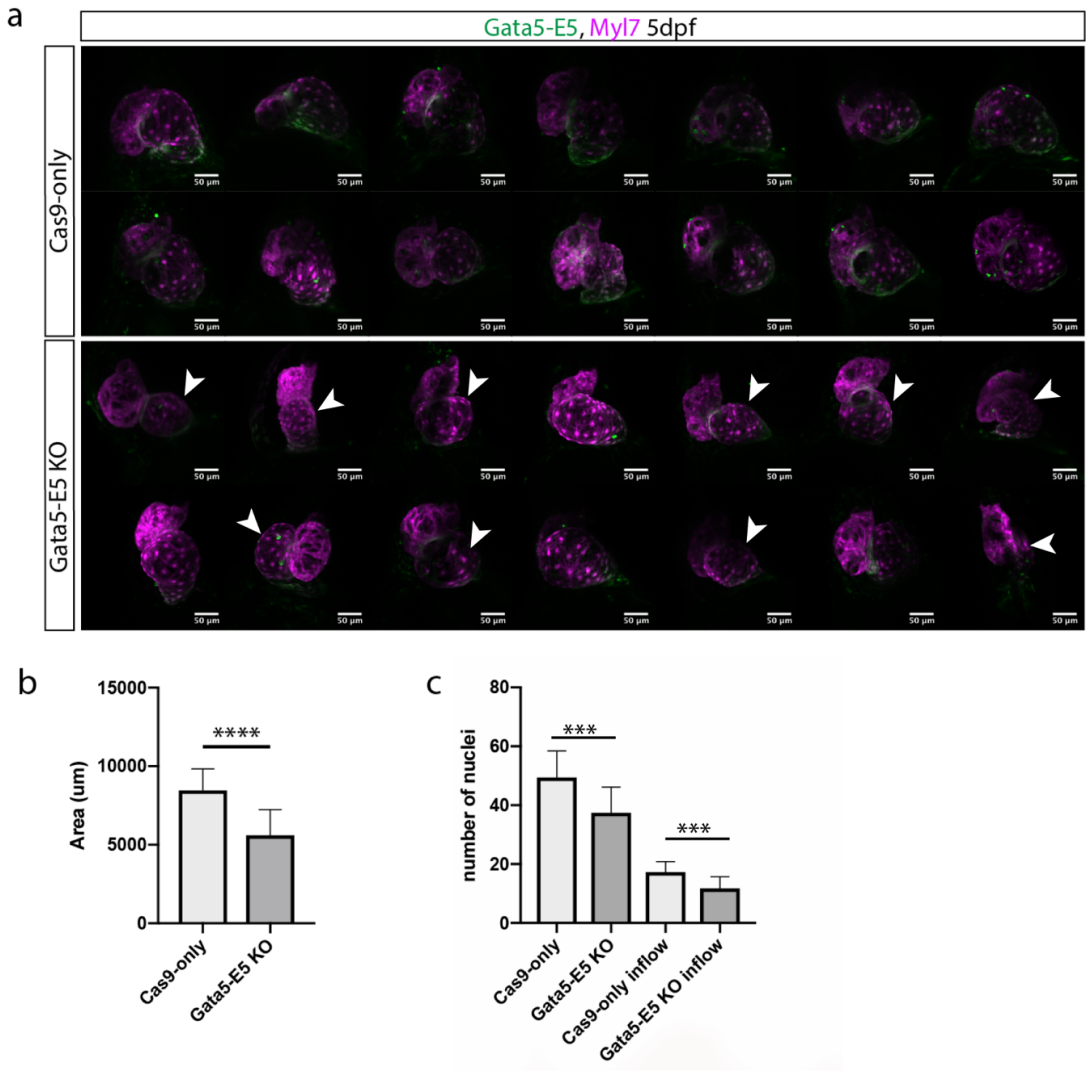


Figure 4-11 Gata5-E5 excision caused a reduction in atrium area and atrial cell numbers by 5dpf.

a) A panel of representative confocal images suggested that the atrium of Gata5-E5-KO embryos was smaller, when compared to Cas9-only controls (white arrowheads). b) The area of the atrium of Gata5-E5-KO embryos was significantly smaller than the Cas9 control ones, as calculated by unpaired Student's t-test. **** = $p<0.0001$ (n=20) c) The number of CMs in both the whole atria and the inflow region alone was significantly lower in the Gata5-E5-KO embryos, compared with Cas9-only controls, as calculated by unpaired Student's t-test. ***= $p<0.0002$, **** = $p<0.0001$.

4.3 Summary

The Tg*(Gata5-E5-mCherry)* reporter showed expression in a very interesting subset of *gata5*-expressing cells. Early in development at the 6-somite stage, Gata5-E5 was active within the LPM. By 48hpf, Gata5-E5 was mainly active within a

subpopulation of CMs in the inflow tract of the developing heart as well as an extracardiac, Myl7-negative population of cells lying just outside the inflow tract.

Tg(*Gata6-E12-mCherry)* was initially expressed throughout the heart tube and early looping heart but by 3dpf its activity became restricted to the ventricle. This line is therefore a very useful tool for the investigation of the early regulation of *gata6* specifically in the ventricle.

HCR confirmed that Gata5-E5-mCherry and Gata6-E12-mCherry were expressed in the same cells as endogenous *gata5* and *gata6* transcripts, respectively, offering further support that they act as enhancers for these GATA genes.

In order to begin to understand the function of these enhancers *in vivo*, I made use of CRISPR-Cas9 technology to excise the endogenous enhancers, while maintaining the enhancer-driven expression of the reporter constructs to enable easy tracking of the relevant cells where the enhancers showed specific activity. When Gata6-E12 was excised, a small decrease in the level of enhancer-driven reporter was observed at 48hpf, however the expression was comparable to that of Cas9-only controls by 5dpf. Also, the level of expression of *gata6* did not appear to be affected by the enhancer knockout, suggesting a level of redundancy at the regulatory level.

By contrast, excision of the Gata5-E5 enhancer led to a decrease in *gata5* endogenous expression in the atrial region of the heart, and a decrease in Tg*(Gata5-E5-mCherry)* reporter fluorescence in the extracardiac population of cells adjacent to the inflow tract.

Although the expression pattern of the Tg*(Gata6-E12-mCherry)* reporter line makes it an extremely useful tool with which to understand the specific upstream regulation

of *gata6* in the ventricle*,* after observing the intriguing phenotype following Gata5-E5-knockout, I decided to proceed to investigate the regulation of Gata5-E5 in depth. I noticed that the expression pattern of the Tg*(Gata5-E5-mCherry)* reporter mimicked that published for Islet1, a marker of the CCS in the atrial myocardium and the SHF progenitor cells outside the atrium. ICC confirmed an overlap between Islet and Gata5-E5-mCherry expression at the venous pole of the atrium, suggesting that these are cells of the early CCS. However, the majority of the Gata5-E5-mCherry-positive cells just outside the heart did not coexpress Islet, but rather lay in very close proximity to Islet-expressing cells. In an effort to understand the relationship between the cells expressing Gata5-E5-mCherry and the Islet-expressing putative SHF progenitor cells, I investigated whether Gata5-E5 knockout affected these SHF progenitor cells. Islet ICC showed a decrease in Islet-positive cells in the region adjacent to the inflow tract after Gata5-E5 knockout. It will be important in future studies to determine whether the cells are lost, or whether Islet stops being expressed upon Gata5-E5 knockout. I also investigated the number of Myl7/Islet/Gata5-E5-positive putative CCS in the inflow tract of these embryos and observed that there were significantly fewer triple-positive cells in Gata5-E5-KO embryos, when compared to Cas9-only controls. When I studied the hearts of Gata5-E5-knockout embryos later in development, I discovered atrial hypoplasia, particularly in the inflow tract (which is also the location of the CCS population), suggesting a defect in SHF-derived growth of the venous pole of the heart. This would be consistent with the hypothesis that the SHF progenitor cells and the nearby Gata5-E5-mCherry-expressing cells are functionally linked and perhaps together are important for SHF cell differentiation and venous pole accretion. It is also possible that these changes arise from an early defect in the Gata5-E5-

mCherry-expressing cells in the ALPM, leading to a loss of these cells. The Gata5-E5-mCherry positive cells perhaps act as support cells for the SHF progenitor cells.

In chapter 5 I strive to fully identify and characterise the populations of cells expressing the Gata5-E5-mCherry reporter in order to begin to understand the upstream regulation of *gata5* in each one of these subpopulations.

5) Single cell transcriptomics identifies the subpopulations in which Gata5-E5 is active, and provides candidate factors for its upstream regulation

5.1 Introduction

Cis-regulatory elements are key intermediate nodes within gene regulatory networks (Long et al., 2016). By studying the relationships between TFs and enhancers and investigating their contribution to gene expression, information flow through these complex networks can begin to be teased apart. As the aim of this project is to identify precise spatiotemporal regulation of the GATA factors during cardiogenesis, it is vital to understand the transcriptome of the cells within each individual population in which the enhancer of interest is active.

The confocal imaging data presented in chapter 4 demonstrated that Gata5-E5 is active within a heterogeneous population of cells (Figure 4-1). Bulk RNA-sequencing would therefore not be informative, as this would obscure the presence of distinct populations and prevent their individual characterisation. The development of next-generation sequencing technologies to analyse single cell transcriptomes has made it possible to really probe cellular diversity within a given

population. The platform I have chosen to use is the Chromium system, 10x Genomics (Pleasanton, CA, USA) (Hernández et al., 2018). This system is a microfluidics device which relies on massive partitioning and barcoding, summarised in Figure 5 -1. Using Gemcode technology, gel beads are infused with millions of barcoded oligonucleotides and contain reverse transcription reagents. The single cell suspension and the gel beads are added to an oil-surfactant solution, creating Gel Beads in Emulsion (GEMs). These GEMs act as isolated vesicles, wherein the gel beads dissolve and the transcripts are labelled with cell-specific barcodes. Barcoded cDNA is then pooled for library preparation. Following sequencing, the cellular barcodes can be used to identify transcripts from the same cell, providing information of the transcriptome of each individual cell. This technology has high cell capture rates (up to 65%) and very low doublet rates (0.8%).

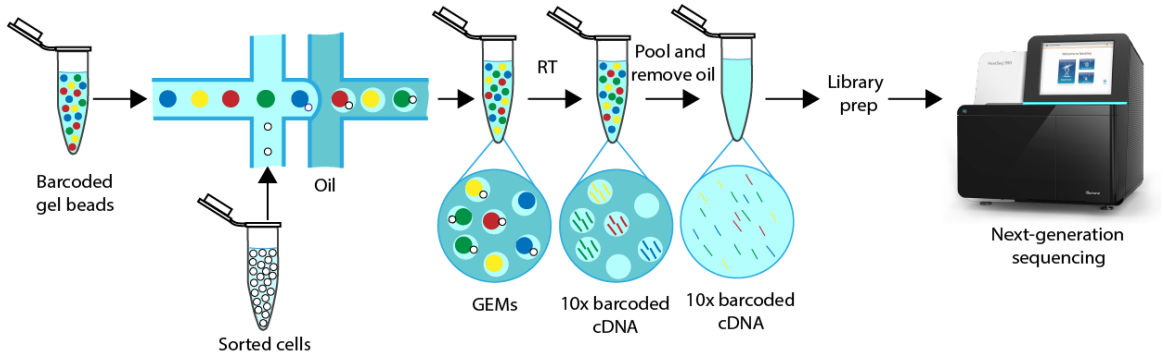


Figure 5 -1 Workflow of 10x Chromium single-cell sequencing

My image analysis suggested that Gata5-E5 was labelling a population of cells of the CCS. The overlap of Isl1 expression with Gata5-E5-driven reporter activity in a population of cells at the venous pole of the Myl7-GFP-expressing heart indicated that these cells represented the sinoatrial node (SAN). Gata5-E5 was also active in a population of cells in the atrioventricular canal (AVC) which are postulated to be

the atrioventricular node (AVN). It is therefore important to corroborate this with the results of the single cell sequencing.

Thus far, imaging analysis and knock-out studies together strongly suggested that Gata5-E5 was also active in a population of cells that resided in the tissue adjacent to the inflow tract of the heart. I hypothesise that these cells may cooperate with the SHF progenitor cells surrounding the venous pole of the heart.

Transcriptional profiling will allow identification of the populations of cells in which Gata5-E5 regulates *gata5* expression and consequent activity. Profiling the transcriptome in each of these populations at the single cell level will uncover the potential factors that can regulate *gata5*, via the Gata5-E5 enhancer activity, specifically in each of these individual subpopulations.

5.2 Results

5.2.1 Single cell transcriptome results

I used fluorescence-activated cell sorting (FACS) to specifically purify the cells of interest from Tg(*Gata5-E5-mCherry;myl7GFP)* embryos (Supplementary Figure 1). Typically, it is prudent to use a cell viability stain to label and exclude dead cells. However, the cell viability stain that we commonly use is 7-AAD, which cannot be used in combination with mCherry due to spectral overlap. As such, this experiment was performed with no cell viability stain, increasing the risk of sequencing material from dead or dying cells. However, strict FACS gating can be used to exclude cells with the dimensions typical of dead or dying cells, which greatly minimises this risk.

I obtained 14,000 cells from FAC-sorting 300 *Tg(Gata5-E5-mCherry)* embryos. From this cell input, the 10x software predicted that 4,548 cells were sequenced. After removing mitochondrial genes and filtering cells with low number of unique

genes detected in each cell or aberrantly high gene count (which could indicate low quality cells, empty droplets, doublets or multiplets), we recovered 3976 cells (87.4% of total number of cells sequenced). The sizes of the remaining single cell libraries were normally distributed, with a mean read count per cell of 161,287. The median number of genes expressed in the libraries that passed quality control was 1342.

5.2.2 Initial dataset clustering

Linear dimensional reduction was carried out and statistically significant principal components determined 12 cluster groups which were plotted using t-distributed stochastic neighbour embedding (t-SNE) (L. van der Maaten, 2008) (Supplementary Figure 2), as well as being presented as a heatmap (Figure 5 -2). This was a much higher number of clusters than expected, and many unexpected genes were revealed, including haematopoietic markers. Upon closer inspection of the data, I noticed that there were no *mCherry* or *gata5* transcripts in the cells of many of these clusters (Supplementary Figure 2). This suggests that the FACS process was not stringent enough and cells were being sorted that did not express the mCherry protein. In particular, there was a large cluster expressing red blood cell markers (such as *hbbe2*), which may have displayed autofluorescence in the mCherry channel.

As such, the data were subclustered, only taking into account the cells that showed mCherry expression (clusters 6, 7, 9, 10 and 11).

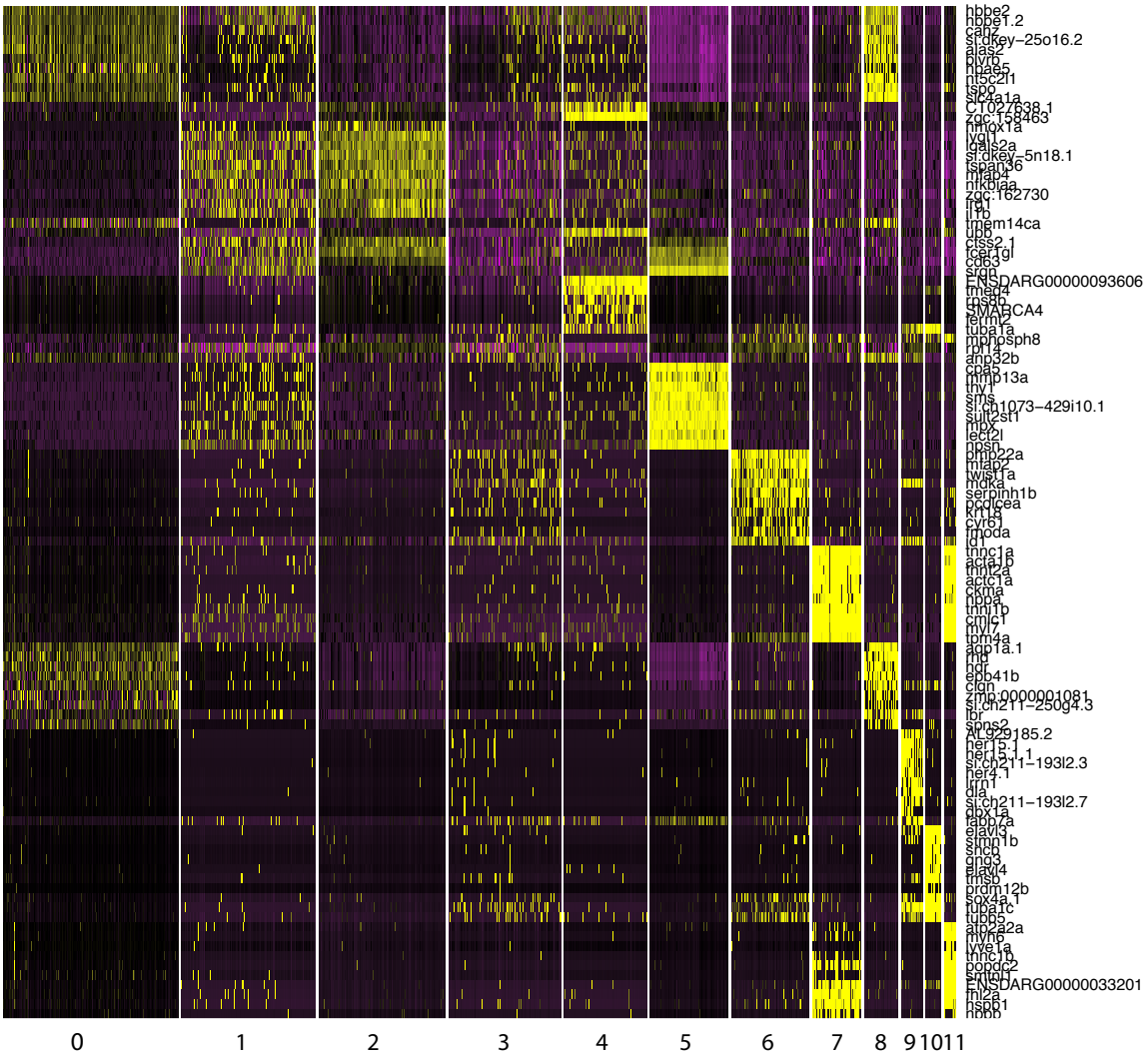


Figure 5 -2 Marker genes identified 12 distinct populations within the single cell dataset.

Heatmap showing the expression of the most differentially expressed genes (rows) within the clustered single cell data (columns). Expression values are fragments per kilobase of transcript per million transcripts (FPKM), yellow colour indicates high expression.

5.2.3 Reclustered results based on *mCherry* transcript expression

Re-running the pipeline again using only the cells present in clusters 6, 7, 9, 10 and 11, we obtained 8 new clusters. The data were plotted on a heatmap (Figure 5 -3) and t-SNE plot (Figure 5-4a), clearly showing these 8 clusters. Based on the imaging data presented in chapters 3 and 4, I expected the data to cluster into fewer subpopulations.

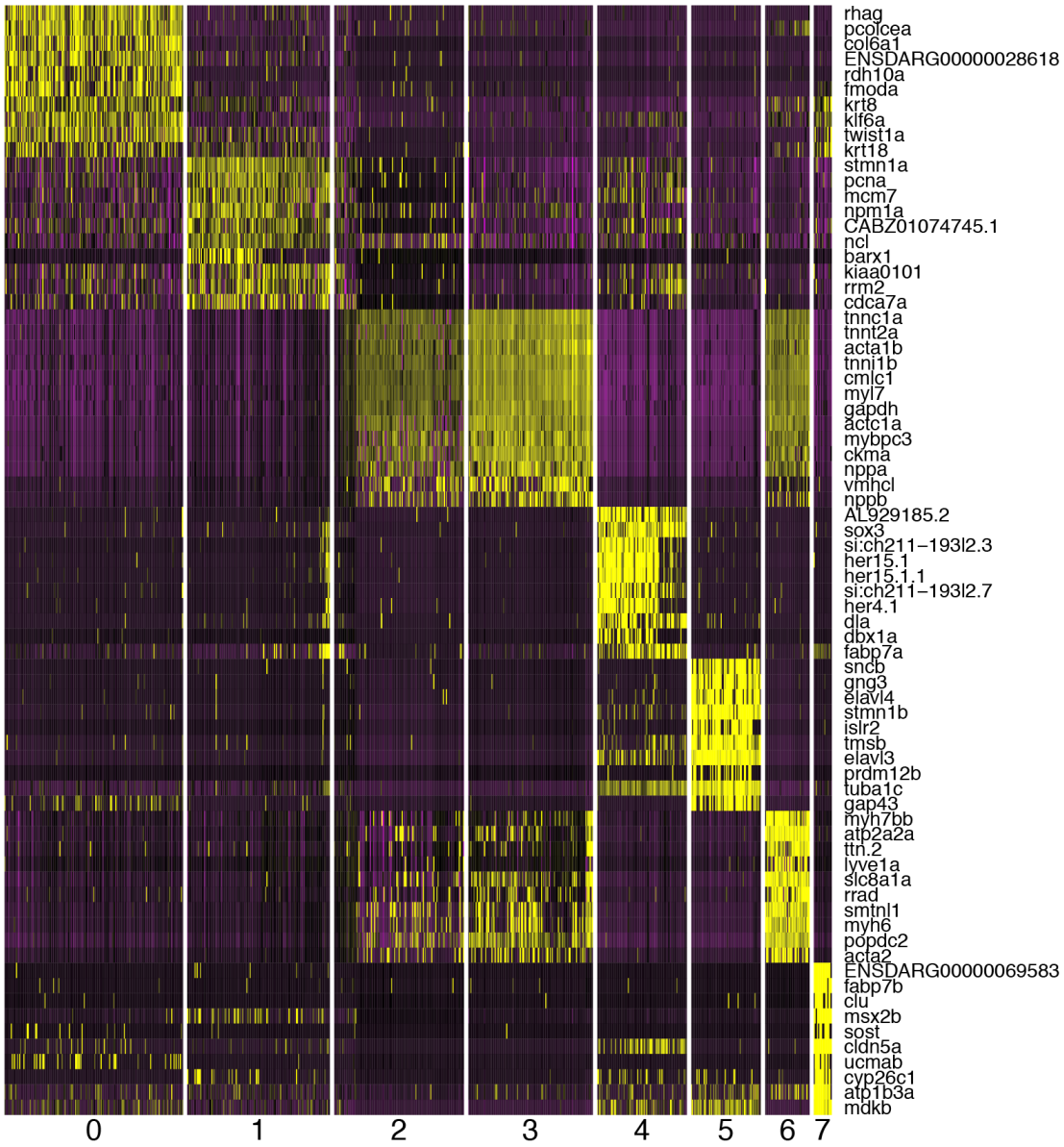


Figure 5 -3 Marker genes identified 8 distinct populations within the single cell dataset.

Heatmap showing the expression of the most differentially expressed genes (rows) within the clustered single cell data (columns). Expression values are FPKM, yellow colour indicates high expression.

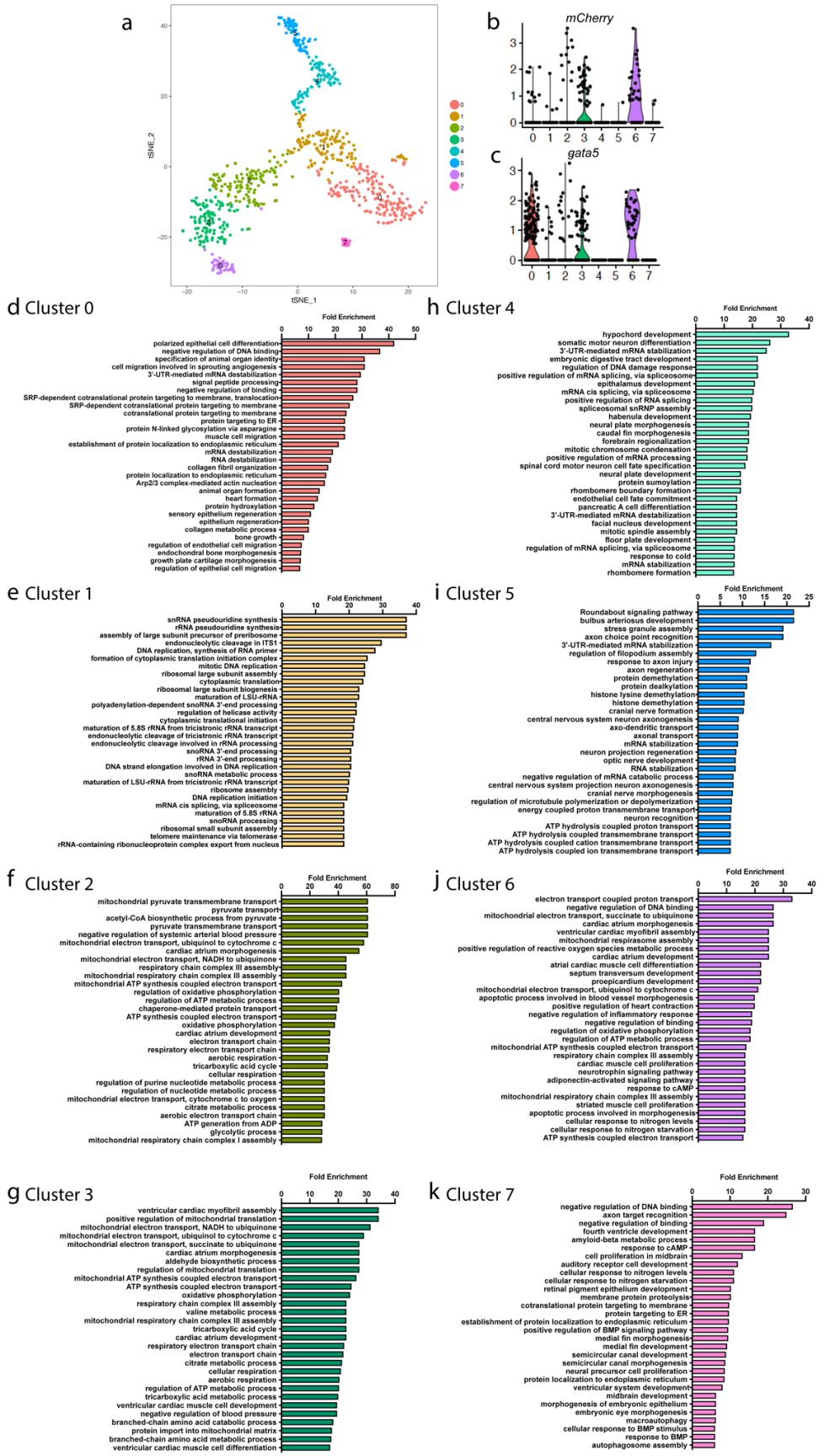

a
tSNE_2
tSNE_1
b
mCherry
c
gata5
d Cluster 0
Fold Enrichment
polarized epithelial cell differentiation
negative regulation of DNA binding
specification of animal organ identity
cell migration involved in sprouting angiogenesis
3'-UTR-mediated mRNA destabilization
signal peptide processing
negative regulation of binding
SRP-dependent cotranslational protein targeting to membrane, translocation
SRP-dependent cotranslational protein targeting to membrane
cotranslational protein targeting to membrane
protein targeting to ER
protein N-linked glycosylation via asparagine
muscle cell migration
establishment of protein localization to endoplasmic reticulum
mRNA destabilization
RNA destabilization
collagen fibril organization
protein localization to endoplasmic reticulum
Arp2/3 complex-mediated actin nucleation
animal organ formation
heart formation
protein hydroxylation
sensory epithelium regeneration
epithelium regeneration
collagen metabolic process
bone growth
regulation of endothelial cell migration
endochondral bone morphogenesis
growth plate cartilage morphogenesis
regulation of epithelial cell migration
e Cluster 1
Fold Enrichment
snRNA pseudouridine synthesis
rRNA pseudouridine synthesis
assembly of large subunit precursor of preribosome
endonucleolytic cleavage in ITS1
DNA replication, synthesis of RNA primer
formation of cytoplasmic translation initiation complex
mitotic DNA replication
ribosomal large subunit assembly
cytoplasmic translation
ribosomal large subunit biogenesis
maturation of LSU-rRNA
polyadenylation-dependent snoRNA 3'-end processing
regulation of helicase activity
cytoplasmic translational initiation
maturation of 5.8S rRNA from tricistronic rRNA transcript
endonucleolytic cleavage of tricistronic rRNA transcript
endonucleolytic cleavage involved in rRNA processing
snoRNA 3'-end processing
rRNA 3'-end processing
DNA strand elongation involved in DNA replication
snoRNA metabolic process
maturation of LSU-rRNA from tricistronic rRNA transcript
ribosome assembly
DNA replication initiation
mRNA cis splicing, via spliceosome
maturation of 5.8S rRNA
snoRNA processing
ribosomal small subunit assembly
telomere maintenance via telomerase
rRNA-containing ribonucleoprotein complex export from nucleus
f Cluster 2
Fold Enrichment
mitochondrial pyruvate transmembrane transport
pyruvate transport
acetyl-CoA biosynthetic process from pyruvate
pyruvate transmembrane transport
negative regulation of systemic arterial blood pressure
mitochondrial electron transport, ubiquinol to cytochrome c
cardiac atrium morphogenesis
mitochondrial electron transport, NADH to ubiquinone
respiratory chain complex III assembly
mitochondrial respiratory chain complex III assembly
mitochondrial ATP synthesis coupled electron transport
regulation of oxidative phosphorylation
regulation of ATP metabolic process
chaperone-mediated protein transport
ATP synthesis coupled electron transport
oxidative phosphorylation
cardiac atrium development
electron transport chain
respiratory electron transport chain
aerobic respiration
tricarboxylic acid cycle
cellular respiration
regulation of purine nucleotide metabolic process
regulation of nucleotide metabolic process
mitochondrial electron transport, cytochrome c to oxygen
citrate metabolic process
aerobic electron transport chain
ATP generation from ADP
glycolytic process
mitochondrial respiratory chain complex I assembly
g Cluster 3
Fold Enrichment
ventricular cardiac myofibril assembly
positive regulation of mitochondrial translation
mitochondrial electron transport, NADH to ubiquinone
mitochondrial electron transport, ubiquinol to cytochrome c
mitochondrial electron transport, succinate to ubiquinone
cardiac atrium morphogenesis
aldehyde biosynthetic process
regulation of mitochondrial translation
mitochondrial ATP synthesis coupled electron transport
ATP synthesis coupled electron transport
oxidative phosphorylation
respiratory chain complex III assembly
valine metabolic process
mitochondrial respiratory chain complex III assembly
tricarboxylic acid cycle
cardiac atrium development
respiratory electron transport chain
electron transport chain
citrate metabolic process
cellular respiration
aerobic respiration
regulation of ATP metabolic process
tricarboxylic acid metabolic process
ventricular cardiac muscle cell development
negative regulation of blood pressure
branched-chain amino acid catabolic process
protein import into mitochondrial matrix
branched-chain amino acid metabolic process
ventricular cardiac muscle cell differentiation
h Cluster 4
Fold Enrichment
hypochord development
somatic motor neuron differentiation
3'-UTR-mediated mRNA stabilization
embryonic digestive tract development
regulation of DNA damage response
positive regulation of mRNA splicing, via spliceosome
epithalamus development
mRNA cis splicing, via spliceosome
positive regulation of RNA splicing
spliceosomal snRNP assembly
habenula development
neural plate morphogenesis
caudal fin morphogenesis
forebrain regionalization
mitotic chromosome condensation
positive regulation of mRNA processing
spinal cord motor neuron cell fate specification
neural plate development
protein sumoylation
rhombomere boundary formation
endothelial cell fate commitment
pancreatic A cell differentiation
3'-UTR-mediated mRNA destabilization
facial nucleus development
mitotic spindle assembly
floor plate development
regulation of mRNA splicing, via spliceosome
response to cold
mRNA stabilization
rhombomere formation
i Cluster 5
Fold Enrichment
Roundabout signaling pathway
bulbus arteriosus development
stress granule assembly
axon choice point recognition
3'-UTR-mediated mRNA stabilization
regulation of filopodium assembly
response to axon injury
axon regeneration
protein demethylation
protein dealkylation
histone lysine demethylation
histone demethylation
cranial nerve formation
central nervous system neuron axonogenesis
axo-dendritic transport
axonal transport
mRNA stabilization
neuron projection regeneration
optic nerve development
RNA stabilization
negative regulation of mRNA catabolic process
central nervous system projection neuron axonogenesis
cranial nerve morphogenesis
regulation of microtubule polymerization or depolymerization
energy coupled proton transmembrane transport
neuron recognition
ATP hydrolysis coupled proton transport
ATP hydrolysis coupled transmembrane transport
ATP hydrolysis coupled cation transmembrane transport
ATP hydrolysis coupled ion transmembrane transport
j Cluster 6
Fold Enrichment
electron transport coupled proton transport
negative regulation of DNA binding
mitochondrial electron transport, succinate to ubiquinone
cardiac atrium morphogenesis
ventricular cardiac myofibril assembly
mitochondrial respirasome assembly
positive regulation of reactive oxygen species metabolic process
cardiac atrium development
atrial cardiac muscle cell differentiation
septum transversum development
proepicardium development
mitochondrial electron transport, ubiquinol to cytochrome c
apoptotic process involved in blood vessel morphogenesis
positive regulation of heart contraction
negative regulation of inflammatory response
negative regulation of binding
regulation of oxidative phosphorylation
regulation of ATP metabolic process
mitochondrial ATP synthesis coupled electron transport
respiratory chain complex III assembly
cardiac muscle cell proliferation
neurotrophin signaling pathway
adiponectin-activated signaling pathway
response to cAMP
mitochondrial respiratory chain complex III assembly
striated muscle cell proliferation
apoptotic process involved in morphogenesis
cellular response to nitrogen levels
cellular response to nitrogen starvation
ATP synthesis coupled electron transport
k Cluster 7
Fold Enrichment
negative regulation of DNA binding
axon target recognition
negative regulation of binding
fourth ventricle development
amyloid-beta metabolic process
response to cAMP
cell proliferation in midbrain
auditory receptor cell development
cellular response to nitrogen levels
cellular response to nitrogen starvation
retinal pigment epithelium development
membrane protein proteolysis
cotranslational protein targeting to membrane
protein targeting to ER
establishment of protein localization to endoplasmic reticulum
positive regulation of BMP signaling pathway
medial fin morphogenesis
medial fin development
semicircular canal development
semicircular canal morphogenesis
neural precursor cell proliferation
protein localization to endoplasmic reticulum
ventricular system development
midbrain development
morphogenesis of embryonic epithelium
embryonic eye morphogenesis
macroautophagy
cellular response to BMP stimulus
response to BMP
autophagosome assembly

Figure 5-4 The single cell data set contained 8 distinct cell clusters.

a) Two-dimensional t-SNE clustering of single cell libraries Numbers indicate assignment to the 8 most significant cell clusters within the dataset. b) violin plot showing *mCherry* transcript levels in each cluster. c) violin plot showing *gata5* transcript levels in each cluster. d-k) GO term analysis showing the biological processes associated with the genes expressed in each cluster; cluster 0 (d), cluster1 (e), cluster 2 (f), cluster 3 (g), cluster 4 (h), cluster 5 (i), cluster 6 (j), cluster 7 (k).

5.2.4 Cluster 0 was enriched in transcripts typical of epithelial cells

169 cells were assigned to cluster 0. GO term analysis associated the gene expression profiles of the cells in cluster 0 with epithelial cell differentiation and migration (Figure 5-4d), for example, the cell adhesion molecules Fibronectin 1a and 1b (*fn1a, fn1b*) and cadherin 2 (*cdh2a*) were highly expressed. Cluster 0 also showed a high expression of collagens (*col6a1, col6a3* and col6a2). Collagen 6s have been investigated previously for their role in the heart. ColVI is secreted by interstitial fibroblasts and has been reported to be overexpressed following myocardial infarction in the mouse. Interestingly, studies in the *Col6a1* knockout mouse showed an improvement in cardiac function, structure and remodelling following myocardial infarction (Luther Daniel et al., 2012).

I also noticed a number of epithelial-to-mesenchymal transition (EMT) markers, including *twist1a, twist1b* and *snai2*. Interestingly, *twist1a* and *twist1b* are both upregulated strongly in the heart following resection injury in the adult zebrafish (Kim et al., 2010a). The extracellular matrix protein Periostin b (Postnb) (Snider et al., 2008) was also expressed in this cluster and this gene has also been associated with EMT. Another gene that was highly expressed in cluster 0 was the platelet-derived growth factor receptor alpha (*pdgfra).* Pdgfra is a highly conserved tyrosine kinase receptor and its signalling is essential for organogenesis and has been implicated in mediating differentiation, migration and function of mesenchymal cells. Knockout of *pdgfra* in zebrafish has been demonstrated to lead to defective medial

migration of myocardial and endocardial progenitors to the midline (El-Rass et al., 2017), indicating a convergence of roles with *gata5*. Knockdown of *PDGFRA* in the chick and mouse has been shown to cause a spectrum of cardiac inflow tract defects, some of which mimic the mutation in humans that causes total anomalous pulmonary venous return (Bleyl et al., 2010). In the chick and mouse, these defects are thought to result from *PDGFRA* dysfunction during pulmonary vein formation in a tissue known as the dorsal mesocardial protrusion (DMP). The DMP is a population of extracardiac mesenchymal cells that arises from the SHF and joins the venous pole of the heart through the dorsal mesocardium (Bleyl et al., 2010). PDGFRA been described as a cardiac progenitor cell marker in both human and mouse (Noseda et al., 2015, Kim et al., 2017, Chong et al., 2013).

aldh1a2, which codes for a retinoic acid (RA) synthesis enzyme was highly expressed in this cluster. *aldh1a2* has been studied with respect to its role in both heart development and regeneration. In the developing embryo, RA is responsible for regulating the size of the cardiac progenitor pool (Stainier and Fishman, 1992) and zebrafish *aldh1a2* mutants generate an excess number of CMs. In the mouse, the loss of RA signalling particularly expanded the posterior SHF region (Sirbu et al., 2008). In the mouse, knockout of the SHF marker *tbx1* led to an expanded expression of *Aldh1a2* in the SHF compared with WT and the inverse was true in the *tbx1* gain-of-function mutant (Liao et al., 2008). It has also been shown in mice that RA restricts the size of the SHF progenitor pool by repressing *Fgf8* expression in the SHF (Sirbu et al., 2008). In the adult zebrafish heart, *aldh1a2* is expressed in the epicardium and is required for myocardial proliferation. Inhibition of RA signalling in the adult after injury impairs the proliferation of CMs, thereby impeding heart repair (Kikuchi et al., 2011).

Cluster 0 also showed enrichment for components of the Hippo signalling pathway, including Yes-associated protein 1 (*yap1*) and TEA domain family member (*tead1b*). The Hippo signalling pathway helps to define the number of cells in a variety of tissues and organs (Zhou et al., 2015). In zebrafish, *yap1* has recently been shown to determine the number of SHF-derived cells in the venous pole by promoting the expression of *bmp* factors which phosphorylate Smads in SHF progenitor cells (Fukui et al., 2018). *yap1* has been shown to regulate the expression of Heart and Neural crest derivatives expressed 2 (*hand2*) *(Fukui et al., 2018)*, a marker of cardiac progenitors that is also present in the cluster 0. *hand2* mutants show a dramatic CM deficit (Yelon et al., 2000), and *hand2* overexpression increases progenitor proliferation within the SHF (Schindler et al., 2014). Hippo signalling may be responsible for the restriction of SHF formation and the subsequent addition of SHF-derived cells to the inflow tract atrial CMs. *yap1* has also been linked to cell migration, with zebrafish expressing a dominant-negative form of *yap1* showing a cardia bifida phenotype (Miesfeld and Link, 2014). In the adult zebrafish, *yap1* deletion downregulates collagen expression and impedes scar formation and immune cell infiltration (Link et al., 2018).

In order to corroborate the single cell sequencing data with the imaging analysis presented in chapter 3, it was important to understand where these cells were located *in vivo*. Therefore, I chose some marker genes expressed in cluster 0 cells and performed HCR to determine their localisation within the developing embryo. The genes chosen for cluster 0 were the paired-like homeobox TF (*prrx1a*), junctional adhesion molecule 2b (*jam2b*), ADAM metallopeptidase with thrombospondin type 1 motif, 15a (*adamts15a*) and collagen, type I, alpha 1b (*col1a1b*). Figure 5 -5b shows that, at 48hpf, all of these chosen markers were

expressed in multiple tissues, with strong expression in the branchial arches. The cells in cluster 0 did not express *myl7*, so excluding the CMs, the region where each of these markers converged with *Gata5-E5-mCherry* expression lay in the extracardiac, venous pole-adjacent region. t-SNE plots (Figure 5 -5c and e) show the proportion of cells that coexpressed the chosen markers, with roughly a third of the cells in the cluster coexpressing both *prrx1a* and *jam2b* and about a quarter expressing both *jam2b* and *adamts15a*. This was confirmed by HCR images, with evidence of some cells coexpressing these markers and some expressing the individual markers alone (Figure 5 -5c, e, and h).

The progenitor cells of the SHF are known to express *isl1.* In chapter 3, I presented data showing Gata5-E5 activity in a population of cells that was interspersed with the Isl1-positive putative SHF cells in the region adjacent to the inflow tract at 48hpf. Upon deletion of the Gata5-E5 enhancer sequence, both of these populations were affected, and I hypothesised that these populations were no longer present. This was supported by the fact that Gata5-E5 knockout embryos developed a hypoplastic atrium, as one would expect if a proportion of the contributing SHF progenitor cells were depleted. These data suggest a connection between the Gata5-E5-mCherry-positive cells and the Isl1-positive putative SHF cells adjacent to the inflow tract, however, the identity of these Gata5-E5-mCherry-positive cells is unknown.

Initial validation by HCR for genes associated with cluster 0 showed colocalisation of marker genes within the extracardiac region adjacent to the venous pole of the heart, suggesting that cluster 0 represents the population of Gata5-E5-mCherry cells that was characterised in chapter 3.

The SHF in the mouse has been characterised as an epithelial sheet that undergoes EMT and tension changes during its accretion to the growing heart tube (Francou et al., 2017). Felker et al. (Felker et al., 2018) recently reported a similar phenomenon in the zebrafish; that the poles of the heart grow by the continuous addition of progenitors from an epithelial sheet. Although the cells in cluster 0 express the progenitor cell marker, *hand2,* and some cells also express the SHF marker *tbx1*, the lack of *isl1* expression suggests that these cells are not the SHF progenitor cells themselves. Rather, the epithelial, junctional adhesion and EMT markers expressed in cluster 0 cells suggest that this population may act as a substrate or as a continuous support tissue for the maintenance and/or migration of SHF progenitor cells. Hippo and RA signalling in the cells in cluster 0 may also act to regulate the proliferation of the SHF cells in this region.

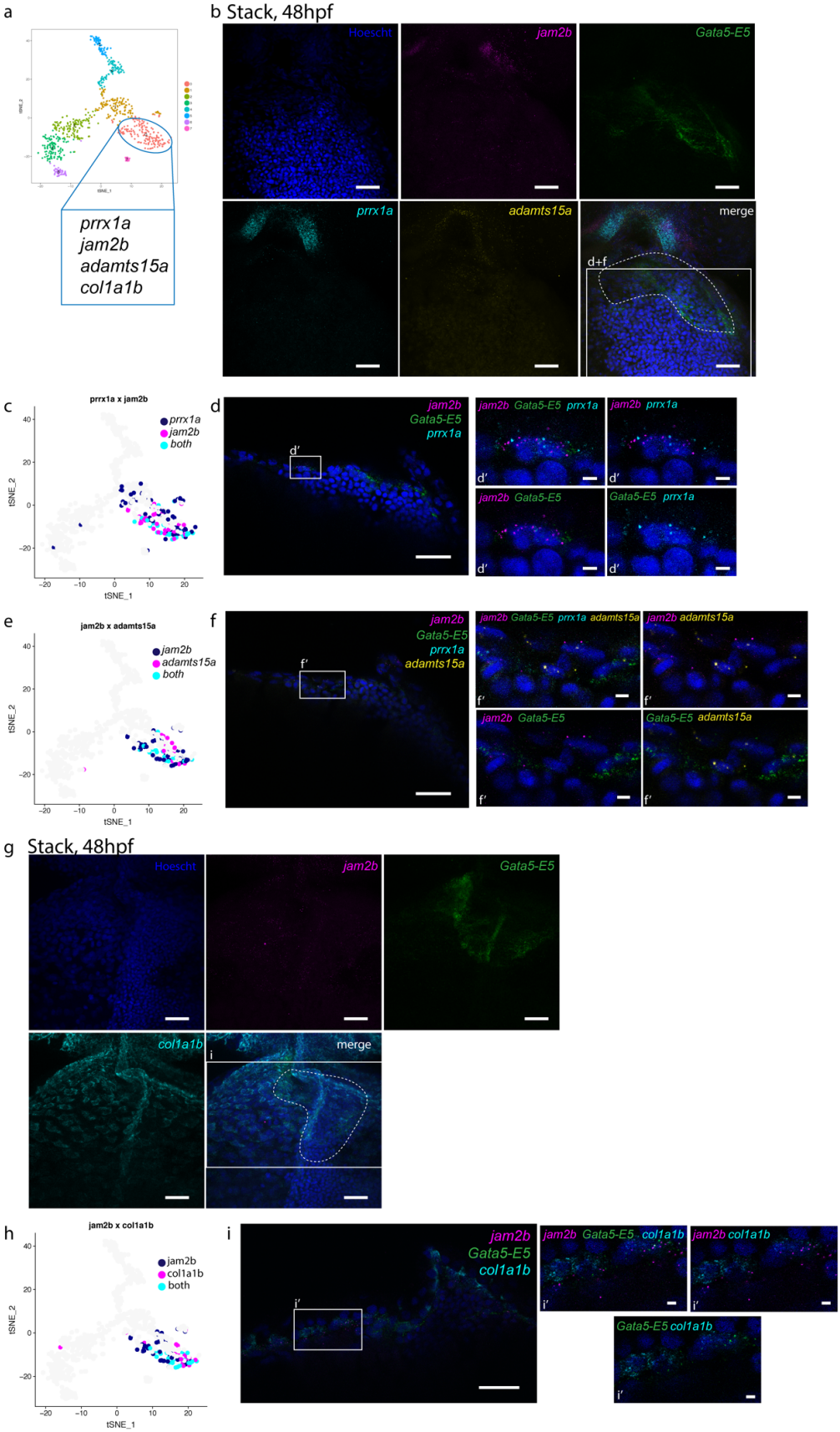

a
b Stack, 48hpf
Hoescht
jam2b
Gata5-E5
prrx1a
adamts15a
merge
d+f
prrx1a
jam2b
adamts15a
col1a1b
c
prrx1a x jam2b
prrx1a
jam2b
both
tSNE_2
tSNE_1
d
jam2b
Gata5-E5
prrx1a
d′
jam2b Gata5-E5 prrx1a
jam2b prrx1a
jam2b Gata5-E5
Gata5-E5 prrx1a
e
jam2b x adamts15a
jam2b
adamts15a
both
f
jam2b
Gata5-E5
prrx1a
adamts15a
f′
jam2b Gata5-E5 prrx1a adamts15a
jam2b adamts15a
jam2b Gata5-E5
Gata5-E5 adamts15a
g
Stack, 48hpf
Hoescht
jam2b
Gata5-E5
col1a1b
merge
i
h
jam2b x col1a1b
jam2b
col1a1b
both
i
jam2b
Gata5-E5
col1a1b
i′
jam2b Gata5-E5 col1a1b
jam2b col1a1b
Gata5-E5 col1a1b

Figure 5 -5 Transcripts encoding the cluster 0 markers prrx1a, jam2b, col1a1b and adamts15a were present in the extracardiac region adjacent to the inflow tract.

(a) t-SNE plot highlighting cluster 0 expression (b) Microscopic expression analysis of *prrx1a, jam2b,* and *adamts15a* in Tg*(Gata5-E5-mCherry*) embryos at 48hpf using hybridisation chain reaction (z-stack). Scale bars= 50µm. c, e and h) t-SNE plots highlighting the expression of marker genes in individual cells within cluster 0. d, f and i) show the extracardiac region in higher magnification single optical section. Scale bars= 50µm. Enlarged images in d', f' and i' visualise the close proximity of *prrx1a, jam2b, col1a1b* and *adamts15a* and *mCherry* transcripts surrounding individual cell nuclei (in blue) Scale bars= 5µm.

5.2.5 Cluster 1 was enriched in transcripts for genes associated with protein synthesis

Analysis of the predicted GO terms (Figure 5-4e) and the raw list of genes expressed by cells grouped into cluster 1 showed a population of 135 transitioning cells. One of the top expressed genes was the proliferation marker *pcna,* and a number of ribosomal protein genes were also expressed, including, 60S acidic ribosomal protein P0 (rplp0), ribosomal protein L3 (*rpl3*). The fact that cluster 1 represents transitioning cells was demonstrated by the population's central positioning in the t-SNE plot (Figure 5-4a), whereby it is clear that these are intermediate cells but the sequencing in the 10x platform is not deep enough to determine the identity of these cells above the fact that they are dividing.

5.2.6 Clusters 2 and 3 were enriched for transcripts inferring CM identity

Clusters 2 and 3 contained 123 and 118 cells, respectively. Figure 5 -3 shows that the gene expression profiles for the cells in clusters 2 and 3 were extremely similar. They both showed many of the hallmark properties of CMs, including *myl7,* cardiac myosin light chain-1 (*cmlc1*), natriuretic peptide A (*nppa*), natriuretic peptide B (*nppb*) and many troponin genes (*tnnc1a, tnnt2a* and *tnn1b*). A larger proportion of cells in cluster 3 expressed the atrium marker, *myh6,* showing that this population

was largely made up of atrial CMs (Figure 5 -3). This is also reflected in the GO term analysis (Figure 5-4). Based on the expression of the ventricular marker *vmhcl* and the atrial marker *myh6*, cluster 2 is a heterogeneous mixture of atrial and ventricular CMs. Some cells in cluster2 expressed *lyve1*. *Lyve1* is best known as a lymphatic marker (Banerji et al., 1999), but it has been shown in chick and mouse to be expressed at a low level in some other cell types, including CMs (Karunamuni et al., 2010). The CM character of clusters 2 and 3 supports the imaging analysis in chapter 2 showing the expression of Gata5-E5-mCherry in subpopulations of atrial and ventricular CMs.

5.2.8 Clusters 4, 5 and 7 were enriched for neural markers

Clusters 4, 5 and 6, comprising 85, 66 and 17 cells respectively, showed expression of predominantly neural markers, such as *tuba1c, sox3,* and *robo1,* which was surprising given the fact that I had not previously seen the expression of Gata5-E5-mCherry in any neural structures. Upon further investigation, I found that only one or two cells in each cluster expressed the *mCherry* transcript and none expressed *gata5* (Figure 5-4b and c) and so these cells are likely not cells of interest and, like many of the clusters that were discarded upon subclustering from the original cohort, were aberrantly sorted by FACs.

5.2.7 Cluster 6 was enriched in transcripts encoding markers of the cardiac conduction system

Although the sinus node was characterised histologically and functionally over a century ago, we are only now beginning to understand the molecular regulators of this cell population. The specialised regions of the heart that make up the CCS possess a molecular signature that can be defined by expression of particular TFs, ion channels and gap junction genes, separating them from the surrounding working

myocardium (Moorman and Christoffels, 2003). In the mouse, TBX3 has been shown to be required for sinus node and AVN development and subsequent homeostasis (Hoogaars et al., 2007). Other TFs including Isl1, Shox2 and Irx2 have also been implicated in cardiac pacemaker cell development.

Although the zebrafish CCS is not as well characterised, several factors have been implicated in its development that give an indication of markers to search for in my single cell datasets. The SAN cells have a CM profile early in development, therefore express the CM marker, *myl7* (Tessadori et al., 2012). Unlike what has been observed in mammals, the adult zebrafish sinus venosus does not express *myl7*. *isl1* is expressed in the SAN throughout development into adulthood. *bmp4* and *tbx2b* are also expressed at the venous pole of the developing heart (de Pater et al., 2009), and as the heart matures, the Bmp4-Tbx2b regulatory axis suppresses *nppa* expression and other chamber differentiation markers. *bmp4* and *tbx2b* expression is maintained in the adult SAN (Tessadori et al., 2012). The adult SAN also expresses the Hcn4 ion channel, a channel responsible for hyperpolarisation-activated currents in the SAN. The developing zebrafish SAN has also been shown to express *shox2*, and MO-mediated knockdown of *shox2* is characterised by severe sinus bradycardia and sinus exit block, demonstrating its importance in SAN formation. Recently, Poon et al. developed an enhancer trap zebrafish line that expressed a fluorescent protein in the pacemaker cells at the SAN (Poon et al., 2016). After mapping the transposon integration site, they deduced that the gene represented by the trapped enhancer was *fhf2a,* based on proximity and *fhf2a* RNA expression pattern at 3dpf. However, as discussed in chapter 1, enhancers can regulate genes over large genomic distances, so proximity does not confer direct regulation. Chromatin conformation capture would be required to directly link the

trapped enhancer to the *fhf2a* gene. The authors described the expression pattern of the newly generated reporter and proposed a function for *fhf2a* in the early conduction system but did not take this study further by sequencing the cells to better understand the transcriptomic profile of the pacemaker cells.

Cluster 6 comprised 42 cells and showed expression of the hallmarks one would expect from a CM, including *myl7,* troponin T type 2a (*tnnt2a),* and myosin light chain kinase *(mylk3)*. *myh6* was one of the top expressed genes in this cluster (as seen in the heatmap in Figure 5 -3) and this is an atrial-specific gene, showing that the cells within this cluster are atrial CMs. Cluster 6 was enriched for genes involved in 'positive regulation of heart contraction', with the top enriched genes being the Calcium-transporting ATPase (*atp2a2a)* and the Solute carrier family 8 (sodium/calcium exchanger), member 1a (*slc8a1a).* This profiling supports the hypothesis that cluster 6 represents the population of putative CCS cells predicted from the imaging analysis presented in chapters 2 and 3.

Further supporting this hypothesis, the Voltage-dependent T-type calcium channel subunit alpha (*cacna1g)* was highly expressed in the cells of cluster 6. T-type channels generate transient currents and are thought to be involved in pacemaker activity. In mice, *CACNA1G* inactivation caused bradycardia and prolonged the SAN recovery time (Mangoni Matteo et al., 2006). The cells in cluster 6 also expressed the genes that have been well characterised in SAN tissue previously; *isl1, shox2, bmp4* and *tbx2b.* Although *hcn4* did not appear to be expressed, *hcn2,* a channel found in human pacemaker cells (Vaccari et al., 1999), was highly expressed. In order to validate this transcriptomic profile, I selected markers expressed in the cluster 6 subpopulation (*shox2, sema3fb, col4a1*) and characterised their expression pattern within the developing embryo by HCR (Figure 5 -6). *shox2,*

sema3fb and *col4a1* expression colocalised with *Gata5-E5-mCherry* in some cells of the cardiac inflow tract, further supporting the hypothesis that cluster 6 represents a population of early CCS cells present in the inflow tract of the heart at 48hpf.

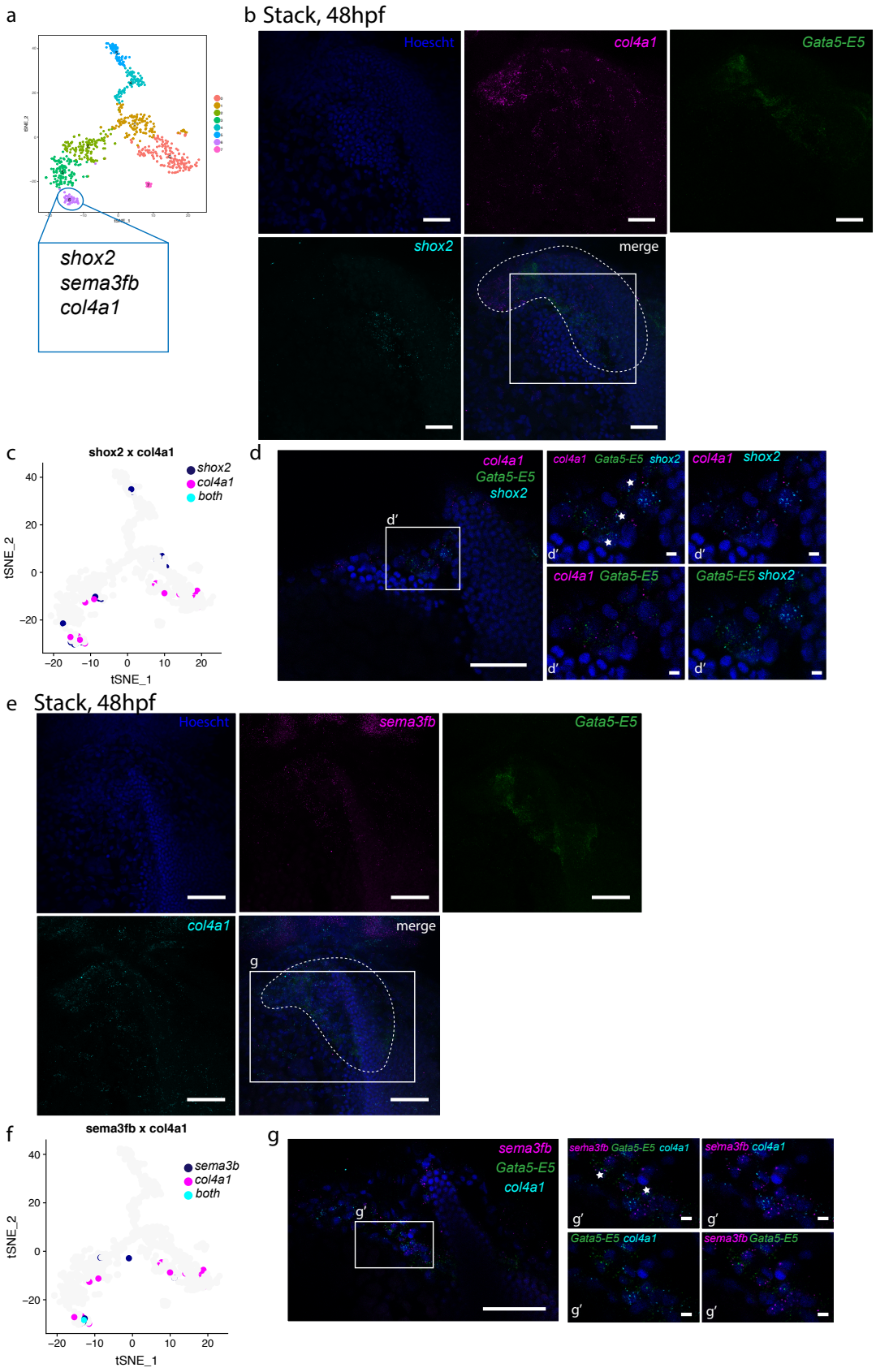

a
b Stack, 48hpf
Hoescht
col4a1
Gata5-E5
shox2
merge
shox2
sema3fb
col4a1
c
shox2 x col4a1
shox2
col4a1
both
tSNE_2
tSNE_1
d
col4a1
Gata5-E5
shox2
d′
e Stack, 48hpf
Hoescht
sema3fb
Gata5-E5
col4a1
merge
g
f
sema3fb x col4a1
sema3b
col4a1
both
tSNE_2
tSNE_1
g
sema3fb
Gata5-E5
col4a1
g′

Figure 5 -6 Transcripts encoding the cluster 6 markers shox2, sema3fb, col14a1 were present in the inflow tract.

a) t-SNE plot highlighting cluster 6 expression (b) Microscopic expression analysis of *shox2, sema3fb* and *col14a1* in Tg(*Gata5-E5-mCherry*) embryos at 48hpf using hybridisation chain reaction (z-stack). Scale bars= 50µm. c and f) t-SNE plots highlighting the expression of marker genes in individual cells within cluster 6. d and g) Show the inflow tract region in a higher magnification single optical section. Scale bars= 50µm. Enlarged images in d' and g' visualise the close proximity of *shox2*, *sema3fb*, *col4a1* and *mCherry* transcripts surrounding individual cell nuclei (in blue, cells coexpressing all markers marked with a white star) Scale bars= 5µm.

5.3 Summary

I performed single cell RNA sequencing to uncover subpopulations of cells in which the Gata5-E5 enhancer was active at 48hpf. After quality control, I found that the Gata5-E5 cherry expression domain comprised 8 transcriptomically distinct cell populations.

The largest cluster, cluster 0, included cells that expressed a panel of epithelial markers (e.g. *cdh2* and *fn1a*), as well as EMT markers (e.g. *twist1a* and *twist1b*). I chose the marker genes *prrx1a*, *jam2b, adamts15a* and *col1a1b* and performed HCR to determine the localisation of cells assigned to cluster 0. These genes colocalised with each other and with *Gata5-E5-mCherry* in the extracardiac region adjacent to the inflow tract, leading me to believe that cluster 0 represents the population of cells described in chapter 3 and 4 that were in close proximity to the *isl1*-positive progenitors of the SHF. Their transcriptional profiles, along with the evidence presented in chapter 4 that deletion of Gata5-E5 caused a reduction in SHF progenitors, suggests that these cells may act as a substrate or support cells for the SHF progenitor cells as they differentiate and migrate to the venous pole.

The population of cells termed cluster 1 was highly enriched in protein synthesis components and proliferation markers. I therefore concluded that cluster 1 is a population of transitioning cells, but unfortunately the sequencing depth permitted

by the 10x platform does not allow sufficient information to be gained about the cell identity of this highly proliferative group of cells.

Clusters 2 and 3 contained cells labelled by known CM marker genes, such as *myl7, cmlc1, nppa* and *tnnc1a*. Based on the expression of the atrial marker, *myh6,* and the ventricular marker, *vmhc*, cluster 2 contained a heterogeneous population of atrial and ventricular CMs, while cluster 3 cells were predominantly atrial CMs. These populations were expected based on the imaging data presented in chapter 4, showing Gata5-E5-mCherry expression in CMs. Future studies should include *in vivo* validation of some markers expressed in each population to understand the precise localisation and character of cells from each cluster.

The populations of cells termed clusters 4, 5 and 7 were all enriched for transcripts inferring neural identity. These cells had been reclustered based on their expression of the *mCherry* transcript, but on closer inspection, mCherry had only been detected in one or two cells in each cluster. This suggests that these cells were aberrantly FAC-sorted, and they do not in fact show activity of the Gata5-E5 enhancer.

Imaging analysis in chapters 3 and 4 suggested that Gata5-E5 was active in a population of *isl1*-positive CMs in the inflow tract of the heart, predicted to be the early SAN of the CCS. Cluster 6 showed enrichment of transcripts identifying the cells as atrial CMs, in addition to a panel of CCS marker genes, including *shox2, bmp4* and *tbx2b*.

As well as aiding to identify and characterise the cell populations within which Gata5-E5 is regulating *gata5* expression, this single cell data provides a detailed

list of candidate TFs for the upstream control of this activity in specific subpopulations of the heart at 48hpf. This upstream regulation is further explored in chapter 6.

6) Binding site perturbation reveals upstream regulation of Gata5-E5

6.1 Introduction

Enhancer sequences contain multiple binding sites for TFs which help to regulate either activation or repression of their associated genes. It is important to highlight that developmental enhancers are not simply transcriptional on/off switches, but they undergo progressive changes throughout development depending on their accessibility and TF occupancy, and these changes are essential for proper spatiotemporal regulation of gene expression (Lee et al., 2015). TF binding sites typically comprise small 6-12bp degenerate DNA sequences. TFs specifically recognise and bind these binding sites, or motifs, and can then recruit co-activators or co-repressors and the regulatory cues of all bound factors contribute to the activity of the enhancer and the functional outcome (Shlyueva et al., 2014b).

When investigating enhancer-TF interactions, two properties are important; motif composition and motif positioning (sometimes known as motif grammar) (Long et al., 2016). TF binding motifs are commonly represented as a position weight matrix, which demonstrates the affinity of a given TF for each base within the motif. The DNA binding activity of a given TF can be investigated *in vivo* using ChIP-Seq methods, or in vitro using methods such as yeast one hybrid screens and protein-

binding microarrays (Geertz and Maerkl, 2010). Details from investigations into TF binding specificity are then systematically collated in motif databases such as the JASPAR database (Anthony Mathelier, 2016). These databases allow motifs to be predicted from candidate DNA sequences. Motif positioning refers to the order and placement of a TF motif within an enhancer and this can also be very important for the functional activity of some enhancers. Other enhancers adhere to the more flexible 'billboard' model, whereby the sequence order of the motifs is less important for the function of the enhancer. In a recent study by Grossman et al. (Grossman et al., 2018), the authors utilised published datasets to probe the positional bias of predicted binding sites for a panel of TFs in 47 different cell types. They discovered that different classes of TFs display different binding preferences, and these arrangements are highly consistent across all cell types studied, suggesting that the position of a binding site within an enhancer may reflect the role of that TF in the activity of the enhancer. For example, TFs with binding sites towards the edges of the enhancers tend to be those known to interact with histone or chromatin remodellers. In developmental enhancers, it is common to find clusters of motifs for the same TF family appearing in tandem, known as homotypic clusters (Valer Gotea, 2010). It is thought that homotypic clusters represent a robust evolutionary mechanism to ensure the binding of those TFs that are essential for enhancer activity.

In order to understand the upstream regulatory activity of *gata5,* it was important to gain an understanding of the TFs that can drive the activity of *Gata5-E5.* To this end, in this chapter I describe both the prediction and perturbation outcome of TF binding sites within Gata5-E5. By understanding the binding interactions necessary for the activity of Gata5-E5, and the spatiotemporal localisation of these

interactions, I can begin to dissect the upstream regulation of *gata5* via Gata5-E5 in specific subpopulations.

6.2 Results

6.2.1 Candidate TF Selection

The active core of enhancers has been shown to be located in the centre of the open chromatin region (Ovcharenko et al., 2004). Using the JASPAR database, I was able to produce a comprehensive list of 656 predicted binding sites within the Gata5-E5 core sequence, identified as the genomic region directly under the ATAC-Seq peak. This list was then used to select candidate TF-enhancer interactions to further study based on the following criteria devised to select for TF motifs most likely to be functional *in vivo*:

- **Expression in the relevant cell type**. Informed by my single cell RNA sequencing data, I excluded any TFs that were not expressed in the appropriate cells, to focus on candidate TFs that are likely to play a functional role *in vivo*.
- **Multiple predicted binding sites for the same TF family**. Predicted binding motifs that appear as homotypic clusters are predicted to be of importance for the activity of the enhancer (Valer Gotea, 2010).
- **Longer TF binding motifs are preferable.** The longer the motif sequence, the more likely that the motif is functional, as there is a reduced probability of the given bases being there by chance.
- **Literature review indicative of a role in heart development**

All TF families that met at least 2 of these criteria were selected for further studies (Table 6).

Table 6 Predicted binding sites within Gata5-E5

TF Family	Number of putative binding sites	TFs present in single cell transcriptomic data set
Gata	6	*gata5, gata6*
Meis	10	*meis1a, meis1b*
Nkx	6	*nkx2.5, nkx2.7*
Tbx	2	*tbx1, tbx5a, tbx2a, tbx2b, tbx3a, tbx20*
Prrx	4	*prrx1a*
Shox	4	*shox2*
Fox	5	*foxp4, foxp1a*
Snai	3	*snai1a, snai2*

It was then important to validate these interactions *in vivo* to identify the TFs that act upstream to regulate *gata5* expression via Gata5-E5. To do this, the same minimal promoter reporter system used for the generation of the enhancer reporter lines was employed. For each TF family predicted to bind to Gata5-E5, I designed mutations in all putative motifs corresponding to each member of that TF family (Figure 6-1). It was important to maintain the genome architecture so, rather than deleting the whole motif, individual nucleotides were substituted, using the PWM for each motif as a guide to target the nucleotides with the highest affinity for the TF. Each motif perturbation was then re-checked using JASPAR to ensure that not only the predicted binding of the TF family was eliminated, but also that no new predicted binding sites for other TFs had been created during the mutagenesis process. A

panel of enhancer sequences, each with perturbations in the binding sites for a different TF family, were ordered as gBlocks (short synthetic double-stranded DNA) and cloned into the Ac/Ds minimal promoter reporter vector. This useful tool could then be used to inform the importance of specific binding events over time during development. If there was no difference between the reporter activity driven by the non-mutated enhancer and the mutated enhancer, then we could conclude that the associated TF-enhancer interaction was not necessary for the activity of the enhancer. On the other hand, if the TF-enhancer interaction was important then we could learn about the precise timing and location of this regulation by observing which cells might lose fluorescence when compared with the non-mutated enhancer-driven reporter. Because of the transient nature of this assay, I expected some variation between embryos. However, by looking at many embryos, across multiple time points, I was able to observe and distinguish patterns of enhancer-driven reporter disruption. The following results inform on potential TF-enhancer interactions, unravelling the upstream regulation of *gata5* during cardiogenesis.

However, all of these results must be further confirmed using ChIP to definitively demonstrate *in vivo* TF binding to Gata5-E5 at the predicted time of development.

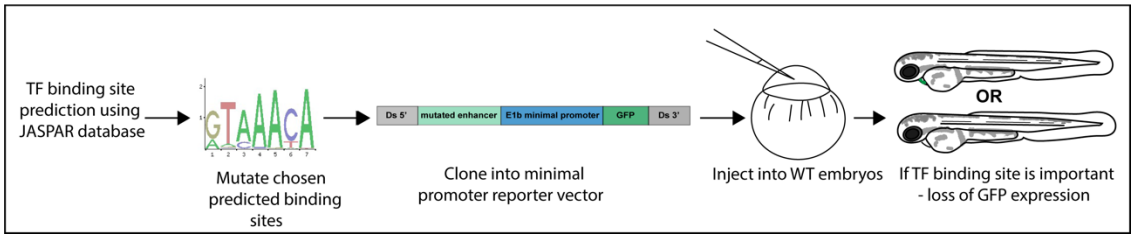


Figure 6-1 Workflow for binding site perturbation studies

Gata5-E5 sequence was analysed using the JASPAR database. Selected binding sites were perturbed, using the TF's PWM at a guideline, and mutated sequences cloned into the AcDs minimal promoter reporter vector and injected into WT embryos for analysis.

6.2.2 Gata5 was necessary for the activity of Gata5-E5 during cardiogenesis from 2dpf

I could identify 6 predicted GATA binding motifs within the Gata5-E5 core sequence. The binding sites predicted by JASPAR were for Gata4 and Gata5. As the JASPAR database is generated mainly from mouse work and from in vitro studies, the exact TF predictions might not always be accurate for zebrafish. According to my single cell RNA-Seq data, *gata4* was not expressed in any of the cells of interest, however both *gata5* and *gata6* were strongly expressed, particularly in clusters 0, 2, 3 and 6, the clusters predicted to be support cells for SHF progenitor cells, CMs, and CCS cells, respectively (as demonstrated using t-SNE plots in Figure 6 -2).

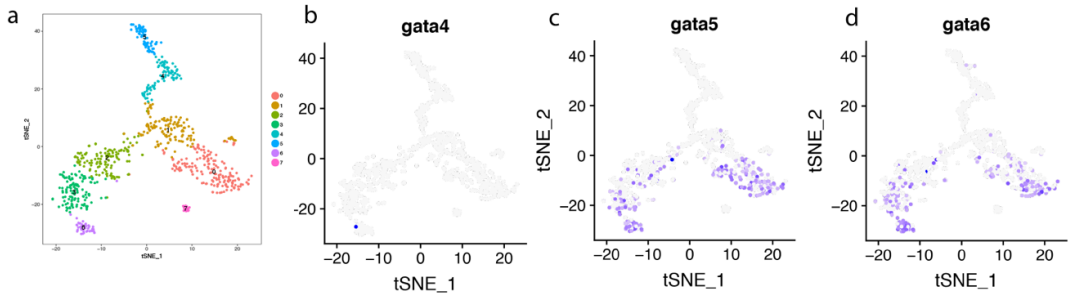


Figure 6 -2 gata5 and gata6 are expressed in clusters 0, 2, 3 and 6.

a) t-SNE plot showing each cluster of interest. b) gata4 was not expressed in any of the Tg(Gata5-E5-mCherry) sorted cells. c) gata5 was expressed in clusters 0, 2, 3, and 6. d) gata6 was expressed in clusters 0, 2, 3 and 6.

Upon substitution of only one or two nucleotides, selected using the GATA motif PWM, enhancer-driven reporter expression was comparable to the non-mutated control at 24hpf. However, by 48hpf, reporter expression was completely abrogated throughout the whole heart (Figure 6 -3a). This was observed in 3 separate experiments, each having 100 analysed embryos. Lack of fluorescence was also confirmed using confocal microscopy (Figure 6 -3a). I could therefore deduce that the binding of a GATA TF was necessary from 48hpf onwards for the activity of Gata5-E5. From the previous observation that deletion of Gata5-E5 led to a decrease in endogenous *gata5* levels and a decrease in Gata5-E5 reporter activity (Figure 4-7), I hypothesise the presence of an autoregulatory loop whereby Gata5 binding is needed for the activity of the Gata5-E5 enhancer.

As part of a strategy to produce a vector whereby Gata5 was tagged with an Avi tag for ChIP-Seq, I designed a vector to express Gata5 under the control of the Gata5-E5 enhancer. It was therefore possible to overexpress Gata5 specifically in the cells that have Gata5-E5 activity. By injecting this construct into Tg(Gata5-E5-mCherry; myl7-GFP) embryos, enhancer activity could be monitored. Gata5 overexpression, confirmed by in situ hybridisation (Figure 6 -3b) led to a significant increase in

enhancer-driven reporter expression in the cells within and adjacent to the inflow tract; the opposite effect to that seen in the same population following Gata5-E5 knockout. This serves as further confirmation of a feedback loop whereby Gata5 maintains its own expression via the Gata5-E5 enhancer in the cells in which Gata5-E5 is active.

The Gata5-E5 non-mutated control drove reporter activity in the atrium, particularly around the inflow tract, in select cells in the AVC and in a number of cells in the ventricle at all of the time points monitored (2, 3 and 5dpf). In these transient experiments, where expression is mosaic, I only saw expression of the enhancer-driven reporter within the heart, and not in the extracardiac region that showed Gata5-E5 activity in the stable *Tg(Gata5-E5-mCherry)* line. For this reason, I focused on the Gata5-E5-GFP-positive CM populations when performing the binding site studies presented in this chapter.

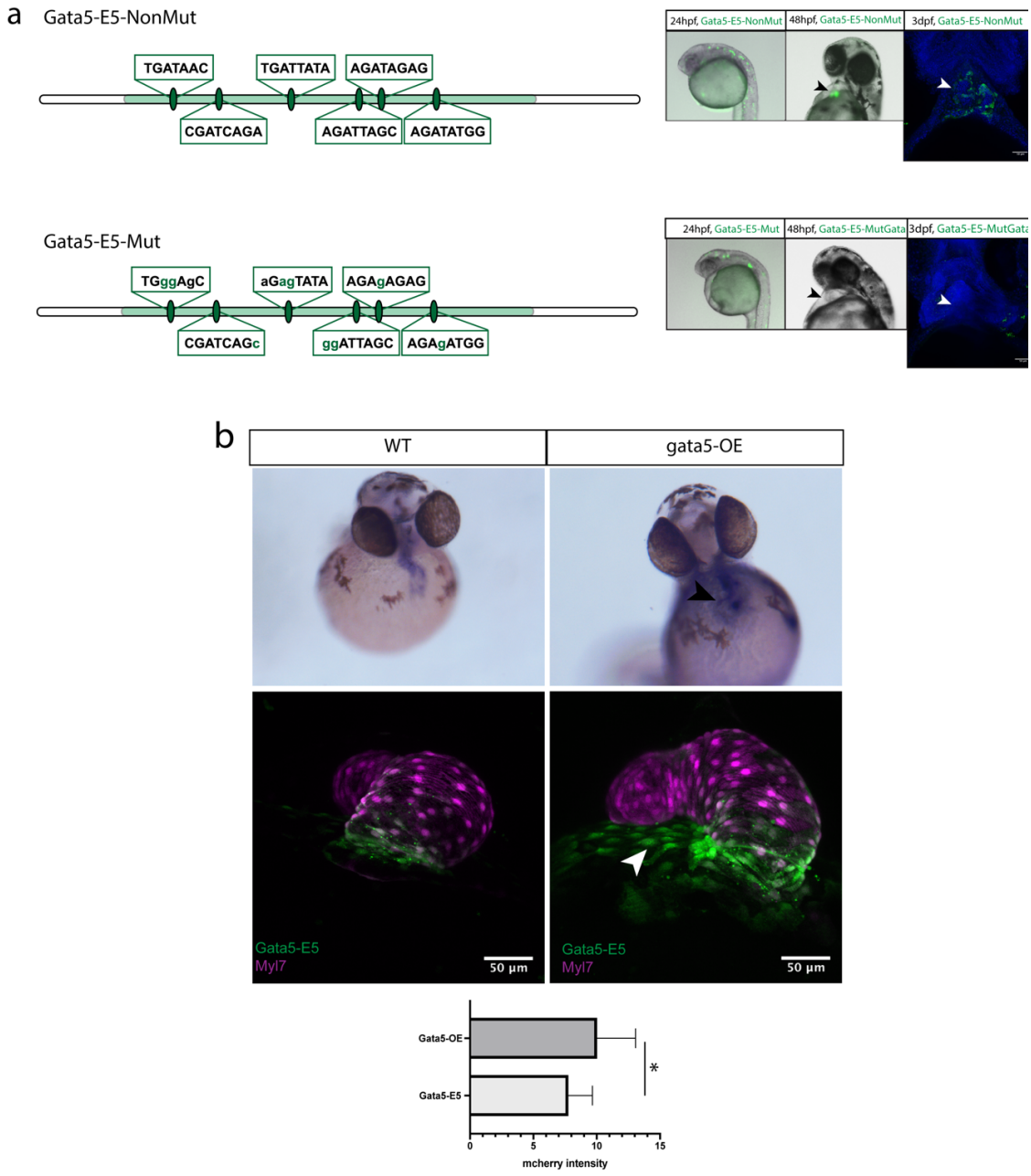


Figure 6 -3 Gata5 binding is necessary for the activity of Gata5-E5 from 2dpf.

a) 1-2 base substitutions were made in the 6 predicted Gata-binding sites (green font). Fluorescence was comparable between Gata5-E5-MutGata and non-mutated control at 24hpf. By 2dpf, heart-specific expression of Gata5-E5-non-mutated-driven fluorescence could be seen in the heart (black arrowhead), but no reporter was expressed in Gata5-E5-MutGata injected embryos. This was confirmed by confocal analysis at 3dpf (white arrowheads). b) Gata5 was overexpressed, driven by the Gata5-E5 enhancer in Tg(Gata5-E5-mCherry; myl7-GFP) embryos. *in situ* analysis confirmed *gata5* overexpression (black arrowhead). Confocal imaging showed an increase in Gata5-E5-driven reporter expression around the inflow tract region (white arrowhead) (n=18). Significance was calculated using an unpaired Student's t-test. *= $p<0.0332$.

6.2.3 Meis-Gata5-E5 binding interactions were required in the ventricle and AVC populations where Gata5-E5 activity was reported

Members of the TALE-class homeodomain TF family, Meis (Meis1 and 2) were predicted by JASPAR to bind to the Gata5-E5 sequence. My single cell transcriptional analysis showed expression of *meis1a* and *meis1b* (Figure 6-4). *meis1a* was particularly prevalent in cluster 0, but was also seen in cluster 1, 2, 3, 6 and 7. *meis1b* was more widely expressed and transcripts were detected in all 8 clusters investigated.

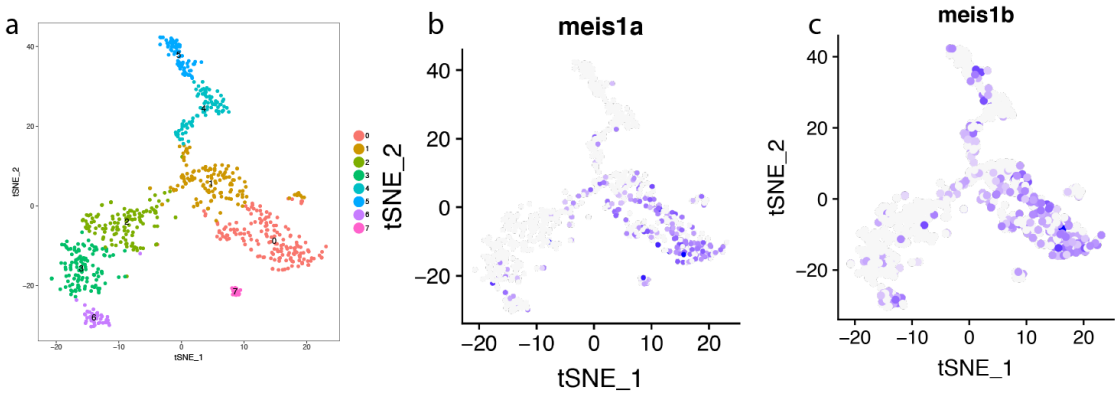


Figure 6-4 meis1a and meis1b were expressed throughout the cell clusters studied.

a) t-SNE plot showing each cluster of interest. b) *meis1a* was predominantly expressed in cells of clusters 0 and 1 c) *meis1b* was expressed throughout the clusters but particularly strongly in clusters 0 and 6.

Previous studies have highlighted the importance of Meis factors in heart development. *Meis1*-knockout mice develop outflow tract abnormalities and ventricular septal defect (Stankunas et al., 2008, Minehata et al., 2008). Overlapping Nkx2.5/Meis1 binding motifs have been recorded in a large number of murine cardiac enhancers, whereby the sequential binding of each TF confers spatiotemporal tuning of enhancer activation (Dupays et al., 2015). In the zebrafish, Meis1 and 2 are essential for proper heart tube formation and subsequent looping

(Paige et al., 2012). *Meis2b* morphant embryos also displayed a reduced heart rate suggesting a potentially unexplored role in cardiac conduction. 10 binding sites for the Meis TF family were predicted within the Gata5-E5 sequence (of note, many were adjacent to predicted binding sites for Nkx factors, akin to what has been described in the literature), I mutated 1-3 nucleotides in each of the Meis motifs in order to disrupt their binding specificity. As reported earlier, the Gata5-E5 non-mutated control drove reporter activity in the atrium, particularly around the inflow tract, in select cells in the AVC and in a number of cells in the ventricle at all of the time points monitored (2, 3 and 5dpf). After disruption of the putative Meis binding sites within Gata5-E5 sequence, significantly fewer fluorescent cells were counted in both the ventricle and the AVC at each time point monitored (Figure 6-5b), when compared to the non-mutated control. Enhancer-driven reporter activity was unaffected in the atrium at 2dpf and 3dpf, but significantly fewer GFP-positive cells were seen in 5dpf embryos. These results suggest that member(/s) of the Meis family of TFs were necessary for the activity of Gata5-E5 in the AVC and ventricular populations throughout early development and in the atrial population at 5dpf. Based on the single cell RNA sequencing results, I hypothesise that the candidate Meis factor is Meis1b as it was shown to be expressed in the CM populations of

cluster 2 and 3 (Figure 6-4). ChIP experiments will test this hypothesis.

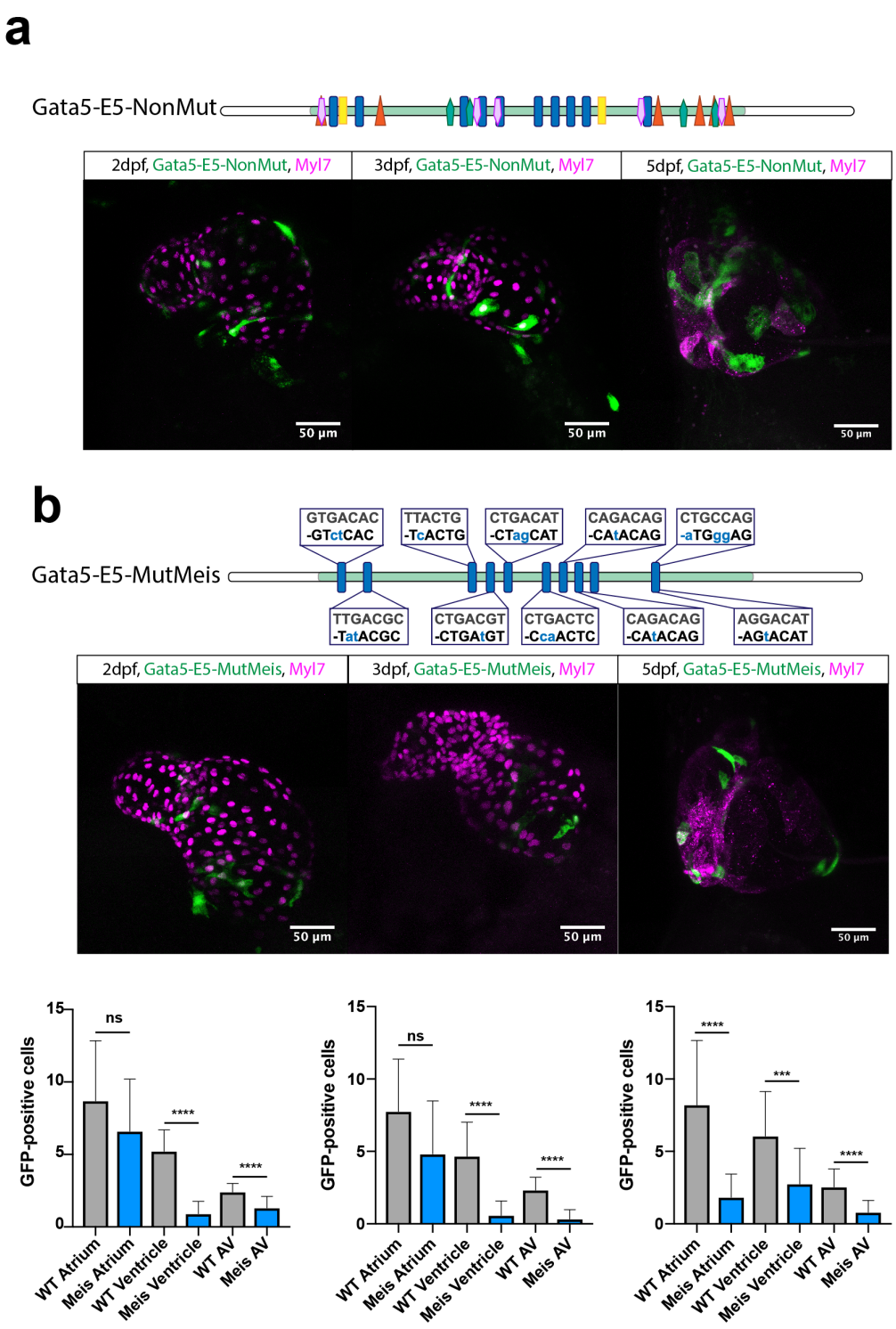


Figure 6-5 Meis binding was necessary for the activity of Gata5-E5 in the ventricle and AVC.

a) Microscopic analysis of Gata5-E5-driven reporter expression, quantified by quantification of GFP fluorescent cells at 2dpf, 3dpf and 5dpf. Expression patterns of

the enhancer-driven reporter are represented in the diagrams below. b) 1-3 base substitutions were made in the 10 predicted Meis-binding sites (blue font). Significantly fewer GFP fluorescent cells were observed in the ventricle and AVC at each time point measured and in the atrium at 5dpf, when compared to non-mutated control embryos. Significance was calculated using an unpaired Student's t-test. *= p<0.0332, **= p<0.0021, ***= p<0.0002, ****= p<0.0001. 2dpf n=20, 3dpf n=12, 5dpf n=24.

6.2.4 NKX-Gata5-E5 binding interactions were required in the atrium and AVC populations

The homeobox protein Nkx2.5 is one of the earliest TFs to be expressed in the cardiac lineage and its role is highly conserved across many species (Lien et al., 1999). The Drosophila homolog of *nkx2.5*, *tinman,* was the first gene of this family to be identified as important for cardiogenesis, as the dorsal vessel (the Drosophila heart-like structure) was not specified in *tinman* mutant embryos (Bodmer, 1993). In zebrafish, *nkx2.5* and *nkx2.7* are expressed in the precardiac mesoderm, cardiac region and developing heart (Chen and Fishman, 1996, Lee et al., 1996). NKX2.5 has an important role in many of the vital processes needed for successful cardiogenesis; it regulates the size and proliferation of cardiac progenitor populations, terminal differentiation of the myocardium and specification of the conduction system. Nkx2.5 regulates the maintenance of working myocardial genes (such as *nppa*), through the repression of *isl1,* thereby regulating the regionalisation of the SAN markers (including *shox2, hcn4, bmp4* and *tbx2b*) to the inflow tract of the developing heart (Colombo et al., 2018). In support of this hypothesis, regionalisation of the SAN is disrupted in *nkx* mutant zebrafish embryos and SAN markers are expressed throughout the working myocardium. Inhibition of *isl1* splicing in these embryos is sufficient to rescue the phenotype and cause downregulation of the SAN genes (Colombo et al., 2018). Overlapping

Nkx2.5/Meis1 binding motifs have also been recorded in many murine cardiac enhancers, whereby the sequential binding of each TF confers spatiotemporal tuning of enhancer activation (Dupays et al., 2015).

The Gata5-E5 sequence comprises 6 putative binding sites for the Nkx family of TFs, including Nkx2.3, Nkx2.5, and Nkx2.7. These binding sites appear in tandem and cluster towards each edge of the core of the enhancer. My single cell RNA sequencing data showed expression of both Nkx2.5 and Nkx2.7 in clusters 2, 3 and 6, which are CMs and putative CCS cells, respectively (Figure 6-6b, Figure 6-6c). Using HCR, I have confirmed that *nkx2.5* is expressed in CMs of the inflow tract, the AVC and the ventricle (Figure 6-6) at 48hpf, and the expression overlaps with that of the Gata5-E5-driven reporter.

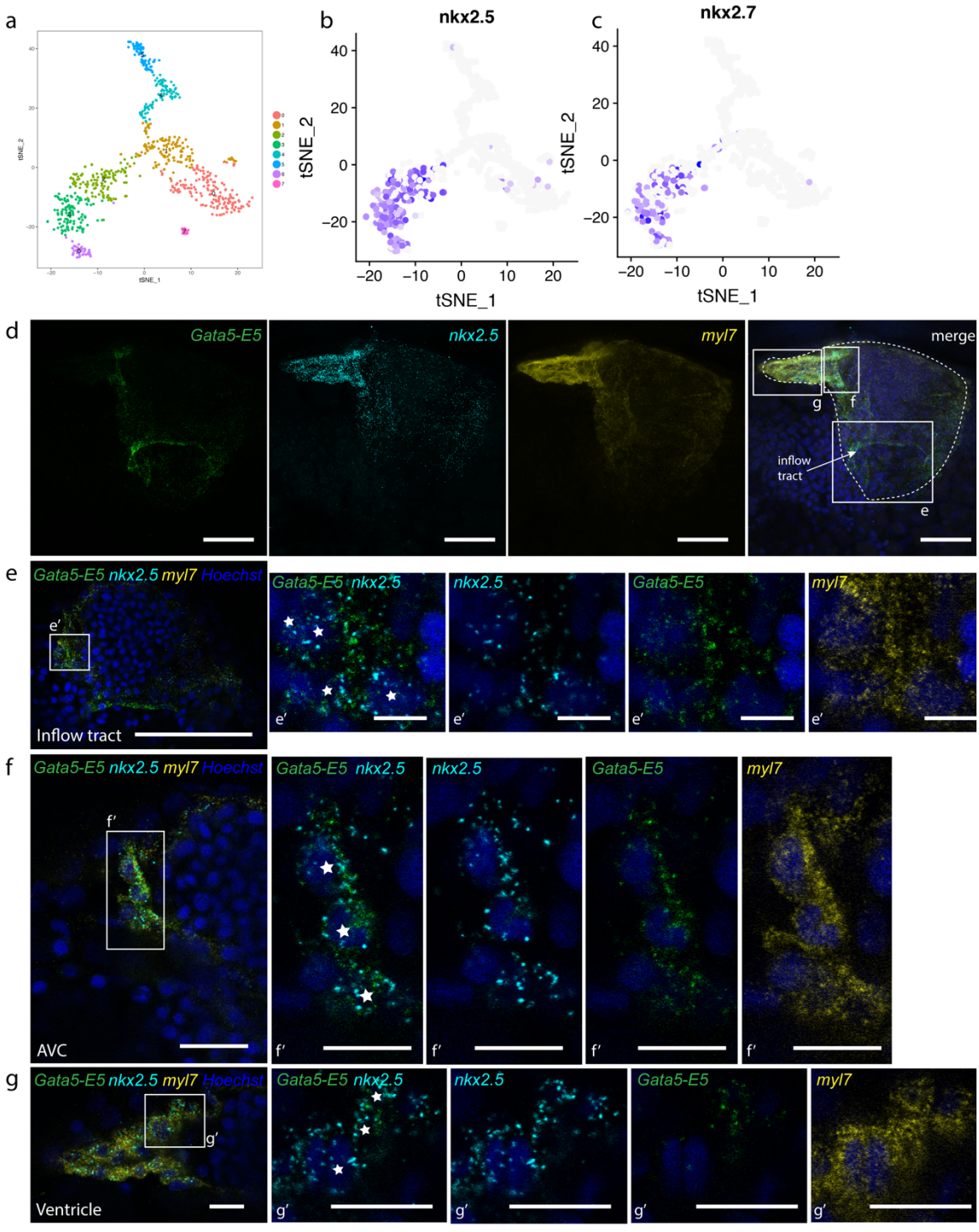

a
b
nkx2.5
c
nkx2.7
tSNE_2
tSNE_1
d
Gata5-E5
nkx2.5
myl7
merge
g
f
inflow
tract
e
e
Gata5-E5 nkx2.5 myl7 Hoechst
e′
Inflow tract
Gata5-E5 nkx2.5
nkx2.5
Gata5-E5
myl7
f
f′
AVC
g
g′
Ventricle

Figure 6-6 Nkx2.5 was strongly expressed in clusters 2, 3 and 6

a) t-SNE plot showing all clusters of cells in which Gata5-E5 is active. b and c) *nkx2.5* and *nkx2.7* were expressed throughout clusters 2, 3 and 6, which are CMs and putative CCS cells, respectively. d) Microscopic expression analysis of *nkx2.5* in Tg(*Gata5-E5-mCherry)* embryos at 48hpf using hybridisation chain reaction (z-stack). Scale bars= 50µm. e, f and g) Show the inflow tract region (Scale bars= 50µm) (e), the AVC (Scale bars= 10µm) (f) and the ventricle (Scale bars= 10µm) (g) in higher magnification single optical sections. Enlarged images in e', f' and g' visualise the close proximity of nkx2.5 and mCherry transcripts surrounding individual cell nuclei (in blue, examples of coexpressing cells marked with a white star). Scale bars= 5µm.

Perturbation of the Nkx binding sites led to a dramatic and significant loss of activity of Gata5-E5 within the atrial cell population (both in the putative CCS and the rest of the atrial CMs) and the AVC population at all time points studied (Figure 6 -7b), suggesting that Nkx factors act upstream of the enhancer during cardiogenesis in these populations. This result warrants further investigation as when the CCS cells begin to mature, they lose expression of *nkx2.5,* although it is not known at which time point this transition occurs. It may be that *nkx2.7* is necessary at these later time points. HCR for *nkx2.5* and *nkx2.7* at 3 and 5dpf would help to clarify which TFs are important in the putative CCS cells. Significantly fewer GFP-positive cells were also present in the ventricle after Nkx binding site perturbation at 2dpf, however, fluorescence seemed to return by 3dpf, suggesting that Nkx binding is only necessary in the ventricular population very early on in development. As Nkx2.5 and Nkx2.7 are both expressed in the subpopulations of interest (Figure 6-6), it will be important to perform ChIP experiments to determine which of these factors might be acting upstream of Gata5-E5.

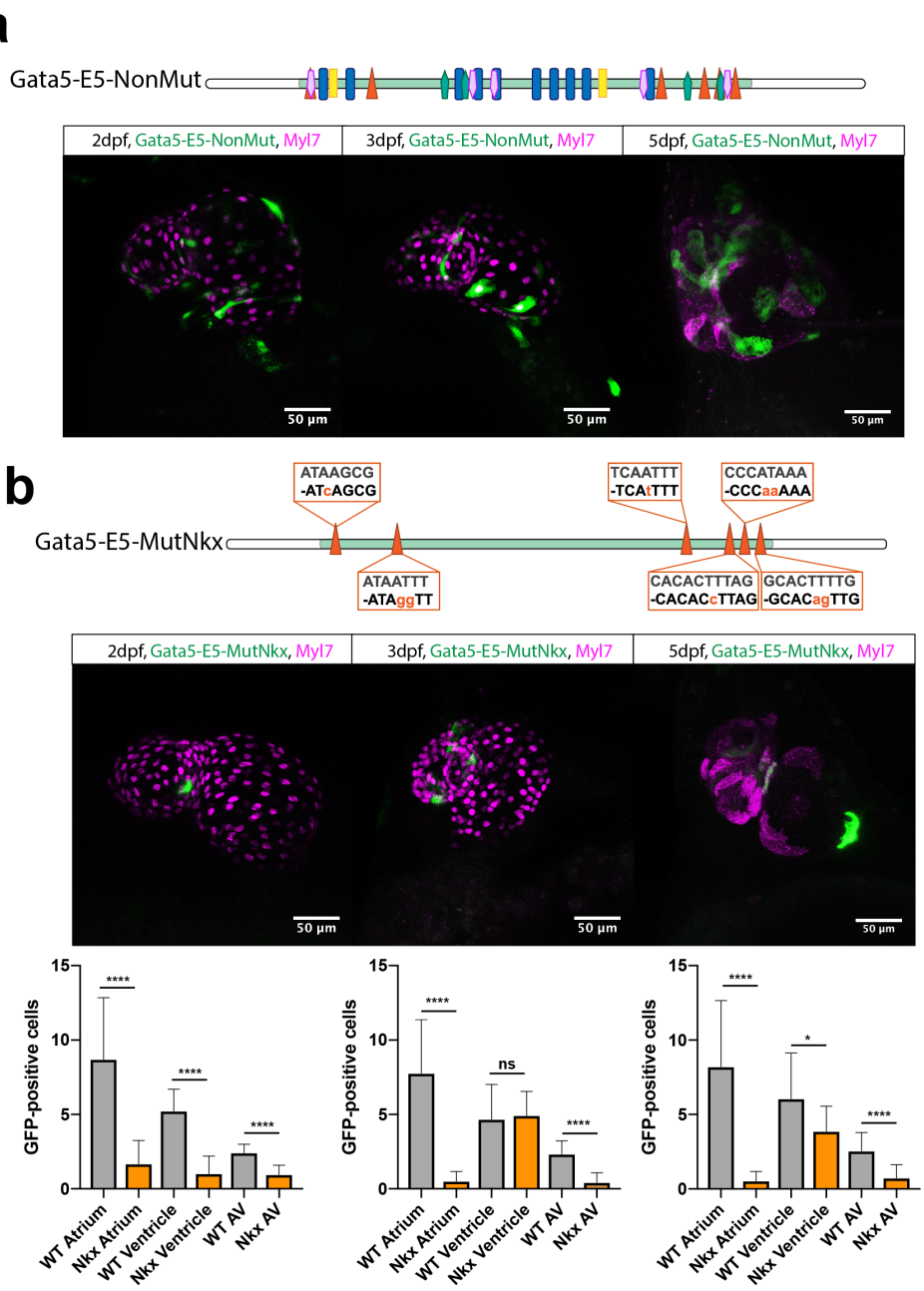


Figure 6 -7 Nkx binding was necessary for the activity of Gata5-E5 in the atrium and AVC.

a) Microscopic analysis of Gata5-E5-driven reporter expression, quantified by counting GFP fluorescent cells at 2dpf, 3dpf and 5dpf. Expression patterns of the enhancer-driven reporter are represented in the diagrams below. b) 1-3 base substitutions were made in the 6 predicted Nkx-binding sites (orange font). Significantly fewer fluorescent cells were observed in the atrium and AVC in the Gata5-E5-MutNkx embryos, compared with non-mutated controls at each time point measured and in the ventricle at 2dpf. Significance was calculated using an unpaired Student's t-test. *= $p<0.0332$, **= $p<0.0021$, ***= $p<0.0002$, ****= $p<0.0001$. 2dpf n=18, 3dpf n=12, 5dpf n=15

6.2.5 TBX-Gata5-E5 binding interactions were required in the atrium and AVC populations

Members of the T-box TF family play critical roles during cardiogenesis; in chamber specification, epicardial development, lineage determination and conduction system regionalisation (Rudat et al., 2011). Tbx2 and 3 are necessary to control cell proliferation in the developing zebrafish AVC by repressing genes required for chamber growth and maturation (Ribeiro et al., 2007). In mice, TBX3 is a regulator of SAN specification as it actively represses factors in the working myocardial gene programme and directs towards a pacemaker phenotype (den Haan et al., 2012). In zebrafish, this role falls to Tbx2b instead of Tbx3 (Tessadori et al., 2012). Tbx1 has been shown to maintain proliferation and suppress differentiation of SHF progenitor cells, as such, the zebrafish *tbx1* mutant, *van gogh*, displays outflow tract defects (Nevis et al., 2013).

Only two putative binding sites were predicted for the Tbx family but because of the importance of numerous Tbx TFs in many aspects of cardiogenesis, I chose to also study these. The binding sites were predicted for Tbx1, 2, 15, 4, 5, and 21 but these were all predicted at the same two regions, showing the similarity of the binding preference for all of these factors. My single cell transcriptome analysis showed expression of *tbx1* in cluster 0 and 1, *tbx5a* in clusters 0, 1, 2, 3 and high expression in cluster 6, *tbx2a* in clusters 0, 1 and 6, *tbx2b* in cluster 1 and 6, tbx3a in clusters 0 and 1 and low expression of *tbx20* in clusters 2, 3 and 6 (Figure 6 -8).

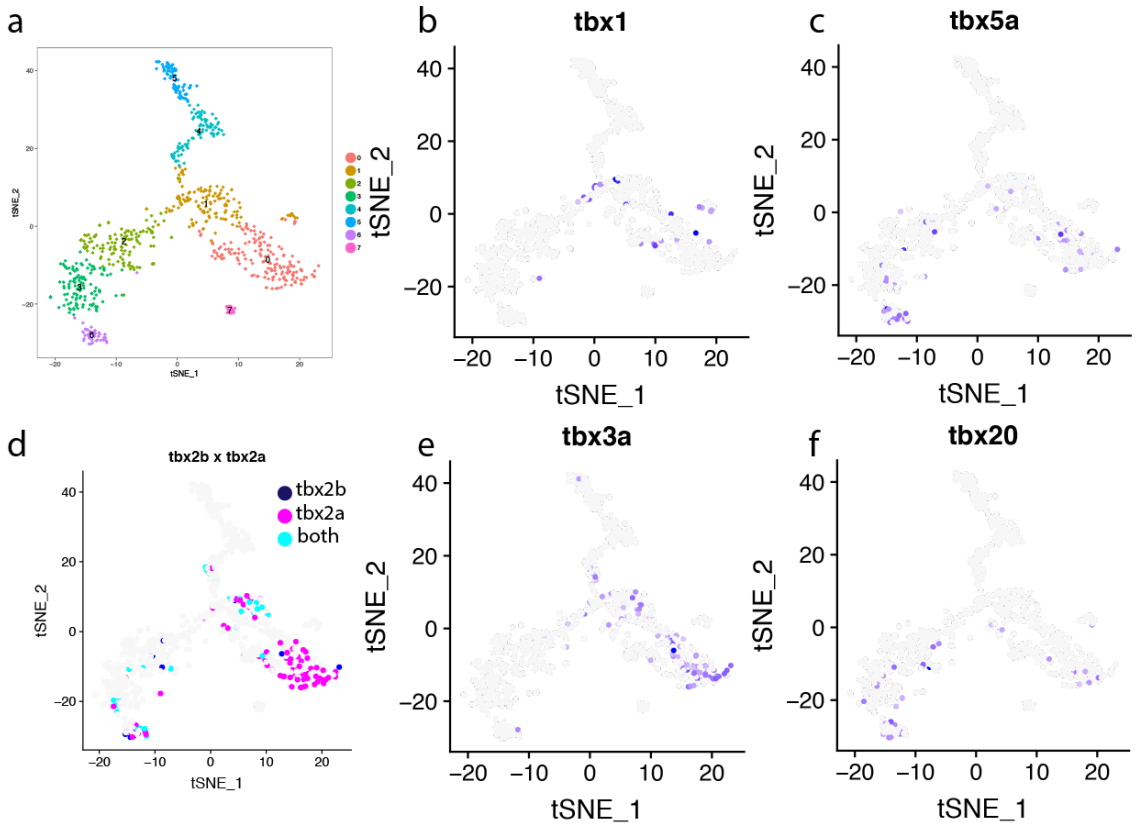


Figure 6 -8 tbx1, 5a, 2a, 2b, 3a and 20 transcripts were enriched in Gata5-E5-mCherry-expressing cells

a) t-SNE plot showing all clusters of cells in which Gata5-E5 is active. b) *tbx1* was predominantly expressed in cells of clusters 0 and 1 c) *tbx5a* was expressed throughout clusters 0, 1, 2 and 3, but particularly strongly in cluster 6. d) *tbx2b* (dark blue) was expressed predominantly in cluster 6, where it overlapped with *tbx2a* in some cells (cyan). Cluster 2 and 3 also contained some cells coexpressing *tbx2a* and *tbx2b. tbx2a* (magenta) was predominantly expressed in clusters 0, 1 and 6. e) *tbx3a* was expressed mainly in clusters 0 and 1. f) tbx20 was expressed in clusters 0, 2, 3 and 6.

At 2, 3 and 5dpf, significantly fewer GFP fluorescent cells were seen in the atrium in Tbx-site mutated embryos, when compared with non-mutated controls (Figure 6 -9), suggesting that Tbx was necessary for Gata5-E5 enhancer activity in the atrial population at these time points. Loss of fluorescence was seen in both the putative CCS and the remaining atrial CMs. The numbers of cells showing enhancer-driven reporter activity in the ventricle were comparable in the non-mutated control and when Tbx sites were mutated at 3dpf and 5dpf, suggesting that enhancer activity is independent of Tbx TF binding in this population of cells. Although there was no

difference in the expression of the enhancer-driven reporter in the AVC upon Tbx site mutation at 3dpf, there were significantly fewer fluorescent cells in the AVC at 5dpf when compared with control-injected embryos. This observation suggests that Tbx binding is important in the AVC at 5dpf. From looking at my transcriptome dataset, the most probable Tbx factor required for the activity of Gata5-E5 in the conduction system cells might be either tbx2a or tbx2b, while the factor acting upstream of Gata5-E5 in the remaining atrial cells might be tbx20. These putative candidates could be confirmed by performing HCR and ChIP experiments in the future.

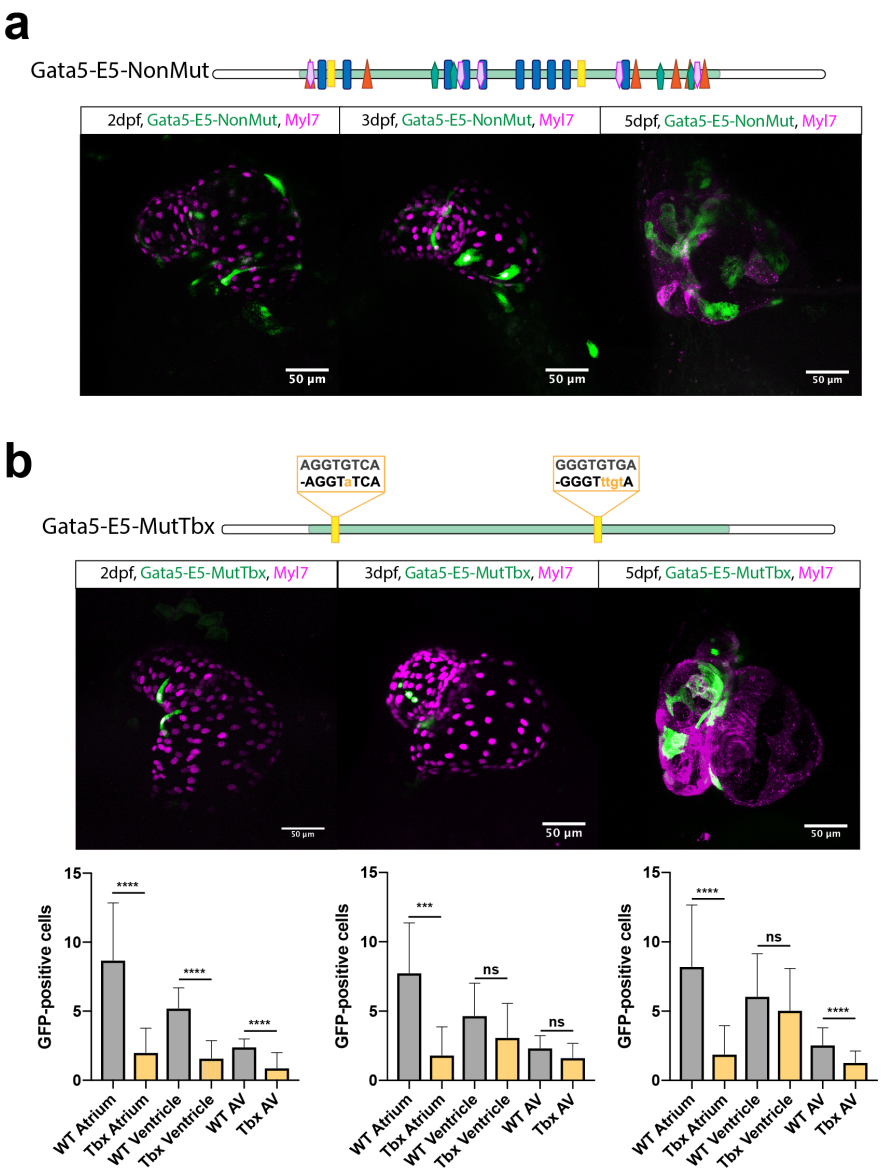


Figure 6 -9 Tbx binding was necessary for the activity of Gata5-E5 in the atrium and AVC.

a) Microscopic analysis of Gata5-E5-driven reporter expression, quantified by counting fluorescent cells at 2dpf, 3dpf and 5dpf. Expression patterns of the enhancer-driven reporter are represented in the diagrams below. b) 1-3 base substitutions were made in the 2 predicted Tbx-binding sites (yellow font). Significantly fewer fluorescent cells were observed in the atrium of Gata5-E5-MutTbx embryos, when compared to the non-mutated control at each time point measured and in the AVC at 5dpf. Significance was calculated using an unpaired Student's t-test. *= $p<0.0332$, **= $p<0.0021$, ***= $p<0.0002$, ****= $p<0.0001$. 2dpf n=18, 3dpf n=12, 5dpf n=15. 2dpf n=17, 3dpf n=12, 5dpf n=22

6.2.6 PRRX/SHOX-Gata5-E5 binding interactions were not required for enhancer activity

Both Prrx and Shox TFs appear in the single cell transcriptomic dataset. *shox2* was expressed in cluster 6, predicted to be the CCS cell population, and *prrx1a* was mainly expressed in clusters 0 and 1. These expression patterns were confirmed by HCR (Figure 6 -10), showing coexpression of *shox2* and *Gata5-E5-mCherry* in the inflow tract region of the atrium and coexpression of *prrx1a* and *Gata5-E5-mCherry* in the region adjacent to the inflow tract. Their binding motifs have extremely similar PWMs meaning that they were predicted at the same loci by the JASPAR software.

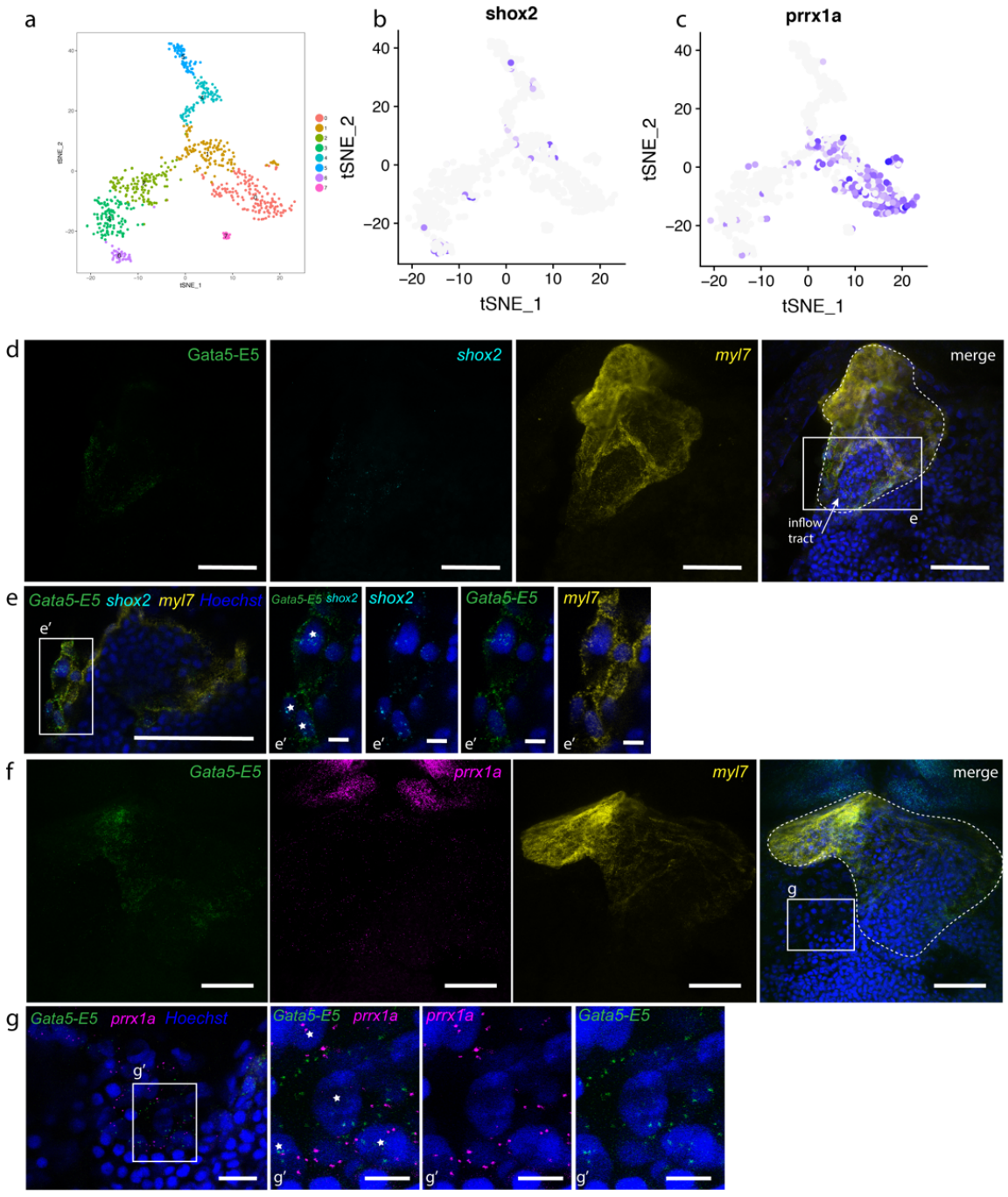

a
b
shox2
c
prrx1a
tSNE_2
tSNE_1
d
Gata5-E5
shox2
myl7
merge
inflow
tract
e
e
Gata5-E5 shox2 myl7 Hoechst
e′
Gata5-E5 shox2
shox2
Gata5-E5
myl7
f
Gata5-E5
prrx1a
myl7
merge
g
g
Gata5-E5 prrx1a Hoechst
g′
Gata5-E5 prrx1a
prrx1a
Gata5-E5

Figure 6 -10 shox2 and prrx1a expression colocalised with regions of Gata5-E5-mCherry expression in the CCS and adjacent to the inflow tract, respectively

a) t-SNE plot showing all clusters of cells in which Gata5-E5 is active. b) shox2 transcripts were expressed predominantly in cluster 6. c) *prrx1a* transcripts were expressed throughout the clusters, but predominantly in clusters 0 and 1 d) Microscopic expression analysis of *shox2* in *Tg(Gata5-E5-mCherry)* embryos at 48hpf using hybridisation chain reaction (z-stack). Scale bars= 50µm. e) Shows the inflow tract region in higher magnification single optical sections. Scale bars= 50µm. Enlarged images in e' visualise the close proximity of *shox2* and *mCherry* transcripts surrounding individual cell nuclei (in blue, examples of coexpressing cells marked with a white star). Scale bars= 5µm. f) Microscopic expression analysis of *prrx1a* in *Tg(Gata5-E5-mCherry)* embryos at 48hpf using hybridisation chain reaction (z-stack). Scale bars= 50µm. g) Shows extracardiac region close to the inflow tract in higher magnification single optical sections. Scale bars= 10µm. Enlarged images in g' visualise the close proximity of *prrx1a* and mCherry transcripts surrounding individual cell nuclei (in blue, examples of coexpressing cells marked with a white star). Scale bars= 5µm.

The paired-like homeobox TF Prrx1a has recently been highlighted as an important factor for effective heart looping during zebrafish cardiogenesis (Ocaña et al., 2017). It is active early on in the LPM, with higher levels on the right side of the embryo, where it drives the differential cell movements necessary for heart looping. MO knockdown, and CRISPR/Cas9 mutations of *prrx1a* both led to a mesocardiac phenotype.

The short stature homeobox gene, Shox2, is expressed in the embryonic pacemaker cells (Hoffmann et al., 2013). When its function was knocked out using MOs, a bradycardia phenotype was observed, suggesting that it plays a vital role in SAN development. Evidence suggests that Shox2 exerts its role by suppressing the working myocardial gene programme through the repression of *nkx2.5* in cells fated to become pacemaker cells (Espinoza-Lewis et al., 2009). Shox2 is also essential for the expression of *isl1* in the SAN (Hoffmann et al., 2013). Because Shox2 has an important role in the CCS, I hypothesised that it might be a strong candidate to act upstream of Gata5-E5 in the putative CCS cells in the atrium, however there was no significant difference in expression between non-mutated control and

mutated Gata5-E5 in the atrial population at any of the monitored time points (Figure 6 -11). I observed some variation between non-mutated controls and mutated Gata5-E5 in the atrium and AVC of embryos at 5dpf, and in the ventricle at 3dpf suggesting a potential interaction at this time point. However, as reflected by the scale bars in Figure 6 -11, there was a large variation in the number of fluorescent cells in these hearts. Increasing the number of embryos tested would help to understand whether the slight decrease in fluorescence observed when Prrx/Shox binding sites were mutated is reliable.

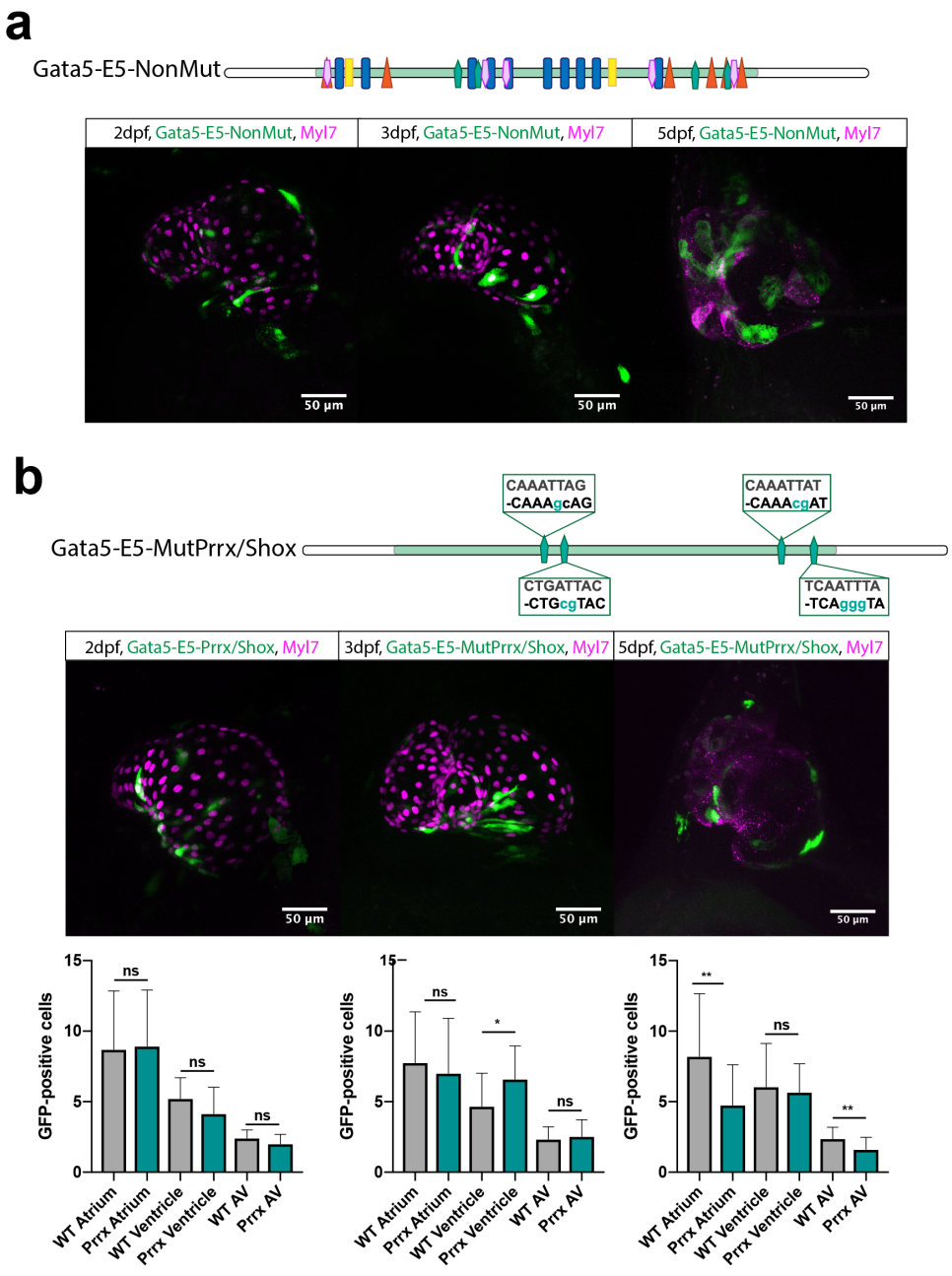


Figure 6 -11 Shox/Prrx binding was not necessary for the activity of Gata5-E5 in any of the populations studied.

a) Microscopic analysis of Gata5-E5-driven reporter expression, quantified by counting fluorescent cells at 2dpf, 3dpf and 5dpf. Expression patterns of the enhancer-driven reporter are represented in the diagrams below. b) 1-3 base substitutions were made in the 6 predicted Shox/Prrx-binding sites (Turquoise font). The expression pattern of the mutated enhancer-driven reporter was mostly comparable to that of the wild type enhancer. Significance was calculated using an unpaired Student's t-test. *= p<0.0332, **= p<0.0021, ***= p<0.0002, ****= p<0.0001. 2dpf n=14, 3dpf n=15, 5dpf n=21

6.2.7 FOX binding interactions are not required for the activity of Gata5-E5

Many members of the forkhead box (Fox) family of TFs have been implicated in cardiogenesis. In a recent study investigating conservation between human and zebrafish enhancers in early cardiogenesis, authors noted that conserved enhancers that were later active in the outflow tract of the heart showed enrichment for Fox binding motifs (FOXO1, FOXP1, FOXK1) (Yuan et al., 2018). In the mouse, FOXH1 is co-expressed with ISL1 in SHF cardiac precursors where they cooperate to activate the transcription of *Mef2c,* placing it at the top of a cascade of cardiac TFs in the SHF lineage (von Both et al., 2004, Evdokia Dodou, 2004). Mouse FOXA2 has been shown to regulate *Tbx1* in the SHF (Maeda et al., 2006). Zebrafish *foxc1a*-knockout embryos have severe heart defects and both Foxp1 and Foxc1a have been shown to regulate *nkx2.5* expression in SHF lineages (Yue et al., 2018, Chang et al., 2013). As Fox factors have been implicated in the transcriptional programme controlling the SHF and its derivatives, it is possible that they may act upstream of *gata5* via Gata5-E5. A number of Fox transcripts were detected in my single cell transcriptome dataset, including *foxp1a,* which was expressed predominantly in clusters 0 and 1 (Figure 6 -12b), and *foxp4,* which was strongly expressed throughout clusters 0, 1, 4 and 5 (Figure 6 -12c). The 5 predicted binding sites for the Fox TF family were spread throughout the Gata5-E5 sequence (Figure 6-13).

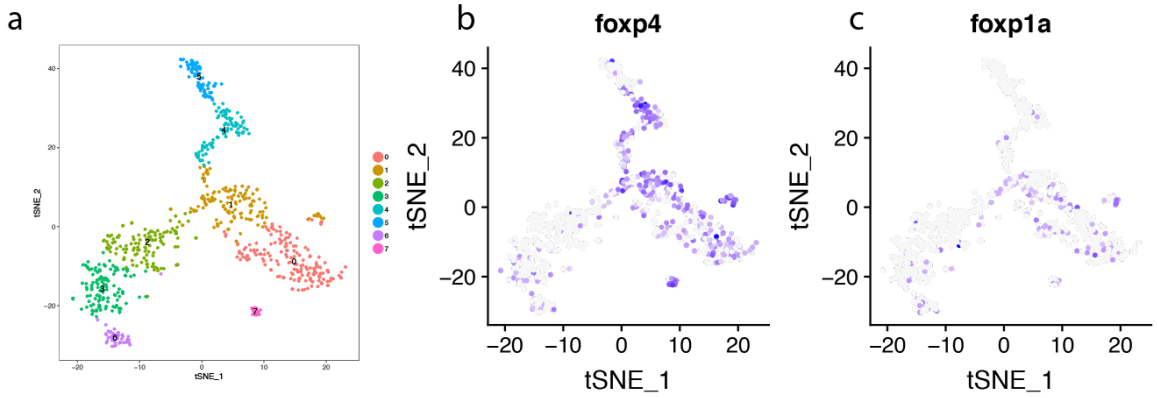


Figure 6 -12 foxp4 and foxp1a were expressed within the single cell dataset

a) t-SNE plot showing each cluster of interest. b) *foxp4* was expressed in all clusters, particularly clusters 0, 1, 4 and 5. c) *foxp1a* was expressed predominantly in clusters 0 and 1, but also in cluster 2, 3 and 6.

After substitution of 1-2 nucleotides within the Fox putative binding sites, activity of the enhancer was almost entirely comparable to non-mutated embryos at 3 and 5dpf, indicating that enhancer activity is independent of Fox TF binding (Figure 6-13). Some small differences were observed at 2dpf in the number of fluorescent cells observed in the atrium and the ventricle, but again these results were variable, so a higher sample size would be needed to verify whether the binding site mutation has a real effect at this timepoint.

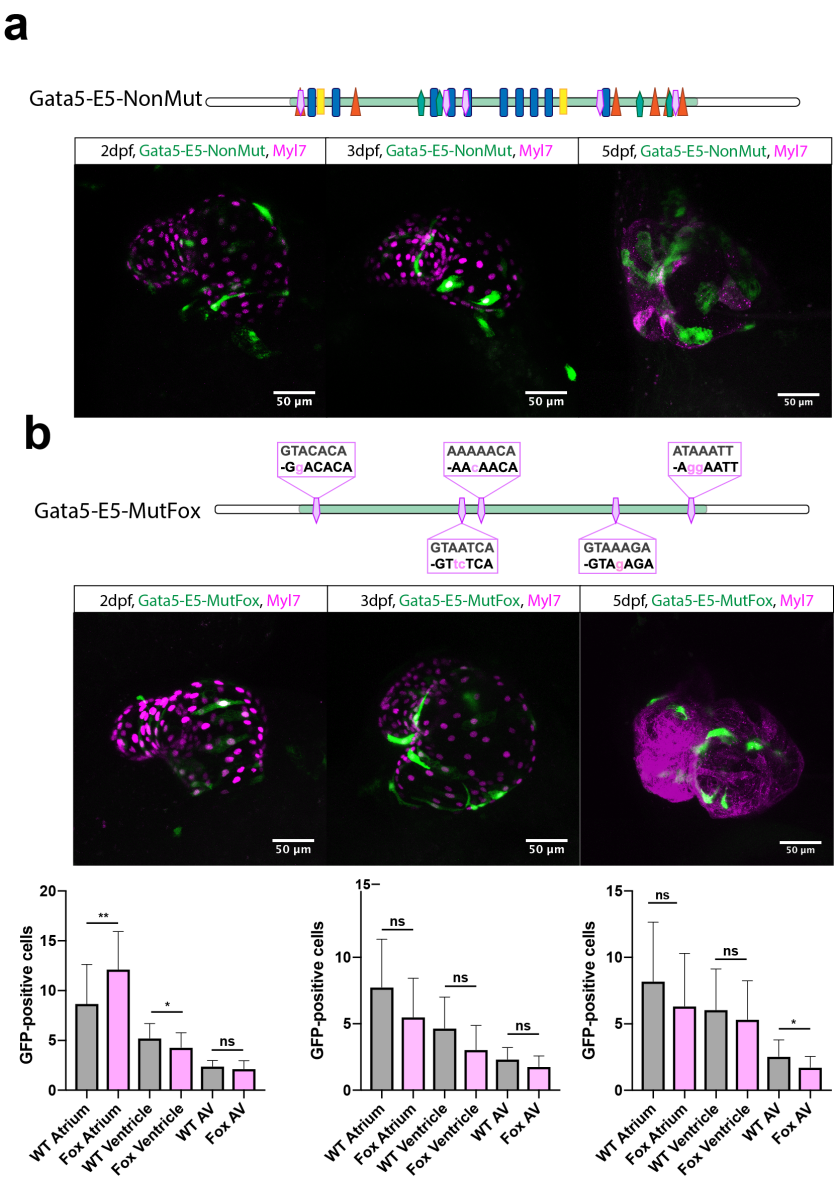


Figure 6-13 Fox binding was not necessary for the activity of Gata5-E5 in any of the populations studied.

a) Microscopic analysis of Gata5-E5-driven reporter expression, quantified by counting GFP fluorescent cells at 2dpf, 3dpf and 5dpf. Expression patterns of the enhancer-driven reporter are represented in the diagrams below. b) 1-2 base substitutions were made in the 5 predicted Fox-binding sites. The expression pattern of the mutated enhancer-driven reporter was comparable to that of the non-mutated enhancer. Significance was calculated using an unpaired Student's t-test. *= p<0.0332, **= p<0.0021, ***= p<0.0002, ****= p<0.0001. 2dpf n=14, 3dpf n=15, 5dpf n=21 2dpf n=23, 3dpf n=14, 5dpf n=21

6.3 Summary

By combining single-cell transcriptomic analysis presented in chapter 5, which informed on the TF expression profile of the subpopulations where Gata5-E5 was active, and binding site prediction software, I was able to select candidate TFs likely to be required for the activity of Gata5-E5 in these particular subpopulations. Using a transient assay, whereby I perturbed specific sites within the Gata5-E5 reporter construct, I tested the requirement for the selected candidate binding sites for the activity of Gata5-E5.

Gata binding to Gata5-E5 was required for the activity of the enhancer in all populations from 48hpf onwards. This observation supports my findings presented in chapter 4, whereby I suggest that Gata5 acts via its own Gata5-E5 enhancer in an autoregulatory loop to regulate its own expression. Meis factor binding to Gata5-E5 was required for Gata5-E5 activity in the ventricular CM population (represented by clusters 2 and 3) and the AVC. Transcriptome data suggest that the required factor is Meis1b. Nkx factor binding was necessary for Gata5-E5 activity throughout the atrial populations and the AVC at all time points measured, and transcriptome data suggest that the candidate TF might be either nkx2.5 or nkx2.7 (represented by clusters 2, 3 and 6). The same requirement was observed for Tbx factors, with transcriptomic data suggesting either tbx2a or tbx2b as the factor required for the activity of gata5-E5 in the conduction system cells (represented by cluster 6), while tbx20 might be the factor acting upstream of Gata5-E5 in the remaining atrial cells (represented by clusters 2 and 3). Although Prrx, Shox and Fox TFs were expressed in the cells where Gata5-E5 was active, Gata5-E5 activity was not affected when

their binding sites were perturbed, suggesting that these TFs do not act via the Gata5-E5 enhancer to regulate *gata5* expression.

The examples shown in this chapter demonstrate the potential to dissect, both in space and time, the upstream regulation of Gata5-E5 by TFs. These findings would have to be confirmed by ChIP studies in order to definitively show that the predicted TFs are indeed binding to the Gata5-E5 sequence at the predicted time. Although the JASPAR database is an incredibly useful tool as a starting point to predict TF binding sites within your enhancer of choice, because it is based on data predominantly from mouse and in vitro studies, it may either miss important binding sites or give false positive 'hits' for binding sites in the zebrafish.

7) Discussion and Future Directions

7.1 Investigating the upstream regulation of the cardiac GATA factors by interrogating the cis-regulatory environment

The cardiac GATA factors, Gata4, Gata5 and Gata6 are zinc finger-containing TFs that have consistently been reported to have an important role in cardiogenesis, as well as playing a role in the adult heart during hypertrophy and, potentially, regeneration (Patient and McGhee, 2002). There have been many recorded cases of GATA mutations causing human congenital heart malformations. Cardiogenesis is an extremely complex biological process, where differentiation of the multiple cell types that contribute to the heart requires the establishment of highly regulated spatio-temporal patterns of gene expression. Previous studies have identified gene regulatory networks containing the cardiac GATA factors, however the precise molecular mechanisms behind their exact spatio-temporal regulation have yet to be described. Information regarding the upstream regulation of a gene can be gained by interrogating its cis-regulatory environment in detail.

Regulation of transcription is largely attained by enhancers; sequences of DNA containing multiple TF-binding sites (Shlyueva et al., 2014b). Enhancers can drive transcription independent of relative distance, location or orientation to their target gene promoter (Long et al., 2016). The identification of active enhancers remains a challenge due to the lack of distinguishing sequence features, the fact that enhancers can be located far from their cognate promoters and a relatively sparse historic understanding of chromatin structure and cis-regulatory element genomic distribution. In recent years, a combination of approaches have proved successful in identifying active enhancers, including cross-species sequence conservation,

genome-wide mapping of epigenomic marks, transcriptional cofactor binding data and mapping of open chromatin regions (outlined in chapter 1).

In this study, I used previously generated ATAC-Seq and p300 ChIP datasets to inform the selection of 4 putative enhancers for each of the cardiac GATA genes in zebrafish. Since this project began, the ATAC-Seq protocol has been optimised to be carried out on single cells (Chen et al., 2018). It would be interesting to perform single-cell ATAC-Seq on sorted adult and embryonic CMs as it may identify discrete subpopulations that may have been masked in the bulk ATAC-Seq datasets. Although we know that enhancers can work from a distance of up to 1Mb, due to the lack of knowledge of 3D chromatin organisation in the regions around the GATA loci, I decided to pick enhancers based on their proximity to the respective GATA gene. It has, however, recently been published that most *cis*-interactions occur within ~300kb of the promoter in question (Hughes et al., 2014). I was able to test these putative enhancers *in vivo*, in transient, using a novel reporter system. This system allowed efficient, high-throughput analysis of multiple putative enhancers. Of the 12 candidate putative enhancers chosen, I observed reproducible heart-specific reporter expression driven by 7: none for *gata4,* 4 for *gata5* and 3 for *gata6.* In future studies, it would be important to return to the preliminary data and select some further putative enhancers for *gata4* to test. The regulation of *gata4* would be particularly interesting in the adult because of its link with regeneration, as discussed in chapter 1.

I generated novel reporter lines, allowing characterisation of the enhancer-driven reporter for many of these putative enhancers (Gata5-E5, Gata5-E6, Gata5-E8, Gata6-E12). Interestingly, although *gata5* and *gata6* are expressed throughout the developing heart, all enhancers tested were active within subpopulations of the

expression domain of their assigned gene. In order to demonstrate the reproducibility of this approach, it would have been important to have multiple founders for each enhancer line, in order to demonstrate a consistent enhancer-driven reporter expression pattern. Unfortunately this was only possible for Gata6-E12, where all offspring of four separate founders showed the same expression pattern. Although Gata5-E5, Gata5-E6 and Gata5-E8 lines were only derived from one founder per line, the fact that such consistency was observed between different Gata6-E12 lines provides some confidence in the reproducibility of this approach. Unfortunately, there were some putative enhancer reporters for which I never identified a founder zebrafish and thereby could not establish a stable line in this time frame (Gata5-E7, Gata6-E9, Gata6-E11). It would be of interest in the future to establish and characterise these lines in order to observe whether the sum of the expression of all putative enhancers broadly reproduced the expression pattern of the associated GATA gene. Although extremely informative, it must be kept in mind that these experiments were performed in an episomal context. In reality, many enhancers work together to fine tune the expression of each gene and so these experiments studying individual enhancers outside of their native context have limited combinatorial power and may not build a full picture of gene regulation. These synthetic enhancers do, however, provide a great deal of information regarding the control of gene expression in particular cell types, and the factors that are upstream of these genes.

Microscopic analysis of the newly generated reporter lines identified particularly interesting expression patterns of the reporters driven by Gata5-E5 and Gata6-E12. I demonstrated by HCR that Gata5-E5-driven and Gata6-E12-driven mCherry was coexpressed with *gata5* and *gata6*, respectively, providing further support that these

enhancers are regulating their predicted genes. Over the course of this project, the HCR protocol, developed by Choi et al., has been dramatically improved (Choi et al., 2014, Choi et al., 2016, Choi et al., 2018). During the protocol, an initiator (I1) tethered to a DNA probe complimentary to the target mRNA triggers the self-assembly of fluorophore-labelled hairpins into tethered fluorescent amplification polymers. In version 2 of HCR (Choi et al., 2014), each probe carried a complete I1, so if this bound non-specifically, HCR was still triggered, thereby decreasing the signal-to-background ration of the image. This method was optimised in 2018 with the introduction of version 3 (Choi et al., 2018), which replaces each probe carrying an I1 with a pair of cooperative 'split-initiator' probes. These pairs must bind specifically to their adjacent binding sites in the target mRNA in order to colocalise the two halves of the initiator and amplify the HCR signal, eliminating the generation of any unspecific background signal. This new version 3 dramatically improved the method and enabled clear imaging of mRNA for all genes tested. One problem with this protocol that we have not yet solved is the lack of ability to mark the cell membrane. To circumvent this problem, we are currently testing antibodies for membrane proteins with a view to combining ICC and HCR experiments to be able to definitively prove that transcripts are expressed within the same cell. Although I am confident of the specificity of the signal observed in the HCR experiments, t would have been to include a negative control, using the hairpins without the probes to confirm that there was no unspecific signal amplification.

Tg(Gata5-E5-mCherry) showed expression very early at the 6-somite stage in the LPM, the location of early cardiac progenitor cell specification (Schoenebeck et al., 2007). In the future, it will be important to investigate the activity of Gata5-E5 at this timepoint in more detail using markers of the populations within the LPM (e.g. the

haemangioblast marker *scl,* and the cardiac progenitor marker *hand2* and *nkx*) in order to understand exactly which populations the enhancer is active in. An interesting recent study has shown that, although heart enhancers appear to be weakly conserved across species after organogenesis, the regulatory elements that establish the conserved cardiac transcriptional programme may be active at the initial stages of heart development (Yuan et al., 2018). Using a reporter line driven by a mouse *Smarcd3* enhancer that is active in early gastrulating mesoderm, the authors were able to isolate early zebrafish cardiac progenitors and profile them to compare with human cardiac progenitor cells to identify many accessible chromatin elements conserved between zebrafish and human. Interestingly, the most enriched TF motif within these conserved enhancers were GATA motifs, consistent with their crucial roles at this time point. Upon morpholino knockdown of either *gata5* or *gata6* in Tg(Smarcd3-EGFP) embryos, fluorescence was lost, showing that the activity of this enhancer requires Gata5 and Gata6 function.

By 48hpf, Gata5-E5-mCherry was expressed in distinct subpopulations of the developing heart: one Myl7-positive population within the inflow tract of the heart, one Myl7-negative population in the extracardiac region just adjacent to the inflow tract as well as a small population in the AVC and ventricular CMs. As the expression pattern of Gata5-E5-mCherry was very similar to that of the LIM homeobox factor Islet1, I used ICC to confirm whether Islet1 was expressed in any of the Gata5-E5-mCherry-positive cells. Cells within the inflow tract that are both Myl7 and Isl1-positive represent the early SAN of the developing CCS. Gata5-E5-mCherry expression colocalised with that of Isl1 in the CMs of the inflow tract, suggesting that this enhancer was active within the early CCS. Isl1 also marks a population of Myl7-negative progenitor cells of the SHF that reside in the region just

adjacent to the venous pole of the developing heart. Gata5-E5-mCherry and Isl1 were not expressed in the same cells in this region, but the cells expressing each of the markers were in very close proximity. It will be important to investigate the identity of this Gata5-E5-only cell population in order to further understand the role of *gata5* in early cardiac development, and its upstream regulation in this particular subpopulation of cells at 48hpf. By 5dpf, expression of the Gata5-E5 enhancer-driven reporter became restricted to the venous pole and a few cells in the AVC and ventricle.

Microscopic analysis showed that Gata6-E12 was active throughout the heart tube at 24hpf and in both atrial and ventricular CMs at 48hpf. However, by 3dpf, Gata6-E12 was active only in the ventricular CMs and the bulbus arteriosus.

Putative enhancers were selected based on their proximity to the cardiac GATA genes. Although the fact that the Gata5-E5-driven and Gata6-E12-driven *mCherry* transcripts were shown to colocalise with endogenous *gata5* and *gata6* transcripts, respectively, supports the hypothesis that they act as enhancers for these GATA genes, this should be definitively confirmed using chromatin conformation capture. An important experiment that must be completed in the future is Capture-C. Capture-C combines 3C and next-generation sequencing with oligonucleotide capture technology (Oudelaar et al., 2017). This enables unbiased analysis of *cis*-interactions at specific loci at high resolution for hundreds of genes in a single experiment and would definitively confirm functional interaction between the putative enhancers studied here and their regulated GATA gene promoters (Hughes et al., 2014). Although this method has been dramatically optimised for lower cell numbers, it still requires at least 150,000 cells per replicate. I have tried to collect enough material to conduct Capture-C in both the embryo at 48hpf and 5dpf, and in

the adult both before and after injury but could not obtain sufficient cell numbers. This was partly due to the fact that the dissociation protocols require optimisation in order to reduce the amount of cell death, to maintain healthy CMs when FAC-sorting. After this failed attempt, and in order to increase the starting material from the embryonic hearts, instead of FAC-sorting, I would extract whole hearts from the embryos using the method outlined in (Burns, 2006).

7.2 Investigating the function of Gata5-E5 and Gata6-E12 by observing the consequences of their genomic deletion

In order to determine the importance of each enhancer for the regulation of *gata5* and *gata6*, respectively, the endogenous enhancer sequences were excised in a transient manner using CRISPR-Cas9 technology. sgRNAs were designed to target Cas9 activity just outside the region cloned into each enhancer-reporter at both the 3' and 5' ends. This strategy allowed for the endogenous enhancer region to be excised in the respective reporter line without disrupting the reporter construct. This meant that the cells of interest where the enhancer has shown activity in the first place could still be monitored by confocal microscopy, even after disruption of the endogenous enhancer region. Although I have demonstrated the benefits of the high-throughput nature of conducting CRISPR knockout of enhancers in transient, in order to definitely assess all of the presented phenotypes, these studies would need to be recapitulated in a stable mutant line, and I am currently screening for founders in order to do this. Due to the nature of the experiment, it possible that, in transient experiments, multiple different mutations are occurring, and this inconsistency would be eliminated by studying the effects of enhancer knockout in a stable line.

Microscopic analysis of *Tg(Gata5-E5-mCherry;myl7-GFP)* embryos following Gata5-E5 knockout revealed a loss of mCherry fluorescence in the region just adjacent to the inflow tract. However, I could not determine whether this reduced fluorescence was attributed to a loss of the cells themselves, or alternatively, was due to a decrease in *gata5* expression that was needed for the activity of Gata5-E5. In situ hybridisation revealed a loss of *gata5* in the atrial region after knockout of Gata5-E5. It would be important to confirm this finding using a more quantitative method like qPCR. As the region that showed a loss of fluorescence was in the cells in the proximity of the Isl1-expressing putative SHF progenitor cells, I tested the effect of enhancer knockout on the Islet1-positive population and saw significantly fewer Islet1-positive cells after Gata5-E5 knockout. It is yet again to be determined whether these cells were lost or whether just the expression of Isl1 in those cells was downregulated. In either case, a defect in the SHF cells would be expected (de Pater et al., 2009). In accordance, I noted a smaller atrium in the Gata5-E5-knockout embryos at 5fpf and, after counting the CMs, I concluded it was due to atrial hypoplasia, particularly in the venous pole. These data suggest that the inflow tract-adjacent population of Gata5-E5-positive cells might be linked to the maintenance and differentiation of the progenitor cells of the SHF. It will be important to further understand the identity and fate of these extracardiac cells. In hindsight, it would also be important to conduct this experiment blinded to the experimental groups. Using the same novel Ac/Ds reporter system, I have constructed a vector with a Dendra2 reporter driven by Gata5-E5. This new reporter line will enable lineage tracing of the cells in which Gata5-E5 is active, allowing investigation of whether this extracardiac population contributes to the inflow tract of the heart, constituting *bona fide* SHF cells, or whether, alternatively, it contributes to the function of the

SHF by acting as a support tissue only. The latter hypothesis would be favoured since, to date, no SHF contribution to the heart has been reported by Isl1-negative cells, which is the case for this extracardiac Gata5-E5 population.

It would be important to investigate the time point in development where Gata5-E5 knockout first has an effect. As Gata5-E5 is already active very early in development in the ALPM, it would be important to determine whether Gata5-knockout has an effect in this tissue, and if so, whether the effect is in a particular population of progenitor cells, such as the anterior haemangioblast, the progenitors of the FHF or the progenitors of the SHF as this is where they are first specified (Felker et al., 2018).

I also observed a decrease in the number of cells in the putative CCS population in the Gata5-E5-knockout embryos, when compared to Cas9-only controls. It would be important to confirm this result by performing HCR for a CCS-specific marker, such as *shox2.* It would also be important to investigate whether Gata5-E5-knockout causes functional defects in this population by monitoring whether embryos show signs of arrhythmias or bradycardia.

When Gata6-E12 was excised, a small decrease in fluorescence was seen at 48hpf, which recovered by 5dpf. In situ hybridisation did not show any discernible differences in the levels of *gata6* between Cas9-only controls and Gata6-E12 knockout embryos, but it may be that this method was not sensitive enough to detect the subtle changes in expression level. In the future, *gata6* expression level should also be measured by qPCR in order to get a more quantitative measurement of potential gene expression changes. Because of the phenotype observed after Gata5-E5 knockout, I chose to concentrate the rest of my studies on this enhancer,

however, Gata6-E12 should be studied further in the future. The fact that Tg(*Gata6-E12-mCherry*) expression becomes specifically restricted to the ventricle provides a very useful tool for dissecting the upstream regulation of *gata6* in regulating the ventricular chamber maintenance and identity, when compared with the atrium. For example, it would be very informative to compare RNA-Seq and ATAC-Seq data obtained from Gata6-E12-mCherry-positive/Myl7-GFP-positive cells versus Gata6-E12-mCherry-negative/Myl7-GFP-positive cells at, 48hpf (when the enhancer is active in atrial and ventricular cells) and at 3dpf (when enhancer activity is restricted to the ventricle) to try to understand what is controlling the upstream regulation of *gata6* during these stages of cardiogenesis.

7.3 Characterising the Gata5-E5-positive subpopulations of cells of the developing heart by single cell transcriptomics

In order to determine which factors were important for the upstream regulation of *gata5* in the particular subpopulations where Gata5-E5 was active, it was important to understand the identity of the cells within this domain of activity, and their transcriptomes to identify important candidate regulators. It was clear from the microscopic analysis that Gata5-E5 was active in a heterogeneous population of cells and that bulk RNA-Seq would have masked this heterogeneity. As such, I used my novel Gata5-E5 reporter line to obtain FACs-purified cells for single-cell transcriptomics using the 10x platform. From microscopic analysis presented in chapter 4, I concluded that performing this analysis at 48hpf was most informative because Gata5-E5 was active in a number of intriguing subpopulations that single-cell RNA-Seq had the potential to tease apart.

The 10x platform is an excellent method to efficiently sequence large numbers of single cells but I also encountered some caveats. It is recommended to load 20,000

cells into the reaction, with the expectation that there will be a drop-out rate of around 50%. However, because of the reagent ratios within the bead kit, it is only possible to start the protocol with 34μl of cells in suspension. Due to the size and shape of the CMs that made up a large proportion of my target cells, we had to use the largest available nozzle for FACS, with each cell sorted carrying a relatively large droplet of PBS with it. This meant that I could load considerably fewer cells than the machine is capable of sequencing. I chose to continue and load as many cells as possible in the given volume rather than spinning the cells down and resuspending in the desired volume based on the experience of others in the lab who saw evidence of cell damage and RNA degradation after trying this method. Another problem I encountered was unspecific sorting as many of the cells that were sequenced did not express *mCherry* transcripts. As the majority of the cells sequenced that were *mCherry* negative were enriched for haematopoietic markers, I suspect that these cells may have been autofluorescent and therefore were aberrantly activated at the same wavelength as mCherry at the time of sorting. More stringent gating should correct this in the future. It may also be possible to treat the cell suspension before sorting with an agent that lyses red blood cells (e.g. RBC Lysis Buffer (Biolegend). However, given the low depth of sequencing reached by this method, there is a small chance that mCherry was actually transcribed in the negative populations of cells, but not enough reads were sequenced to show expression. There is also a chance that Gata5-E5-mCherry may have stopped being transcribed in these cells, but because the mCherry protein was still present, they were still FAC-sorted. Also, the cell viability stain that is routinely used in our lab (7AAD) is not compatible for use with mCherry. As such, for this experiment, I did not use a viability stain, but instead set up the gates based on size (with previous

experience of the size and profile of dead or dying cells). For future experiments using mCherry, I would invest in a cell viability marker that can be used alongside mCherry e.g. Hoechst. Alternatively, I expect that sorting based on GFP may be more accurate (if indeed, red blood cell autofluorescence more closely resembles the mCherry signal). CMs are notoriously difficult to dissociate and they degrade very quickly in suspension, particularly at the 48hpf timepoint. This was reflected in my single cell results as comparatively fewer CMs were sequenced than the microscopic analysis suggested there should have been. It would therefore be important for all next generation sequencing experiments to further optimise the CM dissociation protocol.

After filtering the RNA sequencing results for transcripts in the populations of interest, excluding all cells that did not express mCherry, I identified the following clusters, summarised in Figure 7 -1:

Cluster 0

Cluster 0 was one of the largest subpopulations sequenced, with 169 cells. GO term analysis revealed an association of the predicted genes with epithelial development, cell differentiation and cell migration. I used my novel Gata5-E5 reporter line to validate the findings obtained using single cell transcriptomics to perform microscopy-based validation of subpopulation-specific gene signatures. *prrx1a, jam2b, adam15a* and *col1a1b* were all expressed in cluster 0 so I performed HCR for these markers in order to visualise the spatial location of cells of this cluster at 48hpf in the Tg(*Gata5-E5-mCherry*) line. The expression of these markers converged with that of *mCherry* in the region outside the heart, adjacent to the inflow tract, suggesting that cluster 0 represents the same inflow-adjacent population of cells seen in the microscopic analysis in chapters 3 and 4. It will be important to

further validate the localisation of this cluster by HCR using more markers. This is the population that resides close to the Isl1-positive progenitor cells of the SHF, but that is negative for this marker, as revealed by the 10x and ICC results. Upon Gata5-E5 knockout, this population is either lost or it loses both Gata5-E5-driven fluorescence and Isl1 expression, leading to impaired atrial development. As such, I predicted that this population was important for the function of the SHF, despite the fact that I could not be certain of the identity of these cells.

The observation made during microscopic analysis that the cells within this inflow-tract adjacent population did not express *isl1* was confirmed in the transcriptomic analysis, leading me to believe that the Gata5-E5-mCherry-positive cells that reside outside the heart are not, themselves, SHF progenitors. Despite this observation, some of the cells in this population did express the SHF marker *tbx1* and all cells within the population expressed *hand2,* a gene that has been shown to be expressed in zebrafish SHF progenitor cells and critical for their differentiation into CMs (Schoenebeck et al., 2007).

Recent studies by Felker et al. (Felker et al., 2018) have suggested that, as is the case in the mouse, the SHF comprises an epithelial sheet that undergoes EMT and contributes to the poles of the heart. Interestingly, and consistent with this model, cluster 0 showed enrichment of a number of EMT markers, such as *twist1a, twist1b, snai2* and *postn.* Cluster 0 also showed enrichment for a number of molecules characteristic of epithelial cells, such as fibronectins and cadherins. There was also an enrichment for extracellular matrix components such as the collagen VI family.

Interestingly a number of the genes expressed within cluster 0 have been demonstrated to cause SHF/inflow tract defects when perturbed in animal models,

such as *pdgfra* and *yap1.* This would be consistent with the phenotype observed when I deleted the Gata5-E5 enhancer in the zebrafish embryo.

Both RA and Hippo signalling have been previously shown to regulate the size and positioning of SHF progenitor populations, as well as their migration. Cluster 0 expressed components of both of these signalling mechanisms, including the RA synthesis enzyme, *aldh1a2* and components of the Hippo signalling pathway including *yap1* and *tead1b*. This suggests that perhaps the cells in cluster 0 are supporting the SHF population and providing the external cues necessary for appropriate SHF size and migration.

Cluster 1

Cluster 1 grouped a population of 135 cells that were enriched for signatures of transitioning cells, for example *pcna* and *rplp0*. Because of the limitations in sequencing depth when using the 10x platform, it was only possible to ascertain that these were intermediate proliferating cells but no additional information about their identity could be gained.

Cluster 2 and 3

Clusters 2 and 3 were populations of 123 and 118 cells, respectively. They were both enriched for signatures of mature CMs, including *myl7, cmlc1, nppa* and troponins. Based on the expression of the atrial marker *myh6* and ventricular marker *vmhcl*, cluster 2 was a heterogeneous mixture of atrial and ventricular CMs and cluster 3 comprised mainly atrial CMs. This correlated well with the microscopic analysis presented in chapters 3 and 4. These clusters expressed many of the same genes, but future studies should include validation of these clusters using HCR for any genes that are differentially expressed between the clusters.

Cluster 6

Cluster 6 was a relatively small population, comprising 42 cells. The cells in this cluster were enriched for transcripts typical of atrial CMs, including *myh6, myl7, tnnt2a* and *mylk3*. These cells were also enriched for genes associated in the literature with 'positive regulation of heart contraction', including *isl1, shox2, tbx2b, bmp4, atp2a2a, slc8a1a* and *cacna1g.* When combined with the microscopic analysis in chapter 4 showing that Gata5-E5 was active in an *isl1*-positive population within the inflow tract, this led me to conclude that Gata5-E5 was active within the cells of the SAN of the early CCS and that cluster 6 represented this population. To validate this hypothesis, I performed HCR at 48hpf using for the gene signature characteristic of cluster 6 (*shox2, sema3fb* and *col14a1*) and saw colocalisation of transcripts with Gata5-E5 in the region of the inflow tract predicted to be the SAN of the early CCS.

There is evidence that the GATA factors play an important role in the CCS. Mutations in all 3 cardiac GATA genes have been associated with atrial fibrillation in humans (Posch et al., 2010, Wang XH, 2013, Yang et al., 2012). *gata4* heterozygous mutant mice have shortened PR intervals (Harris et al., 2015) and *gata6* knockout led to proximal atrioventricular conduction defects (Liu et al., 2015). Gata4 and Gata6 also regulate the sodium-calcium exchanger *Slc8a1*, and the *Hcn4* channel, which are important for cardiac pacemaking (Liu et al., 2015). In mice, the cardiac GATAs have been shown to functionally interact with Nkx2.5 and Tbx2/3/5 to modulate their activity on target genes known to regulate the CCS (Nemer and Gharibeh, 2015). GATA factors are known to regulate many conduction-relevant genes, including those for gap junctions and ion channels. There is one other documented GATA enhancer that is active in the CCS. The

chicken Gata6 (cGata6) enhancer drives CCS-specific gene expression (Davis et al., 2001). In transgenic mice, the cGata6 enhancer is active throughout the early heart tube but then becomes restricted to non-chamber myocardium, including the AVC, thus showing a similar expression pattern to the Gata5-E5 enhancer in zebrafish. This perhaps suggests that zebrafish Gata5 shares functional similarity with the chicken Gata6.

Cluster 4, 5 and 7

Upon closer inspection, clusters 4, 5 and 7 did not contain *mCherry* transcripts so should not have been included in the re-clustering of the Gata5-E5-mCherry sequenced cells. These clusters were all enriched for neural signatures, indicating that they may be neural-crest-derived cells. Because these cells did not express *mCherry* or *gata5,* I have not investigated them further and attest them to problems associated with sorting mCherry-positive cells.

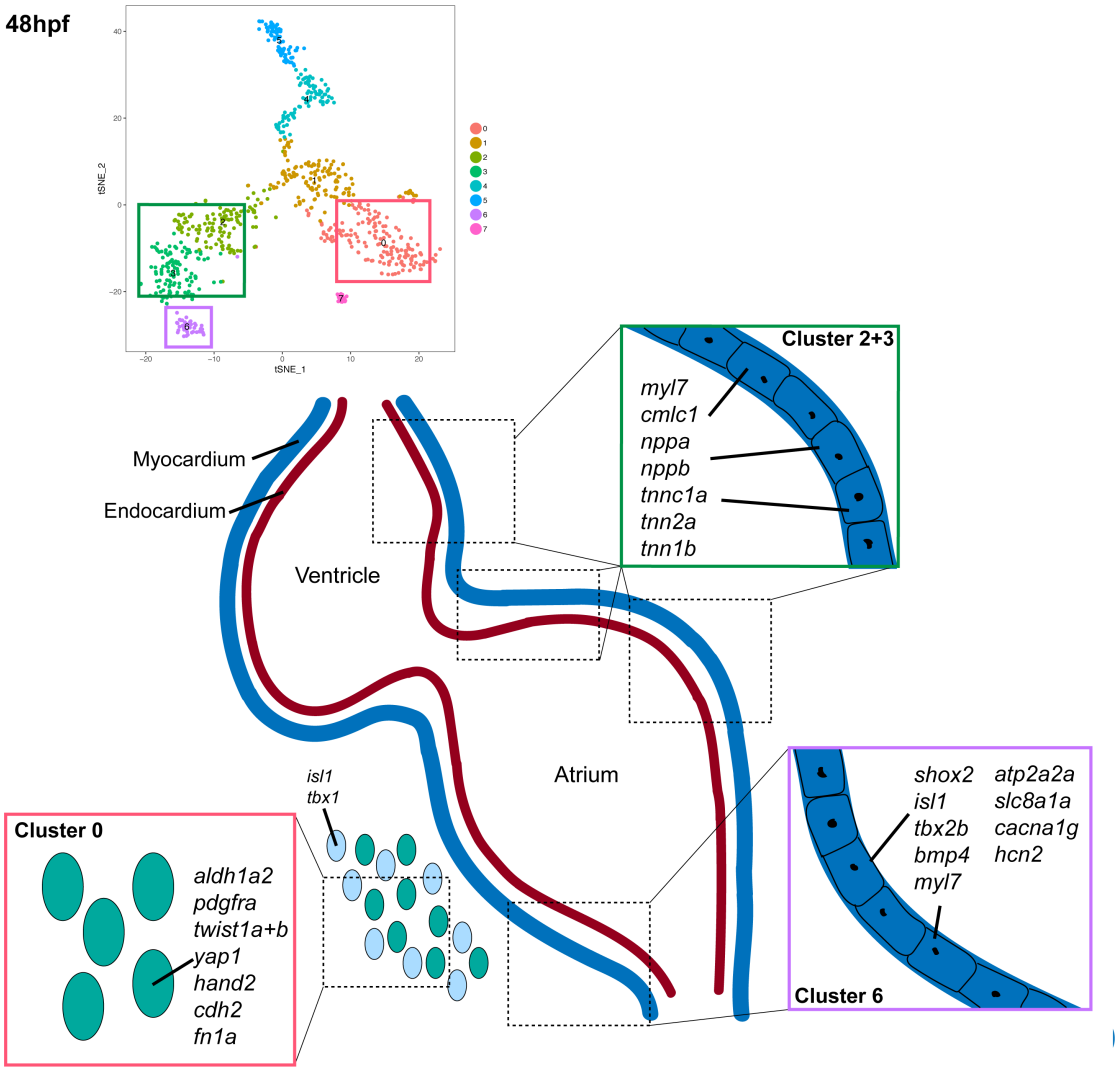


Figure 7 -1 Summary of subpopulations showing Gata5-E5 activity at 48hpf

Cluster 0 showed enrichment for markers of EMT, including twist1a and twist1b and epithelial cell markers, including cdh2 and fn1a. Also enriched, were members of both the Hippo pathway (yap1, tead2b) and the RA signalling pathway (aldh1a2), which have both been shown to contribute to the size regulation of the SHF and to its migration. The cells in this cluster were also enriched for the progenitor cell markers pdgfra and hand2. HCR confirmed that the spatial location of these cells was amongst the isl1-expressing SHF progenitor cells in the region adjacent to the inflow tract of the heart.

Clusters 2 and 3 were enriched for transcripts typical of CM character, including myl7, cmlc1, nppa, nppb, tnnc1a, tnn2a, tnn1b. HCR for specific markers is needed to confirm the spatial location of these clusters but, based on microscopic analysis of the cells in which Gata5-E5 was active, they are predicted to be atrial and ventricular CMs.

Cluster 6 was enriched for markers specific to the early CCS, including shox2, isl1, tbx2b, bmp4, myl7, atp2a2a, slc8a1a, cacna1g and hcn2. HCR confirmed the spatial location of cell cluster 6 in the inflow tract myocardium.

7.4 Identifying the upstream regulation of *gata5* using binding site analysis and perturbation

Enhancers comprise a series of TF binding sites: short degenerate DNA sequences which TFs specifically recognise and bind to in order to contribute to the function of the enhancer (Long et al., 2016). Studying the TF binding sites and interactions within an enhancer can therefore provide invaluable information about the upstream regulation of a gene. In this study, I have interrogated the DNA sequence of the enhancer Gata5-E5, which is active in a subpopulation of the *gata5*-expressing domain in order to understand the upstream regulation of *gata5* in these specific cell populations. I used the TF binding site prediction database, JASPAR, to identify TF binding motifs within the core sequence of the enhancer (defined as the region of the enhancer directly under the ATAC-Seq peak, at the centre of the open chromatin region). This produced a list of 655 putative TF binding sites within the core sequence of Gata5-E5. In order to select candidate TF-enhancer interactions likely to be functional *in vivo*, I used the following criteria: expression in the relevant cell type based on single cell transcriptomics analysis described in chapter 5; the presence of homotypic clusters of predicted binding sites for a particular TF family; longer TF binding motifs; published data showing a role in either heart development or a related biological process. A panel of candidates that satisfied at least two of these criteria were selected to be functionally tested. Despite the JASPAR database being a very useful tool, but the majority of its input data was generated from mouse studies or in vitro investigations, meaning it may not be as accurate at predicting binding within zebrafish DNA sequences. It is also limited to searching for the specific list of TF binding sites for which it has functional input data, removing the opportunity for discovering de novo TF-binding motifs.

PWM-based motif predictions purely consider the DNA sequence and not chromatin architecture and biological context. As such, experimental validation is necessary to determine whether a motif is functional *in vivo*. I designed a transient assay to test the selected TF-binding site interactions within the Gata5-E5 enhancer in the embryo, whereby the binding sites for each candidate TF were perturbed and enhancer-driven fluorescence was used as a read-out for enhancer activity disruption. Perturbation of the binding sites involved making only 1-3 nucleotide substitutions so as to maintain the local genome architecture and ensuring that the predicted binding site was eliminated without creating any new binding sites. This would not show definitively which specific binding site is necessary for the activity of the enhancer, or which specific TF is necessary but it does give a strong indication of which TF families are acting upstream of *gata5* via Gata5-E5.

This method has the benefit of being able to test many different TF-enhancer interactions relatively quickly. However, due to the mosaic nature of these transient experiments, expression of the enhancer-driven reporter was variable, even when driven by the non-mutated enhancer sequence. To bypass this caveat, I screened a large number of embryos to establish general patterns of enhancer-driven reporter expression and quantify changes after binding site perturbation at multiple time points during early development. This system could be improved by using a bidirectional expression system, whereby the non-mutated Gata5-E5 would drive a fluorescent reporter, and in the other direction, the mutated enhancer would drive a different fluorescent reporter. This would serve as an internal control for the injections, making the transient nature of the experiment more reliable. This has been previously attempted in the laboratory; however, it was unsuccessful due to the inability to clone a suitable insulator sequence that would prevent the crosstalk

between the enhancers and the different reporter constructs. It would also be important to repeat these experiments to increase the numbers of embryos studied and test the reproducibility of the results.

gata5 and *gata6* transcripts were enriched in the populations of interest in my single cell transcriptomics data and 6 tandem GATA binding sites were predicted within the Gata5-E5 sequence. Upon perturbation of 1-2 nucleotides within the GATA binding sites, expression of the enhancer-driven reporter was lost by 48hpf. This indicated that the binding of a GATA factor was necessary for the activity of this enhancer. Based on the observations that deletion of the GATA5-E5 enhancer led to a decrease in the expression of *gata5,* and the overexpression of *gata5* caused an increase in the expression of the non-mutated enhancer-driven reporter, I predict that Gata5 regulates its own expression via the Gata5-E5 enhancer throughout all populations where this enhancer is active from 48hpf onwards. This will need to be confirmed by ChIP-Seq. This phenomenon of autoregulation has been observed with GATA factors in other contexts. In the mouse, a distal enhancer for *Gata4* that is sufficient to direct expression throughout the lateral mesoderm has been shown to contain two GATA binding sites that are robustly bound by GATA4 and are required for the function of the enhancer in vivo (Rojas, 2005). A study using a murine erythroid cell line has also demonstrated autoregulation of *Gata1* via two GATA binding sites in a conserved enhancer located approximately 500bp upstream of the *Gata1* TSS (Grass et al., 2003). There is a distinct lack of ChIP-grade antibodies that are suitable for use in zebrafish. As such, I am using a similar biotagging strategy to that described in Figure 3-1. I have constructed an Ac/Ds vector comprising the *myl7* promoter driving Avi-tagged *gata5* expression, upstream of an EGFP reporter. I had initially planned to perform this ChIP-seq experiment in

transient, by injecting the construct directly into the embryos of a zebrafish line expressing biotin ligase (BirA) ubiquitously in every cell, which would allow me to pull-down the Gata5-bound DNA sequences for ChIP-Seq. However, at the moment, ChIP-Seq requires approximately 100,000 cells and as integration of the transgene is mosaic when done in transient, it proved extremely difficult to obtain the high number of cells required for a successful experiment. Instead, I have created a stable line expressing *gata5*-Avi under the control of the *myl7* promoter in order to obtain the necessary number of embryos to perform this ChIP-Seq experiment. These embryos are now growing, and I hope to be able to perform this experiment in the near future.

meis1a and *meis1b* transcripts were enriched in all of the clusters of interest in my single cell transcriptomic data and the Gata5-E5 sequence contained 10 predicted Meis binding sites. When these binding sites were mutated, significantly fewer cells were seen in the ventricle and the AVC at all time points monitored and in the atrium at 5dpf, suggesting that Meis TFs are necessary for the function of Gata5-E5 in these populations. *meis1b* was more highly expressed in the CM populations analysed (clusters 2 and 3), suggesting this may be the factor upstream of Gata5-E5, however this finding must be tested by ChIP-Seq. Overlapping Nkx2.5/Meis1 binding motifs have been reported in many murine cardiac enhancers, whereby the sequential binding of each TF confers spatiotemporal tuning of enhancer activation (Dupays et al., 2015). Zebrafish Meis1 and 2 have been shown to be essential for proper heart tube formation and looping (Paige et al., 2012). *meis2b* morphant embryos also displayed a reduced heart rate suggesting a potentially unexplored role for Meis factors in cardiac conduction. Meis TF binding site perturbation led to

disruption in the Gata5-E5-driven reporter at 5dpf, suggesting that Meis factors might play a role in conduction via *gata5* at this time point.

nkx2.5 and *nkx2.7* transcripts were enriched in clusters 2, 3 and 6 of my single cell RNA-Seq data set. When the 6 putative Nkx binding sites were disrupted, I observed a loss of reporter expression in all cells of the atrium and in the AVC at all time points studied, suggesting that Nkx factors are necessary for the activity of Gata5-E5 in these populations. HCR confirmed that *nkx2.5* was coexpressed with Gata5-E5-mCherry in all subpopulations. Future studies should also include HCR for *nkx2.7.* ChIP-Seq experiments would investigate which Nkx TF is necessary for the activity of Gata5-E5. Gene regulatory networks containing Nkx2.5 and GATA factors are highly conserved across species. In *Drosophila*, the GATA factor, *pannier*, is both upstream and downstream of the *nkx2.5* homolog, *tinman*, suggesting a positive feedback loop for these key regulators (Gajewski et al., 2001). The mouse *Nkx2.5* gene has been shown to be controlled by 2 separate cardiac-specific enhancers that contain essential GATA binding sites (Lien et al., 1999, Searcy et al., 1998) and a cardiac-specific regulatory region driving the expression of *Gata6* has been shown to be a direct target for transcriptional activation by Nkx2.5 *in vivo* (Molkentin et al., 2000). These previous findings, along with the findings shown here in chapter 6 are consistent with the existence of conserved transcriptional regulatory loops between Nkx2.5 and GATA factors that reinforce the expression of both factors during cardiogenesis. *nkx2.5* has also recently been shown to have a crucial role in establishing heart rate variability (a factor that is essential for cardiovascular homeostasis) as the CCS is patterned in the zebrafish embryo (Harrington et al., 2017).

Only 2 Tbx binding sites were predicted within Gata5-E5 core sequence but many of the clusters in my single cell dataset were enriched for Tbx family transcripts, and this family of TFs have been shown to be important in many aspects of cardiogenesis, including formation of the CCS (Poon et al., 2016). As with the perturbation of Nkx sites, perturbation of the Tbx sites within the Gata5-E5 enhancer led to a significant loss of enhancer-driven fluorescence in all populations in the atrium at all time points, and also in the AVC at 5dpf. This suggests that Tbx proteins are important for the activity of Gata5-E5 in both the cells of the CCS and in the atrial CMs. From the single cell data set, I predict that the Tbx factor that acts upstream of *gata5* in the CCS might be *tbx2a, tbx2b* and/or *tbx5a,* while *tbx20* might be the one important in the atrial CMs. This should be tested and further investigated by HCR and ChIP-Seq experiments. Tbx proteins have been shown to act cooperatively with Gata4 and Nkx2.5 in murine models, coordinately controlling cardiac gene expression, differentiation, and morphogenesis (Harrington et al., 2017). Similar to my observations, a 1300bp region upstream of *gata4* in zebrafish has been shown to contain a cluster of consensus binding sites for Tbx family TFs. Mutation of these consensus sequences causes a reduction in reporter gene expression in the heart, suggesting a direct regulation of Gata4 expression by Tbx factors (Heicklen-Klein and Evans, 2004). Gata6 and Tbx5 have been shown to genetically interact *in vivo* (Maitra et al., 2009). As the perturbation of Tbx binding sites in Gata5-E5 disrupted expression of the reporter in the atrium, this may suggest that Tbx and Gata regulation are important for the function of the developing CCS. It has been recorded in the mouse heart that *Gata4, Tbx5* and *Nkx2.5* are expressed in largely overlapping domains in the CCS and co-occupy several CCS

gene loci, suggesting that they may cooperatively activate genes in the CCS (van den Boogaard et al., 2012).

The fact that perturbation of the Gata TF binding sites within Gata5-E5 led to a loss of fluorescence in every cell that initially showed activity of Gata5-E5, but perturbation of Nkx, Tbx and Meis sites only led to a decrease in reporter expression in particular subpopulations could indicate that the regulation is combinatorial between Gata and these other TFs. Gatas have previously been shown to cooperate with many cofactors within these TF families to ensure proper spatiotemporal expression of cardiac genes during normal embryonic development. One well-documented example is the direct and synergistic interactions of Tbx20, Nkx2.5, GATA4 and GATA5 to activate transcription of cardiac genes in mice (Stennard et al., 2003). It may be the case that the regulation of *gata5* via the Gata5-E5 enhancer is combinatorial, and additive by all these factors, but that Gata5 binding is indeed crucial and necessary.

Both the Prrx and Shox TFs have very similar binding site motifs and they were predicted at the same locations within the enhancer sequence. Both *prrx1a* and *shox2* were expressed in the single cell RNA-Seq dataset, which was validated by HCR, indicating that they could be important for the function of Gata5-E5. There was no significant difference in the level of reporter expression from the non-mutated Gata5-E5-driven reporter, suggesting that *prrx1a* and/or *shox2* do not act upstream of *gata5* via Gata5-E5. The same result was seen when Fox TF binding sites were perturbed, showing that Fox factor binding is not required for the activity of the enhancer. Binding site perturbation studies are summarised in Figure 7.2.

These studies demonstrate the power and potential of these newly generated GATA enhancer reporter lines in understanding the upstream regulation of the cardiac GATA genes in a detailed manner. In future experiments, it would be interesting to use the single cell transcriptomics dataset to identify new TF candidates to test their potential enhancer interactions in order to get even more detailed understanding of the upstream regulation of *gata5* in specific subpopulations of the developing heart.

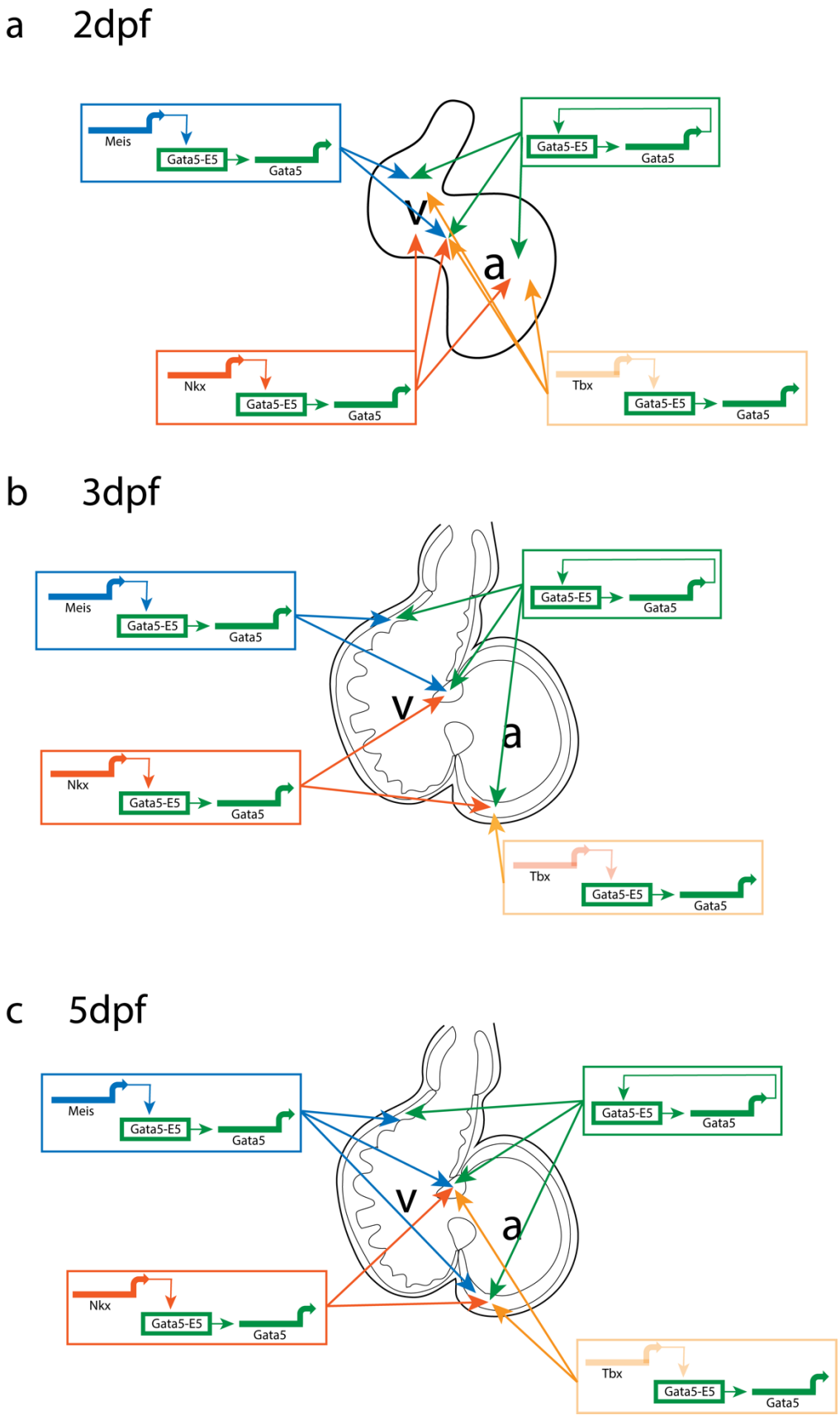


Figure 7 -2 Summary of binding site perturbation studies

a) At 2dpf, Gata, Nkx and Tbx binding are required for the activity of Gata5-E5 in the atrium. Meis, Gata, Nkx and Tbx binding are required for the activity of Gata5-E5 in the ventricle. Meis, Gata, Nkx and Tbx binding are required for the activity of Gata5-E5 in the AVC. b) At 3dpf, Gata, Nkx and Tbx binding are required for the activity of Gata5-E5 in the atrium. Meis and Gata binding are required for the activity of Gata5-E5 in the ventricle. Meis, Gata, and Nkx binding are required for the activity of Gata5-E5 in the AVC. c) At 5dpf, Meis, Gata, Nkx and Tbx binding are required for the activity of Gata5-E5 in the atrium. Meis and Gata binding are required for the activity of Gata5-E5 in the ventricle. Meis, Gata, Nkx and Tbx binding are required for the activity of Gata5-E5 in the AVC.

7.5 Final Remarks

The cardiac GATA genes, *gata4, gata5* and *gata6* are important during all stages of cardiac development, from progenitor specification to CM differentiation. Although the downstream effectors of the GATA factors are well characterised, the processes behind their specific upstream regulation remain an enigma. The aim of this project was to investigate the upstream regulation of the cardiac GATA factors by interrogating their *cis*-regulatory elements.

After putative enhancer prediction and transient *in vivo* activity screening revealed seven putative enhancers for *gata5* and *gata6* were identified as being active in the heart during cardiogenesis (four of which were also validated in newly-generated stable transgenic lines). Two putative enhancers were selected to study further, based on their expression patterns in interesting subpopulations of the *gata5* and *gata6* expression domain: Gata5-E5 and Gata6-E12. Tg(*Gata5-E5-mCherry)* reported activity in the inflow tract CMs, the AVC and some ventricular CMs, as well as a population of extracardiac cells residing adjacent to the inflow tract. *Tg(Gata6-E12-mCherry)* reported activity that became restricted to the ventricular CMs by 3dpf.

After observing that deletion of the endogenous Gata5-E5 enhancer led to a decrease in *gata5* expression and a potential loss of cells in the inflow tract of the

developing heart and the region adjacent to the inflow tract, I decided to focus my efforts on investigating this enhancer in detail.

Using both microscopic analysis and single-cell transcriptomics, I identified a population of extracardiac Gata5-E5 mCherry-positive cells residing in the region adjacent to the inflow tract, termed cluster 0. Based on their proximity to Islet1-positive progenitors of the SHF and the abundance of epithelial markers in their transcriptome, as well as the fact that both these Gata5-E5-mCherry-positive cells and Islet1-positive SHF progenitor cells were lost after deletion of the endogenous Gata5-E5 enhancer, I predict that these cluster 0 cells are acting as support cells for the SHF tissue. Although the Gata5-E5 sequence contained binding sites for TFs expressed in cell cluster 0, the transient and mosaic nature of my binding site perturbation assay meant that TF-Gata5-E5 interactions could not be investigated for this cluster of cells.

Microscopic analysis and single-cell transcriptomics revealed the activity of Gata5-E5 in CMs of the atrium, ventricle and AVC, termed clusters 2 and 3. Binding site analysis revealed a requirement for Gata, Meis, Nkx and Tbx TFs for the function of Gata5-E5 in these CM populations. Transcriptomic analysis suggested that the factors from these TF families that might be important for the upstream regulation of *gata5 via* Gata5-E5 are Gata5, Meis1b, Nkx2.5 and/or Nkx2.7, and Tbx20. These interactions must be tested using ChIP studies.

Similarly, I identified a population of Gata5-E5-mCherry positive cells, termed cluster 6, that showed the localisation and expression pattern characteristic of cells of the early cardiac conduction system. Gata5-E5 contained binding sites for TFs expressed in this cell cluster 6 and binding site perturbation studies revealed a

requirement for Gata, Tbx and Nkx binding sites for the activity of Gata5-E5 in this population. Transcriptomic analysis suggested that the important factors upstream of *gata5* in the early cardiac conduction system might be Gata5, Nkx2.5 and/or Nkx2.7 and Tbx2a/b and/or Tbx5a. These interactions must be further tested by ChIP experiments.

Together, my studies demonstrate the power of probing the *cis*-regulatory elements of a gene in order to understand its detailed upstream regulation. I have generated novel zebrafish transgenic reporter lines and a binding site perturbation assay that can be used for further investigation of the upstream regulation of *gata5* and *gata6* in specific subpopulations of the heart, providing the opportunity to understand the specific spatiotemporal regulation of the GATA genes at an unprecedented resolution. My studies will build into the known regulatory networks driving cardiac development and will enable further understanding of the molecular basis of GATA-related congenital heart defects.

References

Ahuja, P., Perriard, E., Perriard, J.-C. & Ehler, E. 2004. Sequential myofibrillar breakdown accompanies mitotic division of mammalian cardiomyocytes. *Journal of Cell Science,* 117**,** 3295.

Alexander, J., Stainier, D. Y. R. & Yelon, D. 1998. Screening mosaic F1 females for mutations affecting zebrafish heart induction and patterning. *Developmental Genetics,* 22**,** 288-299.

Alexander, J. M. & Bruneau, B. G. 2010. Lessons for cardiac regeneration and repair through development. *Trends in molecular medicine,* 16**,** 426-434.

Anthony Mathelier, O. F., David J. Arenillas, Chih-Yu Chen, Grégoire Denay, Jessica Lee, Wenqiang Shi, Casper Shyr, Ge Tan, Rebecca Worsley-Hunt, Allen W. Zhang, François Parcy, Boris Lenhard, Albin Sandelin, Wyeth W. Wasserman 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research,* 44**,** D110–D115.

Arner, E., Daub, C. O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rönnerblad, M., Hrydziuszko, O., Vitezic, M., Freeman, T. C., Alhendi, A. M. N., Arner, P., Axton, R., Baillie, J. K., Beckhouse, A., Bodega, B., Briggs, J., Brombacher, F., Davis, M., Detmar, M., Ehrlund, A., Endoh, M., Eslami, A., Fagiolini, M., Fairbairn, L., Faulkner, G. J., Ferrai, C., Fisher, M. E., Forrester, L., Goldowitz, D., Guler, R., Ha, T., Hara, M., Herlyn, M., Ikawa, T., Kai, C., Kawamoto, H., Khachigian, L. M., Klinken, S. P., Kojima, S., Koseki, H., Klein, S., Mejhert, N., Miyaguchi, K., Mizuno, Y., Morimoto, M., Morris, K. J., Mummery, C., Nakachi, Y., Ogishima, S., Okada-Hatakeyama, M., Okazaki, Y., Orlando, V., Ovchinnikov, D., Passier, R., Patrikakis, M., Pombo, A., Qin, X.-Y., Roy, S., Sato, H., Savvi, S., Saxena, A., Schwegmann, A., Sugiyama, D., Swoboda, R., Tanaka, H., Tomoiu, A., Winteringham, L. N., Wolvetang, E., Yanagi-Mizuochi, C., Yoneda, M., Zabierowski, S., Zhang, P., Abugessaisa, I., Bertin, N., Diehl, A. D., Fukuda, S., Furuno, M., Harshbarger, J., Hasegawa, A., Hori, F., Ishikawa-Kato, S., Ishizu, Y., Itoh, M., Kawashima, T., Kojima, M., Kondo, N., Lizio, M., Meehan, T. F., Mungall, C. J., Murata, M., Nishiyori-Sueki, H., Sahin, S., Nagao-Sato, S., Severin, J., De Hoon, M. J. L., Kawai, J., Kasukawa, T., Lassmann, T., et al. 2015. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science (New York, N.Y.),* 347**,** 1010-1014.

Arrenberg, A. B., Stainier, D. Y. R., Baier, H. & Huisken, J. 2010. Optogenetic Control of Cardiac Function. *Science,* 330**,** 971.

Bakkers, J. 2011. *Zebrafish as a model to study cardiac development and human cardiac disease*.

Banerji, S., Ni, J., Wang, S. X., Clasper, S., Su, J., Tammi, R., Jones, M. & Jackson, D. G. 1999. LYVE-1, a new homologue of the CD44 glycoprotein, is a lymph-specific receptor for hyaluronan. *The Journal of cell biology,* 144**,** 789-801.

Beis, D., Bartman, T., Jin, S.-W., Scott, I. C., Amico, L. A., Ober, E. A., Verkade, H., Frantsve, J., Field, H. A., Wehman, A., Baier, H., Tallafuss, A., Bally-Cuif, L., Chen, J.-N., Stainier, D. Y. R. & Jungblut, B. 2005. Genetic and cellular analyses of zebrafish atrioventricular cushion and valve development. *Development,* 132**,** 4193.

Benko, S., Fantes, J. A., Amiel, J., Kleinjan, D.-J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C. T., Mcbride, D., Golzio, C., Fisher, M., Perry, P., Abadie,

V., Ayuso, C., Holder-Espinasse, M., Kilpatrick, N., Lees, M. M., Picard, A., Temple, I. K., Thomas, P., Vazquez, M.-P., Vekemans, M., Crollius, H. R., Hastie, N. D., Munnich, A., Etchevers, H. C., Pelet, A., Farlie, P. G., Fitzpatrick, D. R. & Lyonnet, S. 2009. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nature Genetics,* 41**,** 359.

Berdougo E, C. H., Lee Dh, Stainier Dyr, Yelon D 2003. Mutation of weak atrium/atrial myosin heavy chain disrupts atrial function and influences ventricular morphogenesis in zebrafish. *Development,* 130**,** 6121–6129.

Blaschke Rüdiger, J., Hahurij Nathan, D., Kuijper, S., Just, S., Wisse Lambertus, J., Deissler, K., Maxelon, T., Anastassiadis, K., Spitzer, J., Hardt Stefan, E., Schöler, H., Feitsma, H., Rottbauer, W., Blum, M., Meijlink, F., Rappold, G. & Gittenberger-De Groot Adriana, C. 2007. Targeted Mutation Reveals Essential Functions of the Homeodomain Transcription Factor Shox2 in Sinoatrial and Pacemaking Development. *Circulation,* 115**,** 1830-1838.

Bleyl, S. B., Saijoh, Y., Schoenwolf, G. C., Hunter, J., Leppert, M. F., Gittenberger-De Groot, A. C., Wisse, L. J., Bax, N. a. M., Chapman, S. C., Shiratori, H., Hamada, H., Shiota, K., Yamada, S. & Klewer, S. E. 2010. Dysregulation of the PDGFRA gene causes inflow tract anomalies including TAPVR: integrating evidence from human genetics and model organisms. *Human Molecular Genetics,* 19**,** 1286-1301.

Blow, M. J., Mcculley, D. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Bristow, J., Ren, B., Black, B. L., Rubin, E. M., Visel, A. & Pennacchio, L. A. 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nature genetics,* 42**,** 806-810.

Bodmer, R. 1993. The gene tinman is required for specification of the heart and visceral muscles in Drosophila. *Development,* 118**,** 719.

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

Brodsky, V. Y., Arefyeva, A. M., Gvasava, I. G., Sarkisov, D. S. & Panova, N. W. 1994. Polyploidy in cardiac myocytes of normal and hypertrophic human hearts; range of values. *Virchows Archiv,* 424**,** 429-435.

Brown, D. R., Samsa, L. A., Qian, L. & Liu, J. 2016. Advances in the Study of Heart Development and Disease Using Zebrafish. *Journal of cardiovascular development and disease,* 3**,** 13.

Buchmann, K. 2014. Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals. *Frontiers in immunology,* 5**,** 459-459.

Buenrostro, J., Wu, B., Chang, H. & Greenleaf, W. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.],* 109**,** 21.29.1-21.29.9.

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

Burns, C. G., And Macrae, C.A. 2006. Purification of hearts from zebrafish embryos. *Biotechniques,* 40**,** 274–282.

Burns, C. G., Milan, D. J., Grande, E. J., Rottbauer, W., Macrae, C. A. & Fishman, M. C. 2005. High-throughput assay for small molecules that modulate zebrafish embryonic heart rate. *Nature Chemical Biology,* 1**,** 263-264.

Cai, C.-L., Liang, X., Shi, Y., Chu, P.-H., Pfaff, S. L., Chen, J. & Evans, S. 2003. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Developmental cell,* 5**,** 877-889.

Calhoun, V. C., Stathopoulos, A. & Levine, M. 2002. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 9243-9247.

Cavanaugh, A. M., Huang, J. & Chen, J.-N. 2015. Two developmentally distinct populations of neural crest cells contribute to the zebrafish heart. *Developmental biology,* 404**,** 103-112.

Chablais, F., Veit, J., Rainer, G. & Jaźwińska, A. J. B. D. B. 2011. The zebrafish heart regenerates after cryoinjury-induced myocardial infarction. 11**,** 21.

Chang, S.-W., Mislankar, M., Misra, C., Huang, N., Dajusta, D. G., Harrison, S. M., Mcbride, K. L., Baker, L. A. & Garg, V. 2013. Genetic abnormalities in FOXP1 are associated with congenital heart defects. *Human mutation,* 34**,** 1226-1230.

Chen, J. N. & Fishman, M. C. 1996. Zebrafish tinman homolog demarcates the heart field and initiates myocardial differentiation. *Development,* 122**,** 3809.

Chen, J. N., Haffter, P., Odenthal, J., Vogelsang, E., Brand, M., Van Eeden, F. J., Furutani-Seiki, M., Granato, M., Hammerschmidt, M., Heisenberg, C. P., Jiang, Y. J., Kane, D. A., Kelsh, R. N., Mullins, M. C. & Nusslein-Volhard, C. 1996. Mutations affecting the cardiovascular system and other internal organs in zebrafish. *Development,* 123**,** 293.

Chen, X., Miragaia, R. J., Natarajan, K. N. & Teichmann, S. A. 2018. A rapid and robust method for single cell chromatin accessibility profiling. *Nature Communications,* 9**,** 5345.

Choi, H. M. T., Beck, V. A. & Pierce, N. A. 2014. Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. *ACS nano,* 8**,** 4284-4294.

Choi, H. M. T., Calvert, C. R., Husain, N., Huss, D., Barsi, J. C., Deverman, B. E., Hunter, R. C., Kato, M., Lee, S. M., Abelin, A. C. T., Rosenthal, A. Z., Akbari, O. S., Li, Y., Hay, B. A., Sternberg, P. W., Patterson, P. H., Davidson, E. H., Mazmanian, S. K., Prober, D. A., Van De Rijn, M., Leadbetter, J. R., Newman, D. K., Readhead, C., Bronner, M. E., Wold, B., Lansford, R., Sauka-Spengler, T., Fraser, S. E. & Pierce, N. A. 2016. Mapping a multiplexed zoo of mRNA expression. *Development,* 143**,** 3632.

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A. & Pierce, N. A. 2018. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development,* 145**,** dev165753.

Choi, W.-Y., Gemberling, M., Wang, J., Holdway, J. E., Shen, M.-C., Karlstrom, R. O. & Poss, K. D. 2013. In vivo monitoring of cardiomyocyte proliferation to identify chemical modifiers of heart regeneration. *Development (Cambridge, England),* 140**,** 660-666.

Chong, J. J. H., Reinecke, H., Iwata, M., Torok-Storb, B., Stempien-Otero, A. & Murry, C. E. 2013. Progenitor cells identified by PDGFR-alpha expression in the developing and diseased human heart. *Stem cells and development,* 22**,** 1932-1943.

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

Christoffels Vincent, M., Smits Gertien, J., Kispert, A. & Moorman Antoon, F. M. 2010. Development of the Pacemaker Tissues of the Heart. *Circulation Research,* 106, 240-254.

Colombo, S., De Sena-Tomás, C., George, V., Werdich, A. A., Kapur, S., Macrae, C. A. & Targoff, K. L. 2018. Nkx genes establish second heart field cardiomyocyte progenitors at the arterial pole and pattern the venous pole through Isl1 repression. *Development,* 145, dev161497.

Core, L. J., Waterfall, J. J. & Lis, J. T. 2008. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science,* 322, 1845.

Cretekos, C. J., Wang, Y., Green, E. D., Martin, J. F., Rasweiler, J. J. T. & Behringer, R. R. 2008. Regulatory divergence modifies limb length between mammals. *Genes & development,* 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., Boyer, L. A., Young, R. A. & Jaenisch, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences,* 107, 21931.

D. Wei, H. B., N. Zhou, G.F. Zheng, X.Y. Liu, Y.Q. Yang 2013. GATA5 loss-of-function mutation responsible for the congenital ventriculoseptal defect. *Pediatr. Cardiol.,* 34, 504-511.

David, N. B. & Rosa, F. M. 2001. Cell autonomous commitment to an endodermal fate and behaviour by activation of Nodal signalling. *Development,* 128, 3937.

Davis, D. L., Edwards, A. V., Juraszek, A. L., Phelps, A., Wessels, A. & Burch, J. B. E. 2001. A GATA-6 gene heart-region-specific enhancer provides a novel means to mark and probe a discrete component of the mouse cardiac conduction system. *Mechanisms of Development,* 108, 105-119.

Davoli, T. & Lange, T. D. 2011. The Causes and Consequences of Polyploidy in Normal Development and Cancer. 27, 585-610.

De Pater, E., Clijsters, L., Marques, S. R., Lin, Y.-F., Garavito-Aguilar, Z. V., Yelon, D. & Bakkers, J. 2009. Distinct phases of cardiomyocyte differentiation regulate growth of the zebrafish heart. *Development,* 136, 1633.

Den Haan, A. D., Moorman, A. F. M., Verkerk, A. O., Boukens, B. J., Tan, H. L., Van Den Boogaard, M., Bakker, M. L., Christoffels, V. M., Boink, G. J. J., De Bakker, J. M. T., Buermans, H. P., T Hoen, P. a. C., Hoogaars, W. M. H. & Seppen, J. 2012. T-box transcription factor TBX3 reprogrammes mature cardiac myocytes into pacemaker-like cells. *Cardiovascular Research,* 94, 439-449.

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

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. & Ren, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature,* 485, 376.

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B.-K., Pauli,

F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., Shoresh, N., Simon, J. M., Song, L., Trinklein, N. D., Altshuler, R. C., Birney, E., Brown, J. B., Cheng, C., Djebali, S., Dong, X., Dunham, I., Ernst, J., Furey, T. S., Gerstein, M., Giardine, B., Greven, M., Hardison, R. C., Harris, R. S., Herrero, J., Hoffman, M. M., Iyer, S., Kellis, M., Khatun, J., Kheradpour, P., Kundaje, A., Lassmann, T., Li, Q., Lin, X., Marinov, G. K., Merkel, A., Mortazavi, A., Parker, S. C. J., Reddy, T. E., Rozowsky, J., Schlesinger, F., Thurman, R. E., Wang, J., Ward, L. D., Whitfield, T. W., Wilder, S. P., Wu, W., Xi, H. S., Yip, K. Y., Zhuang, J., Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., Snyder, M., Pazin, M. J., Lowdon, R. F., Dillon, L. a. L., Adams, L. B., Kelly, C. J., Zhang, J., Wexler, J. R., Green, E. D., Good, P. J., Feingold, E. A., Bernstein, B. E., Birney, E., Crawford, G. E., Dekker, J., Elnitski, L., Farnham, P. J., Gerstein, M., Giddings, M. C., Gingeras, T. R., Green, E. D., Guigó, R., Hardison, R. C., Hubbard, T. J., Kellis, M., Kent, W. J., Lieb, J. D., Margulies, E. H., Myers, R. M., Snyder, M., Stamatoyannopoulos, J. A., Tenenbaum, S. A., et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57.

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

Dupays, L., Shang, C., Wilson, R., Kotecha, S., Wood, S., Towers, N. & Mohun, T. 2015. Sequential Binding of MEIS1 and NKX2-5 on the Popdc2 Gene: A Mechanism for Spatiotemporal Regulation of Enhancers during Cardiogenesis. *Cell reports,* 13**,** 183-195.

El-Rass, S., Eisa-Beygi, S., Khong, E., Brand-Arzamendi, K., Mauro, A., Zhang, H., Clark, K. J., Ekker, S. C. & Wen, X.-Y. 2017. Disruption of pdgfra alters endocardial and myocardial fusion during zebrafish cardiac assembly. *Biology Open,* 6**,** 348.

Emelyanov, A., Gao, Y., Naqvi, N. I. & Parinov, S. 2006. Trans-kingdom transposition of the maize dissociation element. *Genetics,* 174**,** 1095-1104.

Espinoza-Lewis, R. A., Yu, L., He, F., Liu, H., Tang, R., Shi, J., Sun, X., Martin, J. F., Wang, D., Yang, J. & Chen, Y. 2009. Shox2 is essential for the differentiation of cardiac pacemaker cells by repressing Nkx2-5. *Developmental biology,* 327**,** 376-385.

Evdokia Dodou, M. P. V., Joshua P. Anderson, Shan-Mei Xu, Brian L. Black 2004. Mef2c is a direct transcriptional target of ISL1 and GATA factors in the anterior heart field during mouse embryonic development. *Development,* 131**,** 3931-3942.

Fang, Y., Gupta, V., Karra, R., Holdway, J. E., Kikuchi, K. & Poss, K. D. 2013. Translational profiling of cardiomyocytes identifies an early Jak1/Stat3 injury response required for zebrafish heart regeneration. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 13416-13421.

Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S. & Levine, M. S. 2015. Suboptimization of developmental enhancers. *Science (New York, N.Y.),* 350**,** 325-328.

Felker, A., Prummel, K. D., Merks, A. M., Mickoleit, M., Brombacher, E. C., Huisken, J., Panáková, D. & Mosimann, C. 2018. Continuous addition of progenitors forms the cardiac ventricle in zebrafish. *Nature Communications,* 9**,** 2001.

Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L. & Mccallion, A. S. 2006. Conservation of RET Regulatory Function from Human to Zebrafish Without Sequence Similarity. *Science,* 312**,** 276.

Foglia, M. J. & Poss, K. D. 2016. Building and re-building the heart by cardiomyocyte proliferation. *Development (Cambridge, England),* 143**,** 729-740.
Forouhar, A. S., Liebling, M., Hickerson, A., Nasiraei-Moghaddam, A., Tsai, H.-J., Hove, J. R., Fraser, S. E., Dickinson, M. E. & Gharib, M. 2006. The Embryonic Vertebrate Heart Tube Is a Dynamic Suction Pump. *Science,* 312**,** 751.
Francou, A., De Bono, C. & Kelly, R. G. 2017. Epithelial tension in the second heart field promotes mouse heart tube elongation. *Nature Communications,* 8**,** 14770.
Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F. & Stern, D. L. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature,* 466**,** 490-493.
Fukui, H., Miyazaki, T., Chow, R. W.-Y., Ishikawa, H., Nakajima, H., Vermot, J. & Mochizuki, N. 2018. Hippo signaling determines the number of venous pole cells that originate from the anterior lateral plate mesoderm in zebrafish. *eLife,* 7**,** e29106.
Gajewski, K., Zhang, Q., Choi, C. Y., Fossett, N., Dang, A., Kim, Y. H., Kim, Y. & Schulz, R. A. 2001. Pannier is a Transcriptional Target and Partner of Tinman during Drosophila Cardiogenesis. *Developmental Biology,* 233**,** 425-436.
Garg, V., Kathiriya, I. S., Barnes, R., Schluterman, M. K., King, I. N., Butler, C. A., Rothrock, C. R., Eapen, R. S., Hirayama-Yamada, K., Joo, K., Matsuoka, R., Cohen, J. C. & Srivastava, D. 2003. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature,* 424**,** 443.
Geertz, M. & Maerkl, S. J. 2010. Experimental strategies for studying transcription factor-DNA binding specificities. *Briefings in functional genomics,* 9**,** 362-373.
Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., Yip, K. Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., Alves, P., Chateigner, A., Perry, M., Morris, M., Auerbach, R. K., Feng, X., Leng, J., Vielle, A., Niu, W., Rhrissorrakrai, K., Agarwal, A., Alexander, R. P., Barber, G., Brdlik, C. M., Brennan, J., Brouillet, J. J., Carr, A., Cheung, M.-S., Clawson, H., Contrino, S., Dannenberg, L. O., Dernburg, A. F., Desai, A., Dick, L., Dosé, A. C., Du, J., Egelhofer, T., Ercan, S., Euskirchen, G., Ewing, B., Feingold, E. A., Gassmann, R., Good, P. J., Green, P., Gullier, F., Gutwein, M., Guyer, M. S., Habegger, L., Han, T., Henikoff, J. G., Henz, S. R., Hinrichs, A., Holster, H., Hyman, T., Iniguez, A. L., Janette, J., Jensen, M., Kato, M., Kent, W. J., Kephart, E., Khivansara, V., Khurana, E., Kim, J. K., Kolasinska-Zwierz, P., Lai, E. C., Latorre, I., Leahey, A., Lewis, S., Lloyd, P., Lochovsky, L., Lowdon, R. F., Lubling, Y., Lyne, R., Maccoss, M., Mackowiak, S. D., Mangone, M., Mckay, S., Mecenas, D., Merrihew, G., Miller, D. M., Muroyama, A., Murray, J. I., Ooi, S.-L., Pham, H., Phippen, T., Preston, E. A., Rajewsky, N., Rätsch, G., Rosenbaum, H., Rozowsky, J., Rutherford, K., Ruzanov, P., Sarov, M., Sasidharan, R., Sboner, A., Scheid, P., Segal, E., Shin, H., Shou, C., Slack, F. J., et al. 2010. Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project. *Science,* 330**,** 1775.
Giresi, P. G., Kim, J., Mcdaniell, R. M., Iyer, V. R. & Lieb, J. D. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research,* 17**,** 877-885.
Gittenberger-De Groot, A. C., Bartelings, M. M., Poelmann, R. E., Haak, M. C. & Jongbloed, M. R. M. 2013. Embryology of the heart and its impact on understanding fetal and neonatal heart disease. *Seminars in Fetal and Neonatal Medicine,* 18**,** 237-244.
Glennon, P. E., Sugden, P. H. & Poole-Wilson, P. A. 1995. Cellular mechanisms of cardiac hypertrophy. *British Heart Journal,* 73**,** 496-499.

Gomes, R. S. M., Skroblin, P., Munster, A. B., Tomlins, H., Langley, S. R., Zampetaki, A., Yin, X., Wardle, F. C. & Mayr, M. 2016. "Young at heart": Regenerative potential linked to immature cardiac phenotypes. *Journal of molecular and cellular cardiology,* 92**,** 105-108.

González-Rosa, J. M. & Mercader, N. 2012. Cryoinjury as a myocardial infarction model for the study of cardiac regeneration in the zebrafish. *Nature Protocols,* 7**,** 782.

González-Rosa, J. M., Sharpe, M., Field, D., Soonpaa, M. H., Field, L. J., Burns, C. E. & Burns, C. G. 2018. Myocardial Polyploidization Creates a Barrier to Heart Regeneration in Zebrafish. *Developmental cell,* 44**,** 433-446.e7.

Gove, C., Walmsley, M., Nijjar, S., Bertwistle, D., Guille, M., Partington, G., Bomford, A. & Patient, R. 1997. Over-expression of GATA-6 in Xenopus embryos blocks differentiation of heart precursors. *The EMBO Journal,* 16**,** 355-368.

Grass, J. A., Boyer, M. E., Pal, S., Wu, J., Weiss, M. J. & Bresnick, E. H. 2003. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proceedings of the National Academy of Sciences,* 100**,** 8811.

Grepin, C., Nemer, G. & Nemer, M. 1997. Enhanced cardiogenesis in embryonic stem cells overexpressing the GATA-4 transcription factor. *Development,* 124**,** 2387-2395.

Grépin, C., Robitaille, L., Antakly, T. & Nemer, M. 1995. Inhibition of transcription factor GATA-4 expression blocks in vitro cardiac muscle differentiation. *Molecular and Cellular Biology,* 15**,** 4095-4102.

Grimes, A. C., Durán, A. C., Sans-Coma, V., Hami, D., Santoro, M. M. & Torres, M. 2010. Phylogeny informs ontogeny: a proposed common theme in the arterial pole of the vertebrate heart. *Evolution & development,* 12**,** 552-567.

Grimes, A. C. & Kirby, M. L. 2009. The outflow tract of the heart in fishes: anatomy, genes and evolution. *Journal of Fish Biology,* 74**,** 983-1036.

Grossman, S. R., Engreitz, J., Ray, J. P., Nguyen, T. H., Hacohen, N. & Lander, E. S. 2018. Positional specificity of different transcription factor classes within enhancers. *Proceedings of the National Academy of Sciences,* 115**,** E7222.

Gu, J.-Y., Xu, J.-H., Yu, H. & Yang, Y.-Q. 2012. Novel GATA5 loss-of-function mutations underlie familial atrial fibrillation. *Clinics (Sao Paulo, Brazil),* 67**,** 1393-1399.

Guner-Ataman, B., Paffett-Lugassy, N., Adams, M. S., Nevis, K. R., Jahangiri, L., Obregon, P., Kikuchi, K., Poss, K. D., Burns, C. E. & Burns, C. G. 2013. Zebrafish second heart field development relies on progenitor specification in anterior lateral plate mesoderm and nkx2.5 function. *Development (Cambridge, England),* 140**,** 1353-1363.

Günthel, M., Barnett, P. & Christoffels, V. M. 2018. Development, Proliferation, and Growth of the Mammalian Heart. *Molecular Therapy,* 26**,** 1599-1609.

Gupta, V. & Poss, K. D. 2012. Clonally dominant cardiomyocytes direct heart morphogenesis. *Nature,* 484**,** 479.

Hafez, D., Karabacak, A., Krueger, S., Hwang, Y.-C., Wang, L.-S., Zinzen, R. P. & Ohler, U. 2017. McEnhancer: predicting gene expression via semi-supervised assignment of enhancers to target genes. *Genome biology,* 18**,** 199-199.

Hami, D., Grimes, A. C., Tsai, H.-J. & Kirby, M. L. 2011. Zebrafish cardiac development requires a conserved secondary heart field. *Development (Cambridge, England),* 138**,** 2389-2398.

Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merkenschlager, M. & Lenhard, B. 2017. Topologically associating domains are ancient features that coincide with

Metazoan clusters of extreme noncoding conservation. *Nature communications,* 8**,** 441-441.

Harrington, J. K., Sorabella, R., Tercek, A., Isler, J. R. & Targoff, K. L. 2017. Nkx2.5 is essential to establish normal heart rate variability in the zebrafish embryo. *American journal of physiology. Regulatory, integrative and comparative physiology,* 313**,** R265-R271.

Harris, J. P., Bhakta, M., Bezprozvannaya, S., Wang, L., Lubczyk, C., Olson, E. N. & Munshi, N. V. 2015. MyoR Modulates Cardiac Conduction by Repressing Gata4. *Molecular and Cellular Biology,* 35**,** 649.

Haubner Bernhard, J., Schneider, J., Schweigmann, U., Schuetz, T., Dichtl, W., Velik-Salchner, C., Stein Joerg, I. & Penninger Josef, M. 2016. Functional Recovery of a Human Neonatal Heart After Severe Myocardial Infarction. *Circulation Research,* 118**,** 216-221.

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., Telenius, J., Williams, R., Rode, C., Li, P.-S., Pennacchio, L. A., Sloane-Stanley, J. A., Ayyub, H., Butler, S., Sauka-Spengler, T., Gibbons, R. J., Smith, A. J. H., Wood, W. G. & Higgs, D. R. 2016. Genetic dissection of the α-globin super-enhancer in vivo. *Nature genetics,* 48**,** 895-903.

Heger, P. & Wiehe, T. 2014. New tools in the box: An evolutionary synopsis of chromatin insulators. *Trends in Genetics,* 30**,** 161-171.

Heicklen-Klein, A. & Evans, T. 2004. T-box binding sites are required for activity of a cardiac GATA-4 enhancer. *Developmental Biology,* 267**,** 490-504.

Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. 2015. The selection and function of cell type-specific enhancers. *Nature reviews. Molecular cell biology,* 16**,** 144-154.

Hernández, P. P., Strzelecka, P. M., Athanasiadis, E. I., Hall, D., Robalo, A. F., Collins, C. M., Boudinot, P., Levraud, J.-P. & Cvejic, A. 2018. Single-cell transcriptional analysis reveals ILC-like cells in zebrafish. *Science Immunology,* 3**,** eaau5265.

Hoffmann, S., Berger, I. M., Glaser, A., Bacon, C., Li, L., Gretz, N., Steinbeisser, H., Rottbauer, W., Just, S. & Rappold, G. 2013. Islet1 is a direct transcriptional target of the homeodomain transcription factor Shox2 and rescues the Shox2-mediated bradycardia. *Basic research in cardiology,* 108**,** 339-339.

Hong, J.-W., Hendrix, D. A. & Levine, M. S. 2008. Shadow enhancers as a source of evolutionary novelty. *Science (New York, N.Y.),* 321**,** 1314-1314.

Hoogaars, W. M. H., Engel, A., Brons, J. F., Verkerk, A. O., De Lange, F. J., Wong, L. Y. E., Bakker, M. L., Clout, D. E., Wakker, V., Barnett, P., Ravesloot, J. H., Moorman, A. F. M., Verheijck, E. E. & Christoffels, V. M. 2007. Tbx3 controls the sinoatrial node gene program and imposes pacemaker function on the atria. *Genes & Development,* 21**,** 1098-1112.

Hughes, J. R., Roberts, N., Mcgowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R. & Higgs, D. R. 2014. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics,* 46**,** 205.

Hulsmans, M., Clauss, S., Xiao, L., Aguirre, A. D., King, K. R., Hanley, A., Hucker, W. J., Wülfers, E. M., Seemann, G., Courties, G., Iwamoto, Y., Sun, Y., Savol, A. J., Sager, H. B., Lavine, K. J., Fishbein, G. A., Capen, D. E., Da Silva, N., Miquerol, L., Wakimoto, H., Seidman, C. E., Seidman, J. G., Sadreyev, R. I., Naxerova, K., Mitchell, R. N., Brown, D., Libby, P., Weissleder, R., Swirski, F. K., Kohl, P., Vinegoni, C., Milan, D. J.,

Ellinor, P. T. & Nahrendorf, M. 2017. Macrophages Facilitate Electrical Conduction in the Heart. *Cell,* 169**,** 510-522.e20.
Ishibashi, M., Mechaly, A. S., Becker, T. S. & Rinkwitz, S. 2013. Using zebrafish transgenesis to test human genomic sequences for specific enhancer activity. *Methods,* 62**,** 216-225.
Jiang, X., Rowitch, D. H., Soriano, P., Mcmahon, A. P. & Sucov, H. M. 2000. Fate of the mammalian cardiac neural crest. *Development,* 127**,** 1607.
Jiang, Y. & Evans, T. 1996. TheXenopusGATA-4/5/6 Genes Are Associated with Cardiac Specification and Can Regulate Cardiac-Specific Transcription during Embryogenesis. *Developmental Biology,* 174**,** 258-270.
Jin, S.-W., Beis, D., Mitchell, T., Chen, J.-N. & Stainier, D. Y. R. 2005. Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development,* 132**,** 5199.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science,* 337**,** 816.
Jolma, A., Yin, Y., Nitta, K. R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E. & Taipale, J. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature,* 527**,** 384.
Jopling, C., Sleep, E., Raya, M., Marti, M., Raya, A. & Izpisua Belmonte, J. C. 2010. Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature,* 464.
Juan Manuel González-Rosa, V. M., Marina Peralta, Miguel Torres and Nadia Mercader* 2011. Extensive scar formation and regression during heart regeneration after cryoinjury in zebrafish. *Development,* 138**,** 1663-1674.
Juven-Gershon, T., Hsu, J.-Y., Theisen, J. W. & Kadonaga, J. T. 2008. The RNA polymerase II core promoter - the gateway to transcription. *Current opinion in cell biology,* 20**,** 253-259.
Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., Van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J. & Young, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature,* 467**,** 430.
Karunamuni, G., Yang, K., Doughman, Y. Q., Wikenheiser, J., Bader, D., Barnett, J., Austin, A., Parsons-Wingerter, P. & Watanabe, M. 2010. Expression of lymphatic markers during avian and mouse cardiogenesis. *Anatomical record (Hoboken, N.J. : 2007),* 293**,** 259-270.
Keegan, B. R., Feldman, J. L., Begemann, G., Ingham, P. W. & Yelon, D. 2005. Retinoic Acid Signaling Restricts the Cardiac Progenitor Pool. *Science,* 307**,** 247.
Keegan, B. R., Meyer, D. & Yelon, D. 2004. Organization of cardiac chamber progenitors in the zebrafish blastula. *Development,* 131**,** 3081.
Kikuchi, K. 2014. Advances in understanding the mechanism of zebrafish heart regeneration. *Stem Cell Research,* 13**,** 542-555.
Kikuchi, K., Holdway, J. E., Major, R. J., Blum, N., Dahn, R. D., Begemann, G. & Poss, K. D. 2011. Retinoic acid production by endocardium and epicardium is an injury response essential for zebrafish heart regeneration. *Developmental cell,* 20**,** 397-404.

Kikuchi, K., Holdway, J. E., Werdich, A. A., Anderson, R. M., Fang, Y., Egnaczyk, G. F., Evans, T., Macrae, C. A., Stainier, D. Y. R. & Poss, K. D. 2010. Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes. *Nature,* 464**,** 601-605.

Kikuchi, K. & Poss, K. D. 2012. Cardiac regenerative capacity and mechanisms. *Annu Rev Cell Dev Biol,* 28.

Kim, B.-J., Kim, Y.-H., Lee, Y.-A., Jung, S.-E., Hong, Y. H., Lee, E.-J., Kim, B.-G., Hwang, S., Do, J. T., Pang, M.-G. & Ryu, B.-Y. 2017. Platelet-derived growth factor receptor-alpha positive cardiac progenitor cells derived from multipotent germline stem cells are capable of cardiomyogenesis in vitro and in vivo. *Oncotarget,* 8**,** 29643-29656.

Kim, J., Wu, Q., Zhang, Y., Wiens, K. M., Huang, Y., Rubin, N., Shimada, H., Handin, R. I., Chao, M. Y., Tuan, T.-L., Starnes, V. A. & Lien, C.-L. 2010a. PDGF signaling is required for epicardial function and blood vessel formation in regenerating zebrafish hearts. *Proceedings of the National Academy of Sciences,* 107**,** 17206.

Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G. & Greenberg, M. E. 2010b. Widespread transcription at neuronal activity-regulated enhancers. *Nature,* 465**,** 182.

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. 1995. Stages of embryonic development of the zebrafish. *Developmental Dynamics,* 203**,** 253-310.

Kirby, M. L., Gale, T. F. & Stewart, D. E. 1983. Neural crest cells contribute to normal aorticopulmonary septation. *Science,* 220**,** 1059.

Krasnov, A. N., Mazina, M. Y., Nikolenko, J. V. & Vorobyeva, N. E. 2016. On the way of revealing coactivator complexes cross-talk during transcriptional activation. *Cell & bioscience,* 6**,** 15-15.

Kulkarni, M. M. & Arnosti, D. N. 2003. Information display by transcriptional enhancers. *Development,* 130**,** 6569.

Kuo Ct1, M. E., Anandappa R, Sigrist K, Lu Mm, Parmacek Ms, Soudais C, Leiden Jm. 1997. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev,* 15**,** 1048-60.

Kwasnieski, J. C., Fiore, C., Chaudhari, H. G. & Cohen, B. A. 2014. High-throughput functional testing of ENCODE segmentation predictions. *Genome research,* 24**,** 1595-1602.

L. Van Der Maaten, G. H. 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.,* vol. 9**,** pp. 2579-2605.

Lavine, K. J., Epelman, S., Uchida, K., Weber, K. J., Nichols, C. G., Schilling, J. D., Ornitz, D. M., Randolph, G. J. & Mann, D. L. 2014. Distinct macrophage lineages contribute to disparate patterns of cardiac recovery and remodeling in the neonatal and adult heart. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 16029-16034.

Lazic, S. & Scott, I. C. 2011. Mef2cb regulates late myocardial cell addition from a second heart field-like population of progenitors in zebrafish. *Developmental Biology,* 354**,** 123-133.

Le Douarin, N. K., C. 1999. *The neural crest.*, Cambridge university press.

Lee, H. J., Lowdon, R. F., Maricque, B., Zhang, B., Stevens, M., Li, D., Johnson, S. L. & Wang, T. 2015. Developmental enhancers revealed by extensive DNA methylome maps of zebrafish early embryos. *Nature communications,* 6**,** 6315-6315.

Lee, K.-H., Xu, Q. & Breitbart, R. E. 1996. A Newtinman-Related Gene,nkx2.7,Anticipates the Expression of nkx2.5and nkx2.3 in Zebrafish Heart and Pharyngeal Endoderm. *Developmental Biology,* 180**,** 722-731.

Lee, T. I. & Young, R. A. 2000. Transcription of eukaryotic protein-coding genes. *Annual Review of Genetics,* 34**,** 77-137.

Lemon, B. & Tjian, R. 2000. Orchestrated response: a symphony of transcription factors for gene control. 14**,** 2551-2569.

Lepilina, A., Coon, A. N., Kikuchi, K., Holdway, J. E., Roberts, R. W., Burns, C. G. & Poss, K. D. 2006. A Dynamic Epicardial Injury Response Supports Progenitor Cell Activity during Zebrafish Heart Regeneration. *Cell,* 127**,** 607-619.

Lettice, L. A. 2003. *A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly*.

Li, C., Li, X., Pang, S., Chen, W., Qin, X., Huang, W., Zeng, C. & Yan, B. 2014. Novel and functional DNA sequence variants within the GATA6 gene promoter in ventricular septal defects. *International journal of molecular sciences,* 15**,** 12677-12687.

Li, Q., Ritter, D., Yang, N., Dong, Z., Li, H., Chuang, J. H. & Guo, S. 2010. A systematic approach to identify functional motifs within vertebrate developmental enhancers. *Developmental biology,* 337**,** 484-495.

Li, W., Notani, D. & Rosenfeld, M. G. 2016. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nature Reviews Genetics,* 17**,** 207.

Li, Y.-X., Zdanowicz, M., Young, L., Kumiski, D., Leatherbury, L. & Kirby, M. L. 2003. Cardiac neural crest in zebrafish embryos contributes to myocardial cell lineage and early heart function. *Developmental Dynamics,* 226**,** 540-550.

Liang, Q., De Windt, L. J., Witt, S. A., Kimball, T. R., Markham, B. E. & Molkentin, J. D. 2001. The Transcription Factors GATA4 and GATA6 Regulate Cardiomyocyte Hypertrophy in Vitro and in Vivo. *Journal of Biological Chemistry,* 276**,** 30245-30253.

Liao, J., Aggarwal, V. S., Nowotschin, S., Bondarev, A., Lipner, S. & Morrow, B. E. 2008. Identification of downstream genetic pathways of Tbx1 in the second heart field. *Developmental Biology,* 316**,** 524-537.

Lien, C. L., Wu, C., Mercer, B., Webb, R., Richardson, J. A. & Olson, E. N. 1999. Control of early cardiac-specific transcription of Nkx2-5 by a GATA-dependent enhancer. *Development,* 126**,** 75.

Lin, Q., Schwarz, J., Bucana, C. & Olson, E. N. 1997. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science (New York, N.Y.),* 276**,** 1404-1407.

Lin, X., Huo, Z., Liu, X., Zhang, Y., Li, L., Zhao, H., Yan, B., Liu, Y., Yang, Y. & Chen, Y.-H. 2010. A novel GATA6 mutation in patients with tetralogy of Fallot or atrial septal defect. *Journal of Human Genetics,* 55**,** 662-667.

Link, B. A., Jeffery, B. E., Flinn, M. A. & O'meara, C. C. 2018. Yap is required for scar formation but not myocyte proliferation during heart regeneration in zebrafish. *Cardiovascular Research,* 115**,** 570-577.

Liu, F., Lu Min, M., Patel Neil, N., Schillinger Kurt, J., Wang, T. & Patel Vickas, V. 2015. GATA-Binding Factor 6 Contributes to Atrioventricular Node Development and Function. *Circulation: Cardiovascular Genetics,* 8**,** 284-293.

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

Lopaschuk, G. D. & Jaswal, J. S. 2010. Energy Metabolic Phenotype of the Cardiomyocyte During Development, Differentiation, and Postnatal Maturation. *Journal of Cardiovascular Pharmacology,* 56.

Luther Daniel, J., Thodeti Charles, K., Shamhart Patricia, E., Adapala Ravi, K., Hodnichak, C., Weihrauch, D., Bonaldo, P., Chilian William, M. & Meszaros, J. G. 2012. Absence of Type VI Collagen Paradoxically Improves Cardiac Function, Structure, and Remodeling After Myocardial Infarction. *Circulation Research,* 110**,** 851-856.

Maeda, J., Yamagishi, H., Mcanally, J., Yamagishi, C. & Srivastava, D. 2006. Tbx1 is regulated by forkhead proteins in the secondary heart field. *Developmental dynamics : an official publication of the American Association of Anatomists,* 235**,** 701-710.

Maitra, M., Schluterman, M. K., Nichols, H. A., Richardson, J. A., Lo, C. W., Srivastava, D. & Garg, V. 2009. Interaction of Gata4 and Gata6 with Tbx5 is critical for normal cardiac development. *Developmental biology,* 326**,** 368-377.

Malik, S. & Roeder, R. G. 2010. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature reviews. Genetics,* 11**,** 761-772.

Mangoni Matteo, E., Traboulsie, A., Leoni, A.-L., Couette, B., Marger, L., Le Quang, K., Kupfer, E., Cohen-Solal, A., Vilar, J., Shin, H.-S., Escande, D., Charpentier, F., Nargeot, J. & Lory, P. 2006. Bradycardia and Slowing of the Atrioventricular Conduction in Mice Lacking CaV3.1/α1G T-Type Calcium Channels. *Circulation Research,* 98**,** 1422-1430.

Marín-Juez, R., Marass, M., Gauvrit, S., Rossi, A., Lai, S.-L., Materna, S. C., Black, B. L. & Stainier, D. Y. R. 2016. Fast revascularization of the injured area is essential to support zebrafish heart regeneration. *Proceedings of the National Academy of Sciences of the United States of America,* 113**,** 11237-11242.

Maston, G. A., Evans, S. K. & Green, M. R. 2006. Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics,* 7**,** 29-59.

Mcclintock, B. 1950. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America,* 36**,** 344-355.

Merika, M. & Thanos, D. 2001. Enhanceosomes. *Current Opinion in Genetics & Development,* 11**,** 205-208.

Miesfeld, J. B. & Link, B. A. 2014. Establishment of transgenic lines to monitor and manipulate Yap/Taz-Tead activity in zebrafish reveals both evolutionarily conserved and divergent functions of the Hippo pathway. *Mechanisms of development,* 133**,** 177-188.

Milan, D. J., Giokas, A. C., Serluca, F. C., Peterson, R. T. & Macrae, C. A. 2006. Notch1b and neuregulin are required for specification of central cardiac conduction tissue. *Development,* 133**,** 1125.

Miller, J. A. & Widom, J. 2003. Collaborative Competition Mechanism for Gene Activation In Vivo. *Molecular and Cellular Biology,* 23**,** 1623.

Minehata, K.-I., Kawahara, A. & Suzuki, T. 2008. meis1 regulates the development of endothelial cells in zebrafish. *Biochemical and Biophysical Research Communications,* 374**,** 647-652.

Molkentin, J. D., Antos, C., Mercer, B., Taigen, T., Miano, J. M. & Olson, E. N. 2000. Direct Activation of a GATA6 Cardiac Enhancer by Nkx2.5: Evidence for a Reinforcing

Regulatory Network of Nkx2.5 and GATA Transcription Factors in the Developing Heart. *Developmental Biology,* 217**,** 301-309.
Mongera, A., Singh, A. P., Levesque, M. P., Chen, Y.-Y., Konstantinidis, P. & Nüsslein-Volhard, C. 2013. Genetic lineage labeling in zebrafish uncovers novel neural crest contributions to the head, including gill pillar cells. *Development,* 140**,** 916.
Moorman, A. F. M. & Christoffels, V. M. 2003. Cardiac Chamber Formation: Development, Genes, and Evolution. *Physiological Reviews,* 83**,** 1223-1267.
Morrisey, E. E., Tang, Z., Sigrist, K., Lu, M. M., Jiang, F., Ip, H. S. & Parmacek, M. S. 1998. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes & Development,* 12**,** 3579-3590.
Moskowitz, I. P., Wang, J., Peterson, M. A., Pu, W. T., Mackinnon, A. C., Oxburgh, L., Chu, G. C., Sarkar, M., Berul, C., Smoot, L., Robertson, E. J., Schwartz, R., Seidman, J. G. & Seidman, C. E. 2011. Transcription factor genes Smad4 and Gata4 cooperatively regulate cardiac valve development. [corrected]. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 4006-4011.
N, F. 2014. Trials of autologous bone marrow stem cells for heart disease. *BMJ,* 348.
N. Fedoroff, S. W., And M. Shure. 1983. Isolation of the transposable maize controlling elements Ac and Ds. *Cell,* 35(1)**,** 235–242.
Nechaev, S. & Adelman, K. 2011. Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochimica et biophysica acta,* 1809**,** 34-45.
Nemer, G. & Nemer, M. 2002. Cooperative interaction between GATA5 and NF-ATc regulates endothelial-endocardial differentiation of cardiogenic cells. *Development,* 129**,** 4045.
Nemer, M. & Gharibeh, L. 2015. Guiding Cardiac Conduction With GATA. *Circulation: Cardiovascular Genetics,* 8**,** 247-249.
Nevis, K., Obregon, P., Walsh, C., Guner-Ataman, B., Burns, C. G. & Burns, C. E. 2013. Tbx1 is required for second heart field proliferation in zebrafish. *Developmental dynamics : an official publication of the American Association of Anatomists,* 242**,** 550-559.
Noseda, M., Harada, M., Mcsweeney, S., Leja, T., Belian, E., Stuckey, D. J., Abreu Paiva, M. S., Habib, J., Macaulay, I., De Smith, A. J., Al-Beidh, F., Sampson, R., Lumbers, R. T., Rao, P., Harding, S. E., Blakemore, A. I. F., Jacobsen, S. E., Barahona, M. & Schneider, M. D. 2015. PDGFRα demarcates the cardiogenic clonogenic Sca1+ stem/progenitor cell in adult murine myocardium. *Nature communications,* 6**,** 6930-6930.
Oberpriller, J. O., Oberpriller, J. C., Arefyeva, A. M., Mitashov, V. I. & Carlson, B. M. 1988. Nuclear characteristics of cardiac myocytes following the proliferative response to mincing of the myocardium in the adult newt, Notophthalmus viridescens. *Cell and Tissue Research,* 253**,** 619-624.
Ocaña, O. H., Coskun, H., Minguillón, C., Murawala, P., Tanaka, E. M., Galcerán, J., Muñoz-Chápuli, R. & Nieto, M. A. 2017. A right-handed signalling pathway drives heart looping in vertebrates. *Nature,* 549**,** 86-90.
Ogbourne, S. & Antalis, T. M. 1998. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *The Biochemical journal,* 331 (Pt 1)**,** 1-14.
Ong, C.-T. & Corces, V. G. 2014. CTCF: an architectural protein bridging genome topology and function. *Nature reviews. Genetics,* 15**,** 234-246.

Orr-Weaver, T. L. 2015. When bigger is better: the role of polyploidy in organogenesis. *Trends in Genetics,* 31**,** 307-315.

Ostadal, B. D., N. S. 2012. *Cardiac Adaptations: Molecular Mechanisms.*

Oudelaar, A. M., Downes, D. J., Davies, J. O. J. & Hughes, J. R. 2017. Low-input Capture-C: A Chromosome Conformation Capture Assay to Analyze Chromatin Architecture in Small Numbers of Cells. *Bio-protocol,* 7**,** e2645.

Ovcharenko, I., Stubbs, L. & Loots, G. G. 2004. Interpreting mammalian evolution using Fugu genome comparisons. *Genomics,* 84**,** 890-895.

Paige, S. L., Thomas, S., Stoick-Cooper, C. L., Wang, H., Maves, L., Sandstrom, R., Pabon, L., Reinecke, H., Pratt, G., Keller, G., Moon, R. T., Stamatoyannopoulos, J. & Murry, C. E. 2012. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell,* 151**,** 221-232.

Patient, R. K. & Mcghee, J. D. 2002. The GATA family (vertebrates and invertebrates). *Current Opinion in Genetics & Development,* 12**,** 416-422.

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A. & Rubin, E. M. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature,* 444**,** 499-502.

Peralta M, S. E., Harlepp S, Gonzalez-Rosa Jm, Monduc F, Ariza-Cosano a, Cortes a, Rayon T, Gomez-Skarmeta Jl, Zapata a, Et Al. 2013. Heartbeat-driven pericardiac fluid forces contribute to epicardium morphogenesis. . *Curr. Biol.,* 23**,** 1726–1735.

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

Pertea, M., Shumate, A., Pertea, G., Varabyou, A., Chang, Y.-C., Madugundu, A. K., Pandey, A. & Salzberg, S. 2018. Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise. *bioRxiv***,** 332825.

Peshkovsky, C., Totong, R. & Yelon, D. 2011. Dependence of cardiac trabeculation on neuregulin signaling and blood flow in zebrafish. *Developmental Dynamics,* 240**,** 446-456.

Peterkin, T., Gibson, A., Loose, M. & Patient, R. 2005. The roles of GATA-4, -5 and -6 in vertebrate heart development. *Semin Cell Dev Biol,* 16**,** 83-94.

Peterkin, T., Gibson, A. & Patient, R. 2003. GATA-6 maintains BMP-4 and Nkx2 expression during cardiomyocyte precursor maturation. *The EMBO Journal,* 22**,** 4260-4273.

Peterkin, T., Gibson, A. & Patient, R. 2007a. Redundancy and evolution of GATA factor requirements in development of the myocardium. *Developmental biology,* 311**,** 623-635.

Peterkin, T., Gibson, A. & Patient, R. 2007b. Redundancy and evolution of GATA factor requirements in development of the myocardium. *Dev Biol,* 311**,** 623-35.

Peterkin, T., Gibson, A. & Patient, R. 2009. Common genetic control of haemangioblast and cardiac development in zebrafish. *Development,* 136**,** 1465.

Phillips-Cremins, Jennifer e., Sauria, Michael e. G., Sanyal, A., Gerasimova, Tatiana i., Lajoie, Bryan r., Bell, Joshua s. K., Ong, C.-T., Hookway, Tracy a., Guo, C., Sun, Y., Bland, Michael j., Wagstaff, W., Dalton, S., Mcdevitt, Todd c., Sen, R., Dekker, J., Taylor, J.

& Corces, Victor g. 2013. Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell,* 153**,** 1281-1295.
Poon, K.-L., Liebling, M., Kondrychyn, I., Brand, T. & Korzh, V. 2016. Development of the cardiac conduction system in zebrafish. *Gene Expression Patterns,* 21**,** 89-96.
Porrello, E. R., Mahmoud, A. I., Simpson, E., Hill, J. A., Richardson, J. A., Olson, E. N. & Sadek, H. A. 2011. Transient Regenerative Potential of the Neonatal Mouse Heart. *Science,* 331**,** 1078.
Posch, M. G., Gramlich, M., Sunde, M., Schmitt, K. R., Lee, S. H. Y., Richter, S., Kersten, A., Perrot, A., Panek, A. N., Al Khatib, I. H., Nemer, G., Mégarbané, A., Dietz, R., Stiller, B., Berger, F., Harvey, R. P. & Ozcelik, C. 2010. A gain-of-function TBX20 mutation causes congenital atrial septal defects, patent foramen ovale and cardiac valve defects. *Journal of medical genetics,* 47**,** 230-235.
Poss, K. D., Wilson, L. G. & Keating, M. T. 2002. Heart regeneration in zebrafish. *Science,* 298.
Prazak, L., Fujioka, M. & Gergen, J. P. 2010. Non-additive interactions involving two distinct elements mediate sloppy-paired regulation by pair-rule transcription factors. *Developmental biology,* 344**,** 1048-1059.
Privalsky, M. L. 2004. The Role of Corepressors in Transcriptional Regulation by Nuclear Hormone Receptors. *Annual Review of Physiology,* 66**,** 315-360.
Ptashne, M. & Gann, A. 1997. Transcriptional activation by recruitment. *Nature,* 386**,** 569-577.
Puente, Bao n., Kimura, W., Muralidhar, Shalini a., Moon, J., Amatruda, James f., Phelps, Kate l., Grinsfelder, D., Rothermel, Beverly a., Chen, R., Garcia, Joseph a., Santos, Celio x., Thet, S., Mori, E., Kinter, Michael t., Rindler, Paul m., Zacchigna, S., Mukherjee, S., Chen, David j., Mahmoud, Ahmed i., Giacca, M., Rabinovitch, Peter s., Asaithamby, A., Shah, Ajay m., Szweda, Luke i. & Sadek, Hesham a. 2014. The Oxygen-Rich Postnatal Environment Induces Cardiomyocyte Cell-Cycle Arrest through DNA Damage Response. *Cell,* 157**,** 565-579.
Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A. & Wysocka, J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature,* 470**,** 279-283.
Raya, A., Koth, C. M., Büscher, D., Kawakami, Y., Itoh, T., Raya, R. M., Sternik, G., Tsai, H.-J., Rodríguez-Esteban, C. & Izpisúa-Belmonte, J. C. 2003. Activation of Notch signaling pathway precedes heart regeneration in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America,* 100 Suppl 1**,** 11889-11895.
Reiter, J. F., Alexander, J., Rodaway, A., Yelon, D., Patient, R., Holder, N. And Stainier, D. Y. R. 1999. Gata5 is required for the development of the heart and endoderm in zebrafish. *Genes Dev.,* 13**,** 2983-2995.
Ribeiro, I., Kawakami, Y., Büscher, D., Raya, A., Rodríguez-León, J., Morita, M., Rodríguez Esteban, C. & Izpisúa Belmonte, J. C. 2007. Tbx2 and Tbx3 regulate the dynamics of cell proliferation during heart remodeling. *PloS one,* 2**,** e398-e398.
Ringrose, L. & Paro, R. 2004. Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins. *Annual Review of Genetics,* 38**,** 413-443.
Rohr, S., Otten, C. & Abdelilah-Seyfried, S. 2008. Asymmetric Involution of the Myocardial Field Drives Heart Tube Formation in Zebrafish. *Circulation Research,* 102**,** e12-e19.
Rojas, A. D. V., Sarah; Heidt, Analeah B.; Xu, Shan-Mei; Bristow, James; Black, Brian L. 2005. Gata4 expression in lateral mesoderm is downstream of BMP4 and is

activated directly by Forkhead and GATA transcription factors through a distal enhancer element. *Development* 132**,** 3405-3417.

Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L., Landolin, J. M., Bristow, C. A., Ma, L., Lin, M. F., Washietl, S., Arshinoff, B. I., Ay, F., Meyer, P. E., Robine, N., Washington, N. L., Di Stefano, L., Berezikov, E., Brown, C. D., Candeias, R., Carlson, J. W., Carr, A., Jungreis, I., Marbach, D., Sealfon, R., Tolstorukov, M. Y., Will, S., Alekseyenko, A. A., Artieri, C., Booth, B. W., Brooks, A. N., Dai, Q., Davis, C. A., Duff, M. O., Feng, X., Gorchakov, A. A., Gu, T., Henikoff, J. G., Kapranov, P., Li, R., Macalpine, H. K., Malone, J., Minoda, A., Nordman, J., Okamura, K., Perry, M., Powell, S. K., Riddle, N. C., Sakai, A., Samsonova, A., Sandler, J. E., Schwartz, Y. B., Sher, N., Spokony, R., Sturgill, D., Van Baren, M., Wan, K. H., Yang, L., Yu, C., Feingold, E., Good, P., Guyer, M., Lowdon, R., Ahmad, K., Andrews, J., Berger, B., Brenner, S. E., Brent, M. R., Cherbas, L., Elgin, S. C. R., Gingeras, T. R., Grossman, R., Hoskins, R. A., Kaufman, T. C., Kent, W., Kuroda, M. I., Orr-Weaver, T., Perrimon, N., Pirrotta, V., Posakony, J. W., Ren, B., Russell, S., Cherbas, P., Graveley, B. R., Lewis, S., Micklem, G., Oliver, B., Park, P. J., Celniker, S. E., Henikoff, S., Karpen, G. H., Lai, E. C., Macalpine, D. M., Stein, L. D., White, K. P. & Kellis, M. 2010. Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. *Science,* 330**,** 1787.

Rudat, C., Greulich, F. & Kispert, A. 2011. Mechanisms of T-box gene function in the developing heart. *Cardiovascular Research,* 91**,** 212-222.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J. & Kouzarides, T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature,* 419**,** 407-411.

Sato, M. & Yost, H. J. 2003. Cardiac neural crest contributes to cardiomyogenesis in zebrafish. *Developmental Biology,* 257**,** 127-139.

Schier, A. F., Neuhauss, S. C., Helde, K. A., Talbot, W. S. & Driever, W. 1997. The one-eyed pinhead gene functions in mesoderm and endoderm formation in zebrafish and interacts with no tail. *Development,* 124**,** 327.

Schindler, Y. L., Garske, K. M., Wang, J., Firulli, B. A., Firulli, A. B., Poss, K. D. & Yelon, D. 2014. Hand2 elevates cardiomyocyte production during zebrafish heart development and regeneration. *Development,* 141**,** 3112.

Schoenebeck, J. J., Keegan, B. R. & Yelon, D. 2007. Vessel and Blood Specification Override Cardiac Potential in Anterior Mesoderm. *Developmental Cell,* 13**,** 254-267.

Searcy, R. D., Vincent, E. B., Liberatore, C. M. & Yutzey, K. E. 1998. A GATA-dependent nkx-2.5 regulatory element activates early cardiac gene expression in transgenic mice. *Development,* 125**,** 4461.

Sedmera, D., Reckova, M., Dealmeida, A., Sedmerova, M., Biermann, M., Volejnik, J., Sarre, A., Raddatz, E., Mccarthy, R. A., Gourdie, R. G. & Thompson, R. P. 2003. Functional and morphological evidence for a ventricular conduction system in zebrafish and Xenopus hearts. *American Journal of Physiology-Heart and Circulatory Physiology,* 284**,** H1152-H1160.

Serluca, F. C. 2008. Development of the proepicardial organ in the zebrafish. *Developmental Biology,* 315**,** 18-27.

Shah, A. N., Davey, C. F., Whitebirch, A. C., Miller, A. C. & Moens, C. B. 2015. Rapid reverse genetic screening using CRISPR in zebrafish. *Nature methods,* 12**,** 535-540.

Shlyueva, D., Stampfel, G. & Stark, A. 2014a. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet,* 15**,** 272-286.
Shlyueva, D., Stampfel, G. & Stark, A. 2014b. Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics,* 15**,** 272.
Simões, F. C., Peterkin, T. & Patient, R. 2011. Fgf differentially controls cross-antagonism between cardiac and haemangioblast regulators. *Development,* 138**,** 3235.
Simões, F. C. & Riley, P. R. 2018. The ontogeny, activation and function of the epicardium during heart development and regeneration. *Development,* 145**,** dev155994.
Sirbu, I. O., Zhao, X. & Duester, G. 2008. Retinoic acid controls heart anteroposterior patterning by down-regulating Isl1 through the Fgf8 pathway. *Developmental dynamics : an official publication of the American Association of Anatomists,* 237**,** 1627-1635.
Snider, P., Hinton, R. B., Moreno-Rodriguez, R. A., Wang, J., Rogers, R., Lindsley, A., Li, F., Ingram, D. A., Menick, D., Field, L., Firulli, A. B., Molkentin, J. D., Markwald, R. & Conway, S. J. 2008. Periostin is required for maturation and extracellular matrix stabilization of noncardiomyocyte lineages of the heart. *Circulation research,* 102**,** 752-760.
Song, L. & Crawford, G. E. 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor protocols,* 2010**,** pdb.prot5384-pdb.prot5384.
Soutourina, J. 2017. Transcription regulation by the Mediator complex. *Nature Reviews Molecular Cell Biology,* 19**,** 262.
Spitz, F. & Furlong, E. E. M. 2012. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics,* 13**,** 613.
Stainier, D. Y., Fouquet, B., Chen, J. N., Warren, K. S., Weinstein, B. M., Meiler, S. E., Mohideen, M. A., Neuhauss, S. C., Solnica-Krezel, L., Schier, A. F., Zwartkruis, F., Stemple, D. L., Malicki, J., Driever, W. & Fishman, M. C. 1996. Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Development,* 123**,** 285.
Stainier, D. Y., Lee, R. K. & Fishman, M. C. 1993. Cardiovascular development in the zebrafish. I. Myocardial fate map and heart tube formation. *Development,* 119**,** 31.
Stainier, D. Y. R. 2001. Zebrafish genetics and vertebrate heart formation. *Nat Rev Genet,* 2**,** 39-48.
Stainier, D. Y. R. & Fishman, M. C. 1992. Patterning the zebrafish heart tube: Acquisition of anteroposterior polarity. *Developmental Biology,* 153**,** 91-101.
Stankunas, K., Shang, C., Twu, K. Y., Kao, S.-C., Jenkins, N. A., Copeland, N. G., Sanyal, M., Selleri, L., Cleary, M. L. & Chang, C.-P. 2008. Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circulation research,* 103**,** 702-709.
Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J. L. 2015. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLOS ONE,* 10**,** e0124633.
Stennard, F. A., Costa, M. W., Elliott, D. A., Rankin, S., Haast, S. J. P., Lai, D., Mcdonald, L. P. A., Niederreither, K., Dolle, P., Bruneau, B. G., Zorn, A. M. & Harvey, R. P. 2003. Cardiac T-box factor Tbx20 directly interacts with Nkx2-5, GATA4, and GATA5 in regulation of gene expression in the developing heart. *Developmental Biology,* 262**,** 206-224.

Sun, X., Hoage, T., Bai, P., Ding, Y., Chen, Z., Zhang, R., Huang, W., Jahangir, A., Paw, B., Li, Y.-G. & Xu, X. 2009. Cardiac Hypertrophy Involves Both Myocyte Hypertrophy and Hyperplasia in Anemic Zebrafish. *PLoS ONE,* 4**,** e6596.

Sun, Y., Liang, X., Najafi, N., Cass, M., Lin, L., Cai, C.-L., Chen, J. & Evans, S. M. 2007. Islet 1 is expressed in distinct cardiovascular lineages, including pacemaker and coronary vascular cells. *Developmental biology,* 304**,** 286-296.

Taatjes, D. J., Marr, M. T. & Tjian, R. 2004. Regulatory diversity among metazoan co-activator complexes. *Nature Reviews Molecular Cell Biology,* 5**,** 403.

Tessadori, F., Van Weerd, J. H., Burkhard, S. B., Verkerk, A. O., De Pater, E., Boukens, B. J., Vink, A., Christoffels, V. M. & Bakkers, J. 2012. Identification and Functional Characterization of Cardiac Pacemaker Cells in Zebrafish. *PLOS ONE,* 7**,** e47644.

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

Trinh, L. A., Chong-Morrison, V., Gavriouchkina, D., Hochgreb-Hägele, T., Senanayake, U., Fraser, S. E. & Sauka-Spengler, T. 2017. Biotagging of Specific Cell Populations in Zebrafish Reveals Gene Regulatory Logic Encoded in the Nuclear Transcriptome. *Cell reports,* 19**,** 425-440.

Truett, G. E., Heeger, P., Mynatt, R. L., Truett, A. A., Walker, J. A. & Warman, M. L. 2000. Preparation of PCR-Quality Mouse Genomic DNA with Hot Sodium Hydroxide and Tris (HotSHOT). *BioTechniques,* 29**,** 52-54.

Tsompana, M. & Buck, M. J. 2014. Chromatin accessibility: a window into the genome. *Epigenetics & Chromatin,* 7**,** 33.

Tu, S. & Chi, N. C. 2012. Zebrafish models in cardiac development and congenital heart birth defects. *Differentiation; research in biological diversity,* 84**,** 4-16.

Uren, A. G., Mikkers, H., Kool, J., Van Der Weyden, L., Lund, A. H., Wilson, C. H., Rance, R., Jonkers, J., Van Lohuizen, M., Berns, A. & Adams, D. J. 2009. A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. *Nature protocols,* 4**,** 789-798.

Vaccari, T., Moroni, A., Rocchi, M., Gorza, L., Bianchi, M. E., Beltrame, M. & Difrancesco, D. 1999. The human gene coding for HCN2, a pacemaker channel of the heart1Sequence data reported in this paper have been deposited in GenBank database under accession no. AF065164.1. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* 1446**,** 419-425.

Valer Gotea, A. V., John M. Westlund, Marcelo A. Nobrega, Len A. Pennacchio and Ivan Ovcharenko 2010. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res.,* 20**,** 565–577.

Van Berlo, J. E. A. 2014. c-kit1 cells minimally contribute cardiomyocytes to the heart. . *Nature,* 509.

Van Berlo, J. H., Elrod, J. W., Van Den Hoogenhof, M. M., York, A. J., Aronow, B. J., Duncan, S. A. & Molkentin, J. D. 2010. The transcription factor GATA-6 regulates pathological cardiac hypertrophy. *Circ Res,* 107**,** 1032-40.

Van Den Boogaard, M., Wong, L. Y. E., Tessadori, F., Bakker, M. L., Dreizehnter, L. K., Wakker, V., Bezzina, C. R., T Hoen, P. a. C., Bakkers, J., Barnett, P. & Christoffels, V. M. 2012. Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *The Journal of clinical investigation,* 122**,** 2519-2530.

Varshney, G. K., Pei, W., Lafave, M. C., Idol, J., Xu, L., Gallardo, V., Carrington, B., Bishop, K., Jones, M., Li, M., 2015. High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res.,* 25**,** 1030–1042.

Vietri rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, Duncan t., Tanay, A. & Hadjur, S. 2015. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Reports,* 10**,** 1297-1309.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., Park, T. J., Deaville, R., Erichsen, J. T., Jasinska, A. J., Turner, J. M. A., Bertelsen, M. F., Murchison, E. P., Flicek, P. & Odom, D. T. 2015. Enhancer evolution across 20 mammalian species. *Cell,* 160**,** 554-566.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M. & Pennacchio, L. A. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature,* 457**,** 854.

Visel, A., Prabhakar, S., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E. M. & Pennacchio, L. A. 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nature genetics,* 40**,** 158-160.

Von Both, I., Silvestri, C., Erdemir, T., Lickert, H., Walls, J. R., Henkelman, R. M., Rossant, J., Harvey, R. P., Attisano, L. & Wrana, J. L. 2004. Foxh1 Is Essential for Development of the Anterior Heart Field. *Developmental Cell,* 7**,** 331-345.

Wang Xh, H. C., Wang Q, Li Rg, Xu Yj, Liu X 2013. A novel GATA5 loss-of-function mutation underlies lone atrial fibrillation. *Int J Mol Med.,* 31**,** 43–50.

Waxman, J. S., Keegan, B. R., Roberts, R. W., Poss, K. D. & Yelon, D. 2008. Hoxb5b Acts Downstream of Retinoic Acid Signaling in the Forelimb Field to Restrict Heart Field Potential in Zebrafish. *Developmental Cell,* 15**,** 923-934.

Weake, V. M. & Workman, J. L. 2010. Inducible gene expression: diverse regulatory mechanisms. *Nature Reviews Genetics,* 11**,** 426.

Wei, D., Bao, H., Liu, X.-Y., Zhou, N., Wang, Q., Li, R.-G., Xu, Y.-J. & Yang, Y.-Q. 2012. GATA5 loss-of-function mutations underlie tetralogy of fallot. *International journal of medical sciences,* 10**,** 34-42.

West, A. G., Gaszner, M. & Felsenfeld, G. 2002. Insulators: many functions, many mechanisms. *Genes & Development,* 16**,** 271-288.

Who World Health Statistics 2016. Monitoring Health for the Sustainable Development Goals.

Wills, A. A., Holdway, J. E., Major, R. J. & Poss, K. D. 2008. Regulated addition of new myocardial and epicardial cells fosters homeostatic cardiac growth and maintenance in adult zebrafish. *Development,* 135**,** 183-192.

Witzel, H. R., Cheedipudi, S., Gao, R., Stainier, D. Y. R. & Dobreva, G. D. 2017. Isl2b regulates anterior second heart field development in zebrafish. *Scientific Reports,* 7**,** 41043.

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., Mcewen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y. J. K., Cooke, J. E. & Elgar, G. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS biology,* 3**,** e7-e7.

Wu, C.-C., Kruse, F., Vasudevarao, M. D., Junker, J. P., Zebrowski, D. C., Fischer, K., Noël, E. S., Grün, D., Berezikov, E., Engel, F. B., Van Oudenaarden, A., Weidinger, G. & Bakkers, J. 2016. Spatially Resolved Genome-wide Transcriptional Profiling

Identifies BMP Signaling as Essential Regulator of Zebrafish Cardiomyocyte Regeneration. *Developmental Cell,* 36**,** 36-49.

Xie, X., Mikkelsen, T. S., Gnirke, A., Lindblad-Toh, K., Kellis, M. & Lander, E. S. 2007. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 7145-7150.

Yáñez-Cuna, J. O., Kvon, E. Z. & Stark, A. 2013. Deciphering the transcriptional cis-regulatory code. *Trends in Genetics,* 29**,** 11-22.

Yang, Y.-Q., Li, L., Wang, J., Zhang, X.-L., Li, R.-G., Xu, Y.-J., Tan, H.-W., Wang, X.-H., Jiang, J.-Q., Fang, W.-Y. & Liu, X. 2012. GATA6 loss-of-function mutation in atrial fibrillation. *European Journal of Medical Genetics,* 55**,** 520-526.

Yelon, D. 2001. Cardiac patterning and morphogenesis in zebrafish. *Developmental Dynamics,* 222**,** 552-563.

Yelon, D., Ticho, B., Halpern, M. E., Ruvinsky, I., Ho, R. K., Silver, L. M. & Stainier, D. Y. 2000. The bHLH transcription factor hand2 plays parallel roles in zebrafish heart and pectoral fin development. *Development,* 127**,** 2573.

Yuan, X., Song, M., Devine, P., Bruneau, B. G., Scott, I. C. & Wilson, M. D. 2018. Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. *Nature Communications,* 9**,** 4977.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B. D., Shen, Y., Pervouchine, D. D., Djebali, S., Thurman, R. E., Kaul, R., Rynes, E., Kirilusha, A., Marinov, G. K., Williams, B. A., Trout, D., Amrhein, H., Fisher-Aylor, K., Antoshechkin, I., Desalvo, G., See, L.-H., Fastuca, M., Drenkow, J., Zaleski, C., Dobin, A., Prieto, P., Lagarde, J., Bussotti, G., Tanzer, A., Denas, O., Li, K., Bender, M. A., Zhang, M., Byron, R., Groudine, M. T., Mccleary, D., Pham, L., Ye, Z., Kuan, S., Edsall, L., Wu, Y.-C., Rasmussen, M. D., Bansal, M. S., Kellis, M., Keller, C. A., Morrissey, C. S., Mishra, T., Jain, D., Dogan, N., Harris, R. S., Cayting, P., Kawli, T., Boyle, A. P., Euskirchen, G., Kundaje, A., Lin, S., Lin, Y., Jansen, C., Malladi, V. S., Cline, M. S., Erickson, D. T., Kirkup, V. M., Learned, K., Sloan, C. A., Rosenbloom, K. R., Lacerda De Sousa, B., Beal, K., Pignatelli, M., Flicek, P., Lian, J., Kahveci, T., Lee, D., James Kent, W., Ramalho Santos, M., Herrero, J., Notredame, C., Johnson, A., Vong, S., Lee, K., Bates, D., Neri, F., Diegel, M., Canfield, T., Sabo, P. J., Wilken, M. S., Reh, T. A., Giste, E., Shafer, A., Kutyavin, T., Haugen, E., Dunn, D., Reynolds, A. P., Neph, S., Humbert, R., Scott Hansen, R., De Bruijn, M., et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature,* 515**,** 355.

Yue, Y., Jiang, M., He, L., Zhang, Z., Zhang, Q., Gu, C., Liu, M., Li, N. & Zhao, Q. 2018. The transcription factor Foxc1a in zebrafish directly regulates expression of nkx2.5, encoding a transcriptional regulator of cardiac progenitor cells. *The Journal of biological chemistry,* 293**,** 638-650.

Zaret, K. 2005. Micrococcal Nuclease Analysis of Chromatin Structure. *Current Protocols in Molecular Biology,* 69**,** 21.1.1-21.1.17.

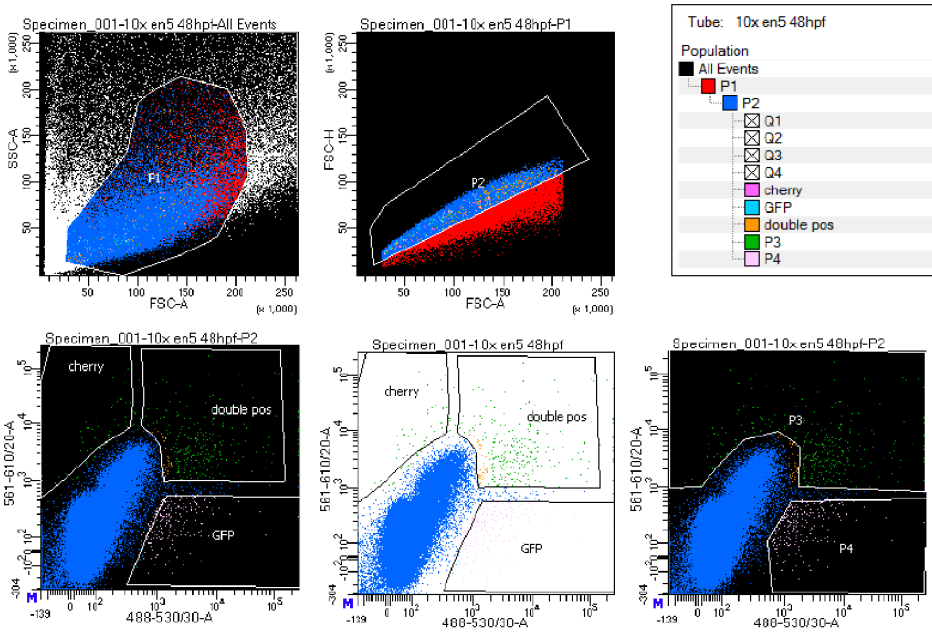
Zhang, Z. & Gerstein, M. 2003. Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements. *Journal of biology,* 2**,** 11-11.

Zhao, L., Borikova, A. L., Ben-Yair, R., Guner-Ataman, B., Macrae, C. A., Lee, R. T., Burns, C. G. & Burns, C. E. 2014. Notch signaling regulates cardiomyocyte proliferation during zebrafish heart regeneration. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 1403-1408.

Zhou, Q., Li, L., Zhao, B. & Guan, K.-L. 2015. The Hippo Pathway in Heart Development, Regeneration, and Diseases. *Circulation Research,* 116**,** 1431-1447.

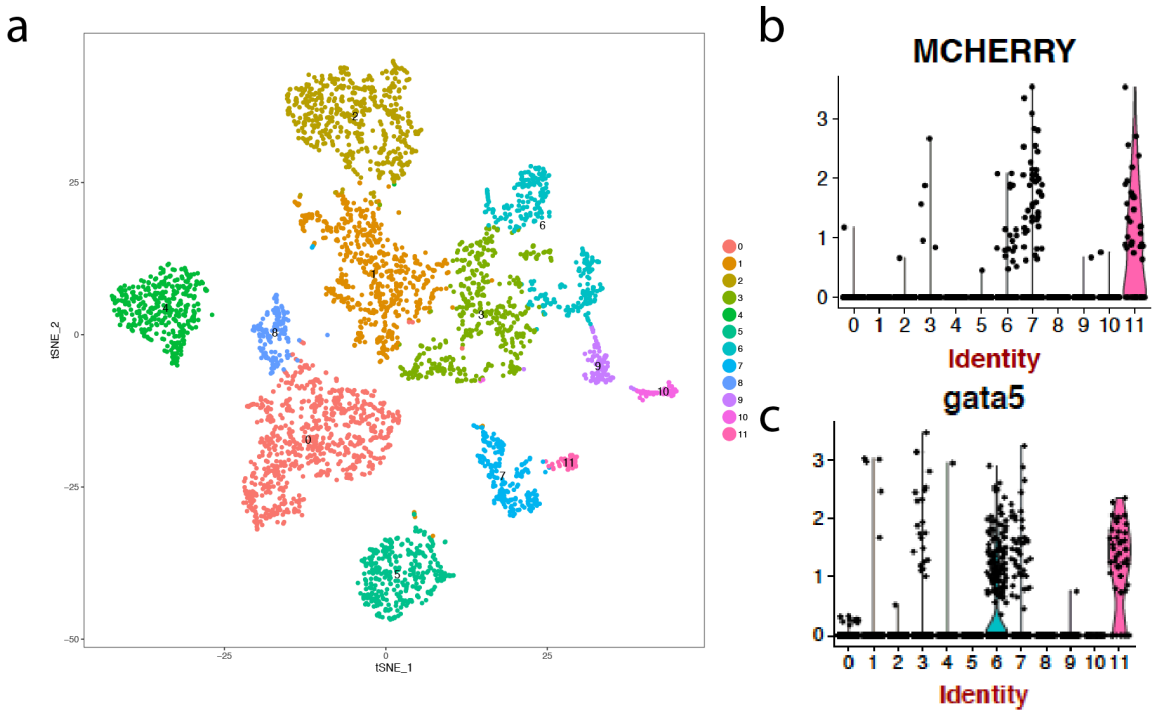
Zhou, Y., Cashman, T. J., Nevis, K. R., Obregon, P., Carney, S. A., Liu, Y., Gu, A., Mosimann, C., Sondalle, S., Peterson, R. E., Heideman, W., Burns, C. E. & Burns, C. G. 2011. Latent TGF-β binding protein 3 identifies a second heart field in zebrafish. *Nature,* 474**,** 645-648.

Supplementary Figures



Supplementary Figure 1

FACs gating strategy to isolate mCherry-positive cells from 48hpf Tg(Gata5-E5-mCherry; myl7-GFP) embryos for 10x Chromium single cell analysis.



Supplementary Figure 2

a) Two-dimensional t-SNE clustering of single cell libraries before filtering. Numbers indicate assignment to the 12 most significant cell clusters within the dataset. b) violin plot showing *mCherry* transcript levels in each cluster. c) violin plot showing *gata5* transcript levels in each cluster.