

# **Epicardial heterogeneity during zebrafish heart development**

Trinity Term 2017

Michael Weinberger

DPhil

BHF 4-Year Course in Cardiovascular Science

Supervisors: Paul Riley, Filipa Simoes, Roger Patient

Pembroke College

## Abstract

**Epicardial heterogeneity during zebrafish heart development, Michael Weinberger,  
Pembroke College, DPhil Cardiovascular Science, TT 2017**

The epicardium, a cell layer enveloping the heart muscle, drives embryonic heart development and heart repair in the adult zebrafish. Previous studies found the epicardium to consist of multiple cell populations with distinct phenotypes and functions. Here, I investigated epicardial heterogeneity in the developing zebrafish heart, focusing on the developmental gene program that is also reactivated during adult heart regeneration. Transcription factor 21 (*Tcf21*), T-box 18 (*Tbx18*) and Wilms' tumor suppressor 1b (*Wt1b*) are often used interchangeably to identify the zebrafish epicardium. Analyzing newly generated reporter lines and endogenous gene expression, I showed that the epicardial expression of *tcf21*, *tbx18* and *wt1b* during development is heterogeneous. I then collected epicardial cells from newly generated reporter lines at 5 days-post-fertilization and performed single-cell RNA sequencing. I identified three distinct epicardial subpopulations with specific gene expression profiles. The first subpopulation expressed *tcf21*, *tbx18* and *wt1b* and appeared to represent the main epicardial layer. The second subpopulation expressed *tbx18*, but not *tcf21* or *wt1b*. Instead, it expressed smooth muscle markers and seemed restricted to the bulbus arteriosus. The third epicardial subpopulation only expressed *tcf21* and resided within the epicardial layer. I compared the single-cell subpopulations with transcriptomic bulk data and visualized the expression of marker genes to investigate their spatial distribution. Using ATAC sequencing, I additionally identified open regulatory regions located in proximity to subpopulation-specific marker genes and showed subpopulation-specific activity *in vivo*. My results detail distinct cell

populations in the developing zebrafish epicardium, likely to fulfil distinct and specific cellular functions. Future experiments will involve targeting signature genes enriched within each epicardial subpopulation, such as those encoding Adrenomedullin a (first subpopulation), Alpha Smooth Muscle Actin (second subpopulation) and Claudin 11a (third subpopulation), employing cell type-specific genome editing to test whether and how the identified heterogeneity underlies distinct epicardial cell fates and functions. Taken together, my work adds significantly to the understanding of the cellular and molecular basis of epicardial development and can offer novel insights in the context of heart regeneration.

## Acknowledgements

Above all, my thanks go to Filipa Simoes, Paul Riley and Roger Patient for their continued supervision, motivation and advice. I owe a lot to their incredible job at sharing knowledge and experience whenever I needed it, from handling everyday issues in the lab to shaping the project as a whole. Importantly, Filipa, Paul and Roger also gave me the freedom and independence to approach experiments in my own way and pace, trusting me to come ask for help should I need it. This is something I have always appreciated very much. Last, but not least, I would like to thank the three for improving this written work a good deal by proofreading and providing insightful comments.

I could not have accomplished this project without Tatjana Sauka-Spengler's inspiring and intense support regarding all things HCR and sequencing. She has always been happy to share her immense knowledge and has continuously encouraged me to go ahead and try new things, for which I am very grateful. Furthermore, working in her lab and with her group has introduced me to a great working environment with a sense for fun and community, as well as a shared drive for cutting-edge research.

I would further like to acknowledge the staff at the various facilities I used for their invaluable work, including the BMS and BSB aquatics staff, Christoffer Lagerholm at the WIMM imaging facility, Kevin Clark at the WIMM FACS facility, Neil Ashley at the WIMM single cell facility and the computational biology research group at the WIMM for providing me with much needed server capacity. Also thanks to Professor Paul Martin, who kindly provided his custom-made Lcp1 antibody.

I am very grateful for all the help and advice I got from various people in Roger's and Paul's groups, in particular for Jana sharing her vast expertise on zebrafish work,

immunofluorescence and microscopy, for Caroline teamworking so kindly and capably on cloning enhancer constructs and screening fish, for Mukesh discussing both biological and technical questions and for Aldo pointing out publications that I would surely have missed otherwise. Not to forget Florian and Tomek, who have shared the highs and lows of being a doctoral student in Oxford.

Thank you also to Tatjana's bunch for both lab advice and lunchtime conversation. Especially to Ruth for teaching me what there is to know about the Sauka-Spengler lab, sequencing, etc, to Daria for the crash course in data analysis, to Upeka and Vanessa for their cloning wisdom and to Georgia, who managed to make ATAC sequencing fun.

Phil has done an excellent job at managing finances and my various orders, and of course at keeping a keen eye on health and safety in the lab. Dee has done an equally excellent job at managing things down the hill at DPAG. Katie and Linda have always been a great help at organizing meetings and generally keeping things together.

I would like to say special thanks to my family and friends for their support. Many thanks to my parents, who have truly done their best to support this endeavor, not least by providing me with a safe haven in which to write this thesis. Thanks also to my brother, to Daniel and Dennis for their Skype mediated support and thanks to Christiane, Naoko and Sai for taking my mind off things here in Oxford.

Regarding funding, I owe it all to the British Heart Foundation and their generous 4-year studentship, as well as to Roger, who provided for animal costs.

## Table of contents

<b>Table of figures .....</b>	<b>12</b>
<b>Table of tables .....</b>	<b>15</b>
<b>List of abbreviations.....</b>	<b>16</b>
<b>Chapter 1 - Introduction.....</b>	<b>17</b>
1.1 - The zebrafish as a model organism to study heart development and regeneration . .....	17
1.2 - Heart development in the zebrafish .....	19
1.3 - The epicardium during heart development.....	24
1.4 - The epicardium during heart regeneration .....	28
1.5 - Heterogeneity in the epicardium.....	31
1.6 - Revealing cellular heterogeneity through single cell RNA sequencing .....	32
1.7 - Aims.....	35
1.8 - Summary of findings .....	36
<b>Chapter 2 - Material and methods.....</b>	<b>39</b>
2.1 - Zebrafish lines .....	39
2.2 - Immunocytochemistry .....	41
2.3 - Hybridization chain reaction (HCR) .....	42
2.4 - Microscopy and digital image analysis.....	43

2.5 - Larval heart injury .....	44
2.6 - Lightsheet imaging .....	44
2.7 - Purification of embryonic hearts and dissociation for FACS.....	45
2.8 - cDNA synthesis from single cell samples .....	46
2.9 - cDNA synthesis from bulk samples .....	48
2.10 - Illumina library preparation from cDNA .....	49
2.11 - Single cell RNA sequencing data analysis .....	51
2.12 - Bulk RNA sequencing data analysis.....	54
2.13 - Assay for Transposase Accessible Chromatin (ATAC) sequencing.....	54
2.14 - ATAC sequencing data analysis and enhancer studies .....	56
<b>Chapter 3 - Microscopic analysis of epicardial heterogeneity in the zebrafish embryo .</b>	
<b>.....</b>	<b>59</b>
3.1 - Introduction .....	59
3.2 - Early morphogenesis of the <i>tbx18</i> positive embryonic zebrafish epicardium matches previous findings obtained studying <i>wt1a</i> .....	60
3.3 - The analysis of <i>Tg(tcf21:dsRed2; wt1b:GFP)</i> and <i>Tg(tbx18:dsRed2; wt1b:GFP)</i> reveals heterogeneity in the developing zebrafish epicardium .....	63
3.4 - Novel transgenic reporter lines enable simultaneous visualization of the expression patterns of <i>tcf21</i> , <i>tbx18</i> and <i>wt1b</i> .....	67
3.5 - A novel <i>wt1b:H2B-Dendra2</i> single positive epicardial cell population envelops the bulbus arteriosus.....	71
3.6 - The response of the developing zebrafish epicardium to cardiac injury is not equivalent to epicardial reactivation in the adult.....	74

3.7 - Imaging the developing zebrafish epicardium in the living embryo through lightsheet microscopy and computational image alignment .....	78
3.8 - Summary .....	82
<b>Chapter 4 - Single cell transcriptomics identify three distinct cell populations in the embryonic zebrafish epicardium that are likely to have specific functions.....</b>	<b>84</b>
4.1 - Introduction .....	84
4.2 - Quality metrics of the single cell data set.....	85
4.3 - Distinct epicardial clustering within the single cell data set.....	89
4.4 - Epicardial cell cluster 1 is enriched in transcripts encoding the signaling peptide Adma, as well as in transcripts encoding the adhesion molecules Podxl and Jam2b and the epicardial marker Aldh1a2.....	92
4.5 - Epicardial cell cluster 2 is enriched in transcripts encoding the smooth muscle cell markers Mylka, Acta2 and Myh11a as well as in transcripts encoding the extracellular matrix factors Lox and Elnb.....	96
4.6 - Epicardial cell cluster 3 is enriched in transcripts encoding the tight junction marker Cldn11a as well as in transcripts encoding the signaling molecule Cxcl12a .....	98
4.7 - The expression of CD248a distinguishes epicardial cell clusters 2 and 3 from epicardial cell cluster 1.....	101
4.8 - Marker gene expression in the non-epicardial cell clusters .....	102
4.8.1 - Cluster 3 - myelinating cells, neural progenitors and orphan <i>Tg(tcf21:H2B-Dendra2)</i> derived cells .....	102
4.8.2 - Cluster 4 - neural cells.....	103
4.8.3 - Cluster 5 - mesenchymal cells.....	104

4.8.4 - Clusters 6 and 7 - haematopoietic cells .....	105
4.8.5 - Cluster 8 - cardiomyocytes.....	107
4.9 - Summary of cell clusters in the single cell transcriptomic data set.....	107
4.10 - t-SNE clustering confirms the presence of distinct epicardial cell populations in the single cell data set.....	108
4.11 - The presence of <i>tcf21</i> , <i>tbx18</i> and <i>wt1b</i> in the epicardial cell clusters is heterogeneous .....	112
4.12 - Presence of <i>adma</i> in epicardial cell cluster 1 might drive endothelial differentiation of mesenchymal cells in the developing zebrafish heart .....	120
4.13 - Cells in epicardial cell cluster 2 transcriptionally resemble smooth muscle cells and might be regulated cell-autonomously by Wnt11 Related.....	124
4.14 - Presence of <i>cxc12a</i> in epicardial cell cluster 3 might guide myeloid haematopoietic cells into the developing zebrafish heart .....	130
4.15 - Summary .....	135

**Chapter 5 - Validation of the cell populations identified via single cell transcriptomics .....**

<b>.....</b>	<b>138</b>
5.1 - Introduction .....	138
5.2 - Transcriptomic analysis of epicardial bulk samples derived from the developing zebrafish heart .....	139
5.3 - Epicardial bulk transcriptomic data support the single cell RNA sequencing results .....	146
5.4 - Epicardial cell cluster 1 represents most of the epicardial cell layer that envelops the developing zebrafish heart .....	152

5.5 - Epicardial cell cluster 2 resides within the bulbus arteriosus of the developing zebrafish heart .....	154
5.6 - Epicardial cell cluster 3 is located within the epicardial cell layer that envelops the developing zebrafish heart.....	157
5.7 - The developing zebrafish heart is in close proximity to tuba1c positive neurons .....	159
5.8 - The mesenchymal cell marker Cyp26b1 labels a large portion of the developing zebrafish myocardium.....	161
5.9 - Myeloid haematopoietic cells are in direct contact with the developing zebrafish epicardium .....	163
5.10 - Putative enhancers in genomic proximity to markers of epicardial cell cluster 1 and 2 show epicardial subpopulation-specific in vivo activity.....	165
5.10.1 - A putative enhancer element located upstream of <i>podxl</i> is active in the developing zebrafish epicardial cell layer .....	166
5.10.2 - A putative enhancer element located in intron 3 of <i>lox</i> is active in the bulbus arteriosus of the developing zebrafish heart.....	170
5.10.3 - A putative enhancer element located upstream of <i>elnb</i> is active in the bulbus arteriosus of the developing zebrafish heart.....	175
5.11 - Summary .....	178
<b>Chapter 6 - General discussion, final conclusions and future work .....</b>	<b>180</b>
6.1 - Investigating the developing zebrafish epicardium using novel <i>pcf21</i> , <i>tbx18</i> and <i>wt1b</i> reporter lines .....	180
6.2 - Epicardial subpopulations in the developing zebrafish heart .....	185

6.2.1 - <i>wt1b</i> :H2B-Dendra2 single positive cells cover the bulbus arteriosus .....	185
6.2.2 - <i>Adma</i> , <i>Podxl</i> , <i>Tcf21</i> , <i>Tbx18</i> , <i>Wt1b</i> expressing cells form the epicardial sheet and might regulate mesenchymal cells in the developing zebrafish heart .....	186
6.2.3 - <i>Lox</i> , <i>Elnb</i> , <i>Mylka</i> , <i>Acta2</i> , <i>Tbx18</i> expressing smooth muscle cells in the bulbus arteriosus are potentially regulated by <i>Wnt11r</i> signaling .....	189
6.2.4 - <i>Cldn11a</i> , <i>Cxcl12a</i> , <i>Nr1h5</i> , <i>Tcf21</i> expressing cells in the epicardial sheet might attract myeloid haematopoietic cells into the developing zebrafish heart.....	193
6.3 - Non-epicardial cell populations in the developing zebrafish heart .....	197
6.3.1 - Mesenchymal cells .....	197
6.3.2 - Neural cells.....	201
6.4 - Future studies .....	201
6.5 - Final remarks.....	204
<b>References.....</b>	<b>205</b>
<b>Appendix A - HCR probe sequences .....</b>	<b>238</b>
<b>Appendix B - Illumina Nextera indexing primer sequences.....</b>	<b>240</b>
<b>Appendix C - R scripts.....</b>	<b>242</b>

## Table of figures

Figure 1-1 - Stages of cardiac development in the zebrafish embryo .....	23
Figure 3-1 - Epicardial development in the zebrafish embryo between 2dpf and 14dpf... .....	62
Figure 3-2 - The expression of <i>tcf21</i> , <i>tbx18</i> and <i>wt1b</i> in the developing zebrafish epicardium is heterogeneous.....	65
Figure 3-3 - Simultaneous expression analysis of <i>tcf21</i> , <i>tbx18</i> and <i>wt1b</i> details epicardial heterogeneity and reveals a novel <i>wt1b</i> :H2B-Dendra2 single positive epicardial cell population.....	69
Figure 3-4 - The number of <i>wt1b</i> :GFP labelled epicardial cells does not increase in response to larval heart injury.....	77
Figure 3-5 - Live imaging of the developing zebrafish epicardium using lightsheet microscopy and a custom-made alignment algorithm.....	81
Figure 4-1 - Single cell RNA sequencing to study cell populations within the developing zebrafish epicardium.....	88
Figure 4-2 - Novel marker genes identify distinct epicardial cell populations within the single cell data set.....	91
Figure 4-3 - The single cell data set contains three distinct epicardial cell clusters.....	111
Figure 4-4 - Cells within the epicardial cell clusters express epicardial gene signatures ... .....	118
Figure 4-5 - Cells within epicardial cell cluster 1 might regulate mesenchymal cells expressing <i>calcr1a</i> , <i>ramp2</i> and endothelial markers.....	122

Figure 4-6 - Wnt11r might regulate cells in epicardial cell cluster 2 and cells within the mesenchymal cell cluster. .... 128

Figure 4-7 - Cells within epicardial cell cluster 3 might guide myeloid haematopoietic cells via Cxcl12a and Ccl25b mediated chemokine signaling..... 133

Figure 5-1 - Epicardial bulk RNA sequencing to confirm single cell transcriptomics in the embryonic zebrafish heart. .... 144

Figure 5-2 - Epicardial bulk transcriptomics support the validity of the epicardial subpopulations identified through single cell transcriptomics. .... 150

Figure 5-3 - Transcripts encoding the epicardial cell cluster 1 marker *Adma* label the epicardial cell layer..... 153

Figure 5-4 - The epicardial cluster 2 markers *elnb* and *Mylka* label cells within the bulbus arteriosus..... 155

Figure 5-5 - Transcripts encoding the epicardial cluster 3 marker *Cldn11a* are present in the epicardial cell layer. .... 158

Figure 5-6 - The embryonic zebrafish heart is located adjacent to *Tuba1c* positive neuronal structures..... 160

Figure 5-7 - The embryonic zebrafish heart contains *myl7*:GFP positive myocardial cells that co-express *Cyp26b1*..... 162

Figure 5-8 - Myeloid haematopoietic cells are in direct contact with the developing zebrafish epicardium..... 164

Figure 5-9 - An enhancer element located upstream of *podxl* is active in the developing zebrafish epicardium..... 169

Figure 5-10 - An enhancer element located in intron 3 of *lox* is active in the bulbus arteriosus of the developing zebrafish heart..... 173

Figure 5-11 - An enhancer element located upstream of *elnb* is active in the bulbus arteriosus of the developing zebrafish heart..... 177

Figure 6-1 - Subpopulations in the developing zebrafish heart, as identified in this study, and their possible regulation. .... 195

## Table of tables

Table 2-1 - Primers used for BAC entry cassette amplification. .... 40

Table 4-1 - Composition of the single cell data set before and after quality control (QC).

..... 89

## List of abbreviations

A - atrium

ATAC - assay for transposase accessible chromatin

BA - bulbus arteriosus

BAC - bacterial artificial chromosome

CM - cardiomyocytes

DMSO - dimethyl sulfoxide

dpf - days post fertilization

ECM - extracellular matrix

EMT - epithelial-to-mesenchymal transition

EPDC - epicardium derived cell

Epi - epicardial cell cluster

FACS - fluorescence assisted cell sorting

FPKM - fragments per kilobase of transcript per million transcripts

GFP - green fluorescent protein

H2B - histone 2b

HC - haematopoietic cells

HCR - hybridization chain reaction

hpf - hours post fertilization

hpi - hours post injury

L - liver

MC - mesenchymal cells

myr - myristoylation

NC - neural cells

PBS - phosphate buffered saline

PEO - proepicardial organ

PFA - paraformaldehyde

SSC - sodium chloride sodium citrate

Tg - transgenic

t-SNE - t-distributed stochastic neighbor embedding

V - ventricle

## **Chapter 1**

### **Introduction**

#### **1.1 The zebrafish as a model organism to study heart development and regeneration**

In recent years, the zebrafish has become increasingly popular as a vertebrate model organism that allows the study of both developmental and regenerative processes in various organs (Bakkers, 2011; Kroeger and Wingert, 2014; Kroehne et al., 2011). The zebrafish offers unique technical advantages, as it can be housed in large numbers, can be used in large scale forward genetic screens and is very amenable towards genetic manipulation (Lawson and Wolfe, 2011). Additionally, zebrafish embryos develop externally, are laid in large quantities and are completely transparent during the first days of development (Kimmel et al., 1995). In consequence, a large variety of transgenic zebrafish reporter lines has been generated that enable the analysis of tissues, cells and molecules in the living animal, based on the expression of fluorescent proteins such as GFP (Burns et al., 2005; Jin, 2005; Kikuchi et al., 2011a; Perner et al., 2007; Traver et al., 2003). Zebrafish heart development has been studied intensively using reporter lines like the ones listed above, as the heart develops early during embryogenesis and starts beating by 24 hours post fertilization (hpf), yet embryos with impaired cardiac function survive and thus can be studied until later stages of development due to oxygen delivery via diffusion (Bakkers, 2011; Mably et al., 2003; Stainier, 2001). Another reason the zebrafish heart has received scientific attention is that it is able to regenerate following the resection of the ventricular apex (Poss, 2002; Raya et al., 2003). More recently,

studies have demonstrated that the zebrafish heart is also able to regenerate following genetic ablation of cardiomyocytes (Wang et al., 2011) and following injury with a cryo-probe (Chablais et al., 2011; Gonzalez-Rosa et al., 2011; Schnabel et al., 2011). This regenerative capacity persists throughout life (Itou et al., 2012a).

Myocardial infarction is a major source of morbidity and mortality in humans (Ziaeiian and Fonarow, 2016). Cryoinjury in the zebrafish is the injury model that most closely resembles a myocardial infarction in humans as both elicit extensive cardiac scarring. Scarring however is permanent in humans (Sutton and Sharpe, 2000), while in the zebrafish the scarred tissue is replaced with functional muscle tissue (Gonzalez-Rosa et al., 2011). There is capacity for heart regeneration in mammals, but it is lost shortly after birth (Porrello et al., 2011). Furthermore, the postnatal mammalian heart contains progenitor cells that have the potential to give rise to cardiomyocytes *in vitro* (Laugwitz et al., 2005). Studying zebrafish heart regeneration might reveal ways to stimulate the regenerative potential of mammalian cardiac progenitor cells *in vivo*, which might enable adult mammalian heart regeneration.

The cellular and molecular processes driving heart regeneration are thought to recapitulate those regulating heart development (Xin et al., 2013). Investigating embryonic heart development can thus serve as a proxy for the study of adult heart regeneration. Therefore, much effort is being put into elucidating both the mechanisms that drive embryonic heart development in the zebrafish and those that regulate adult heart regeneration.

## 1.2 Heart development in the zebrafish

The zebrafish heart is two-chambered and consists of a single atrium and ventricle, as well as an inflow tract, the sinus venosus, and a prominent outflow tract, the bulbus arteriosus (BA) (Hu et al., 2000) (figure 3-1a). The BA is an evolutionary novelty of the teleost lineage and is composed of smooth muscle, while in other vertebrate species the outflow tract is composed of cardiomyocytes (Moriyama et al., 2016). The BA functions as an elastic reservoir to absorb the rapid rise of pressure during ventricular contraction (Hu et al., 2001). Smooth muscle formation in the zebrafish BA is dependent on the zebrafish specific extracellular matrix (ECM) component Elastin b (Elnb), which exclusively labels smooth muscle cells in the BA (Moriyama et al., 2016), although these results require further validation. Transcriptomic profiles of the adult zebrafish ventricle, atrium and BA have been generated recently (Singh et al., 2016), shedding some light on the transcriptional programs that drive chamber identity.

Although the morphology of the zebrafish heart differs from that of the four-chambered mammalian heart, the molecular mechanisms governing heart development appear largely conserved across vertebrate evolution (Olson, 2006). In the early zebrafish embryo, cardiac progenitors can be identified from the early blastula stage (Stainier et al., 1993) (figure 1-1a). During early and mid somite stages, cardiac progenitor cells are located in the anterior lateral plate mesoderm adjacent to myeloid haematopoietic and endothelial cell progenitors, in a region known as the first heart field (Peterkin et al., 2009; Schoenebeck et al., 2007) (figure 1-1b-d). Several factors regulate cardiac progenitor specification and differentiation, among them retinoic acid, which restricts the ability of cells to adopt a cardiogenic fate (Keegan, 2005) as well as Fibroblast Growth

Factors (Fgfs) and Heart And Neural Crest Derivatives Expressed 2 (Hand2), which promote the emergence of cardiac progenitors (Schoenebeck et al., 2007; Simoes et al., 2011). The GATA motif binding transcription factors Gata4, Gata5 and Gata6 also play important roles during early heart development, both in the zebrafish and in other vertebrates (Stefanovic and Christoffels, 2015). Gata4 was shown to be redundant during early zebrafish heart development (Peterkin et al., 2007). In contrast, Gata5 is required during cardiac differentiation and morphogenesis (Reiter et al., 1999) and its expression is induced by Bone Morphogenetic Protein (Bmp) signaling (Reiter et al., 2001). Gata6 is important for early cardiogenesis as well, regulating cardiomyocyte precursor maturation (Peterkin et al., 2003).

Cells derived from the first heart field form the linear heart tube (figure 1-1e), which subsequently loops to structurally mature into atrium and ventricle (figure 1-1f). In addition, cells from the anterior or second heart field make important contributions to the heart (Zhou et al., 2011b). In mammals, the second heart field gives rise to the right ventricle and large parts of the atria and the outflow tract (Buckingham et al., 2005) in a process that is regulated by retinoic acid (Li et al., 2010). Despite the absence of a right ventricle, the second heart field is present in the zebrafish and second heart field precursors are specified in the anterior lateral plate mesoderm (Guner-Ataman et al., 2013; Hami et al., 2011), expressing Gata4 and the homeobox transcription factor Nkx2.5 (Guner-Ataman et al., 2013). At later stages, Latent-TGF $\beta$  Binding Protein 3 (Ltbp3) positive cells contribute to the arterial pole of the ventricle and the BA, giving rise to myocardial, endothelial and smooth muscle cells (de Pater et al., 2009; Zhou et al., 2011b).

Another contribution to the developing zebrafish heart comes from cardiac neural crest cells that invade the heart in multiple waves to give rise to both cardiomyocytes and BA tissue (Cavanaugh et al., 2015; Li et al., 2003). These neural crest cells were additionally found to be necessary for the recruitment of cells from the second heart field into the developing zebrafish heart (Cavanaugh et al., 2015).

By 24hpf, the linear heart tube is contracting and initially functions as a suction pump (Forouhar, 2006). Following looping, cardiac valves begin to form to prevent retrograde blood flow (Pestel et al., 2016) (figure 1-1g). In the zebrafish, formation of the atrio-ventricular valves has been studied best. This process initiates with morphological changes in endocardial cells at the atrio-ventricular boundary (Beis, 2005). Valve formation in the developing mammalian heart involves the formation of mesenchymal endocardial cushions, but this appears not to be the case in the embryonic zebrafish. Instead, valve formation in the zebrafish heart was found to be directly mediated by endocardial invagination to form immature valve leaflets that consist of an atrial and a ventricular endocardial layer and subsequently differentiate into mature valves (Pestel et al., 2016; Scherz et al., 2008). Transcription factors regulating this process include Cysteine Rich Protein 2 (Crip2) (Kim et al., 2014) and, in the mouse, Nuclear Factor of Activated T-Cells 1 (Nfatc1), SRY-box 9 (Sox9) and T-box 2 (Tbx2) (Akiyama et al., 2004; Lange and Yutzey, 2006; Singh et al., 2012).

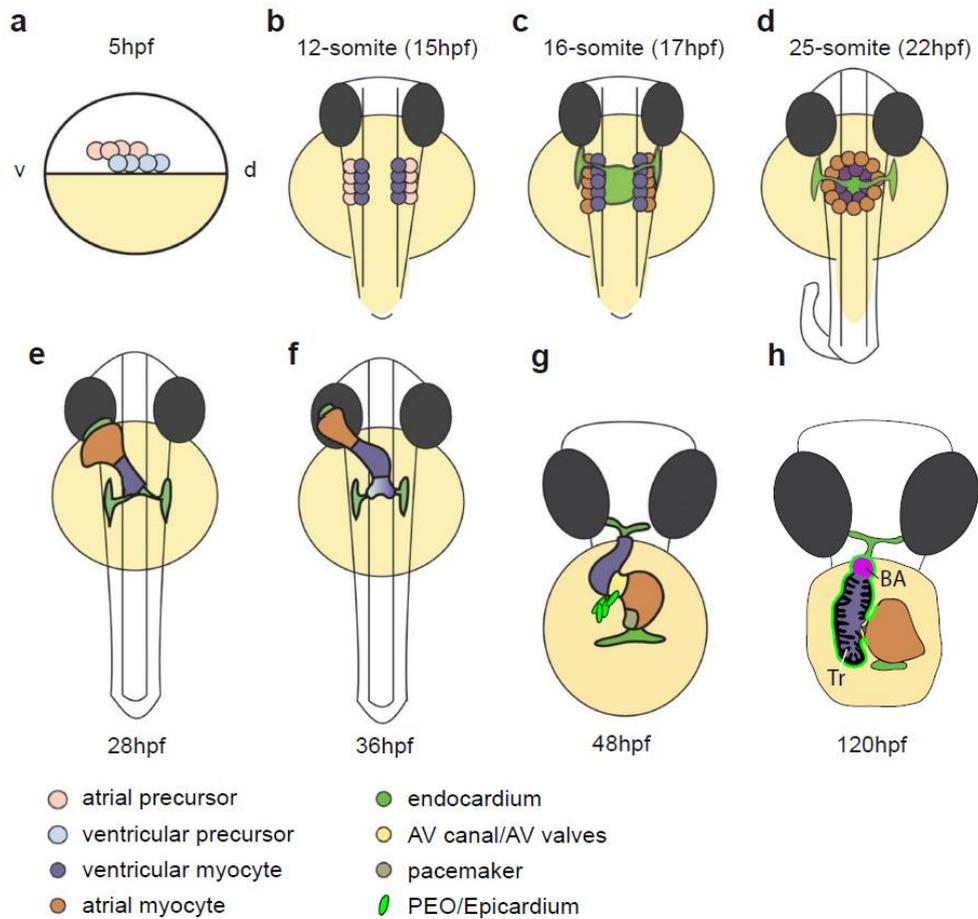
Another process that takes place following looping of the developing zebrafish heart is trabeculation. Trabeculae are myocardial protrusions extending into the ventricular lumen that increase the ventricle's contractile strength and modulate the haemodynamic forces within the ventricle (Sedmera et al., 2000) (figure 1-1h). Trabeculation begins with the delamination of ventricular cardiomyocytes (Staudt et al., 2014). These protrude and

connect to form a radial pattern of ridges, followed by the detachment of cardiomyocyte bundles that then connect to form a trabecular network (Peshkovsky et al., 2011). This process is dependent on blood flow and Neuregulin signaling (Liu et al., 2010; Peshkovsky et al., 2011).

Myocardial differentiation in the zebrafish continues into the juvenile stage, when a subset of cardiomyocytes induces the expression of Gata4 and expands on the surface of the ventricular myocardium to form a cortical muscle layer (Gupta et al., 2013).

The coronary vasculature in the zebrafish forms much later during development than in mammals (Harrison et al., 2015; Tomanek, 2005). First angiogenic sprouts emerge from the endocardium at the atrio-ventricular boundary at 37 days post fertilization (dpf) and vascularize the ventricle in response to myocardial chemokine signaling (Harrison et al., 2015).

The formation of the cardiac conduction system is another important aspect of heart development. The cardiac conduction system coordinates the contraction of the heart chambers and consists of specialized cardiomyocytes (Gourdie et al., 1995). The endocardium regulates cardiac conduction system development, for example through Neuregulin signaling (Milan et al., 2006). The first conduction signals travel the linear heart tube from sinus venosus to BA. Subsequently, the conduction system differentiates to include an atrio-ventricular conduction delay and a fast conduction network within the ventricle (Chi et al., 2008; Jensen et al., 2012). There is very little knowledge about when and how the developing zebrafish heart is innervated by extra-cardiac neurons belonging to the autonomous nervous system. However, in zebrafish embryos, between 5dpf and 7dpf, the heartbeat responds to autonomic stimuli and cardiac innervation is hence likely to be present at this stage (Mann et al., 2010).



**Figure 1-1: Stages of cardiac development in the zebrafish embryo.** (a) At 5 hours post fertilization (hpf), the blastula (white) covers approximately 50% of the large yolk cell (yellow) and cardiac progenitor cells are located bilaterally in the lateral marginal zone. Atrial progenitor cells (pink) are located more ventrally than the ventricle progenitor cells (light blue). (b) In the anterior lateral plate mesoderm, cardiogenic differentiation is initiated by the expression of cardiac myosins in the future ventricle myocardial cells (purple) at the 12-somite stage. (c) During mid- and late-somite stages, the myocardial tissue expands by the cardiogenic differentiation of future atrial myocytes (orange; venous differentiation). Whilst the endocardial cells (light green) have already migrated towards the mid-line, myocardial cells follow this behavior slightly later. (d) At the mid-line, the bilateral heart fields fuse to form a cardiac disc structure with the endocardial

cells within the hole at the center, ventricular myocytes at the circumference and atrial myocytes at the periphery of the disc. Cardiac morphogenesis transforms the cardiac disc into a cardiac tube, with the endocardium forming the inner lining. **(e)** At 28hpf, the linear heart tube has formed, with the venous pole located at the anterior left and the arterial pole fixed at the mid-line. **(f)** Cardiogenic differentiation continues at the arterial pole as new cardiomyocytes are added to this region (purple gradient). At 36hpf, cardiac looping has started, with a displacement of the ventricle towards the mid-line. **(g)** The heart tube continuous to loop until 48hpf. Ellipsoid pro-epicardial cells (green) are located near the atrio-ventricular boundary (yellow), from where they translocate to envelop the myocardium with an epicardial layer. The pacemaker is present in the inner curvature of the atrium near the venous pole (dark green). **(h)** At 120hpf, myocardial trabeculae have formed and expand into the lumen of the ventricle. The bulbus arteriosus (BA) (magenta) is present at the arterial pole and the epicardium (green) completely covers the ventricle. At the atrio-ventricular boundary, valves have formed (yellow). PEO = proepicardial organ, AV = atrio-ventricular, Tr = trabeculae. Figure adapted from (Bakkers, 2011).

### **1.3 The epicardium during heart development**

The epicardium is a mesothelial outer cell layer enveloping the heart muscle and an important signaling hub during heart development, also contributing to other cardiac cell lineages (Acharya et al., 2012; Greulich et al., 2012; Masters and Riley, 2014). The presence of the epicardium and its origin in the proepicardial organ (PEO) is conserved across vertebrate species (Hirakow, 1992; Komiyama et al., 1987; Serluca, 2008). At

around 55hpf, the zebrafish PEO emerges as cell clusters located close to the sinus venosus and the atrio-ventricular boundary (Peralta et al., 2013). Between 55hpf and 72hpf, proepicardial cells in the zebrafish translocate to the ventricle via pericardial fluid movements, where they spread and form a continuous cell layer (Peralta et al., 2013). The epicardial cell layer initially only envelops the ventricle, starting to colonize the atrium at 5dpf (Plavicki et al., 2014).

Proepicardial cells are thought to arise from Nkx2.5 and Islet 1 expressing cardiac progenitors in the anterior lateral plate mesoderm (Zhou et al., 2008a). In the chick embryo, fate mapping showed that proepicardial progenitors are located adjacent to the posterior end of the cardiogenic domain (Bressan et al., 2013). Factors that specify proepicardial cells in the zebrafish are Hand2 and Tbx5 (Liu and Stainier, 2010). It has been proposed that the PEO evolved from a pronephric external glomerulus in early vertebrates (Pombal et al., 2008). This is supported by the epicardial expression of genes that are also expressed in the kidney, such as the genes encoding the basic helix-loop-helix transcription factor Tcf21 (also known as Pod1, Epicardin or Capsulin), the zinc finger transcription factor Wilms' Tumor Suppressor 1 (Wt1, zebrafish orthologues Wt1a and Wt1b), and the T-box transcription factor Tbx18 (Airik et al., 2006; Buckler et al., 1991; Hidai et al., 1998; Kraus et al., 2001; Moore et al., 1999; Quaggin et al., 1998; Robb et al., 1998).

In frog embryos lacking Tcf21, proepicardial cells proliferate excessively and fail to form a coherent cell layer (Tandon et al., 2013). Here, Tcf21 was further found to associate with co-factors that mediate transcriptional repression and to downregulate the expression of several proepicardial factors, such as Tbx18. Tcf21 thus promotes the maturation and

epithelization of proepicardial cells. Wt1 might have a similar role, since it promotes cell adhesion through induction of Integrin  $\alpha 4$  (*Itga4*) (Kirschner et al., 2006).

After establishing a coherent cell layer, a subset of epicardial cells undergoes epithelial-to-mesenchymal transition (EMT) and migrates into the underlying myocardium. Subsequently, these epicardium derived cells (EPDCs) transdifferentiate into a variety of cardiac cell types, such as smooth muscle cells and fibroblasts (Kikuchi et al., 2011a). In the mouse, Tcf21 is crucial for this process, both inhibiting the transdifferentiation of EPDCs into smooth muscle (Braitsch et al., 2012) and driving the formation of cardiac fibroblasts (Acharya et al., 2012). Wt1 is another driver of epicardial EMT, both *in vitro* and *in vivo* (von Gise et al., 2011; Takeichi et al., 2013). Retinoic acid, produced by the enzyme Retinaldehyde Dehydrogenase (*Raldh2*, *Aldh1a2* in zebrafish), is also essential for epicardial EMT and is regulated by Wt1 (Guadix et al., 2011). Retinoic acid, in turn, regulates the expression of Tcf21 and Wt1 (Braitsch et al., 2012). Tbx18 regulates epicardial EMT *in vitro* by inducing Snail Family Transcriptional Repressor 2 (*Snai2*) (Takeichi et al., 2013), but was found to be redundant *in vivo* (Greulich et al., 2012). However, another study showed Tbx18 deficient mouse hearts to have a defective coronary vasculature (Wu et al., 2013). Epicardial Notch signaling mediates EMT by activating the expression of *Snai1* (Del Monte et al., 2011). Other factors possibly involved in epicardial EMT are the transcription factors *Twist1* and *Hand2*. In the chick embryo, *Twist1* is expressed in EPDCs and activates the expression of *Tbx20* during endocardial cushion formation (Lee and Yutzey, 2011; Shelton and Yutzey, 2008). A similar mechanism might be involved in epicardial EMT. *Hand2* activates *Platelet derived growth factor  $\alpha$*  (*Pdgfra*), which is required for epicardial EMT (Smith et al., 2011). Once

epicardial cells have undergone EMT, *Nfatc1* promotes their migration into the myocardium via the ECM remodeling enzyme Cathepsin k (*Ctsk*) (Combs et al., 2011).

Lineage tracing studies in the mouse have proposed that EPDCs give rise to cardiomyocytes as well (Cai et al., 2008; Zhou et al., 2008b). However, these results remain controversial, for example because of the expression of *Tbx18* in parts of the murine myocardium (Christoffels et al., 2009; Rudat and Kispert, 2012). In the zebrafish, lineage tracing using *pcf21* driven transgenes revealed an EPDC contribution to smooth muscle and fibroblast lineages only (Kikuchi et al., 2011a).

The developing epicardium furthermore stimulates myocardial proliferation through paracrine factors such as retinoic acid (Masters and Riley, 2014). Epicardial responsiveness to retinoic acid is required for myocardial growth (Merki et al., 2005). Additionally, retinoic acid was found to induce Erythropoietin production in the liver, which stimulates the expression of *Igf2* in the epicardium (Brade et al., 2011). Indeed, *Igf* signaling promotes myocardial proliferation in mouse and zebrafish (Huang et al., 2013; Li et al., 2011b). In the zebrafish, inhibition of retinoic acid signaling did not affect cardiomyocyte proliferation, suggesting alternative ways of *Igf* signaling activation (Choi et al., 2013). Epicardial development in human tissue samples was found to recapitulate the findings obtained in animal models in that human epicardial cells expressed *Tcf21*, *Tbx18* and *Wt1* and transdifferentiated into smooth muscle cells (Risebro et al., 2015).

## 1.4 The epicardium during heart regeneration

Heart regeneration in the adult zebrafish relies on the concerted action of several cell types and molecular factors (Kikuchi, 2014). The mechanisms driving heart regeneration appear to be largely similar in adult zebrafish and neonatal mice (Porrello and Olson, 2014). Myocardial regeneration in the zebrafish was proposed to be mediated mostly by pre-existing cardiomyocytes that dedifferentiate and proliferate (Jopling et al., 2010), accompanied by the expression of Gata4 (Kikuchi et al., 2010). However, the cardiac expression of the *gata4* based reporter line used here is not exclusive to dedifferentiating cardiomyocytes, lowering its validity as a tool to show injury-induced myocardial dedifferentiation. Furthermore, a non-myocardial contribution to the regenerating myocardium cannot be excluded based on the results presented by Jopling et al. and by Kikuchi et al., since lineage tracing was either cardiomyocyte specific (*myl7:Cre* based) or used a *beta-actin* driven labeling construct which is not expressed in tissues such as the epicardium (personal observation).

Alongside myocardial regeneration, revascularization of the injured area from pre-existing coronary vessels (Marín-Juez et al., 2016; Zhao et al., 2014), nerve regeneration (Mahmoud et al., 2015), recruitment of immune cells (Aurora et al., 2014) and scar resolution (Gonzalez-Rosa et al., 2011) are crucial morphological elements of heart regeneration.

Cardiac injury triggers the reactivation of the epicardium from a quiescent state it assumes in the intact adult heart (Lepilina et al., 2006; Zhou et al., 2011a). This is underlined by the re-expression of developmental epicardial markers following cardiac injury, such as Tbx18 and Aldh1a2 (Lepilina et al., 2006). The ATPase Brg1, which is

involved in chromatin remodeling, regulates the transcriptional activity of the *Wt1* locus and thereby the reactivation of the murine epicardium (Vieira et al., 2017). Another factor involved in this process is Thymosin  $\beta$ 4 (Tmsb4), an actin organizing protein that is part of the same chromatin remodeling complex as Brg1, and treatment with Tmsb4 prior to cardiac injury enhances the regenerative capacity of the epicardium (Smart et al., 2010). Additionally, EMT markers are upregulated in the regenerating zebrafish heart and epicardial cells were found to proliferate (Kim et al., 2010). The epicardial reactivation initially is organ wide, but is only sustained around the injured area (Gonzalez-Rosa et al., 2011; Lepilina et al., 2006). However, the mechanism by which an initial organ wide activation is observed as a response to a local injury is not yet understood.

The epicardial functions following adult heart injury match those during development in that they comprise both cellular and paracrine contributions to the heart. Transplantation experiments in the zebrafish showed that adult EPDCs give rise to smooth muscle cells and fibroblasts, comparable to their embryonic counterparts (González-Rosa et al., 2012). In the mouse, EPDCs were found to additionally contribute endothelial cells and cardiomyocytes to the regenerating heart (Smart et al., 2007, 2011). However, epicardial transdifferentiation into cardiomyocytes was dependent on treatment with Tmsb4 before injury (Zhou et al., 2011a, 2012).

In addition to a cellular contribution to the regenerating heart, the epicardium has been described as required to stimulate myocardial proliferation following injury (Wang et al., 2015). However, since this study used genetic ablation of epicardial cells in addition to myocardial injury, factors such as a large amount of debris from ablated epicardial cells at the wound area might in itself have had a negative impact on the regenerative process. Signaling through the Neuregulin co-receptor ERBB2 has been found to promote

cardiomyocyte proliferation following injury in the neonatal mouse heart (D'Uva et al., 2015). Neuregulin 1 (Nrg1) was proposed to mediate these effects in the zebrafish heart (Gemberling et al., 2015). However, a heart regeneration requirement for Nrg1 specifically was not shown in this study, and other results rather point towards Nrg2a promoting regeneration following heart injury in the zebrafish (personal communication Didier Stainier).

Epicardial contributions to heart regeneration via retinoic acid and Fgf signaling (Kikuchi et al., 2011b; Lepilina et al., 2006) as well as via Tgfb1, which might regulate the formation of an initial Collagen matrix in the wound area and its subsequent replacement with myocardium (Chablais and Jazwinska, 2012), have also been described.

Additionally, epicardium derived Fibronectin and other ECM components are required for heart regeneration and appear to promote the migration of cardiomyocytes (Mercer et al., 2013; Wang et al., 2013). Furthermore, chemokine signaling emanating from the epicardium was shown to guide the migration of cardiomyocytes into the injured area (Itou et al., 2012b).

Finally, proper orchestration of the immune response following cardiac injury is required for heart regeneration (Aurora et al., 2014; Lai et al., 2017) and might be dependent on the epicardium as well, if recapitulating the epicardial function during development in recruiting immune cells to the heart (Stevens et al., 2016).

## 1.5 Heterogeneity in the epicardium

The diversity of paracrine and cellular epicardial contributions to the developing and regenerating heart raises the question whether there are distinct epicardial cell populations with differential fate and/or signaling potential. Indeed, Tcf21 is only present in a subset of epicardial cells in the developing mouse and chick heart (Braitsch et al., 2012). Additionally, Wt1 and Tbx18 label subsets of epicardial cells that only partially overlap with the epicardial cell population expressing Tcf21 (Braitsch et al., 2012). Furthermore, the formation of fibroblasts and smooth muscle cells from embryonic EPDCs in the mouse is depending on the presence or absence, respectively, of Tcf21 (Acharya et al., 2012; Braitsch et al., 2012). This fate determination of EPDCs appears to take place before they undergo EMT and leave the epicardial sheet (Acharya et al., 2012; Braitsch et al., 2012). The expression of key epicardial markers, such as Tcf21, thus specifies the fate of epicardial cells before they become EPDCs, suggesting the presence of functionally distinct cell populations within the epicardial cell layer.

Epicardial heterogeneity might be rooted in the fact that multiple tissues contribute to the epicardium. While most ventricular epicardial cells originate from the PEO, the epicardium covering the BA was found to originate from the pericardial sac (Peralta et al., 2014; Pérez-Pomares et al., 2003). In the mouse, haematopoietic cells have been identified as an additional source of epicardial cells (Balmer et al., 2014). The PEO itself was shown to be heterogeneous (Plavicki et al., 2014). A subset of murine proepicardial cells that expresses the transcription factor Scleraxis and the chemokine Semaphorin 3d gives rise to endocardium and coronary endothelium. Most of these cells do not express Wt1 or Tbx18 (Katz et al., 2012). Another study found that Wt1 expressing EPDCs from

the injured mouse heart are heterogeneous as well (Bollini et al., 2014). The expression levels of epicardial markers, such as Tbx18, in the adult EPDCs were different from those in embryonic Wt1 expressing EPDCs, suggesting that the composition of the adult epicardium differs from that of its embryonic counterpart. Recently, single cell transcriptomics revealed heterogeneity in the adult zebrafish epicardium (Cao et al., 2015). However, this analysis was focused on the intact heart and did not investigate the regenerating epicardium.

In summary, collective evidence from multiple vertebrate species found the epicardium to be composed of multiple cell populations with distinct origins and functional potentials. However, this evidence originates in large parts from studies focusing exclusively on a few epicardial markers and not characterizing epicardial cell populations on a whole-transcriptome level. To date, no single cell RNA sequencing study of epicardial heterogeneity in the developing or regenerating zebrafish heart has been published. Consequently, there is the need for a more complete understanding of epicardial heterogeneity, achieved by characterizing subpopulations through transcriptomic and genomic techniques in addition to cellular microscopy studies.

## **1.6 Revealing cellular heterogeneity through single cell RNA sequencing**

In recent years, the development of next-generation sequencing technologies that enable the analysis of single cell transcriptomes have revolutionized our ability to query the cellular diversity of superficially homogeneous cell populations in a high-throughput and high-resolution manner (Liu and Trapnell, 2016; Tang et al., 2009). Single cell approaches avoid the averaging artifacts associated with traditional bulk population data

by isolating single cells and preparing separate cDNA libraries for each of them (Kolodziejczyk et al., 2015). Thereby, single cell transcriptomics can reveal both the presence of distinct cell populations and the genetic associations between them, which would be obscured in bulk experiments (Wills et al., 2013).

However, single cell sequencing methods face novel challenges when it comes to the statistical processing of acquired sequencing data (Stegle et al., 2015). Since only minute amounts of RNA are captured from each cell, cDNA amplification is necessary to yield enough material for subsequent sequencing (Wang and Navin, 2015). This amplification step can introduce bias and limits the quantitative accuracy of single cell sequencing (Kolodziejczyk et al., 2015). Exogenous RNA standards such as those from the External RNA Control Consortium (ERCC) can be “spiked in” with cellular RNA to facilitate the normalization of read counts following sequencing (Jiang et al., 2011). An equal amount of spike-in RNA is added to each cell prior to library preparation and variations in the spike-in read counts between samples can therefore be attributed to capture efficiency, amplification bias and sequencing depth (Jiang et al., 2011). Spike-in counts are then equalized across samples to normalize the endogenous gene read counts. However, the performance of spike-in based normalization is dependent on the statistical approach used (Risso et al., 2014). Also, spike-in RNA might not behave similarly to endogenous mRNA during single cell lysis and cDNA preparation, as it is not in complex with ribosomes and RNA binding proteins (Liu and Trapnell, 2016). An alternative approach is to normalize single cell read counts using endogenous gene counts as a reference (Anders and Huber, 2010). Here, the assumption is that most genes are more or less uniformly expressed between single cell samples and that normalization can therefore be

performed by scaling the endogenous read counts so that, on average, there is no difference in the expression of most genes between samples.

Beyond read count normalization, a common step of single cell transcriptome analysis is to identify genes that are highly variable in their expression level and use such genes to assign transcriptional relationships between samples (Grün and Van Oudenaarden, 2015). Software packages have been developed that provide a framework for these analysis steps (Fan et al., 2016; Satija et al., 2015). Landmark studies have employed single cell transcriptomics this way to identify the presence of subpopulations and rare cell types within complex tissues, such as lung (Treutlein et al., 2014), brain (Zeisel et al., 2015) and intestine (Grün et al., 2015). These studies not only revealed novel cell populations, but also gene expression profiles indicating fate and function of the identified populations as well as transcriptional regulatory networks connecting the different populations.

In addition to transcriptomic studies, it is now possible to perform chromatin immunoprecipitation (ChIP) (Rotem et al., 2015), which maps histone modifications and genomic binding of transcription factors, bisulfite sequencing (Guo et al., 2013), which measures DNA methylation, and assay for transposase-accessible chromatin (ATAC) sequencing (Buenrostro et al., 2015), which measures chromatin accessibility, in single cells. Bulk ATAC sequencing is widely used to identify accessible regions in the genome that might provide a binding platform for transcription factors (Buenrostro et al., 2013). These so-called enhancers mediate robust transcription of target genes over long genomic distances in a cell type specific manner (Long et al., 2016). Furthermore, multiple enhancers might form a regulatory landscape surrounding a target gene and cooperatively finetune its expression. Therefore, identifying enhancers is crucial to

elucidate the transcriptional regulation of genes of interest and aids in the identification of transcriptional networks that might drive lineage differentiation (Lavin et al., 2014).

In conclusion, single cell transcriptomics offer an unprecedented capacity to elucidate the cellular composition of complex tissues and to characterize subpopulations therein as to their fate and function. Applying single cell transcriptomics to the developing and regenerating epicardium might allow for new insights into the regulation of cardiac development and repair. Combining single cell transcriptomics and bulk ATAC sequencing might furthermore facilitate the analysis of transcriptional regulation within identified epicardial subpopulations.

## **1.7 Aims**

### **Working hypothesis:**

The embryonic zebrafish epicardium contains distinct subpopulations of cells that possess specific lineage potentials and/or functions during cardiac development.

### **Aims:**

1. To characterize epicardial heterogeneity in the zebrafish heart using newly generated, bacterial artificial chromosome (BAC) based, transgenic reporter lines that allow for the simultaneous expression analysis of *tcf21*, *tbx18* and *wt1b*.
2. To profile heterogeneity among epicardial cells isolated from *tcf21*, *tbx18* or *wt1b* based reporter lines using single cell RNA sequencing.

3. To corroborate the validity of the epicardial cell populations identified through single cell transcriptomics using bulk RNA sequencing of epicardial cells isolated from *tcf21*, *tbx18* or *wt1b* based reporter lines.
4. To analyze the spatial localization of the epicardial subpopulations identified through single cell transcriptomics using visualization of marker gene expression.
5. To identify enhancer elements located in genomic proximity to marker genes expressed in the epicardial subpopulations identified through single cell transcriptomics.

## 1.8 Summary of findings

I investigated the presence of distinct cell populations in the developing zebrafish epicardium, using both microscopic analysis of existing and newly generated epicardial reporter lines as well as single cell transcriptomics. Using the transparent zebrafish embryo allowed for microscopic analysis of the heart and the surrounding tissues. It also provided small cell populations that could efficiently be queried via medium scale single cell sequencing. The pre-existing reporter lines *Tg(tcf21:dsRed2; wt1b:GFP)* and *Tg(tbx18:dsRed2; wt1b:GFP)* showed *tcf21*, *tbx18* and *wt1b* driven fluorophores to label partially overlapping subsets of epicardial cells. The endogenous expression of *tcf21*, *tbx18* and *wt1b* in the developing zebrafish epicardium was also heterogeneous in nature. To analyze the expression patterns of *tcf21*, *tbx18* and *wt1b* simultaneously, I generated the triple reporter line *Tg(tcf21:myr-tdTomato; tbx18:myr-GFP; wt1b:H2B-Dendra2)* and refined the results obtained using *Tg(tcf21:dsRed2; wt1b:GFP)* and *Tg(tbx18:dsRed2; wt1b:GFP)*. Additionally, I identified a *wt1b:H2B-Dendra2* single

positive epicardial cell population covering the developing BA. Single cell transcriptomic analysis at 5dpf, including epicardial reporter cells and cells from other cardiac lineages, identified three distinct epicardial cell populations, termed epicardial cell cluster 1, 2 and 3. Single cell sequencing additionally identified populations of mesenchymal and neural cells that have not been studied in the developing zebrafish heart before. The epicardial subpopulations expressed epicardial marker genes such as *tcf21*, *tbx18*, *wt1b* and *aldh1a2*, albeit in a differential manner. Furthermore, epicardial cell cluster 1 strongly expressed the signaling peptide gene *adma*, which might regulate a subset of mesenchymal cells expressing Adma receptors and endothelial marker genes. The transcriptomic profile of epicardial cell cluster 2 indicated it was a smooth muscle cell population. Additionally, the cells in epicardial cell cluster 2 expressed the non-canonical Wnt ligand gene *wnt11 related (wnt11r)* which, based upon the expression patterns of other Wnt signaling components and the downstream target gene *alcama*, might regulate both epicardial and mesenchymal cells in the developing zebrafish heart. Epicardial cell cluster 3 expressed *cxc12a* and its transcriptomic profile suggested it might guide immune cells in the developing zebrafish heart through the receptor Cxcr4b. Transcriptomic bulk data from epicardial cells supported the presence of epicardial subpopulations matching those characterized by single cell transcriptomics. Via microscopic analysis of cluster specific marker gene expression, I confirmed epicardial cell cluster 1 and 3 to be located in the epicardial cell layer, while epicardial cell cluster 2 resides in the BA. To elucidate the transcriptional regulation in the different epicardial subpopulations, I generated and analyzed ATAC sequencing data and identified putative enhancer elements located in proximity to marker genes of epicardial cell clusters 1 and 2. An *in vivo* reporter assay revealed that the activity of the newly identified enhancer

elements recapitulated the location of the respective epicardial subpopulation. Furthermore, the identified enhancers contained several binding sites for transcription factors expressed in epicardial cell clusters 1 and 2, such as Gata4, Hand2 and Foxc1a. Collectively, these findings reveal the presence of multiple epicardial subpopulations in the developing zebrafish heart and generate numerous hypotheses as to how interactions between epicardial and non-epicardial cell populations might be orchestrated to tightly regulate cardiovascular development. Additionally, similar epicardial and non-epicardial cell populations should be mostly present in the mammalian system, as genes and biological circuits have been conserved during evolution. At the same time, this presents an exciting opportunity for looking into the adult zebrafish heart, which might recapitulate the various developmental cellular interactions revealed herein to promote cardiac repair upon injury.

## Chapter 2

### Material and methods

#### 2.1 Zebrafish lines

All lines were maintained at 28.5°C on a 14 hours light / 10 hours dark cycle. Fish were fed three times daily with flake food, high protein pellets and enriched brine shrimp. The transgenic reporter lines *Tg(tcf21:dsRed2)<sup>pd37</sup>* (Kikuchi et al., 2011a), *Tg(tbx18:dsRed2)<sup>pd22</sup>* (Kikuchi et al., 2011a), *Tg(wt1b:GFP)<sup>li1</sup>* (Perner et al., 2007), *Tg(myf7:GFP)* (Burns et al., 2005), *Tg(myf7:dsRed2-NLS)* (Mably et al., 2003), *Tg(kdrl:GFP)* (Jin, 2005), *Tg(gata1a:dsRed)* (Traver et al., 2003) and *Tg(ptprc:dsRed)* (Bertrand et al., 2008) have been described before. *Tg(tcf21:dsRed2; wt1b:GFP)*, *Tg(tbx18:dsRed2; wt1b:GFP)* and *Tg(tcf21:dsRed2; myf7:GFP)* and *Tg(kdrl:GFP; gata1a:dsRed)* double transgenic lines have been established previously. To generate *Tg(tcf21:H2B-Dendra2)*, *Tg(tcf21:myr-GFP)*, *Tg(tcf21:myr-tdTomato)*, *Tg(tbx18:H2B-Dendra2)*, *Tg(tbx18:myr-GFP)*, *Tg(tbx18:myr-Citrine)*, *Tg(wt1b:H2B-Dendra2)* and *Tg(wt1b:myr-tdTomato)*, a bacterial artificial chromosome (BAC) recombineering approach was used (Bussmann and Schulte-Merker, 2011). Fluorophore sequences were cloned into a pGEM vector backbone (provided by Shankar Srinivas/Tatjana Sauka-Spengler), covalently linked to an upstream *histone 2b* (*H2B*) or *myristoylation* (*myr*) sequence that targets the translated protein to the cell nucleus or the plasma membrane, respectively (Dempsey et al., 2012; Zacharias, 2002). Fluorophore sequences were accompanied downstream by an SV40 late polyA sequence and a kanamycin resistance gene flanked by FRT sites. The fluorophore sequences including FRT-Kan-FRT were amplified via PCR, using primers with 5' homology arms

targeting the first coding exon of the gene contained on the recipient BAC. Primers used are listed in table 2-1.

Primer name	Primer sequence 5' -> 3'
BAC Tcf21 H2BDendra2 fkf for	CATCTCCTCAAGAAGTCCTTTTCTCCACTCCACCCTTGCTCCAGCCAACat gccagagccagcgaagtctg
BAC Tcf21 myrtdTomato fkf for	CATCTCCTCAAGAAGTCCTTTTCTCCACTCCACCCTTGCTCCAGCCAACat gggctgcatcaagagcaagc
BAC Tcf21 fkf rev	CCAAACAACATTAGATTAACCGAATCGGAAAACCAAATGAATTTATGAA AACTCAATATTAATTCTGATTGCAAGAAGTGTCTCACTccaGAAGTAGTGA GGAGGCTTT
BAC Tbx18 H2BDendra2 fkf for	TTCTGGTGAACCTCTCTTTCTCGGCCAATCTGTCTTCTCGGTCGGTAACCat gccagagccagcgaagtctg
BAC Tbx18 myrCitrine/GFP fkf for	TTCTGGTGAACCTCTCTTTCTCGGCCAATCTGTCTTCTCGGTCGGTAACCat gggctgcatcaagagcaagc
BAC Tbx18 fkf rev	TACTTACGGGATTCGTCGCTGGTCAATCTATCTCGCAACTCCTGGTGCTt ccaGAAGTAGTGAGGAGGCTTT
BAC Wt1b H2BDendra2 fkf for	GTGTTTTGCAACCCAGAAAATCCGTCTAAATGCTGACAGAGCCGTGCGG CCCgatgccagagccagcgaagtctg
BAC Wt1b myrtdTomato fkf for	GTGTTTTGCAACCCAGAAAATCCGTCTAAATGCTGACAGAGCCGTGCGG CCCgatgggctgcatcaagagcaagc
BAC Wt1b fkf rev	GACCACATTGAGAGAGATTTTGAGGCGAGATTGTAAGGACGGGATGGTT TTCTCACTccaGAAGTAGTGAGGAGGCTTT

**Table 2-1: Primers used for BAC entry cassette amplification.**

The amplified donor cassettes were electroporated into electrocompetent SW105 cells harboring BACs that contained the loci of *tcf21* (DKEYP 79F12, pINDIGO), *tbx18* (DKEYP 117G5, pINDIGO) or *wt1b* (CH73 157N22, pTAR) (provided by Filipa Simoes). Successful replacement of the first coding exon of *tcf21*, *tbx18* and *wt1b* with the donor cassette was confirmed by kanamycin resistance screening and sequencing. Subsequently, the FRT-Kan-FRT was excised by arabinose-induced Flp recombinase. DNA sequences containing Tol2 recognition sites were electroporated into cells harboring BACs

containing the fluorophore cassettes. Final BAC constructs were validated by sequencing. Wildtype embryos were injected at the 1 cell stage using between 85pg and 200pg of the BAC DNA and 100pg or 200pg of *in vitro* transcribed Tol2 RNA to enable integration into the genome. Injected embryos were screened for fluorescence at 5dpf and grown to adulthood. Fluorescent offspring were grown to adulthood and crossed to generate *Tg(tcf21:myr-tdTomato; wt1b:H2B-Dendra2)*. Subsequently, double fluorescent offspring were crossed to *Tg(tbx18:myr-GFP)* to generate *Tg(tcf21:myr-tdTomato; tbx18:myr-GFP; wt1b:H2B-Dendra2)*.

## 2.2 Immunocytochemistry

Embryos were collected following natural spawning and incubated at 28.5°C. Embryos were euthanized in 5g/L ethyl-3-amino benzoate methanesulfonate salt (MS222) (in H<sub>2</sub>O, pH7.0) and fixed in 4% paraformaldehyde (PFA) (in phosphate buffered saline (PBS)) for 45 minutes at room temperature. Afterwards, embryos were washed 3 times for 5 minutes in PBS-T (PBS, 0.4% Triton, 2% dimethyl sulfoxide (DMSO)) and blocked by adding 5% goat serum for 45 minutes at room temperature. Primary antibodies used were: chicken  $\alpha$  GFP (1:500), rabbit  $\alpha$  dsRed2 (1:500), mouse  $\alpha$  Mlck (1:500, M7905, Sigma-Aldrich), rabbit  $\alpha$  Tuba1 (1:500, GTX124965, Genetex), rabbit  $\alpha$  Cyp26b1 (1:500, GTX127919, Genetex) and rabbit  $\alpha$  Lcp1 (1:400, generated by Prof Paul Martin, Bristol) in PBS-T, 2.5% goat serum. Embryos were incubated with primary antibodies overnight at 4°C. The next day, embryos were washed 6 times for 15 minutes in PBS-T at room temperature. Secondary antibodies used were: goat  $\alpha$  chicken-488nm (1:1000, ThermoFisher), goat  $\alpha$  rabbit-555nm (1:1000, ThermoFisher) (used with rabbit  $\alpha$  dsRed2),

goat  $\alpha$  rabbit-633nm (1:1000, ThermoFisher), goat  $\alpha$  mouse-633nm (1:1000, ThermoFisher) and Hoechst reagent 33258 (1:1000) in PBS-T, 2.5% goat serum, added for 2 hours at room temperature. Embryos were washed in PBS-T 3 times for 15 minutes at room temperature and overnight at 4°C, and subsequently mounted onto glass slides using VectaShield mounting medium.

### **2.3 Hybridization chain reaction (HCR)**

HCR was performed following a modified protocol from (Choi et al., 2014a). Embryos were collected at 5dpf, euthanized in 5g/L MS222 and fixed in 4% PFA (PBS) overnight at 4°C. Subsequently, embryos were washed into methanol and stored at -20°C for at least 2 hours. On the first day of staining, embryos were rehydrated into PBS-T (PBS, 0.1% Triton), permeabilized with 30 $\mu$ g/ml proteinase K (PBS-T) for 45 minutes at room temperature, post-fixed in 4% PFA, pre-hybridized with probe hybridization buffer (50% formamide, 5x sodium chloride sodium citrate (SSC), 9mM Citric acid (pH 6.0), 0.1% Tween20, 50 $\mu$ g/ml heparin, 500 $\mu$ g/ml tRNA, 1x Denhardt's solution, 10% dextran sulfate) for 30 minutes at 65°C and hybridized with DNA probes against the desired target mRNAs (each probe at 2nM in hybridization buffer, incubation overnight at 45°C). Probes against *tcf21* (amplifier B1), *tbx18* (amplifier B2), *wt1b* (amplifier B3), *myl7* (amplifier B4) and *mCherry* (amplifier B5) were ordered from Molecular Instruments directly, probes against *adma* (amplifier B5), *cldn11a* (amplifier B4) and *elnb* (amplifier B4) were custom-designed and ordered from IDT. These probe sequences were designed to target 50bp sequences, to have a melting temperature of around 71°C, to have a GC content of 50% - 60%, to not span exon boundaries and to be spaced at least 25bp apart. Probe sequences

were then flanked with 41bp amplifier sequences to create 132bp final probes. Custom-designed probe sequences are listed in Appendix A. Excess probes were removed by washing embryos from wash buffer (50% formamide, 5x SSC, 9mM Citric acid (pH 6.0), 0.1% Tween20, 50µg/ml heparin) into 5x SSC-T (5x SSC, 0.1% Triton) at 45°C. Subsequently, embryos were pre-amplified in amplification buffer (5x SSC, 0.1% Tween20, 10% dextran sulfate) for 2 hours 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 over night 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.

## **2.4 Microscopy and digital image analysis**

Embryos were either euthanized in 5g/L MS222 and mounted into plastic dishes using 1% low melting agarose or stained and mounted onto glass slides using VectaShield mounting medium. Imaging was performed on an inverted LSM510 or an upright LSM780 confocal microscope (Zeiss). Embryos mounted in agarose were imaged with a 20x dipping objective, zoom 1, scan speed 7, averaging 1. Embryos mounted on glass slides were imaged with a 40x water objective, an overview image was taken at zoom 1, scan speed 6, averaging 4 and regions of interest were imaged at zoom 3, scan speed 8, averaging 4. To quantify cell numbers, labelled cell nuclei were marked using the ZEN software (Zeiss) and counted. Cells labelled by plasma membrane-targeted fluorescence were counted using Imaris software (bitplane). Data was visualized in R and significance

was determined by an unpaired Student's *t*-test. To compare the total number of epicardial cells labelled by *tcf21*, *tbx18* or *wt1b* driven fluorescence to the number of double or triple labelled cells, a one-way *t*-test was used, since the number of double or triple labelled cells could only be lower than the total number of cells labelled and not higher. To compare control and injured conditions following heart injury, a two-way *t*-test was used, because injury might affect cell numbers either way. If multiple comparisons were possible, the p-value was corrected according to Bonferroni (\* 0.05/(n possible comparisons), \*\* 0.01/(n possible comparisons), \*\*\* 0.001/(n possible comparisons)).

## 2.5 Larval heart injury

A 1460nm infrared laser (class 1) coupled to a 40x magnification objective (XYClone) was used for ventricular ablation at 4dpf (Matrone et al., 2013). Embryos were mounted into glass bottom dishes and a single pulse (100% laser power, 2ms) was aimed at the apex of the ventricle. In control embryos, a similar pulse was aimed at the distal tail fin. Following injury, embryos were either collected for immunocytochemistry or incubated at 28.5°C until collection at 2hpi, 15hpi or 24hpi.

## 2.6 Lightsheet imaging

*Tg(myI7:GFP; tcf21:dsRed2)* embryos at 4dpf were mounted onto 1% agarose rods made in 1ml syringes with their tip cut off. Longitudinally, a piece of the rod was sliced off, a single embryo was placed onto the agarose and fixed with 1% low melting agarose. Then,

the rod was pulled back into the syringe. Mounted embryos were anaesthetized with 0.5g/L MS222 and imaged on a Z1 lightsheet microscope (Zeiss) using a W Plan-Apochromat 20x/1.0 UV-VIS objective and 488nm as well as 561nm laser lines. 190 zstacks (44 frames per zstack, covering the rear half of the ventricle) were taken with an exposure time of 20ms per frame. Green and red fluorescence was imaged in a single channel. Subsequently, a custom-made Matlab algorithm was used to correct for the movements of the beating heart. The design of this algorithm followed the basic concept of code that has been published (Liebling et al., 2005, 2006; Staudt et al., 2014), but differed in its strategy: First, a reference image in the middlemost frame of the zstacks was identified that showed the ventricle at maximum dilation. Starting from the reference image, the most similar image in each successive frame towards both ends of the stacks was identified and incorporated into a final reconstructed zstack. First, the alignment algorithm was run only using the myocardial signal. In a second approach, myocardial and epicardial signals were processed separately and the channels were subsequently re-combined.

## **2.7 Purification of embryonic hearts and dissociation for FACS**

Transgenic reporter embryos were screened for fluorescence at 4dpf and euthanized in 5g/L MS222 at 5dpf. Hearts were isolated following the protocol published by Burns & MacRae (Burns and MacRae, 2006). About 150 embryos were processed at a time, washed 3 times in ice-cold medium (L-15 medium, 10% FBS) and dissociated using a 21-gauge needle: Embryos were drawn into a syringe and expelled back into the tube 14 times. Subsequently, large fragments were filtered off using a 100µm cell strainer

(Corning). Then, hearts were retained on a 40 $\mu$ m cell strainer. Hearts were washed off the strainer into a dish with medium and a p20 pipette was used to manually collect hearts on ice. Typically, 150 - 200 hearts were collected from 250 - 350 embryos. These were then spun down at 800g for 10 minutes at room temperature and dissociated using 6mg Collagenase (C8176, Sigma Aldrich) in 300 $\mu$ l Trypsin solution (0.05% trypsin/0.53mM EDTA in HBSS (with sodium bicarbonate, without calcium and magnesium)). Samples were incubated 3 times for 4 minutes at 30°C and pipetted up and down 10 times between incubations. Samples were then diluted in 5ml Hank's buffer (37.5mg BSA in 1.25ml 10x HBSS, 125 $\mu$ l 1M HEPES (pH 8.0), 8.625ml H<sub>2</sub>O), pipetted up and down again 10 times and spun down at 500g for 10 minutes at room temperature. Samples were then resuspended in 5ml Hank's and filtered using 100 $\mu$ m cell strainers, followed by centrifugation at 750g for 10 minutes at room temperature. Samples were then resuspended in 500-1000 $\mu$ l of Hank's, 15 $\mu$ l of the cell viability dye 7-AAD was added and fluorescent cells as well as non-fluorescent control cells were purified using fluorescence-activated cell sorting (FACS), gating for 7-AAD negative singlet events. Cell bulks for cDNA synthesis were purified into PCR tubes, single cells were purified into 96-well plates and cell bulks for ATAC were purified into 1.5ml Eppendorf tubes.

## **2.8 cDNA synthesis from single cell samples**

FACS-purified single cells were processed following a modified Smart-seq2 protocol (Picelli et al., 2014) to reverse transcribe poly-adenylated RNA. All volumes are per well. In the clean room, 1 $\mu$ l RNase inhibitor (40U/ $\mu$ l, Clontech) was added to 19 $\mu$ l 0.4% Triton. 2 $\mu$ l of the RNase inhibitor/Triton mix were then added to 1 $\mu$ l dNTPS (10mM,

ThermoFisher), 1µl Oligo-dT30VN primer (10µM, IDT, sequence: 5'AAGCAGTGGTATCAACGCAGAGTACT<sub>30</sub>VN-3') and 0.0375µl ERCC synthetic spike in RNA (ThermoFisher) to obtain 1x lysis buffer. The lysis buffer was then dispensed into a 96-well plate and single cells were FACS-purified into individual wells. 288 fluorescent cells were purified from *Tg(tcf21:H2B-Dendra2)*, 96 fluorescent cells from *Tg(tbx18:myr-GFP)* and 96 fluorescent cells from *Tg(wt1b:H2B-Dendra2)*. Additionally, 31 cells were collected from *Tg(myf7:GFP)*, 40 GFP fluorescent cells from *Tg(kdrl:GFP; gata1a:dsRed)*, 20 dsRed fluorescent cells from *Tg(kdrl:GFP; gata1a:dsRed)*, 96 non-fluorescent cells from *Tg(tcf21:dsRed2; myf7:GFP) x Tg(kdrl:GFP; gata1a:dsRed)* and 96 cells from wildtype hearts. Collected cells were frozen at -80°C until further processing (maximum storage 16 days). In the clean room, lysed cells were thawed, incubated at 72°C for 3 minutes to hybridize the oligo-dT primer and then cooled down to 4°C. Subsequently, 2µl 5x SuperScript II first strand buffer, 0.5µl dithiothreitol (DTT, 100mM), 2µl betaine (5M, Sigma-Aldrich), 0.1µl MgCl<sub>2</sub> (1M, Sigma-Aldrich), 0.25µl RNase inhibitor, 0.085µl template-switching oligo (TSO, 100µM, sequence: 5'-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3', Exiqon), 0.8165µl H<sub>2</sub>O and 0.25µl Superscript II reverse transcriptase (200U/µl, Invitrogen) were added. Reverse transcription was carried out by incubation at 42°C for 90 minutes, 10 cycles of 50°C for 2 minutes and 42°C for 2 minutes, followed by 70°C for 15 minutes and 4°C hold. cDNA was amplified by adding 12.5µl KAPA Hifi HotStart ReadyMix (2x, KAPA Biosystems), 0.2µl ISPCR primer (10µM, sequence: 5'-AAGCAGTGGTATCAACGCAGAGT-3', IDT) as well as 2.3µl H<sub>2</sub>O and PCR (98°C 3min, 20 cycles (98°C 20sec, 67°C 15sec, 72°C 6min), 72°C 5min). Amplified cDNA was purified using Agencourt AMPure XP beads (Beckman Coulter). 15.4µl beads were added to each sample, incubated for 10 minutes and placed on a

magnetic separation device for 5 minutes. Beads were then washed twice with 200 $\mu$ l 80% ethanol, air dried for 5 minutes, taken off the magnetic separator and resuspended in 15 $\mu$ l H<sub>2</sub>O. Following 5 minutes of incubation, samples were placed back on the magnetic separator for 2 minutes, 13 $\mu$ l purified cDNA was taken off and stored at -20°C. The amount of synthesized cDNA in each well was quantified using the Quant-iT PicoGreen dsDNA Assay (ThermoFisher). Afterwards, cDNA from 11 wells sampling the whole range of the obtained PicoGreen results was analyzed on an Agilent 2100 Bioanalyzer (Agilent Technologies). The Bioanalyzer results were used to determine a quality cut-off PicoGreen value. Only wells with a PicoGreen value above the cut-off were taken further to library preparation. Since a primer dimer peak at 92bp was visible on the Bioanalyzer results, samples were re-purified using 11 $\mu$ l sample and 7.7 $\mu$ l AMPure beads, eluting into 10 $\mu$ l H<sub>2</sub>O. This eliminated the 92bp peak. Purified cDNA was stored at -20°C.

## 2.9 cDNA synthesis from bulk samples

FACS-purified cell bulks were processed using the SMART-Seq<sup>™</sup> v4 Ultra<sup>™</sup> Low Input RNA Kit for Sequencing (Takara Clontech) to reverse transcribe poly-adenylated RNA. In a clean room and under a flow hood, 10x Reaction Buffer was prepared by mixing 1 $\mu$ l RNase inhibitor with 19 $\mu$ l 10x lysis buffer. Then, 1 $\mu$ l 10x Reaction Buffer was added to 6 $\mu$ l H<sub>2</sub>O to obtain 1x lysis buffer. Subsequently, cardiac cells were FACS-purified into 1x lysis buffer. Numbers of fluorescent cells purified from *Tg(tcf21:H2B-Dendra2)* were 1000 (bulk 1), 1000 (bulk 2) and 676 (bulk 3). Cell numbers purified from *Tg(tbx18:H2B-Dendra2)* were 350 (bulk 1), 345 (bulk 2) as well as 809 cells from *Tg(tbx18:myr-GFP)* (bulk 3). Cell numbers purified from *Tg(wt1b:H2B-Dendra2)* were 690 (bulk 1), 442 (bulk

2) and 600 (bulk 3). Additionally, 1000 non-fluorescent control cells were purified during each sort. Collected cells were frozen at  $-80^{\circ}\text{C}$  until further processing (maximum storage 6 days). In the clean room, sample volume was adjusted to  $10.5\mu\text{l}$  with  $\text{H}_2\text{O}$ . Then,  $2\mu\text{l}$  3' SMART-Seq CDS Primer II A ( $12\mu\text{M}$ ) were added and samples were incubated at  $72^{\circ}\text{C}$  for 3 minutes, followed by incubation at  $4^{\circ}\text{C}$  for 2 minutes. Afterwards,  $4\mu\text{l}$  5x Ultra Low First-Strand Buffer,  $1\mu\text{l}$  SMART-Seq v4 Oligonucleotide ( $48\mu\text{M}$ ),  $0.5\mu\text{l}$  RNase inhibitor and  $2\mu\text{l}$  SMARTScribe Reverse Transcriptase were added and samples were incubated at  $42^{\circ}\text{C}$  for 90 minutes,  $70^{\circ}\text{C}$  for 10 minutes and  $4^{\circ}\text{C}$  hold. Following reverse transcription, cDNA was amplified by adding  $25\mu\text{l}$  2x SeqAmp PCR Buffer,  $1\mu\text{l}$  PCR Primer II A ( $12\mu\text{M}$ ),  $1\mu\text{l}$  SeqAmp DNA Polymerase and  $3\mu\text{l}$   $\text{H}_2\text{O}$ . In the general lab, cDNA was amplified via PCR ( $95^{\circ}\text{C}$  1min, x cycles ( $98^{\circ}\text{C}$  10sec,  $65^{\circ}\text{C}$  30sec,  $68^{\circ}\text{C}$  3min),  $72^{\circ}\text{C}$  10min). 10 cycles were used for control samples, 11 cycles for samples with more than 700 fluorescent cells and 12 cycles for samples with less than 700 fluorescent cells. Samples were purified as described for single cell samples, except that  $50\mu\text{l}$  AMPure beads and  $1\mu\text{l}$  10x lysis buffer were used, samples were air dried for 2 minutes and resuspended in  $17\mu\text{l}$  elution buffer. The purified cDNA was quantified using Qubit Fluorometric Quantitation (ThermoFisher) and the quality was verified on an Agilent 2100 Bioanalyzer (Agilent Technologies).

## **2.10 Illumina library preparation from cDNA**

Final sequencing libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina). For cDNA made from single cell samples, a fixed input volume of  $1.25\mu\text{l}$  was used as input and added to  $2.5\mu\text{l}$  TD buffer. Next,  $1.25\mu\text{l}$  ATM was added and samples were tagmented at  $55^{\circ}\text{C}$  for 8 minutes. Immediately after, samples were put on ice and

1.25µl NT buffer was added to stop tagmentation. Tagmented cDNA was amplified and double indexes added (Forward: S517, S502-S507 and reverse: N701-N712). To this end, 3.25µl NPM, 0.6µl forward primer (10µM, IDT), 0.6µl index reverse primer (10µM, IDT) and 1.8µl H<sub>2</sub>O were added, followed by PCR (72°C 3min, 95°C 30sec, 12 cycles (95°C 10sec, 55°C 30sec, 72°C 30sec), 72°C 5min). All index primer sequences are listed in Appendix B. Amplified libraries were purified adding 6µl AMPure beads and incubating for 5 minutes. Afterwards, samples were placed on a magnetic separator, incubated for 2 minutes and washed twice with 200µl 80% ethanol. Beads were air dried for 15 minutes, resuspended in 12.5µl resuspension buffer, incubated for 5 minutes and placed on the magnetic separator for 2 minutes. Then, 10µl purified library was taken off and stored at -20°C. The libraries were quantified using the Quant-iT PicoGreen dsDNA Assay (ThermoFisher) and the quality of 10 random libraries was verified on a 2200 TapeStation system (Agilent). To reduce batch effects due to technical sequencing variability, single cell libraries of all sorting conditions were shuffled, pooled and diluted to a final concentration of 3.3nM, yielding three 96 sample library pools. The pools were sequenced to a depth of 130\*10<sup>6</sup> reads, 140\*10<sup>6</sup> reads and 180\*10<sup>6</sup> reads on a NextSeq500 machine (Illumina, 150 Cycle Mid Output Kits). In a pilot experiment, two pools that exclusively contained libraries derived from *Tg(tcfd21:H2B-Dendra2)* were sequenced to a depth of 100\*10<sup>6</sup> reads and 125\*10<sup>6</sup> reads.

Bulk cell cDNA libraries were prepared similarly, except that all volumes used were 4 times as high (the volumes stated in the original protocol). Also, 1ng cDNA in 5µl H<sub>2</sub>O was used as input for all bulk cDNA samples. Samples were tagmented for 5 minutes and 10 seconds at 55°C and amplified using 15µl NPM per sample, 1µl non-index forward primer (Ad1\_noMX), 1µl index reverse primer as well as 12 PCR cycles. Index primers used were:

*tcf21* bulk 1 N714, *tcf21* bulk 2 N716, *tcf21* bulk 3 N721, *tbx18* bulk 1 N701, *tbx18* bulk 2 N703, *tbx18* bulk 3 N709, *wt1b* bulk 1 N705, *wt1b* bulk 2 N711, *wt1b* bulk 3 N719, *tcf21* control bulk 1 N715, *tcf21* control bulk 2 N718, *tcf21* control bulk 3 N722, *tbx18* control bulk 1 N702, *tbx18* control bulk 2 N704, *tbx18* control bulk 3 N710, *wt1b* control bulk 1 N706, *wt1b* control bulk 2 N712, *wt1b* control bulk 3 N720. All index primer sequences are listed in Appendix B. For purification, 30µl AMPure beads were used per sample and purified libraries were eluted into 52.5µl elution buffer. The libraries were quantified using Qubit Fluorometric Quantitation (ThermoFisher) and their quality checked on a 2200 TapeStation system (Agilent). All bulk cDNA libraries were pooled, diluted to a final concentration of 4nM and sequenced to a depth of  $355 \times 10^6$  reads on a NextSeq500 machine (Illumina, 150 Cycle High Output Kit).

## 2.11 Single cell RNA sequencing data analysis

Sequencing data was transferred to a Unix server environment and an initial quality check was performed with FastQC (Babraham Institute, Cambridge). Transcriptome data was subsequently mapped to the zebrafish reference genome (GRCz10) using the STAR gapped aligner (Dobin et al., 2013). Duplicate reads were removed and reads were summarized using featureCounts (Liao et al., 2014). The resulting read count matrix was further analyzed in R. Quality control of the single cell transcriptome data was done using the scater package (McCarthy et al., 2016). Libraries that had a size more than 3 median absolute deviations (MADs) below the median of the whole data set were excluded from the analysis. This was true for 28 libraries. Furthermore, libraries were excluded from the analysis if the number of expressed genes was more than 3 MADs below the median of

the whole data set (true for 60 libraries), or if the percentage of counts representing mitochondrial genes or spike in features was more than 3 MADs above the median of the whole data set. This was true for 58 libraries (mitochondrial gene fraction) and for 96 libraries (spike in fraction). A single library might be excluded in multiple of the above categories. In total, 108 of 460 libraries were excluded from analysis during the quality control process. Before quality control, the data set contained 228 fluorescent cells from *Tg(tcf21:H2B-Dendra2)*, 52 fluorescent cells from *Tg(tbx18:myr-GFP)*, 50 fluorescent cells from *Tg(wt1b:H2B-Dendra2)*, 14 fluorescent cells from *Tg(myf7:GFP)*, 15 GFP fluorescent cells from *Tg(kdrl:GFP; gata1a:dsRed)*, 2 dsRed fluorescent cells from *Tg(kdrl:GFP; gata1a:dsRed)*, 46 non-fluorescent cells from *Tg(tcf21:dsRed2; myf7:GFP) x Tg(kdrl:GFP; gata1a:dsRed)* and 50 wildtype cells, as well as 3 control libraries generated from empty wells. Following quality control, the data set contained 137 fluorescent cells from *Tg(tcf21:H2B-Dendra2)*, 51 fluorescent cells from *Tg(tbx18:myr-GFP)*, 46 fluorescent cells from *Tg(wt1b:H2B-Dendra2)*, 8 fluorescent cells from *Tg(myf7:GFP)*, 15 GFP fluorescent cells from *Tg(kdrl:GFP; gata1a:dsRed)*, 2 dsRed fluorescent cells from *Tg(kdrl:GFP; gata1a:dsRed)*, 45 non-fluorescent cells from *Tg(tcf21:dsRed2; myf7:GFP) x Tg(kdrl:GFP; gata1a:dsRed)* and 48 wildtype cells, a total of 352 cells. Genes with an average expression across all cells of below 0.1 counts as well as mitochondrial genes were excluded from analysis, which gave a total of 20946 features. The cleaned data set was further processed using Pagoda routines in the SCDE package (Fan et al., 2016; Kharchenko et al., 2014). Spike in features were excluded from the analysis, which reduced the total number of genes analyzed to 20862. First, error models were computed for the individual cells using k-nearest neighbor model fitting and estimating 9 subpopulations. The variance in expression of each gene derived from the error models

was then normalized against the expression magnitude and the individual drop-out probability to determine genes with highly variable expression across the data set. The data set was then normalized against gene coverage in the individual cells and highly variable genes were grouped, both according to the correlation of their expression profiles in the data set and according to their gene ontology annotations. Then, genes annotated to cell cycle related gene ontology terms were excluded from the data set and the remaining highly variable genes were grouped a second time. Final gene sets were ranked according to the significance of their overall variability and, based on the similarity of their patterns, further grouped into top aspects of heterogeneity. Cells were then clustered according to the top heterogeneity aspects and assigned to the 9 most significant clusters. Subsequently, differential gene expression analysis was performed between each cluster and the rest of the data set to determine enriched genes. Counts were transformed into FPKM expression values with the `rpkm()` command in the `edgeR` package (Robinson et al., 2010) and plotted in a heatmap using `pheatmap`. The `Rtsne` package (van der Maaten, 2014) was used to plot t-SNE representations of the data set that were based on 2376 highly variable genes. To ensure reproducibility, the seed was set to zero and the perplexity to 10. The overlay of expression values and t-SNE was adapted from (Grün et al., 2015). To compare single cell data to bulk data, counts of all cells assigned to a cluster were summed up to create pseudo- bulks. The first two principal components of pseudo- bulk data as well as bulk data were then plotted.

## 2.12 Bulk RNA sequencing data analysis

To process bulk transcriptomic data, the DESeq2 package (Love et al., 2014) was used. Only genes with more than 1 count assigned in the entire data set were considered for analysis. Differential gene expression analysis was performed using a generalized linear model based on a negative binomial distribution that considered library size and gene expression magnitude. Bulk data from fluorescent cells was first compared to the respective non-fluorescent control bulk data. Then, experimental bulk data was subset to only contain genes that were significantly enriched in at least one experimental bulk data set and experimental bulk data subsets were compared to each other. Counts were transformed into FPKM expression values with the `rpkm()` command in the edgeR package and plotted in a heatmap using pheatmap.

## 2.13 Assay for Transposase Accessible Chromatin (ATAC) sequencing

ATAC was performed following a protocol modified from Buenrostro et al. (Buenrostro et al., 2013). First, cells were FACS-purified into 400 $\mu$ l Hank's buffer. 700 (bulk 1) and 568 (bulk 2) fluorescent cells were collected as technical replicates from hearts isolated from *Tg(tcf21:H2B-Dendra2)* embryos. Additionally, two bulks of 700 fluorescent cells were collected from non-cardiac tissues derived from the same embryos. Two bulks of 2300 fluorescent cells as well as two bulks of 2300 non-fluorescent cells were collected as technical replicates from hearts isolated from *Tg(tbx18:myr-GFP)* and *Tg(tbx18:myr-Citrine)* embryos. In addition, two bulks of 2300 fluorescent as well as non-fluorescent cells were collected from non-cardiac tissues derived from the same embryos. Collected

cells were pelleted at 600g for 7.5 minutes at 4°C, washed with 50µl ice-cold PBS, pelleted again and resuspended in 50µl ice-cold lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl<sub>2</sub>, 0.1% IGEPAL CA-630 (nonionic detergent, Sigma-Aldrich) in H<sub>2</sub>O). Cells were pelleted immediately at 600g for 10 minutes at 4°C. Samples collected from *Tg(tcf21:H2B-Dendra2)* embryos were then tagmented using 0.35µl Tn5 transposase (Nextera DNA Library Preparation Kit, Illumina) in 10µl TD buffer (Illumina) and 9.65µl H<sub>2</sub>O for 20 minutes at 37°C. Samples collected from *Tg(tbx18:myr-GFP)* and *Tg(tbx18:myr-Citrine)* embryos were tagmented using 0.25µl Tn5 in 5µl TD buffer and 4.75µl H<sub>2</sub>O for 30 minutes at 37°C (bulk 1) or using 0.5µl Tn5 in 5µl TD buffer and 4.5µl H<sub>2</sub>O for 28 minutes at 37°C (bulk 2). Then, EDTA was added to a concentration of 50mM and samples were incubated for 30 minutes at 50°C. MgCl<sub>2</sub> was added to a concentration of 50mM and 16µl of each sample were PCR amplified. To this end, 7.75µl H<sub>2</sub>O, 0.625µl non-index Illumina forward primer (Ad1\_noMX), 0.625µl index Illumina reverse primer and 25µl 2x NEB Next HiFi PCR mix (New England BioLabs) were added and amplification was run as follows: 72°C 5min, 98°C 30sec, x cycles of (98°C 10sec, 63°C 30sec, 72°C 1min). 15 PCR cycles were used for samples collected from *Tg(tcf21:H2B-Dendra2)* embryos, 16 cycles were used for samples collected from *Tg(tbx18:myr-GFP)* and *Tg(tbx18:myr-Citrine)* embryos. Index primers used for samples collected from *Tg(tcf21:H2B-Dendra2)* embryos were N711 (fluorescent cardiac bulk 1), N712 (fluorescent cardiac bulk 2), N714 (fluorescent non-cardiac bulk 1) and N715 (fluorescent non-cardiac bulk 2). Index primers used for samples collected from *Tg(tbx18:myr-GFP)* and *Tg(tbx18:myr-Citrine)* embryos were N706 (fluorescent cardiac bulks 1 and 2), N707 (non-fluorescent cardiac bulks 1 and 2), N708 (fluorescent non-cardiac bulks 1 and 2) and N709 (non-fluorescent non-cardiac bulks 1 and 2). All index primer sequences are listed

in Appendix B. Following PCR, samples were purified using 45µl AMPure beads. Beads were incubated with samples for 10 minutes, separated, washed and air-dried for 6 minutes, before eluting DNA into 14µl 0.1x TE buffer. The quality of the ATAC libraries was checked on a 2200 TapeStation system (Agilent) and libraries were quantified integrating Qubit Fluorometric Quantitation (ThermoFisher) with the relative areas covered by peaks of different sizes as determined from the TapeStation profiles. The formula used to determine the volume of each library in 20µl of a 4nM pool was  $(1/(\text{sum relative peak areas}) * (52.8/\text{Qubit reading}))$ . Libraries from *Tg(tcf21:H2B-Dendra2)* embryos as well as bulk 1 and 2 samples from *Tg(tbx18:myr-GFP)* and *Tg(tbx18:myr-Citrine)* embryos were pooled separately, diluted to a final concentration of 4nM and sequenced to a depth of  $540 * 10^6$ ,  $360 * 10^6$  and  $560 * 10^6$  reads on a NextSeq500 machine (Illumina, 75 Cycle High Output Kits).

## 2.14 ATAC sequencing data analysis and enhancer studies

Sequencing data was transferred to a Unix server environment and an initial quality check was performed with FastQC. Reads were then aligned to the reference genome (Zv10) using BWA-MEM (Li and Durbin, 2009). Duplicate reads were removed, data transformed into the BigWig format and visually inspected in the UCSC genome browser. Genomic regions chosen for further analysis were: chr4:11,758,003-11,758,871 (6.5kb upstream of *podxl*), chr10:7,790,365-7,791,160 (in intron 3 of *lox*) and chr21:40,880,540-40,881,245 (25kb upstream of *elnb*). The region on chromosome 4 was chosen due to its high accessibility in cardiac experimental samples from *Tg(tcf21:H2B-Dendra2)* embryos as well as from *Tg(tbx18:myr-GFP)* and *Tg(tbx18:myr-Citrine)* embryos. The regions on

chromosomes 10 and 21 were chosen due to their high accessibility in cardiac experimental samples from *Tg(tbx18:myr-GFP)* and *Tg(tbx18:myr-Citrine)* embryos. The regions were amplified from 5dpf genomic DNA using the primers tcgagtttacgtaccgctagTGTTTCTGCCTCATCATGTGT (*podxl* forward), TATCGCCGCAAGCTTgctagGTTTCAGTCTCACTTTACGTTCA (*podxl* reverse), tcgagtttacgtaccgctagATCTTGAGAAGGTCCAGTTTCTTGAAC (*lox* forward), TATCGCCGCAAGCTTgctagAATGGAAAACACTGCAGCACTG (*lox* reverse), tcgagtttacgtaccgctagAAGTGGAGACCTAACATAACTCC (*elnb* forward) and TATCGCCGCAAGCTTgctagATATATAAAGCTAAATCATATTCCTTGCAGC (*elnb* reverse).

The underlined parts of the primer sequences are homologous to sequences adjacent to an NheI cutting site on the recipient DNA vector. The restriction site was located upstream of an E1b-minimal-promoter-mCherry reporter sequence that was flanked by recognition sites of the Ac transposase (Emelyanov et al., 2006). The amplified PCR fragments were integrated into the vector using infusion technology (Takara Clontech). The vector was subsequently injected into 1-cell stage embryos together with *in vitro* transcribed Ac mRNA to enable integration into the genome. Transcription factor binding sites were identified querying the Jaspas2014 database with the TFBSTools package (Tan and Lenhard, 2016), using a minimum score of 87%. The core sequence of enh15-podxl was:

TCTCTAAGTCCGAGATAAATATATGTGTGTGTGTGTGTGTGTGTATCGTTGGGAATGTGTG  
 ATTAGTAGTTGACAGGACACAGAACACTCTGCACCTCTTTATCAGAAGCCTGGTACCAAGCGGA  
 GTCCATCCCCAGAGAAAAACAGCAGCTCACATATATATATATGTGTGTGTGTGTGT

The core sequence of enh17-lox was:

TTACGCCCCATTTATGTGCAAGACCTAAAGTCATTGTATTCACTGGAGCTCGTCCAAATGTACGCC  
ATAGTTGGTGGTGAGACATGGCTTACTTCAACACAAAACCTATGCAGTGAGCCGTTACTGAGTCAA  
CTGAAGTATATGCAGCAGTAGTGGGCCTCTGCAAAGTGTTTTCATTTAGACAAGAATACAGACCT  
TTCCCTTCTAAAAACAGCACCAGGATTCAGATATCACTCAGATACAACATTCAAAGCACGATAAA  
CGCACAAGTAAGCCTTGCAAGTCTTTCTTTTAGCGAAGGAATTCAGTTCATTGCATCATGTCTGCTG  
TCTTTAGAATTCTTAAGAGGAACATAAACTCTCAAAAATGTTATCTCATAAGTTTCTGAAAAATTT  
AAACATGCTGATAATTGACTCACTATCAGGTCATA

The core sequence of enh20-eInb was:

ATCACGGACCCATATCGCTGTAGAACGAGCATGTGCTGAATAAGTTACATGCCATAGTCATGAGC  
AAATTATATATAACAACAGTGGAGCTTACAGTGTGAATCACTGAACAAACAATGCCGGGGAATTTA  
ATCAGAGCACAACGGAGCAAAGAAAAAAAAAAGTGGCGGGGGGGTGAAGCTGTCTGTCAGCAT  
TCTGGCCAAGGCCAAGTCTTTCTTTCTGGGCACATACCTGTGAGCTGTTACCATACATAGAATCAG  
CACTTATCACAGAAGAGGGATATAAATATTGA

## Chapter 3

### Microscopic analysis of epicardial heterogeneity in the zebrafish embryo

#### 3.1 Introduction

Of the many studies dealing with the epicardium, a small number have provided indications for the presence of distinct epicardial subpopulations. It was shown that the epicardial expression of markers such as *Tcf21* and *Tbx18* is heterogeneous and that the presence of these markers affects the lineage potential of embryonic EPDCs (Acharya et al., 2012; Bollini et al., 2014; Braitsch et al., 2012). However, these studies did not perform *in vivo* microscopic expression analysis of more than two epicardial marker genes and did not investigate epicardial heterogeneity over a developmental time course. Therefore, we currently have a limited understanding of exactly how the epicardial expression patterns of commonly used marker genes compare to each other and whether epicardial heterogeneity remains static throughout development or if it is a dynamic phenomenon. Furthermore, there is no *in vivo* data on epicardial heterogeneity following zebrafish heart injury. Analysis in the embryo might give indications as to the presence of epicardial heterogeneity in the adult heart following injury, as heart regeneration largely recapitulates heart development (Xin et al., 2013).

To address these issues, I microscopically compared the expression patterns of *tcf21*, *tbx18* and *wt1b* driven fluorophores in existing reporter lines and in a newly generated reporter line that for the first time enabled the simultaneous expression analysis of all three markers. Furthermore, I visualized and compared the endogenous expression of *tcf21*, *tbx18* and *wt1b*. I found that the embryonic zebrafish epicardium was indeed

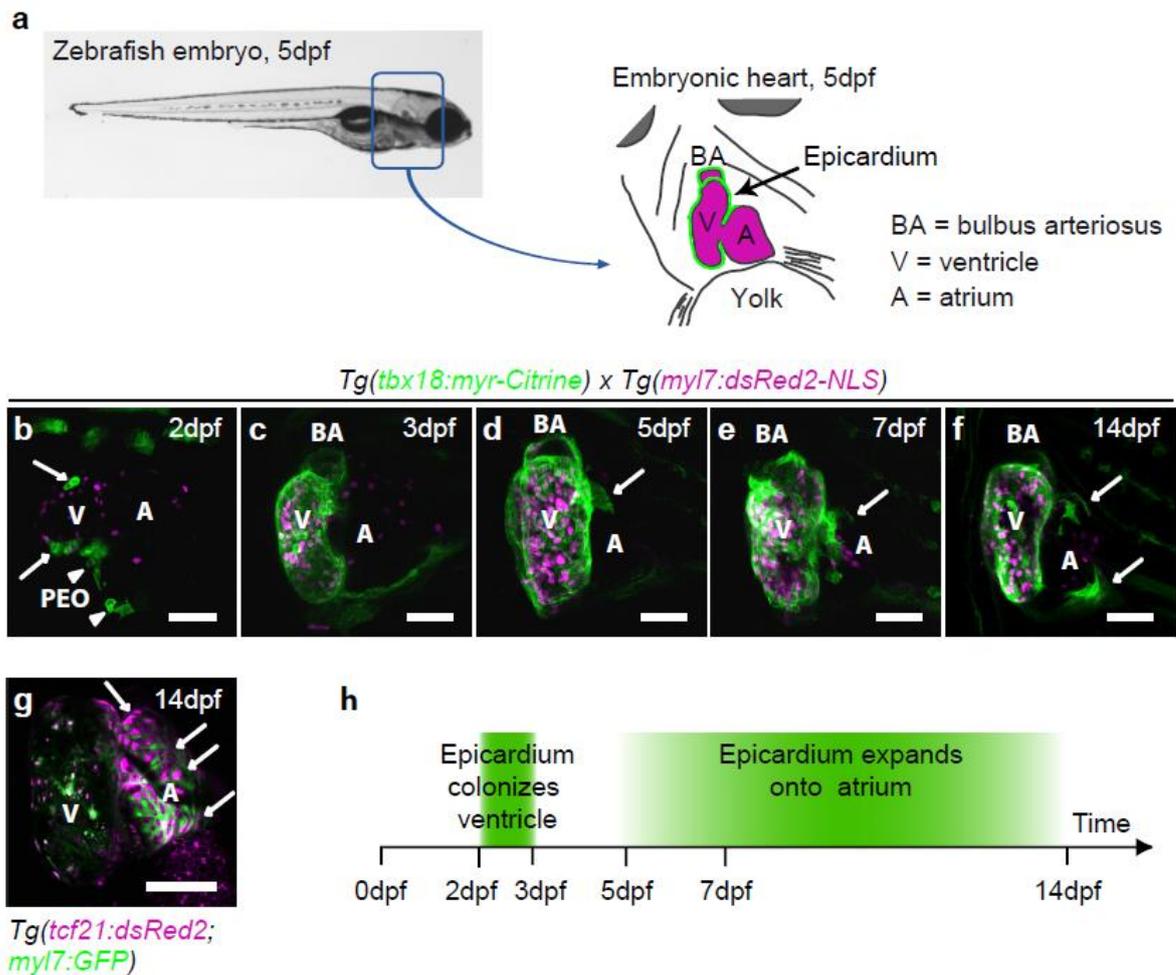
heterogeneous and that this heterogeneity developed dynamically between 3 days post fertilization (dpf) and 7dpf. Fluorophores driven by *tcf21*, *tbx18* and *wt1b* labelled multiple epicardial cell populations, including a prominent *wt1b*:H2B-Dendra2 single positive population covering the bulbus arteriosus. I also analyzed the epicardial response to embryonic cardiac injury, but did not identify changes in epicardial heterogeneity, as indicated by the lack of an increase in the number of epicardial cells expressing *wt1b*:GFP.

### **3.2 Early morphogenesis of the *tbx18* positive embryonic zebrafish epicardium matches previous findings obtained studying *wt1a***

Epicardial cells translocate from the proepicardial organ (PEO) to the surface of the heart muscle (Serluca, 2008). Subsequently, epicardial cells spread to form a continuous cell sheet enveloping the ventricle (schematically depicted in figure 3-1a). The study of epicardial sheet formation in the zebrafish embryo has been focused on cells expressing a *wt1a* driven reporter (Peralta et al., 2013). Here, epicardial cells colonize the heart muscle as single cells, detaching from the PEO and translocating to the ventricle via pericardiac fluid convections. To investigate if there is heterogeneity during the early morphogenesis of the epicardium, I asked if *tbx18*:myr-Citrine positive epicardial cells behaved the same way as *wt1a* positive epicardial cells. I crossed the newly generated line *Tg(tbx18:myr-Citrine)* to the existing line *Tg(myf7:dsRed2-NLS)* (Mably et al., 2003), which visualized the myocardium, and microscopically analyzed double positive embryos (Figure 3-1b-f). Epicardial cells expressing *tbx18*:myr-Citrine were first evident at 2dpf

and were located in the PEO and as single cells on the surface of the ventricle, consistent with the previous finding that proepicardial cells are released into the pericardial cavity before adhering to the myocardium (Peralta et al., 2013) (Figure 3-1b). The PEO appeared to consist of two distinct proepicardial cell clusters, one residing close to the atrio-ventricular boundary and the other at the venous pole of the heart, again matching the findings obtained studying *wt1a* expressing proepicardial cells. By 3dpf, a continuous epicardial sheet enveloped the ventricle (Figure 3-1c). By 5dpf, first epicardial cells were visible on the surface of the atrium, close to the atrio-ventricular boundary (Figure 3-1d). Until 14dpf, the coverage of *tbx18:myr-Citrine* positive epicardial cells on the atrium kept expanding, but *tbx18:myr-Citrine* positive cells never covered the entire atrial surface (Figure 3-1e,f). In contrast, epicardial cells labelled by *tcf21:dsRed2* completely covered the atrial surface at 14dpf (Figure 3-1g). Figure 3-1h summarizes the morphogenesis of the *tbx18:myr-Citrine* positive epicardium between 2dpf and 14dpf.

In summary, the translocation of *tbx18:myr-Citrine* positive epicardial cells onto the ventricle matched the results obtained by Peralta et al. (Peralta et al., 2013), suggesting that *tbx18* and *wt1a* label the same proepicardial cells or that the proepicardial cell populations labelled by *tbx18* and *wt1a* behave similarly during early epicardial morphogenesis. Later during epicardial morphogenesis however, epicardial cells covering the developing zebrafish atrium showed a widespread expression of *tcf21:dsRed2*, while they rarely expressed *tbx18:myr-Citrine*. The atrial epicardium thus provided a first indication of epicardial heterogeneity in the zebrafish embryo.



**Figure 3-1: Epicardial development in the zebrafish embryo between 2dpf and 14dpf.**

(a) The zebrafish epicardium envelops the developing heart muscle, which consists of a single atrium (A), a single ventricle (V) and a prominent outflow tract, the bulbus arteriosus (BA). (b-f) Microscopic analysis of epicardial development in *Tg(tbx18:myr-Citrine) x Tg(myI7:dsRed2-NLS)* embryos. (b) At 2dpf, *tbx18:myr-Citrine* positive epicardial cells were located in two cell clusters close to the heart (arrowheads). Single *tbx18:myr-Citrine* positive cells were also located on the surface of the ventricle (arrows). (c) At 3dpf, an epicardial sheet enveloped the ventricle. (d-f) Starting from 5dpf, epicardial cells could be seen on the atrium (arrow), with coverage increasing until 14dpf. (g) In *Tg(tcf21:dsRed2; myI7:GFP)* embryos, *tcf21:dsRed2* positive cells completely covered the

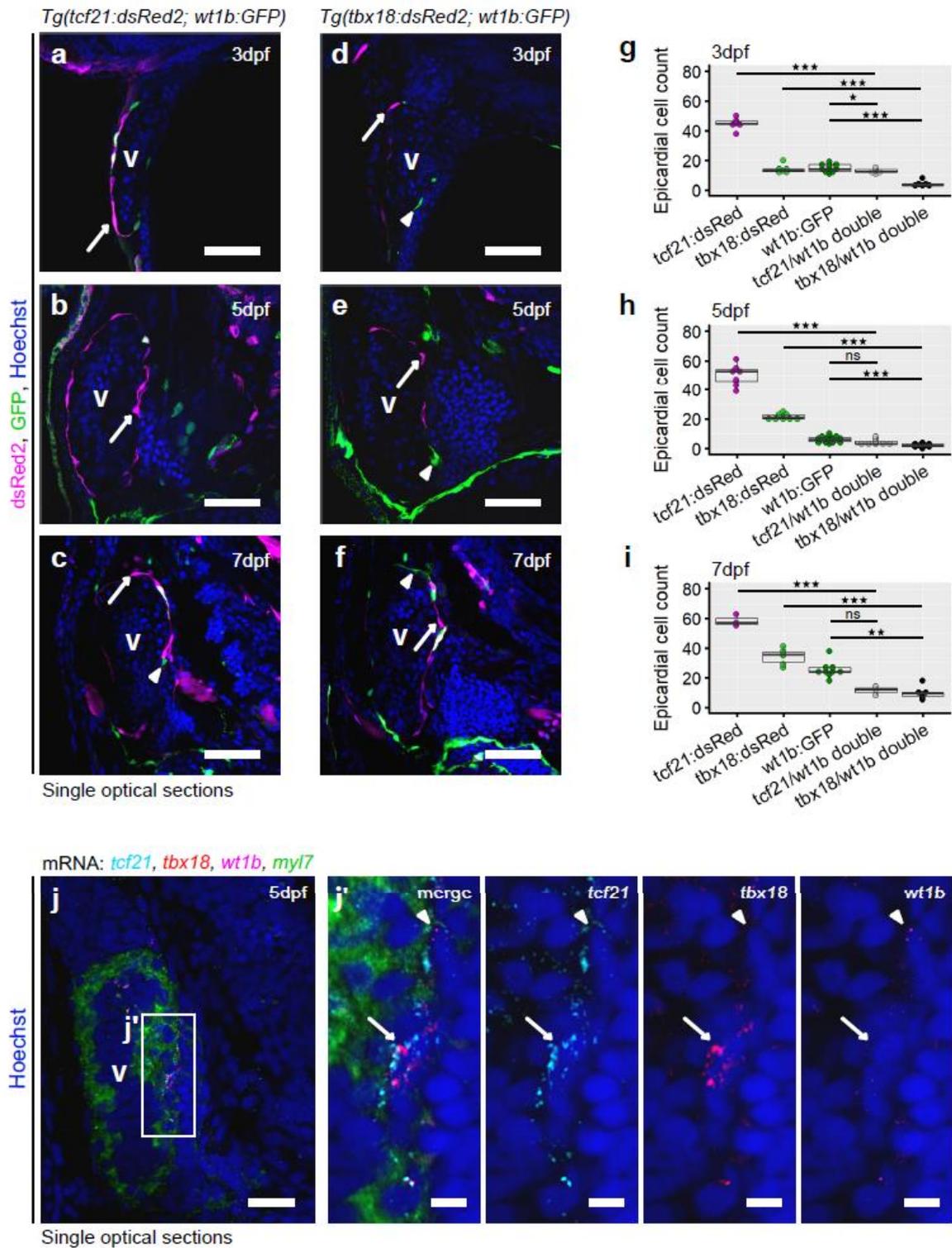
atrium at 14dpf (arrows). **(h)** Graphical representation of epicardial morphogenesis in the zebrafish embryo along a developmental time scale until 14dpf. Scale bars in b-g are 50µm.

### **3.3 The analysis of *Tg(tcf21:dsRed2; wt1b:GFP)* and *Tg(tbx18:dsRed2; wt1b:GFP)* reveals heterogeneity in the developing zebrafish epicardium**

To analyze epicardial heterogeneity in more depth, I investigated the epicardial fluorophore expression in *Tg(tcf21:dsRed2; wt1b:GFP)* and *Tg(tbx18:dsRed2; wt1b:GFP)* (Kikuchi et al., 2011a; Perner et al., 2007) embryos at 3dpf, 5dpf and 7dpf. Epicardial cells expressing *tcf21:dsRed2* were abundant at all three time points and a large fraction did not co-express *wt1b:GFP* (arrows in figure 3-2a-c). Epicardial cells expressing *wt1b:GFP* were sparse and most co-expressed *tcf21:dsRed2* (arrowhead in figure 3-2c). In contrast, most *wt1b:GFP* expressing epicardial cells did not co-express *tbx18:dsRed2* (arrowheads in figure 3-2d-f). At 5dpf and at 7dpf, there were more epicardial cells that expressed *tbx18:dsRed2* than epicardial cells that expressed *wt1b:GFP*. However, there were less epicardial cells that expressed *tbx18:dsRed2* than epicardial cells that expressed *tcf21:dsRed2*. A major fraction of *tbx18:dsRed2* expressing epicardial cells did not co-express *wt1b:GFP* (arrows in figure 3-2d-f). Quantification showed that indeed there was a significant number of epicardial cells expressing *tcf21:dsRed2* (3dpf  $p=4.889e-07$ ,  $n=6$ ; 5dpf  $p=4.204e-09$ ,  $n=9$ ; 7dpf  $p=8.329e-05$ ,  $n=3$ ) or *tbx18:dsRed2* (3dpf  $p=4.654e-05$ ,  $n=6$ ; 5dpf  $p=1.708e-15$ ,  $n=10$ ; 7dpf  $p=3.782e-06$ ,  $n=6$ ) that did not express *wt1b:GFP* (Figure 3-2g-i). Conversely, the number of *wt1b:GFP* expressing epicardial cells that did not express *tcf21:dsRed2* was only significant at 3dpf (3dpf  $p=0.004044$ ,  $n=6$ ; 5dpf  $p=0.0115$ ,  $n=9$ ;

7dpf  $p=0.008315$ ,  $n=3$ ) (Figure 3-2g). However, the number of *wt1b*:GFP expressing epicardial cells that did not express *tbx18*:dsRed2 was significant at all three time points (3dpf  $p=9.301e-06$ ,  $n=6$ ; 5dpf  $p=3.573e-05$ ,  $n=10$ ; 7dpf  $p=0.0001089$ ,  $n=6$ ) (Figure 3-2g-i). Generally, *tcf21*:dsRed2, *tbx18*:dsRed2 and *wt1b*:GFP expressing epicardial cell populations all increased in numbers from 3dpf until 7dpf. Only at 5dpf, epicardial cells expressing *wt1b*:GFP were decreased in numbers, as compared to 3dpf. This might point towards fluctuations in the expression of *wt1b*:GFP in epicardial cells, with a decreased expression between 3dpf and 5dpf.

To investigate the endogenous epicardial expression of *tcf21*, *tbx18* and *wt1b*, I performed hybridization chain reaction (HCR), an *in situ* staining technique used to study multiple gene expression patterns simultaneously (Choi et al., 2014a). I found that the expression patterns of *tcf21*, *tbx18* and *wt1b* in the epicardium overlapped only partially (Figure 3-2j). Expression of all three genes was evident at the atrio-ventricular boundary, however a large area in which *tcf21* and *tbx18* transcripts were present lacked *wt1b* transcripts (arrow in figure 3-2j'). Conversely, another area contained *tcf21* and *wt1b* transcripts, but lacked *tbx18* transcripts (arrowhead in figure 3-2j'). Taken together, these results demonstrate that the epicardial expression of *tcf21*, *tbx18* and *wt1b* in the developing zebrafish embryo is markedly heterogeneous, both in terms of a transgenic reporter read-out and at the level of endogenous gene expression.



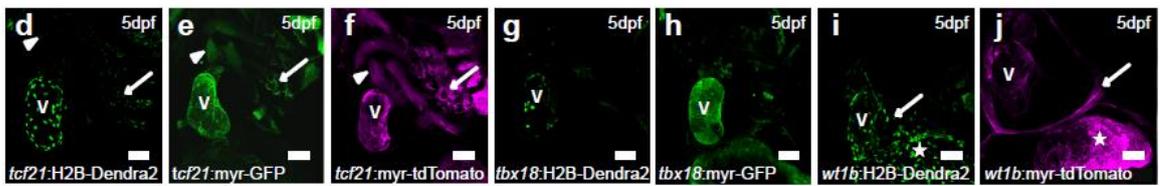
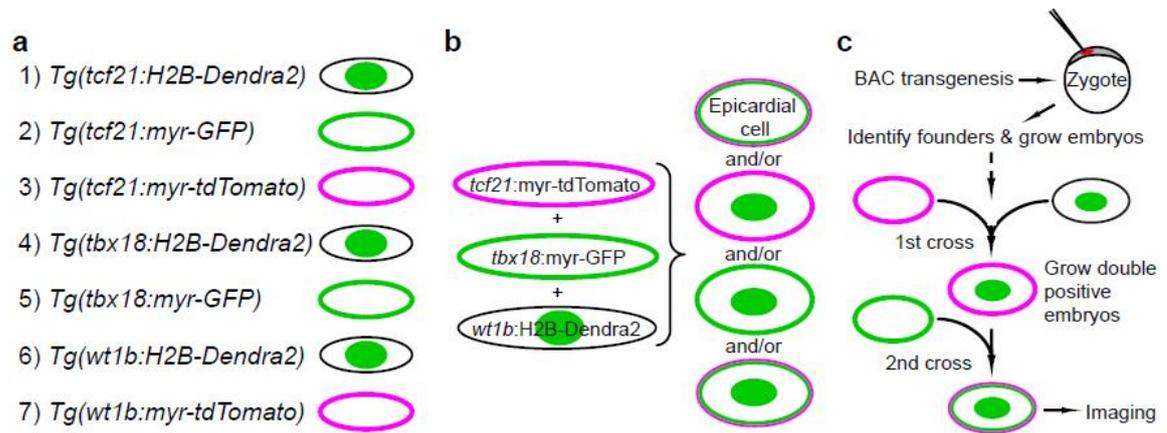
**Figure 3-2: The expression of *tcf21*, *tbx18* and *wt1b* in the developing zebrafish epicardium is heterogeneous.** (a-c) Microscopic analysis of *Tg(tcf21:dsRed2; wt1b:GFP)* hearts at 3dpf, 5dpf and 7dpf. At all three time points, *tcf21:dsRed2* positive, *wt1b:GFP* negative epicardial

cells were present (arrows), while *wt1b*:GFP positive, *tcf21*:dsRed2 negative epicardial cells were only seen at 7dpf (arrowhead). (d-f) Microscopic analysis of *Tg(tbx18:dsRed2; wt1b:GFP)* hearts at 3dpf, 5dpf and 7dpf. At all three time points, both *tbx18*:dsRed2 positive, *wt1b*:GFP negative epicardial cells (arrows) and *wt1b*:GFP positive, *tbx18*:dsRed2 negative epicardial cells (arrowheads) were present. (g-i) Quantitative analysis of the data shown in a-f. Plotted are raw counts of the total number of *tcf21*:dsRed2, *tbx18*:dsRed2 and *wt1b*:GFP positive epicardial cells as well as counts of *tcf21*:dsRed2/*wt1b*:GFP (*tcf21*/*wt1b* double) and *tbx18*:dsRed2/*wt1b*:GFP (*tbx18*/*wt1b* double) double positive cells. Boxplots show the median and upper/lower quartiles. (j) Visualization of endogenous *tcf21*, *tbx18*, *wt1b* and *myl7* transcripts at 5dpf using hybridization chain reaction. Enlarged images in j' show heterogeneous expression in the epicardial layer, such as a *tcf21*, *tbx18* positive area without *wt1b* expression (arrows) and a *tcf21*, *wt1b* positive area without *tbx18* expression (arrowheads). Scale bars in a-f are 50µm, scale bar in j is 20µm and scale bars in j' are 5µm. V = ventricle. Number of embryos analyzed in *Tg(tcf21:dsRed2; wt1b:GFP)*: 6 at 3dpf, 9 at 5dpf and 3 at 7dpf. Number of embryos analyzed in *Tg(tbx18:dsRed2; wt1b:GFP)*: 6 at 3dpf, 10 at 5dpf and 6 at 7dpf. Significance was calculated using one-way Bonferroni-corrected Student's *t*-test. \* =  $p < 0.005$ , \*\* =  $p < 0.001$ , \*\*\* =  $p < 0.0001$ .

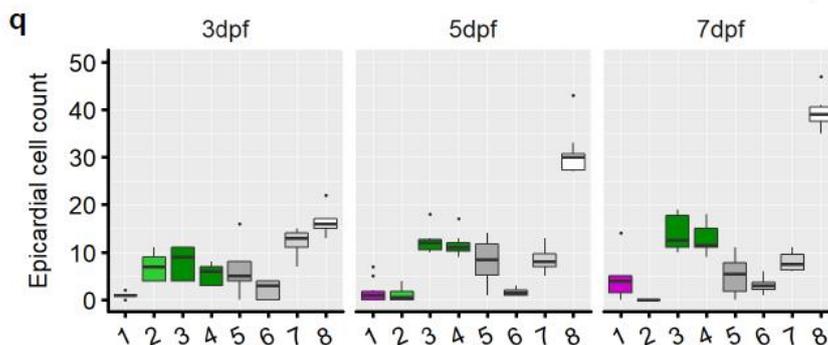
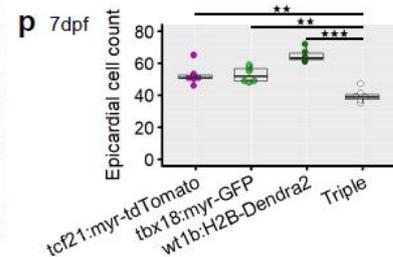
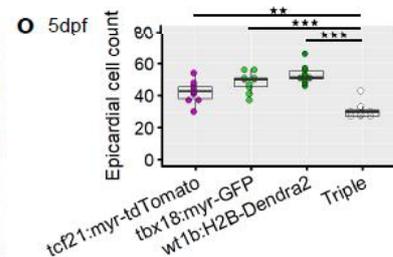
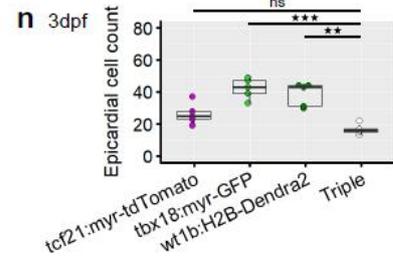
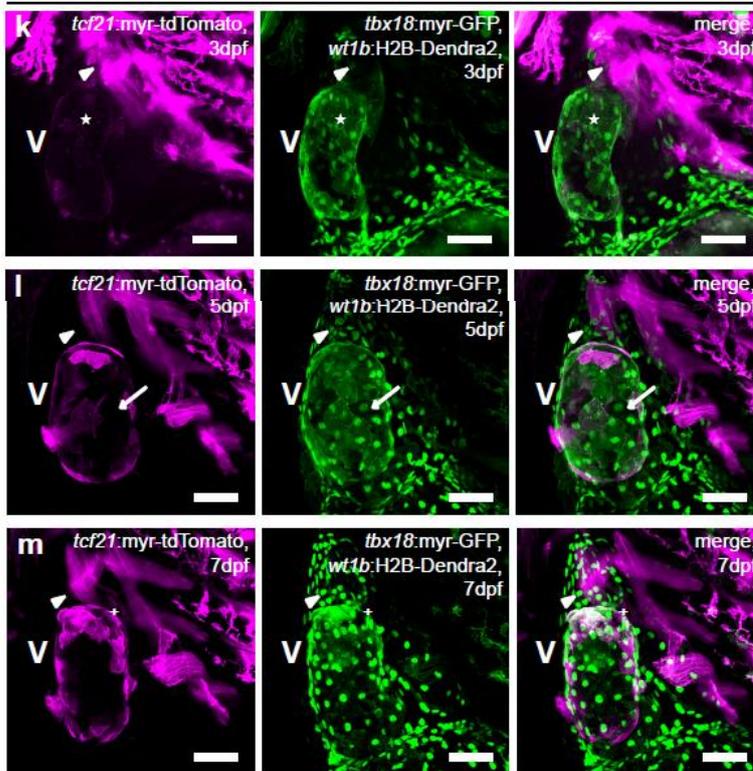
### **3.4 Novel transgenic reporter lines enable simultaneous visualization of the expression patterns of *tcf21*, *tbx18* and *wt1b***

The epicardial reporter lines described above did not allow me to compare the expression patterns of *tcf21* and *tbx18* simultaneously, nor was it possible to relate the expression of all three epicardial marker genes *tcf21*, *tbx18* and *wt1b*. To address this issue, I generated the transgenic zebrafish reporter lines *Tg(tcf21:H2B-Dendra2)*, *Tg(tcf21:myr-tdTomato)*, *Tg(tcf21:myr-GFP)*, *Tg(tbx18:H2B-Dendra2)*, *Tg(tbx18:myr-GFP)*, *Tg(wt1b:H2B-Dendra2)* and *Tg(wt1b:myr-tdTomato)* (Figure 3-3a). The design incorporated fluorophores with two non-overlapping fluorescence spectra. GFP and Dendra2 have absorption peaks in the blue spectrum and emission peaks in the green spectrum (Gurskaya et al., 2006; Heim et al., 1994). tdTomato has an absorption peak in the red spectrum and an emission peak close to the far-red spectrum (Shaner et al., 2004). Fluorescence from GFP and Dendra2 was made distinguishable by localizing GFP to the plasma membrane with a myristoylation (myr) tag (Zacharias, 2002) and Dendra2 to the nucleus with a Histone 2b (H2B) tag (Dempsey et al., 2012). tdTomato was combined with a myr tag too, as described previously (Trichas et al., 2008). Combining the three fluorophores, driven by *tcf21*, *tbx18* and *wt1b* regulatory elements, made it possible to quantify epicardial cell populations that expressed any combination of the three marker genes (Figure 3-3b). In two subsequent breeding cycles three single reporter lines were combined to generate the triple reporter line *Tg(tcf21:myr-tdTomato; tbx18:myr-GFP; wt1b:H2B-Dendra2)* (Figure 3-3c). Microscopic analysis confirmed that the fluorophore expression patterns within the single reporter lines matched *tcf21*, *tbx18* and *wt1b* driven fluorophore expression in existing transgenic reporter lines (Figure 3-3d-

j). In addition to the epicardium, *tcf21* driven fluorophores consistently labelled the branchial arches (arrows in figure 3-3d-f) and craniofacial muscles (arrowheads in figure 3-3d-f), comparable to *Tg(tcf21:dsRed2)*. Fluorophores driven by *tbx18* were consistently expressed in epicardium, head structures, fins and somites (Figure 3-3g-h and not shown), comparable to *Tg(tbx18:dsRed2)*. Fluorophores driven by *wt1b* were consistently expressed in epicardium, pericardial sac (arrows in figure 3-3i-j), liver (asterisks in figure 3-3i-j) and kidney. The newly generated reporter lines thus faithfully recapitulated the expression of *tcf21*, *tbx18* and *wt1b*. Importantly, *tbx18* and *wt1b* driven fluorescence was not detected within non-epicardial tissues in the ventricle, unlike what has been reported for *Tg(tbx18:dsRed2)* and *Tg(wt1b:GFP)* (Kikuchi et al., 2011a). This might be because the newly generated lines are based on bacterial artificial chromosomes (BACs), which are over 100kb in size and likely to contain a large number of regulatory elements for the respective gene (Bussmann and Schulte-Merker, 2011). *Tg(wt1b:GFP)* is a plasmid based line and thus might not faithfully recapitulate the endogenous expression of *wt1b*. *Tg(tbx18:dsRed2)* is BAC based, but the BAC is different from the one used in the newly generated *tbx18* reporter lines. These differences in the design of the reporter lines might explain the differences in the fluorophore expression.



$Tg(tcf21:myr-tdTomato; tbx18:myr-GFP; wt1b:H2B-Dendra2)$



**Figure 3-3: Simultaneous expression analysis of *tcf21*, *tbx18* and *wt1b* details epicardial heterogeneity and reveals a novel *wt1b*:H2B-Dendra2 single positive epicardial cell population.** (a) Novel zebrafish reporter lines to visualize the expression of *tcf21*, *tbx18* and *wt1b* simultaneously. Appended diagrams indicate the color of the respective fluorophore and whether it is located in the nucleus (H2B) or at the cell membrane (myr). (b) Design of a *tcf21*, *tbx18*, *wt1b* triple reporter line to visualize epicardial heterogeneity. Double and triple positive cells can be distinguished by fluorophore color and location. (c) Workflow to generate novel *tcf21*, *tbx18* and *wt1b* single and triple reporter lines. Two sequential crosses combined three *tcf21*, *tbx18* and *wt1b* single reporter lines into the triple reporter line. (d-j) Microscopic overview of the fluorophore expression patterns within the generated single reporter lines. Reporters for *tcf21* showed expression in epicardium, branchial arches (arrows in d-f) and craniofacial muscles (arrowheads in d-f). Reporters for *tbx18* showed expression in the epicardium (g-h). Reporters for *wt1b* showed expression in epicardium, pericardial sac (arrows in i-j) and liver (asterisks in i-j). (k-m) Microscopic analysis of epicardial heterogeneity in *Tg(tc21:myr-tdTomato; tbx18:myr-GFP; wt1b:H2B-Dendra2)* embryos at 3dpf, 5dpf and 7dpf. Epicardial cells located on the bulbus arteriosus expressing only *wt1b* are indicated by arrowheads, an epicardial cell located on the ventricle expressing only *wt1b* is indicated by an arrow in l. A *tbx18* single positive cell at 3dpf is indicated by an asterisk and a *tcf21* single positive cell at 7dpf is indicated by a cross (n-p) Quantitative analysis of the data shown in k-m. Plotted are raw counts of the total numbers of *tcf21:myr-tdTomato*, *tbx18:myr-GFP* and *wt1b:H2B-Dendra2* positive epicardial cells as well as counts of *tcf21:myr-tdTomato*, *tbx18:myr-GFP*, *wt1b:H2B-Dendra2* triple positive cells (Triple). (q) Detailed quantification of *tcf21:myr-tdTomato*, *tbx18:myr-GFP* and *wt1b:H2B-Dendra2* single,

double and triple positive cell populations. Boxplots show the median and upper/lower quartiles. Scale bars are 50µm. V = ventricle. Number of embryos analyzed: 5 at 3dpf, 10 at 5dpf and 6 at 7dpf. Significance was calculated using one-way Bonferroni-corrected Student's *t*-test. \* =  $p < 0.0083$ , \*\* =  $p < 0.0017$ , \*\*\* =  $p < 0.00017$ .

### **3.5 A novel *wt1b*:H2B-Dendra2 single positive epicardial cell population envelops the bulbus arteriosus**

I found *Tg(wt1b:myr-tdTomato)* to show weak fluorescence and thus combined *Tg(tcf21:myr-tdTomato)*, *Tg(tbx18:myr-GFP)* and *Tg(wt1b:H2B-Dendra2)* to generate *Tg(tcf21:myr-tdTomato; tbx18:myr-GFP; wt1b:H2B-Dendra2)*. Subsequently, I microscopically analyzed triple transgenic reporter embryos at 3dpf, 5dpf and 7dpf (Figure 3-3k-m). Interestingly, at all three time points I noticed epicardial cells covering the surface of the bulbus arteriosus that exclusively expressed *wt1b*:H2B-Dendra2 (arrowheads in figure 3-3k-m), in addition to rare *wt1b*:H2B-Dendra2 single positive epicardial cells located on the ventricular surface (arrows in figure 3-3l). The expression of *wt1b*:H2B-Dendra2 in epicardial cells located on the bulbus arteriosus became more prominent with age. In addition to *wt1b*:H2B-Dendra2 single positive epicardial cells, I noticed *tbx18*:myr-GFP single positive epicardial cells at 3dpf (asterisks in figure 3-3k) and *tcf21*:myr-tdTomato single positive epicardial cells at 7dpf (+ sign in figure 3-3m).

Quantification showed that at all three time points the total numbers of epicardial cells labelled by *tbx18*:myr-GFP and *wt1b*:H2B-Dendra2 were significantly higher than the number of *tcf21*:myr-tdTomato, *tbx18*:myr-GFP, *wt1b*:H2B-Dendra2 triple positive

epicardial cells (3dpf  $p(tbx18)=0.0001054$ ,  $p(wt1b)=0.0005462$ ,  $n=5$ ; 5dpf  $p(tbx18)=6.34e-07$ ,  $p(wt1b)=1.095e-08$ ,  $n=10$ ; 7dpf  $p(tbx18)=0.0002383$ ,  $p(wt1b)=4.538e-07$ ,  $n=6$ ) (figure 3-3n-p). The total number of *tcf21:myr-tdTomato* expressing epicardial cells was significantly higher than that of the triple positive cells at 5dpf and at 7dpf (3dpf  $p=0.01404$ ,  $n=5$ ; 5dpf  $p=0.0001775$ ,  $n=10$ ; 7dpf  $p=0.001412$ ,  $n=6$ ) (figure 3-3o,p).

The relative number of triple positive epicardial cells out of the total number of epicardial cells labelled by the different fluorophores increased between 3dpf and 7dpf. At 3dpf, the relative number of triple positive cells was around 50%, while at 7dpf it was around 67%. All labelled cell populations increased in number between 3dpf and 7dpf, which suggests some level of proliferation throughout the epicardium. The relative growth in the number of triple positive epicardial cells may be due to an elevated level of proliferation in these cells. Alternatively, new triple positive cells might emerge from single and double positive cell populations by activation of additional fluorophore expression. The fact that I noticed triple positive cells at 5dpf and 7dpf which resided in a neighborhood of non-triple positive cells suggests that the latter might occur to some degree. However, bromodeoxyuridine/ethynyldeoxyuridine (BrdU/EdU) labelling experiments (Chehrehasa et al., 2009) will be necessary to analyze epicardial proliferation levels in detail.

The *tcf21:myr-tdTomato* expressing epicardial cell population increased rapidly in number, particularly between 3dpf and 5dpf (figure 3-3q). Increased proliferation in the *tcf21:myr-tdTomato* expressing population might underlie this observation, but it might also be explained by a delay in fluorophore maturation. tdTomato structurally matures within 1 hour at 37°C (Shaner et al., 2004), but the presence of the myr tag, the resulting spatial localization close to the plasma membrane and the incubation of zebrafish

embryos at 29°C might lead to a prolonged maturation time of myr-tdTomato. A delay in the maturation of myr-tdTomato could also explain the fact that I observed low tdTomato fluorescence levels at 3dpf and a pronounced variation in fluorescence levels between epicardial cells at 5dpf and 7dpf (figure 3-3k-m). In a previous study (Trichas et al., 2008), the fluorescence levels of myr-tdTomato were rather uniform in multiple tissues, raising the possibility that the varying fluorescence levels I observed were not connected to the fluorophore directly, but to e.g. transcriptional interference by regulatory elements at the genomic locus the BAC integrated in. Also, myr-tdTomato fluorescence levels were uniform in the craniofacial muscles and the branchial arches, suggesting there might be an epicardium specific mechanism affecting *tcf21:myr-tdTomato*. This implies that the number of *tcf21:myr-tdTomato* expressing epicardial cells, particularly at 3dpf, might be higher than the numbers in figure 3-3n-p. However, myr tagged GFP showed readily detectable and uniform fluorescence levels from 3dpf, which indicates a potential myr tag induced prolonged maturation time would be specific to tdTomato, possibly because of its tandem repeat structure.

The very prominent decrease in *tbx18:myr-GFP* single positive cell numbers, particularly between 3dpf and 5dpf, might be attributed to an increase in *tcf21:myr-tdTomato* fluorescence in these cells. Indeed, the number of *tcf21:myr-tdTomato*, *tbx18:myr-GFP* double positive cells increased between 3dpf and 5dpf.

The quantification in figure 3-3q furthermore shows that at all time points analyzed, 80% to 90% of the *wt1b:H2B-Dendra2* single positive cells resided on the bulbus arteriosus, while most *wt1b:H2B-Dendra2* expressing epicardial cells on the ventricle also expressed *tcf21:myr-tdTomato* and/or *tbx18:myr-GFP*.

In summary, analysis of the novel epicardial reporter line *Tg(tcf21:myr-tdTomato; tbx18:myr-GFP; wt1b:H2B-Dendra2)* revealed heterogeneity in the epicardial expression of *tcf21*, *tbx18* and *wt1b* and revealed a *wt1b* single positive epicardial cell population that envelops the bulbus arteriosus.

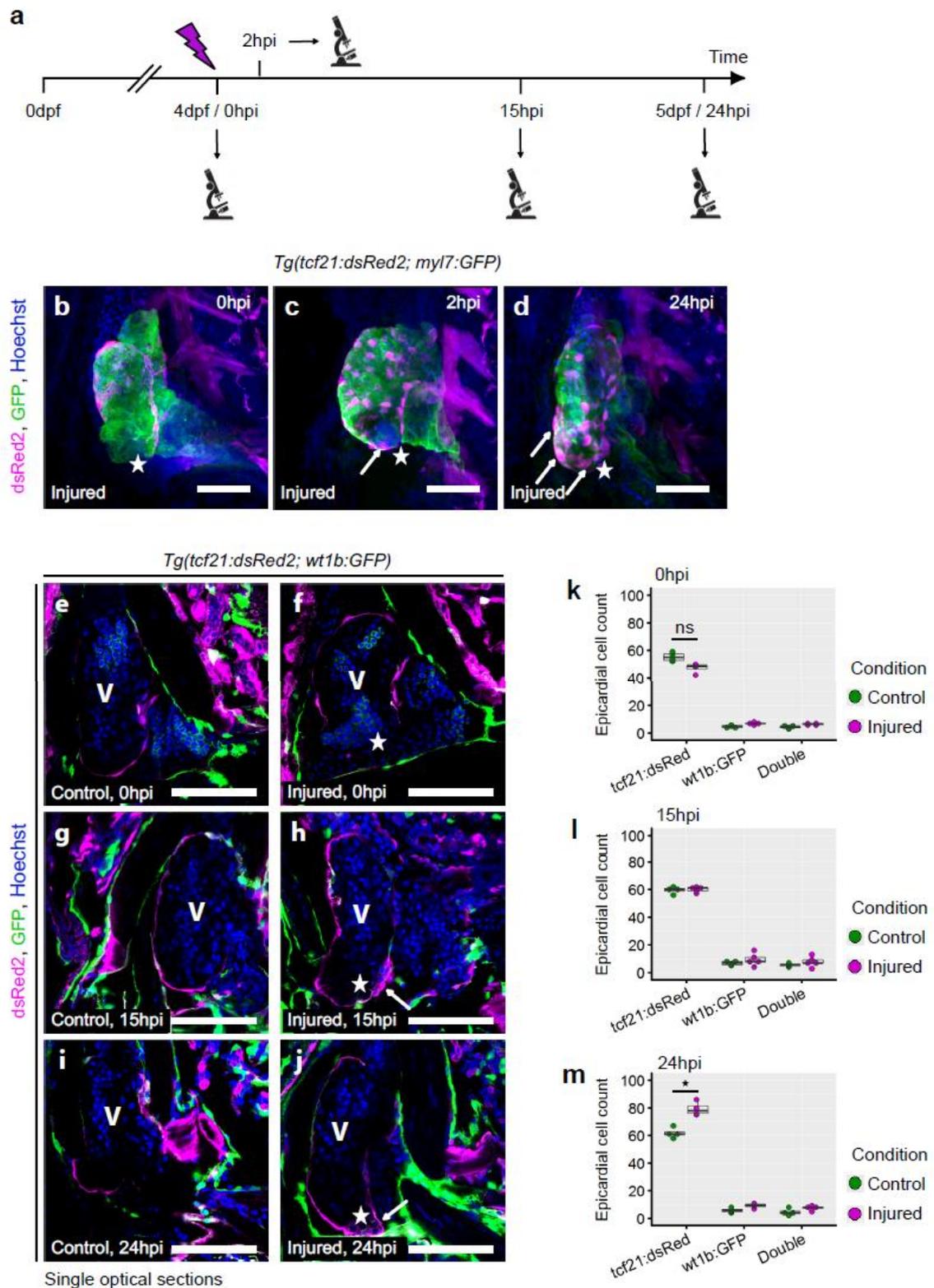
### **3.6 The response of the developing zebrafish epicardium to cardiac injury is not equivalent to epicardial reactivation following injury in the adult heart**

It is assumed that adult heart regeneration largely recapitulates the mechanisms acting during heart development (Alexander and Bruneau, 2010). However, it is unclear if the developing epicardium is able to respond to heart injury in a way that is comparable to the response of its adult counterpart. To characterize the response of the embryonic epicardium to cardiac injury, I ablated the ventricular apex in *Tg(tcf21:dsRed2; myl7:GFP)* embryos at 4dpf, mediated by a far-red laser pulse (Matrone et al., 2013) (workflow in figure 3-4a). By this time point, the epicardium has enveloped the entire ventricle (figure 3-1). Microscopic analysis immediately following injury (0 hours post injury (hpi)) as well as 2 hours (2hpi) and 24 hours (24hpi) later revealed a pronounced repair process that was mostly completed within 24 hours, matching previously published results on laser-mediated ablation of the embryonic zebrafish ventricle (Matrone et al., 2013) (figure 3-4b-d). At 0hpi, the impact of ablation could clearly be seen with considerable damage to both epicardium and myocardium (figure 3-4b). As soon as 2 hours later, epicardial cells were present within the site of injury (figure 3-4c) and at 24hpi the former site of injury

was completely covered by a large number of *tcf21:dsRed2* expressing epicardial cells (figure 3-4d). These observations might be explained by embryonic epicardial cells from non-injured parts of the ventricle migrating to colonize the site of injury.

I further tested whether the response to injury in the developing epicardium involves the activation of epicardial marker expression, as observed in the adult epicardium following injury (Itou et al., 2012a; Lepilina et al., 2006; Peralta et al., 2014; Schnabel et al., 2011). I injured *Tg(tcf21:dsRed2; wt1b:GFP)* embryos and analyzed them at 0hpi, 15hpi and 24hpi (figure 3-4e-j). While *tcf21:dsRed2* expressing cells formed a thickened epicardial layer at the site of injury from 15hpi, hardly any epicardial cells expressing *wt1b:GFP* were present in this area (figure 3-4h,j). Quantification of the entire epicardium showed that the *wt1b:GFP* positive epicardial cell population indeed did not significantly increase in number following injury, as compared to control embryos (0hpi  $p=0.01219$ ,  $n=4$ ; 15hpi  $p=0.2671$ ,  $n(ctr)=4$ ,  $n(inj)=5$ ; 24hpi  $p=0.0274$ ,  $n=4$ ) (figure 3-4k-m). Also, the *tcf21:dsRed2*, *wt1b:GFP* double positive cell population did not increase significantly in number (0hpi  $p=0.01038$ ,  $n=4$ ; 15hpi  $p=0.2434$ ,  $n(ctr)=4$ ,  $n(inj)=5$ ; 24hpi  $p=0.1033$ ,  $n=4$ ). Only the number of *tcf21:dsRed2* expressing epicardial cells was increased significantly at 24hpi (0hpi  $p=0.01705$ ,  $n=4$ ; 15hpi  $p=0.7576$ ,  $n(ctr)=4$ ,  $n(inj)=5$ ; 24hpi  $p=0.001763$ ,  $n=4$ ). In conclusion, the developing zebrafish epicardium did respond to injury by increasing the number of *tcf21:dsRed2* expressing cells, possibly through proliferation, which would need to be confirmed via BrdU or EdU staining. As dsRed2 matures in around 10 hours (Shaner et al., 2004), activation of *tcf21:dsRed2* in previously negative epicardial cells might also explain the increase in *tcf21:dsRed2* positive epicardial cell number at 15hpi and at 24hpi. However, the increase in epicardial cell number was confined to *tcf21:dsRed2* expressing cells, while the number of *wt1b:GFP* positive epicardial cells did

not increase. This argues against an activation of *wt1b* comparable to what is seen following adult heart injury and suggests that cardiac injury during embryonic development cannot be used as a proxy to study the epicardial response to adult heart injury. Instead, the developing uninjured epicardium, which expresses markers associated with adult epicardial activation such as *tbx18* and *wt1b*, already appears to be in a fully activated state that may be studied as a proxy for the regenerating adult epicardium, in addition to gaining fundamental insights into its role during heart development.



**Figure 3-4: The number of *wt1b:GFP* labelled epicardial cells does not increase in response to larval heart injury. (a) Workflow to analyze the epicardial response to laser mediated ablation of the embryonic heart. Embryos injured at 4dpf were processed for**

imaging at 0 hours post injury (hpi), 2hpi, 15hpi and 24hpi. **(b-d)** Microscopic analysis of larval heart regeneration in *Tg(tcf21:dsRed2; myl7:GFP)* embryos. The site of injury (asterisks) was soon repopulated by *tcf21:dsRed2* positive epicardial cells (arrows). **(e-j)** Microscopic analysis of epicardial heterogeneity following larval heart injury in *Tg(tcf21:dsRed2; wt1b:GFP)* embryos. Un-injured control hearts are shown in e, g and i, injured hearts are shown in f, h and j. *tcf21:dsRed2* single positive cells (arrows) accumulated at the site of injury (asterisks), while the numbers of *wt1b:GFP* positive cells remained low. **(k-m)** Quantitative analysis of the data shown in e-j. Plotted are raw counts of the total numbers of *tcf21:dsRed2* and *wt1b:GFP* positive epicardial cells as well as counts of *tcf21:dsRed2*, *wt1b:GFP* double positive cells (Double). Boxplots show the median and upper/lower quartiles. Scale bars are 50 $\mu$ m. V = ventricle. Number of injured embryos analyzed: 4 at 0hpi, 5 at 15hpi and 4 at 24hpi. Number of control embryos analyzed: 4 at 0hpi, 4 at 15hpi and 4 at 24hpi. Significance was calculated using two-way Bonferroni-corrected Student's *t*-test. \* =  $p < 0.0033$ .

### **3.7 Imaging the developing zebrafish epicardium in the living embryo through lightsheet microscopy and computational image alignment**

Several studies have used lightsheet microscopy to image the beating heart as it allows for rapid image acquisition and low photo damage, crucial to image moving structures over prolonged amounts of time (Arrenberg et al., 2010; Mickoleit et al., 2014; Scherz et al., 2008). However, the development of the epicardium has not been imaged using lightsheet microscopy, therefore I tested the Zeiss Z1 microscope using *Tg(tcf21:dsRed2;*

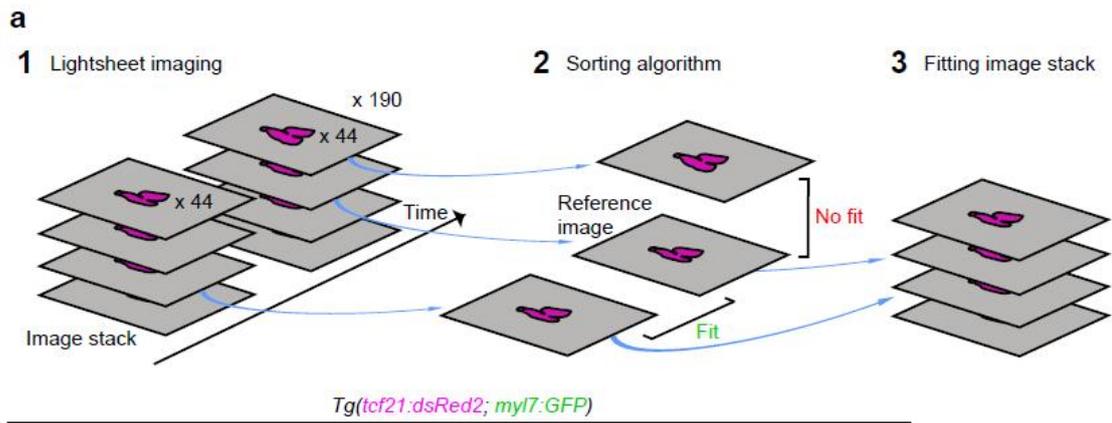
*myl7:GFP*) embryos at 4dpf and acquired 190 image stacks with 44 images each (workflow in figure 3-5a). The image stacks spanned the posterior half of the embryonic heart. Subsequently, I realigned the image stacks using a custom-made Matlab algorithm that followed the rationale of previously published code (Liebling et al., 2005, 2006; Staudt et al., 2014), yet deployed an independent coding strategy. First, a reference image in the middlemost plane of each stack was chosen automatically so that the ventricle was depicted at maximum dilation. Then, the code iterated bidirectionally through the remaining planes, starting adjacent to the reference plane and progressing towards the outer ends of the stack. In each plane, the code identified the image most similar to the reference image or the image identified in the previous plane. Finally, the identified images were reassembled into an image stack in which the contraction of the heart was compensated for.

In the original image stacks, the *tcf21:dsRed2* signal was neither aligned within itself nor with the *myl7:GFP* signal (figure 3-5b). I first tested the alignment algorithm on the more abundant, and therefore easier to work with, *myl7:GFP* signal. This worked well for the *myl7:GFP* signal, however the *tcf21:dsRed2* signal only showed slightly better alignment than in the non-aligned image stacks (figure 3-5c). Therefore, I optimized the algorithm to align *tcf21:dsRed2* and *myl7:GFP* signals separately and to combine the signals in the final image stack. This time, the *tcf21:dsRed2* signal indeed aligned to form a single layer outlining the ventricle (figure 3-5d). However, *tcf21:dsRed2* and *myl7:GFP* signals still did not align to each other, even though the algorithm contained a cross-channel alignment procedure to choose reference images matching each other.

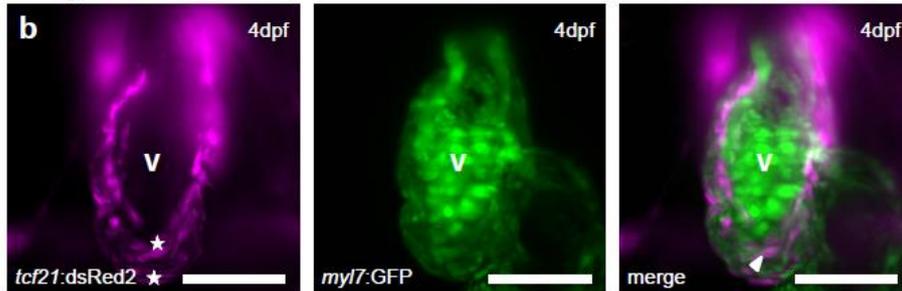
Finally, I tested if it was possible to apply lightsheet imaging to the newly generated reporter lines described in figure 3-3 and imaged *Tg(tcf21:myr-tdTomato; wt1b:H2B-*

*Dendra2*) embryos at 5dpf. However, I found that the fluorescence levels of both fluorophores were too weak for lightsheet microscopy with the Z1 (figure 3-5e).

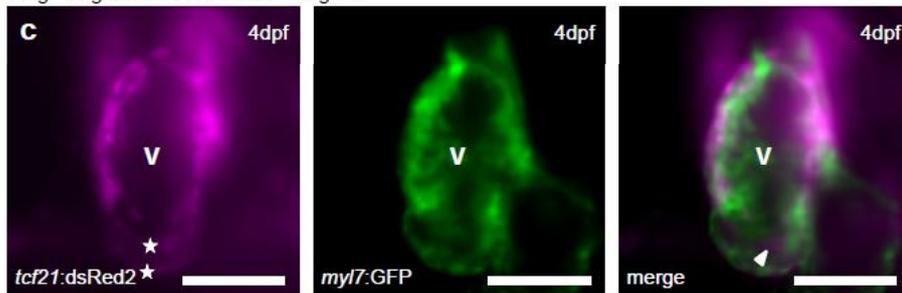
In summary, it is feasible to use lightsheet microscopy to image the epicardium in the living embryo. However, the fluorescent signal within the imaged epicardial cells needs to be strong. This was an issue for the newly generated reporter line *Tg(tcf21:myr-tdTomato; wt1b:H2B-Dendra2)*, which might be because mostly there is only a single integration event generating a BAC based line, while there often are multiple integration events in plasmid based lines (Bussmann and Schulte-Merker, 2011). *Tg(tcf21:dsRed2)* is also BAC based, however fluorescence in this line is present in the entire cell and consequently the signal is stronger, while in *Tg(tcf21:myr-tdTomato)* it is confined to the plasma membrane. Therefore, one needs to test whether a reporter line generates suitable fluorescence levels for lightsheet imaging. Also, further optimization of the setup described here is necessary to guarantee proper alignment of multiple fluorescent signals.



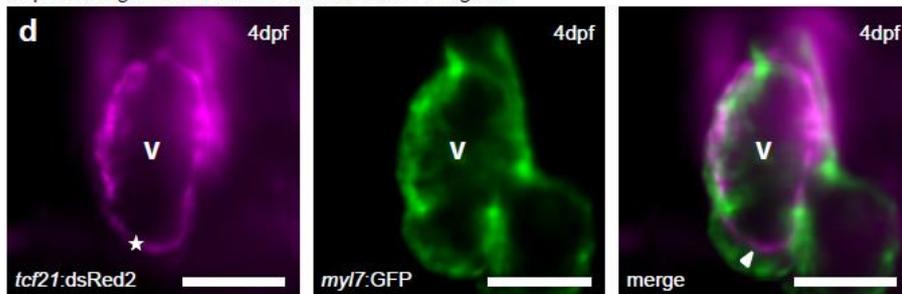
Not aligned



Single alignment based on GFP signal

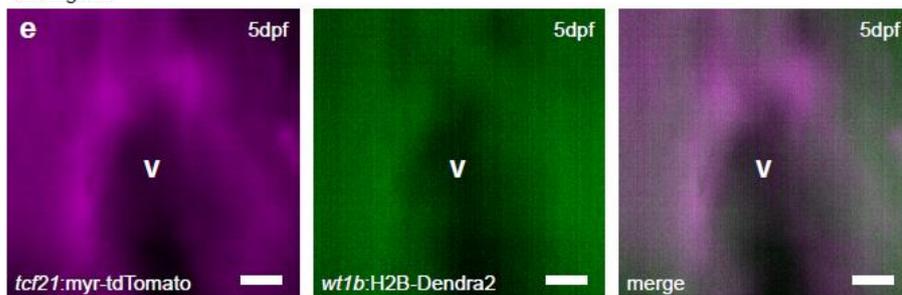


Separate alignments based on dsRed and GFP signals



*Tg(tcf21:myr-tdTomato; wt1b:H2B-Dendra2)*

Not aligned



**Figure 3-5: Live imaging of the developing zebrafish epicardium using lightsheet microscopy and a custom-made alignment algorithm.** (a) Workflow to image the epicardium in the living embryo. 190 image stacks of 44 image each were taken using the Zeiss Z1 lightsheet microscope, images fitting each other were identified using a custom-made alignment algorithm and combined to generate an image stack that was corrected for the movement of the beating heart. (b-d) Analysis using *Tg(tcf21:dsRed2; myl7:GFP)* embryos. (b) In the non-aligned image stack, the *tcf21:dsRed2* signal formed multiple layers (asterisks) and did not align with the *myl7:GFP* signal (arrowhead). (c) When images were aligned only using the *myl7:GFP* signal, the *tcf21:dsRed2* signal still formed multiple layers (asterisks) and did not align with the *myl7:GFP* signal (arrowhead). (d) When images were aligned using both the *tcf21:dsRed2* and the *myl7:GFP* signal, the *tcf21:dsRed2* signal formed a single layer (asterisk), but still did not align with the *myl7:GFP* signal (arrowhead). (e) Non-aligned image stack using *Tg(tcf21:myr-tdTomato; wt1b:H2B-Dendra2)* embryos. The fluorescent signals were too weak for proper alignment. Scale bars in b-c are 50 $\mu$ m, scale bars in d are 20 $\mu$ m. V = ventricle.

### 3.8 Summary

Taken together, the data presented in this chapter suggest that there are several epicardial cell populations in the embryonic zebrafish heart, as judged by the differential expression of *tcf21*, *tbx18* and *wt1b* both in transgenic reporter lines and on the endogenous gene expression level. I analyzed the existing reporter lines *Tg(tcf21:dsRed2; wt1b:GFP)* and *Tg(tbx18:dsRed2; wt1b:GFP)*, finding that the number of *wt1b:GFP* expressing epicardial cells was much smaller than the numbers of *tcf21:dsRed2* or

*tbx18:dsRed2* expressing epicardial cells. Also, the expression patterns of the different fluorophores overlapped only partially.

Additionally, I generated the BAC based triple reporter line *Tg(tcf21:myr-tdTomato; tbx18:myr-GFP; wt1b:H2B-Dendra2)* that for the first time allowed simultaneous expression analysis of *tcf21*, *tbx18* and *wt1b*. Microscopic analysis of this line showed that a major fraction of epicardial cells was triple positive for *tcf21*, *tbx18* and *wt1b*, but it also revealed the presence of single and double positive epicardial cell populations. This expanded on the results obtained analyzing the existing reporter lines, however there were discrepancies between existing and newly generated reporters. For example, *wt1b:GFP* labelled far less epicardial cells than *wt1b:H2B-Dendra2*. Also, *wt1b:GFP* was not expressed in epicardial cells that covered the BA and strongly expressed *wt1b:H2B-Dendra2*. These differences might reflect the fact that *Tg(wt1b:GFP)* is plasmid based, while *Tg(wt1b:H2B-Dendra2)* is BAC based and thus likely to contain more regulatory elements controlling the *wt1b* locus. *Tg(wt1b:H2B-Dendra2)* should therefore be superior to *Tg(wt1b:GFP)* in recapitulating the endogenous expression pattern of *Wt1b*. The discrepancies between existing and newly generated reporter lines are further discussed in Chapter 6.

I furthermore found that the number of *tcf21:dsRed2* positive epicardial cells increased following cardiac injury in the embryo, but that was not the case for *wt1b:GFP* expressing epicardial cells. As discussed above, this suggests that embryonic heart injury does not generate the same epicardial response as adult heart injury.

## Chapter 4

### Single cell transcriptomics identify three distinct cell populations in the embryonic zebrafish epicardium that are likely to have specific functions

#### 4.1 Introduction

Recent studies have characterized the cellular composition of the developing mouse heart at single cell resolution (DeLaughter et al., 2016; Li et al., 2016). Single cell transcriptomics have further been applied to study the adult zebrafish epicardium (Cao et al., 2015). The results presented in these studies demonstrate the unprecedented capacity of single cell transcriptomics to identify distinct cell populations within the heart. After having established that the embryonic zebrafish epicardium is heterogeneous, I therefore applied single cell RNA sequencing to uncover sub-populations of epicardial cells and found that there are three distinct epicardial cell clusters present in the data set I generated. These epicardial sub-populations expressed specific marker genes, such as *adrenomedullin a*, *myosin light chain kinase a* and *claudin 11a*. Importantly, transcripts encoding *tcf21*, *tbx18* and *wt1b* were not present in all three sub-populations, but their expression was mostly confined to one (*wt1b*) or two (*tcf21* and *tbx18*) of the epicardial cell clusters. The differential expression of Tcf21, Tbx18 and Wt1b might regulate the fate of cells within the epicardial cell clusters, similar to what has been shown previously (Braitsch et al., 2012).

## 4.2 Quality metrics of the single cell data set

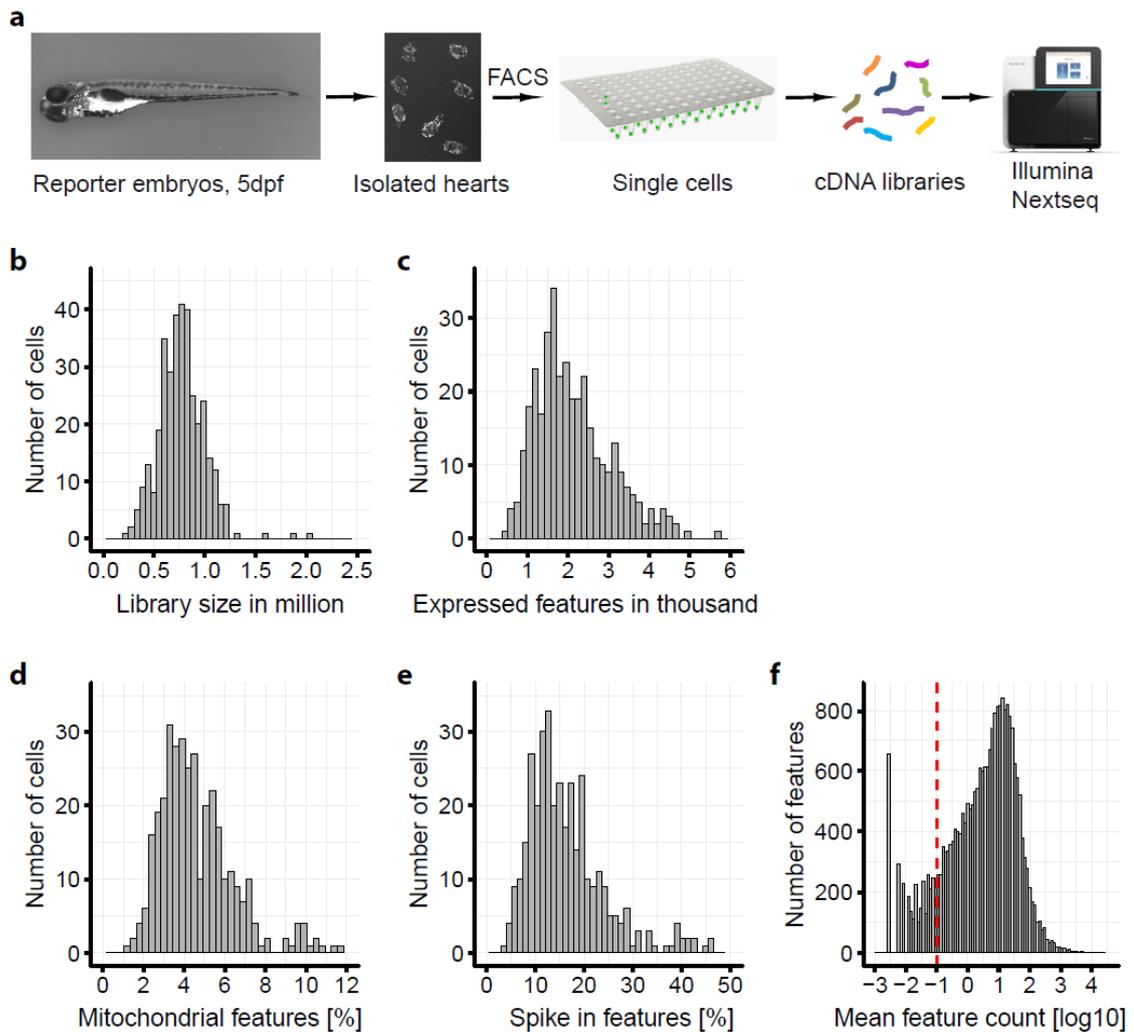
To isolate and amplify single cell cDNA I used the Smartseq2 technology (Picelli et al., 2014), which allows for a higher sequencing depth than the more recent droplet-based single cell processing methods do (Ziegenhain et al., 2017). However, it is not readily feasible to process more than a few hundred cells using Smartseq2 due to a cost factor of around 15 GBP per processed cell. I therefore used fluorescence-activated cell sorting (FACS) to specifically purify cells from cardiac tissues of interest, as determined by reporter fluorescence (workflow in figure 4-1a). As described in chapter 3, both cardiac and non-cardiac expression domains of *Tcf21*, *Tbx18* and *Wt1b* were fluorescently labelled in the respective reporter lines. To exclude non-cardiac fluorescence from FACS, I thus isolated hearts (mostly ventricle and BA, while atrial tissue was lost) from 5dpf embryos of the newly generated lines *Tg(tcf21:H2B-Dendra2)*, *Tg(tbx18:myr-GFP)* and *Tg(wt1b:H2B-Dendra2)* (figure 3-3). Subsequently, I FACS-purified fluorescent cells from the collected hearts and generated single cell cDNA libraries to be sequenced on an Illumina Nextseq machine. To obtain non-epicardial reference samples, I additionally generated libraries from *Tg(myf7:GFP)* hearts and from *Tg(kdrl:GFP; gata1a:dsRed)* hearts. In an attempt to include epicardial cells that did not possess transcripts encoding *tcf21*, *tbx18* or *wt1b* I furthermore processed non-fluorescent cells from *Tg(tcf21:dsRed2; myf7:GFP) x Tg(kdrl:GFP; gata1a:dsRed)* hearts as well as cells from wildtype hearts.

Following sequencing of the single cell libraries, I first performed a quality control step using the scater package (McCarthy et al., 2016), removing libraries of low complexity and insufficient size. During quality control, 108 out of 460 libraries were excluded from further analysis. The sizes of the remaining 352 single cell libraries were normally

distributed on a range between 250,000 and 1,250,000 read counts, with a mean library size of 774,000 read counts (figure 4-1b). The number of genes expressed in the libraries that passed quality control ranged from 500 to 5,000, with a mean complexity of 2,100 expressed genes (figure 4-1c). The proportion of reads in these libraries that mapped to mitochondrial genes ranged from 1% to 12%, with a mean of 4.6% (figure 4-1d). Finally, the proportion of reads in these libraries that mapped to synthetic spike in features ranged from 3% to 45%, with a mean of 16.4% (figure 4-1e). These quality metrics are within the range of those obtained in previous studies (Lun et al., 2016). To focus the analysis on genes with actual expression in the data set, only genes with an average expression level of more than 0.1 counts were included in the downstream analysis (figure 4-1f). This retained 21,000 of 32,000 genes with a mean average count of 37, among them 37 mitochondrial genes and 84 spike in features that were subsequently excluded.

As detailed in table 4-1, the proportion of libraries retained after quality control was greater than 90% for most lines that I isolated cells from. However, only 60% of libraries generated from *Tg(tcf21:H2B-Dendra2)* cells passed the quality control. This can be explained by the fact that about 70% of these cells were processed in a pilot experiment that did not include an additional early quality control step directly following cDNA preparation. Only 50% of these pilot libraries passed the final quality control step following sequencing. In contrast, over 90% of the *Tg(tcf21:H2B-Dendra2)* libraries that had been screened for good cDNA quality also passed the quality control step following sequencing. This underlines that the quality of RNA sequencing data is heavily dependent on efficient cDNA preparation. In addition, only 57% of the libraries prepared from *Tg(myf7:GFP)* cells passed the quality control step following sequencing, even though

these had been screened for good cDNA quality. I furthermore found that cardiomyocytes rarely yielded high quality cDNA to begin with. The most likely reason is that the dissociation protocol I used to prepare samples for FACS was not optimized for cardiomyocytes. Hence, cardiomyocytes were not in a healthy condition when being FACS-purified even if they were not labelled by 7-AAD, a cell viability stain that I used to label and exclude dead cells. Dissociation protocols have been optimized to isolate cardiomyocytes (Sander et al., 2013) and using these to process *Tg(myI7:GFP)* hearts before FACS should improve cDNA and library quality.



**Figure 4-1: Single cell RNA sequencing to study cell populations within the developing zebrafish epicardium. (a)** Workflow to generate single cell cDNA libraries from FACS-purified cardiac reporter derived cells at 5dpf. **(b)** Library size distribution within the single cell data set following quality control. **(c)** Distribution of the number of expressed genes per library following quality control. **(d)** Distribution of the proportion of reads mapping to mitochondrial genes following quality control. **(e)** Distribution of the proportion of reads mapping to synthetic spike in features following quality control. **(f)** Distribution of the average gene expression levels in the data set. Genes with an average

expression level below 0.1 read counts per cell (dashed red line) were excluded from further analysis.

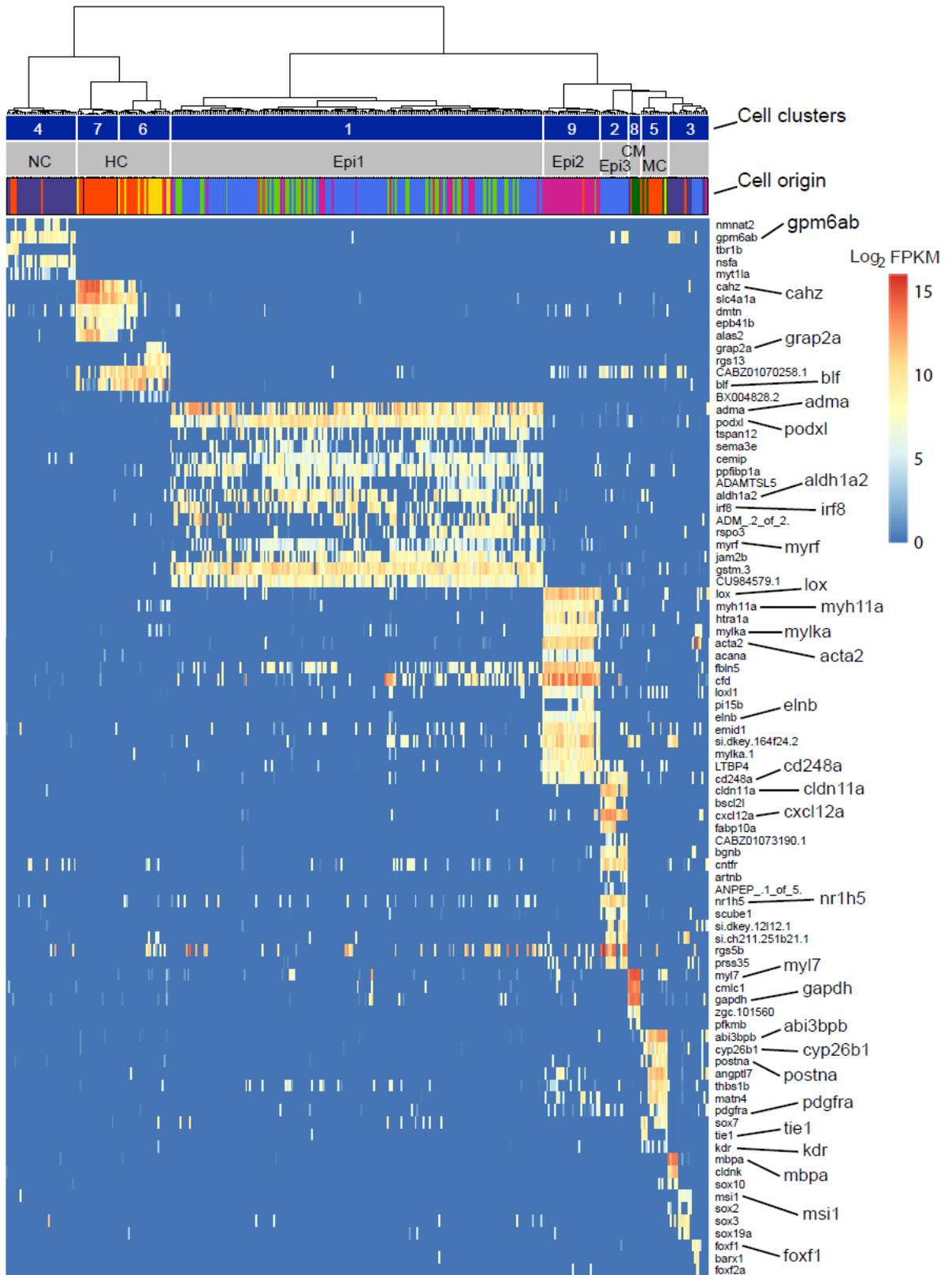
Line sorted from	Fluorescent cells sorted	Number of cells sequenced	Number of cells after QC	Proportion passed [%]
Wildtype	No	50	48	96
<i>Tg(tc21:H2B-Dendra2)</i>	Yes	228	137	60
<i>Tg(tbx18:myr-GFP)</i>	Yes	52	51	98
<i>Tg(wt1b:H2B-Dendra2)</i>	Yes	50	46	92
<i>Tg(my17:GFP)</i>	Yes	14	8	57
<i>Tg(kdrl:GFP)</i>	Yes	15	15	100
<i>Tg(gata1a:dsRed)</i>	Yes	2	2	100
<i>Tg(tc21:dsRed2; my17:GFP) x Tg( kdrl:GFP; gata1a:dsRed)</i>	No	46	45	98

**Table 4-1: Composition of the single cell data set before and after quality control (QC).**

### 4.3 Distinct epicardial clustering within the single cell data set

To identify the cell populations contained within the single cell data set, I first determined genes that were highly variable in their expression level using the pagoda package (Fan et al., 2016). Subsequently, the highly variable genes were grouped into gene sets based on the similarity of their expression patterns. The cells in the data set were then clustered according to the expression of the identified sets of highly variable genes (phylogenetic tree in figure 4-2). I focused my analysis on the nine most significant

cell clusters, as these turned out to delineate the major cardiac cell types that were contained within the data set (cell clusters are indicated as white numbers in figure 4-2). Interestingly, the large cell cluster 1 as well as the smaller cell clusters 9 and 2 were all epicardial in origin, as they were almost exclusively comprised of cells isolated from *Tg(tcf21:H2B-Dendra2)*, *Tg(tbx18:myr-GFP)* and *Tg(wt1b:H2B-Dendra2)* hearts. I therefore labelled cell clusters 1, 9 and 2 as epicardial clusters 1, 2 and 3, respectively. The presence of three epicardial cell clusters suggested epicardial cells to be considerably heterogeneous on a whole-transcriptome level, with the phylogenetic bifurcation event that separated epicardial cluster 1 from epicardial clusters 2 and 3 being the second most significant in the entire data set. Epicardial clusters 2 and 3 were more closely related and were separated by the fifth most significant bifurcation event. Furthermore, the epicardial clustering within the single cell data set appeared to be linked to whether the cells originated from *Tg(tcf21:H2B-Dendra2)*, *Tg(tbx18:myr-GFP)* or *Tg(wt1b:H2B-Dendra2)* embryos. While cells within epicardial cluster 1 originated from all three lines, cells within epicardial cluster 2 almost exclusively originated from *Tg(tbx18:myr-GFP)*, with only a single cell originating from *Tg(wt1b:H2B-Dendra2)* and none from *Tg(tcf21:H2B-Dendra2)*. In contrast, all cells contained within epicardial cluster 3 originated from *Tg(tcf21:H2B-Dendra2)*. This suggests that the heterogeneity in the epicardial expression of Tcf21, Tbx18 and Wt1b, as described in chapter 3, is not confined to these factors but translates to the whole-transcriptome.



- Cell origin:
- Wildtype
  - Tg(tcf21:H2B-Dendra2)* (+)
  - Tg(tbx18:myr-GFP)* (+)
  - Tg(wt1b:H2B-Dendra2)* (+)
  - Tg(myf7:GFP)* (+)
  - Tg(kdr1:GFP)* (+)
  - Tg(gata1a:dsRed)* (+)
  - Tg(tcf21:dsRed2; myf7:GFP) x Tg(kdr1:GFP; gata1a:dsRed)* (-)

**Figure 4-2: Novel marker genes identify distinct epicardial cell populations within the single cell data set.** Heatmap showing the expression of the most differentially expressed genes (rows) within the clustered single cell data set (columns). Expression values are FPKM, red color indicates high expression. The phylogenetic tree at the top summarizes the transcriptomic relationships of the single cells within the data set. The significance of each cell cluster is indicated by the height of the respective bifurcation event. White numbers above the heatmap indicate the nine most significant cell clusters. The annotation underneath indicates the cell type each cell cluster contains, as well as the color-coded origin of each single cell. The key to the color code is located at the bottom of the heatmap and indicates the reporter line each cell was isolated from and if it was a fluorescent cell (+) or a non-fluorescent cell (-). NC = neural cells, HC = haematopoietic cells, Epi = epicardial cell cluster, CM = cardiomyocytes, MC = mesenchymal cells, FPKM = fragments per kilobase of transcript per million transcripts.

#### **4.4 Epicardial cell cluster 1 is enriched in transcripts encoding the signaling peptide *Adma*, as well as in transcripts encoding the adhesion molecules *Podxl* and *Jam2b* and the epicardial marker *Aldh1a2***

Differential gene expression analysis between each of the three epicardial cell clusters and the entire data set identified several genes that were enriched within epicardial cells. Importantly, most of these genes were specific to cells in one epicardial cluster and showed little or no expression in cells in the remaining two. The most enriched gene in cells within epicardial cell cluster 1 was *adrenomedullin a (adma)*. In humans, elevated

plasma levels of ADM are associated with a variety of cardiovascular disease states (Eto et al., 1999). Adrenomedullin signals through the Calcitonin Receptor-Like Receptor (Calcrl), a G-protein coupled receptor. For binding and signaling to occur, Calcrl needs to associate to the receptor activity-modifying proteins Ramp2 or Ramp3 (McLatchie et al., 1998). In the mouse, loss of Ramp3 does not affect embryogenesis, however ablation of Adm, Calcrl or Ramp2 all lead to midgestational lethality due to interstitial edema and cardiovascular defects, such as reduced myocardial proliferation (Caron and Smithies, 2001; Dackor et al., 2006; Ichikawa-Shindo et al., 2008). Additionally, Calcrl and Ramp2 show strong expression in the vasculature and Tie2-Cre mediated endothelial-specific knockout of Calcrl recapitulates the phenotype of the global knockouts (Fritz-Six et al., 2008). Adm mediated signaling was also found to stimulate proliferation and migration of lymphatic endothelial cells (Jin et al., 2008) as well as to decrease lymphatic permeability (Dunworth et al., 2008). In the zebrafish, Calcrl is present as two paralogues (Calcrla and Calcrlb). However, only Calcrla is expressed during embryonic development, in vasculature and heart tube (Nicoli et al., 2008). Here, morpholino mediated knockdown of Calcrla was found to perturb arterial differentiation from angioblasts expressing the Vascular Endothelial Growth Factor (Vegf) receptor Kdr, associated with decreased levels of Vegfa. Similar results were obtained in murine embryonic stem cell cultures (Yurugi-Kobayashi et al., 2006). Thus, Adrenomedullin-mediated signaling plays conserved roles during vascular development, however its role during heart development is unknown.

Cells in epicardial cell cluster 1 also expressed cell adhesion molecules such as Podocalyxin Like (Podxl) and Junctional Adhesion Molecule 2b (Jam2b). Podxl is a member of the CD34 family of type I transmembrane glycoproteins and a close relative of Podocalyxin (Podx), which was originally identified as a podocyte marker (Kerjaschki et

al., 1984). However, Podx and Podxl are also expressed in hematopoietic and endothelial cells (Kershaw, D. B., Beck, S.G., Wharram, B.L.Wiggins, J.E., Goyal, M., Thomas, P.E., Wiggins, 1997; Kershaw et al., 1997; Sasseti et al., 1998). In the zebrafish embryo, Podxl is expressed in the heart tube as well as in the developing kidney (ZFIN). In the endothelium, Podxl acts as a pro-adhesive factor (Debruin et al., 2014), however in podocytes it rather functions as a regulated blocker of adhesion (Takeda et al., 2000). Furthermore, Podxl has been found necessary for Tgfb induced epithelial to mesenchymal transition (Meng et al., 2011) and is a marker of poor prognosis in multiple types of cancer (Binder et al., 2013; Lin et al., 2014). Podxl might thus be both involved in cell adhesion of cells within epicardial cell cluster 1 and in the mobilization of cells within this cluster towards a more mesenchymal phenotype. Jam2b (also named Jam-b2) is a zebrafish specific member of the Jam protein family, which in mammals consists of the three members F11r/Jam-A, Jam2/Jam-B and Jam3/Jam-C. In the zebrafish, each of the corresponding genes has been duplicated. However, there is high structural conservation and all family members are type I cell surface receptors with two immunoglobulin domains, a single transmembrane domain and a short cytoplasmic region (Powell and Wright, 2012). Jam2 is mediating cell-cell adhesion by interacting with itself (Cunningham et al., 2000) and with Jam3 (Arrate et al., 2001). Endothelial Jam2 directs the migration of leucocytes via integrin interactions (Cunningham et al., 2002) and zebrafish Jam2a was found to facilitate muscle development (Powell and Wright, 2011). However, Jam2b is not expressed in developing zebrafish blood vessels or trunk muscles (Powell and Wright, 2012) and has not yet been assigned a function. Taken together, the enrichment of the adhesion molecules Podxl and Jam2b in epicardial cell cluster 1 suggests that the cells within this cluster might be connected to form a continuous cell layer.

In addition to signaling mediators and adhesion molecules, epicardial cell cluster 1 was enriched in transcripts encoding the transcription factors Interferon Regulatory Factor 8 (Irf8) and Myelin Regulatory Factor (Myrf). Irf8, also known as interferon consensus sequence-binding protein, encodes a member of the Irf family and contains a highly conserved N-terminal DNA-binding domain and a less conserved C-terminal Irf association domain (Sharf et al., 1995). Irf8 is necessary for the differentiation of myeloid haematopoietic cells and has been used as a macrophage marker (Li et al., 2011a). In mouse cardiomyocytes, Irf8 was found to prevent the hypertrophic response to pressure overload by sequestering Nfatc1 in the cytoplasm (Jiang et al., 2014). Myrf is a membrane-associated transcription factor that undergoes an activating proteolytic cleavage to release its nuclear-targeted N-terminal region (Bujalka et al., 2013) and is essential for the myelination of axons (Emery et al., 2009). Neither Irf8 nor Myrf have been studied during heart development. In summary, epicardial cell cluster 1 is enriched in signaling and adhesion molecules as well as in transcription factors that have not been implicated with the epicardium before.

Many cells in epicardial cell cluster 1 furthermore showed expression of CU984579.1, which encodes the primary mesothelial cell marker Leucine Rich Repeat Neuronal 4 (Lrrn4) (Kanamori-Katayama et al., 2011). This and the fact that epicardial cell cluster 1, unlike the two other epicardial cell clusters, contained cells purified from *tcf21*, *tbx18* and *wt1b* reporter lines suggests that this cell population forms the main epicardial mesothelium. Supporting this hypothesis, a subset of cells in epicardial cluster 1 was labelled by the expression of the well-known epicardial marker Retinaldehyde Dehydrogenase (Aldh1a2). This enzyme is the main producer of retinoic acid in the heart and is expressed in epicardial and endocardial cells (Gupta et al., 2013). Retinoic acid is

essential for normal heart development as it regulates heart looping and cardiomyocyte differentiation (Niederreither et al., 1999), and activates the expression of Tcf21 and Wt1 (Braitsch et al., 2012). The fact that transcripts encoding Aldh1a2 were enriched in cells in epicardial cell cluster 1 underlines the epicardial identity of the cells contained within this population. Furthermore, this finding points towards a functional relevance of epicardial heterogeneity, as cells in epicardial cell clusters 2 and 3 mostly did not possess transcripts encoding Aldh1a2 and therefore are unlikely to mediate epicardial retinoic acid signaling.

#### **4.5 Epicardial cell cluster 2 is enriched in transcripts encoding the smooth muscle cell markers Mylka, Acta2 and Myh11a as well as in transcripts encoding the extracellular matrix factors Lox and Elnb**

Differential gene expression analysis identified several genes that are part of the contractile machinery in smooth muscle cells as enriched in epicardial cell cluster 2, most prominently *myosin light chain kinase (mylka)* (Scholey et al., 1980), *alpha smooth muscle actin (acta2)* (Duband et al., 1993) and *myosin heavy chain 11a (myh11a)* (Wallace et al., 2005). Mylka (also named Mlck1a) facilitates acto-myosin contractions by phosphorylating the regulatory light chain of myosin on the residue Ser19 (Scholey et al., 1980) and is one of two zebrafish paralogues of the mammalian Mlck1. Mlck1 is part of a protein family that further includes Mlck2 and Mlck3 (Kamm and Stull, 2001). While Mlck3 is expressed exclusively in the myocardium and Mlck2 in skeletal muscle and myocardium, Mlck1 has been detected in various tissues. In the zebrafish, Mylka is

expressed in thrombocytes, the gut, the optic fissure and the bulbus arteriosus (Moriyama et al., 2016; Tournoij et al., 2010). *Acta2* is an early marker of smooth muscle cell differentiation and required for contractile activity (Duband et al., 1993). *Acta2* is mostly expressed in smooth muscle, although transient expression has been detected in other types of muscle and in myofibroblasts (Darby et al., 1990; Woodcock-Mitchell et al., 1988). In the zebrafish, *Acta2* is expressed in gut, swim bladder, optic fissure, aortic arches, ventral aorta and bulbus arteriosus (Whitesell et al., 2014). Like *Acta2*, *Myh11a* is part of the acto-myosin complex enabling smooth muscle contraction and is present in gut smooth muscle as well as in vascular smooth muscle cells (Wallace et al., 2005). Thus, epicardial cell cluster 2 might represent a smooth muscle cell population.

Smooth muscle cells are known to produce extracellular matrix (ECM) components (Johnson et al., 2000) and the cells in epicardial cell cluster 2 indeed expressed ECM factors, among them Lysyl Oxidase (*Lox*) and Elastin b (*Elnb*). *Lox* is a copper-containing amine oxidase that facilitates extracellular crosslinking of fibrillar collagens and elastins (Molnar et al., 2003). In the zebrafish, eight *Lox* and *Lox*-Like protein family members have been identified, some of which regulate notochord development (Gansner et al., 2007). *Lox* itself plays a role during zebrafish muscle development and neurogenesis (Reynaud et al., 2008). Elastins are a major component of the ECM and provide its elastic properties. Two Elastin genes, *elna* and *elnb*, have been identified in the zebrafish (Miao et al., 2007). *Elnb* is specifically expressed in the bulbus arteriosus (Moriyama et al., 2016). Other ECM components with enriched transcripts in cells in epicardial cell cluster 2 include Aggrecan a (*Acana*), Fibulin 5 (*Fbln5*) and *Lox*-Like 1 (*Loxl1*). In summary, the genes enriched in epicardial cell cluster 2 suggest a smooth muscle phenotype of cells in this cluster and point towards an association with the bulbus arteriosus.

#### **4.6 Epicardial cell cluster 3 is enriched in transcripts encoding the tight junction marker *Cldn11a* as well as in transcripts encoding the signaling molecule *Cxcl12a***

In contrast to cells in epicardial cell cluster 2, cells in epicardial cell cluster 3 was not enriched in marker genes of a specific cell type. Instead, they expressed a diverse set of genes, including those encoding Claudin 11a (*Cldn11a*), Chemokine (C-X-C motif) Ligand 12a (*Cxcl12a*) and Nuclear Receptor Subfamily 1, Group H, Member 5 (*Nr1h5*). *Cldn11a* is an orthologue of the mammalian four-transmembrane tight junction molecule *Cldn11*, which is important for the creation of diffusion barriers in the myelin sheaths of the central nervous system and in Sertoli cells in the testis (Denninger et al., 2015; Gow et al., 1999; Morita et al., 1999). In contrast, *Cldn11a* was reported to have a vascular expression pattern in the zebrafish embryo (Cannon et al., 2013).

*Cxcl12a* (also named Stromal Cell-Derived Factor 1a, *Sdf1a*) and its mammalian orthologue *Cxcl12/Sdf1* belong to a chemoattractant cytokine family named chemokines, which contain four conserved cysteines forming two disulfide bonds (Wang and Knaut, 2014). *Cxcl12a* and *Cxcl12* have been well studied as they coordinate the migration of both tissues and single cells in various scenarios. This includes guiding blood vessel formation (Bussmann et al., 2011), migration of the lateral line primordium (David et al., 2002), neural development (Belmadani, 2005; Klein et al., 2001; Schwarting et al., 2006; Tiveron et al., 2006) and germ cell migration (Knaut et al., 2003). Additionally, *Cxcl12* signaling attracts lymphocytes (Bleul et al., 1996) and stimulates the proliferation of

blood cells (Lataillade et al., 2002). Cxcl12a mainly signals through the receptors Cxcr4a and Cxcr4b (Chong et al., 2001), but also acts via Cxcr7 (Valentin et al., 2007). However, another important function of Cxcr7 is to sequester Cxcl12a and act as a clearance receptor (Boldajipour et al., 2008; Naumann et al., 2010). The sequestering function of Cxcr7 has been shown to locally decrease the levels of Cxcl12a around migrating cells, creating a Cxcl12a gradient that enables self-directed migration (Donà et al., 2013). Interestingly, Cxcl12 has been shown to be expressed in the developing mouse epicardium, guiding the formation of the coronary vasculature (Cavallero et al., 2015) and Cxcl12a is upregulated in the adult zebrafish epicardium following cardiac injury, guiding the migration of Cxcr4b expressing cardiomyocytes into the wound area (Itou et al., 2012b; Lien et al., 2006).

The Farnesoid X Receptor Beta gene *nr1h5* encodes a class 1 nuclear receptor and is present as a pseudogene in humans. It is however expressed in the mouse where the encoded receptor is activated by lanosterol, a cholesterol precursor (Otte et al., 2003). Nr1h5 is further transactivated by retinoic acid and heterodimerizes with the Retinoid X Receptor Rxra (in zebrafish Rxraa and Rxrab) to regulate target gene transcription, however a specific function of Nr1h5 has not been reported. In the zebrafish, Nr1h5 has been found to be expressed in the liver (ZFIN). The enrichment of transcripts in epicardial cluster 3 encoding other retinoic acid activated receptors, such as Nr2f5 and Rxraa, suggests that the differentiation and/or function of these cells is regulated by retinoic acid.

Thus, the gene expression profile of cells in epicardial cell cluster 3 suggests that these cells might form tight junctions similar to those found in myelin sheaths and that they might guide Cxcl12a-responsive cells into or within the heart. Furthermore, these

processes might be regulated by retinoic acid, acting via *Rxraa* and nuclear receptors such as *Nr1h5*.

Interestingly, a subset of cells within epicardial cell cluster 3 prominently expressed the adipocyte differentiation marker Seipin-Like (*Bscl2l*), which is also involved in the formation of lipid droplets (Payne et al., 2008), and the hepatocyte marker Fatty Acid Binding Protein 10a (*Fabp10a*) (Choi et al., 2014b). Therefore, a subpopulation of cells within epicardial cell cluster 3 might fulfill lipid-related functions, potentially giving rise to epicardial adipose tissue comparable to what has been shown in the mouse (Yamaguchi et al., 2015).

Another explanation for the presence of transcripts encoding *Fabp10a* in a subset of cells within epicardial cell cluster 3 would be the presence of a population of *tcf21:H2B-Dendra2* expressing liver cells accidentally FACS-purified together with the epicardial cells. However, *Tcf21* has not been reported to be expressed in the liver and I did not detect fluorescence in liver cells of *Tg(tcf21:H2B-Dendra2)* embryos (figure 3-3d-f). This and the fact that care was taken not to purify cell doublets makes it unlikely that liver cells from *Tg(tcf21:H2B-Dendra2)* were FACS-purified, even if liver tissue was accidentally co-isolated when collecting isolated hearts. Furthermore, transcripts encoding other hepatocyte markers, such as *Hnf4 $\alpha$*  (Field et al., 2003), were absent from the data set. To investigate further, it would be interesting to analyze if there is fluorescence in the epicardium of the *fabp10a* reporter lines used previously to identify hepatocytes (Choi et al., 2014b).

## **4.7 The expression of CD248a distinguishes epicardial cell clusters 2 and 3 from epicardial cell cluster 1**

The C-type lectin transmembrane receptor Endosialin a (CD248a) was highly enriched in cells in both epicardial cell cluster 2 and epicardial cell cluster 3. Additionally, transcripts encoding CD248a were not present in cells in epicardial cell cluster 1 and CD248a might therefore be a key factor in discriminating cells in epicardial cell clusters 2 and 3 from those in epicardial cell cluster 1. In the mouse, CD248 is expressed in stromal cells and pericytes (Bagley et al., 2008). Expression of CD248 in pericytes promotes angiogenesis by acting on Platelet Derived Growth Factor b (Pdgfb) mediated signaling downstream of the receptor Pdgfrb, but is dispensable for pericyte differentiation and survival (Naylor et al., 2014; Tomkowicz et al., 2010). By enhancing Pdgfb signaling, CD248 is also able to increase cell proliferation (Wilhelm et al., 2015). Indeed, Pdgfrb was expressed in a subset of cells located within epicardial cell clusters 2 and 3 (data not shown), making it possible for CD248a to play a similar role in these epicardial subpopulations. However, the precise function of CD248 is unknown, as is the expression pattern of CD248a in the zebrafish embryo.

## 4.8 Marker gene expression in the non-epicardial cell clusters

### 4.8.1 Cluster 3 - myelinating cells, neural progenitors and orphan

#### ***Tg(tcf21:H2B-Dendra2)* derived cells**

In addition to cluster 1 (termed epicardial cell cluster 1), cluster 9 (termed epicardial cell cluster 2) and cluster 2 (termed epicardial cell cluster 3), the data set contained several non-epicardial cell clusters with specific marker gene expression. Cluster 3 was heterogeneous and mostly contained quadruple negative cells derived from *Tg(tcf21:dsRed2; myl7:GFP) x Tg(kdrl:GFP; gata1a:dsRed)* and cells derived from *Tg(tcf21:H2B-Dendra2)*. A large subset of quadruple negative cells expressed the myelinating protein Myelin Basic Protein a (Mbpa), a marker of oligodendrocytes and Schwann cells (Aggarwal et al., 2011). These cells furthermore expressed Myelin-Associated Glycoprotein (Mag) (Quarles, 2007), the myelin-associated Claudin k (Cldnk) (Münzel et al., 2012) and the neural crest transcription factor Sox10 (Dutton et al., 2001). In conclusion, this subpopulation of cells within cluster 3 likely comprises myelinating cells.

Another subset of quadruple negative cells within cluster 3 expressed the neural proteoglycan Brevican (Bcan) (Frischknecht and Seidenbecher, 2012), the neural progenitor marker Musashi 1 (Msi1) (MacNicol et al., 2015) as well as the B1-sox transcription factors Sox2, Sox3, Sox19a and Sox19b, involved in neural patterning (Okuda et al., 2010). Taken together, this suggests a subpopulation of neural progenitor cells.

A third subpopulation within cluster 3 consisted of *tcf21*:H2B-Dendra2 positive cells and these cells expressed the transcription factors Foxf1, Foxf2a and Barx1, which label the splanchnic mesoderm and regulate craniofacial, lung and gut development (Hellqvist et al., 1996; Nichols et al., 2013; Ormestad, 2006; Tissier-Seta et al., 1995; Xu et al., 2016). Given the transcriptomic profile and the fact that *tcf21*:H2B-Dendra2 is expressed in craniofacial muscles, branchial arches and foregut, the cells within this sub-cluster might be derived from these non-cardiac tissues instead of the epicardium.

#### **4.8.2 Cluster 4 - neural cells**

Interestingly, the quadruple negative cell cohort purified from *Tg(tcf21:dsRed2; myl7:GFP) x Tg(kdrl:GFP; gata1a:dsRed)* ventricles was almost entirely neural in nature (dark purple colored cell origin in figure 4-2). Almost all of these cells were found in cluster 4 and expressed Glycoprotein M6ab (Gpm6ab), a neuronal proteolipid protein which induces neurite outgrowth (Huang et al., 2011). Additionally, cells in cluster 4 were labelled by the expression of N-Ethylmaleimide Sensitive Factor a (Nsfa), an ATPase that plays a role in organizing myelinated axons (Woods et al., 2006) and Myelin Transcription Factor 1-Like a (Myt1la), a neuron-specific transcription factor that safeguards neuronal identity (Mall et al., 2017). This shows that the embryonic zebrafish heart is in close contact with a neuronal network by 5dpf that might also contain the neuronal support cells in cluster 3. These neuronal cells are likely to be located within the ventricle or adherent to the ventricular surface, as they were FACS purified from isolated hearts. Additionally, the high transcriptional similarity between the neuronal cells argues against a random origin from contaminating non-cardiac tissue. This is a significant finding since

the processes of neuronal development in the embryonic zebrafish heart are largely unknown.

#### **4.8.3 Cluster 5 - mesenchymal cells**

While most of the cells within cluster 5 were derived from wildtype hearts (orange colored cell origin in figure 4-2), it also contained cells from *Tg(myf7:GFP)* hearts (dark green colored cell origin in figure 4-2), some of which contained transcripts encoding *Myf7*, and a single cell from *Tg(kdrl:GFP)* hearts (yellow colored cell origin in figure 4-2). These cells were clustered together due to the shared expression of the ECM adaptor protein ABI Family, Member 3 (NESH) Binding Protein b (*Abi3bbp*) (Iriyama et al., 2001). The exact biological function of *Abi3bbp* is unknown. However, *Abi3bp* was found to promote the differentiation of mesenchymal stem cells and c-Kit expressing cardiac progenitor cells in the mouse, thereby improving recovery following myocardial infarction (Hodgkinson et al., 2013, 2014). Cells in cluster 5 were furthermore labelled by the expression of the retinoic acid metabolizing enzyme *Cyp26b1*, which regulates the development of cranial neural crest derived tissues as well as osteogenesis from mesenchymal progenitors (Laue et al., 2008; Maclean et al., 2009; Reijntjes et al., 2007; Spoorendonk et al., 2008). The presence of *Abi3bbp* and *Cyp26b1* in cells in cluster 5 thus indicates that these might be mesenchymal cells, potentially acting as a cardiac progenitor population.

In agreement with this hypothesis, a large subset of the cells in cluster 5 expressed Platelet-Derived Growth Factor Receptor Alpha (*Pdgfra*), which has been described as a cardiac progenitor gene in human and mouse (Chong et al., 2013; Kim et al., 2017;

Nosedá et al., 2015; Prall et al., 2007). *Pdgfra* has also been studied as a marker of cardiac fibroblasts (Moore-Morris et al., 2014; Rudat et al., 2013). The *Pdgfra* expressing subset of cells within cluster 5 co-expressed multiple ECM regulatory proteins, such as Periostin a (*Postna*) (Snider et al., 2008), Thrombospondin 1b (*Thbs1b*) (Gonzalez-Quesada et al., 2013), Matrilin 4 (*Matn4*) (Deák et al., 1999) and Angiopoietin like 7 (*Angptl7*) (Comes et al., 2011).

Another subset of cells within cluster 5 did not express *Pdgfra*, but the F-Sox transcription factor *Sox7*, which is important during vascular development (Behrens et al., 2014), as well as the endothelial receptors *Tie1* (Lyons et al., 1998) and *Kdr* (Terman et al., 1992), suggesting an endothelial cell fate.

In summary, the transcriptomic profile of cells in cluster 5 indicates that they comprise a heterogeneous mesenchymal cell population, with one subset of cells possibly functioning in organizing the cardiac extracellular matrix and another subset potentially having an endothelial fate.

#### **4.8.4 Clusters 6 and 7 - haematopoietic cells**

Clusters 6 and 7 were largely comprised of cells isolated from wildtype hearts. Also, cluster 6 contained almost all of the cells sorted due to GFP fluorescence from *Tg(kdrl:GFP; gata1a:dsRed)* hearts, while cluster 7 contained the two cells sorted due to dsRed fluorescence from the same reporter hearts (black colored cell origin in figure 4-2). Their transcriptomic profiles identified both clusters as haematopoietic, marked by the widespread presence of transcripts encoding the zebrafish zinc finger protein Bloody Fingers (*Blf*) (Sumanas et al., 2005a). During gastrulation and neurulation, *Blf* was found

to regulate morphogenetic movements, however from mid-somitogenesis Blf becomes restricted to haematopoietic cells. Cells in cluster 7 were marked by the prominent expression of the zebrafish Carbonic Anhydrase (Cahz), an enzyme that catalyzes the conversion between carbon dioxide and bicarbonate. Cahz is predominantly expressed in erythroid lineage cells (Qian et al., 2005). Thus, cells in cluster 7 appeared to be derived from the erythroid lineage, which was underlined by the expression of other erythroid markers, such as Solute Carrier Family 4 (Anion Exchanger), Member 1a (Slc4a1a) (Paw et al., 2003) and Erythrocyte Membrane Protein Band 4.1b (Epb41b) (Nunomura et al., 2007). Cells in cluster 7 also expressed several Alpha and Beta Globins as well as Gata1a (data not shown). In contrast, cells in cluster 6 expressed the leukocyte-specific GRB2-Related Adaptor Protein 2a (Grap2a) (Qiu et al., 1998) as well as Regulator of G Protein Signaling 13 (Rgs13), which constrains murine lymphocyte migration by repressing chemokine signaling and the activity of the transcription factor CREB (Shi et al., 2002). This suggests that cluster 6 does not contain erythrocytes, but leukocytes. Surprisingly, the above expression profile extended to the large number of *Tg(kdrl:GFP)* derived cells in cluster 6, although *Kdrl* is an endothelial marker (Liao et al., 1997). At 5dpf, there hence appears to be a significant number of *kdrl:GFP* labelled cells in the embryonic zebrafish heart that belong to the leukocyte lineage rather than being endocardial. This might point towards a differentiation of these cells from *Kdrl* expressing haematopoietic progenitor cells, similar to what has been proposed in the mouse (Lugus et al., 2009). The lack of an additional *kdrl:GFP* expressing endocardial cell population in the data set might be due to the dissociation protocol prior to FACS, which might not be suited for endocardial cells.

#### 4.8.5 Cluster 8 - cardiomyocytes

Cluster 8 was phylogenetically positioned close to the epicardial cell clusters and the cells contained in it were labelled by a strong expression of *Myl7*, suggesting they were cardiomyocytes. This matched the origin of the cells contained within cluster 8, as most of them were derived from *Tg(myl7:GFP)* embryos. Additionally, cluster 8 was enriched in transcripts encoding the cardiomyocyte-specific myosin light chain *Cmlc1* (Chen et al., 2008). Interestingly, the glycolytic enzyme Glyceraldehyde-3-Phosphate Dehydrogenase (*Gapdh*) was equally enriched in cluster 8. *gapdh* is commonly used as a housekeeping gene and as such is supposed to be expressed at a constant level across different tissues. However, this appears not to be the case in the embryonic zebrafish heart. In contrast, the family member *gapdhs*, although identified as specific to spermatogenic cells in mammals (Welch et al., 1992), as well as the housekeeping genes *beta actin 2* (*actb2*), *ubiquitin c* (*ubc*) and *eukaryotic translation initiation factor 1B* (*eif1b*) showed a largely uniform expression across the data set (data not shown).

#### 4.9 Summary of cell clusters in the single cell transcriptomic data set

In summary, the single cell data set represents transcriptomic data characterizing several cardiac cell types, including multiple epicardial, neural and haematopoietic cell populations as well as cardiomyocytes and mesenchymal progenitor cells. However, there was no endocardial cell population and most cardiac cells isolated due to *kdrl:GFP* fluorescence expressed markers of myeloid haematopoietic cells. Also, there were only few cells derived from *Tg(myl7:GFP)* hearts in the data set and half of these did not

express markers of mature cardiomyocytes, but mesenchymal or even endothelial marker genes. Although cells processed from wildtype hearts were present in all major cell populations except epicardial cell cluster 3, most of them were located in the haematopoietic clusters 6 and 7 as well as in cluster 5. This might reflect the cellular contributions of different tissues to the developing heart, underlining the need for reporter-based tissue enrichment when studying tissues that are low in cell numbers, such as the epicardium. Additionally, wildtype cells were probably underrepresented in the cardiomyocyte cluster, as the tissue digestion protocol used appeared not to be well suited for cardiomyocytes. A non-optimized digestion protocol might be the reason for the absence of an endocardial cell population as well.

#### **4.10 t-SNE clustering confirms the presence of distinct epicardial cell populations in the single cell data set**

To visualize the transcriptomic complexity within the data set in a two-dimensional way, cells were clustered and plotted using t-distributed stochastic neighbor embedding (t-SNE) (van der Maaten, 2014). This statistical method enables the reduction of multi-dimensional data into representative principal components that are subsequently used to cluster the individual samples according to their similarity. Overlaying the resulting plot with the most significant pagoda-identified cell clusters (figure 4-3a) and with the transgenic origin of each individual cell (figure 4-3b) showed that the t-SNE based clustering was very similar to the hierarchical clustering depicted in figure 4-2. Cells that were assigned to different epicardial cell clusters by the hierarchical clustering also formed distinct clusters in the t-SNE plot. Similarly, cells within hierarchical clusters of

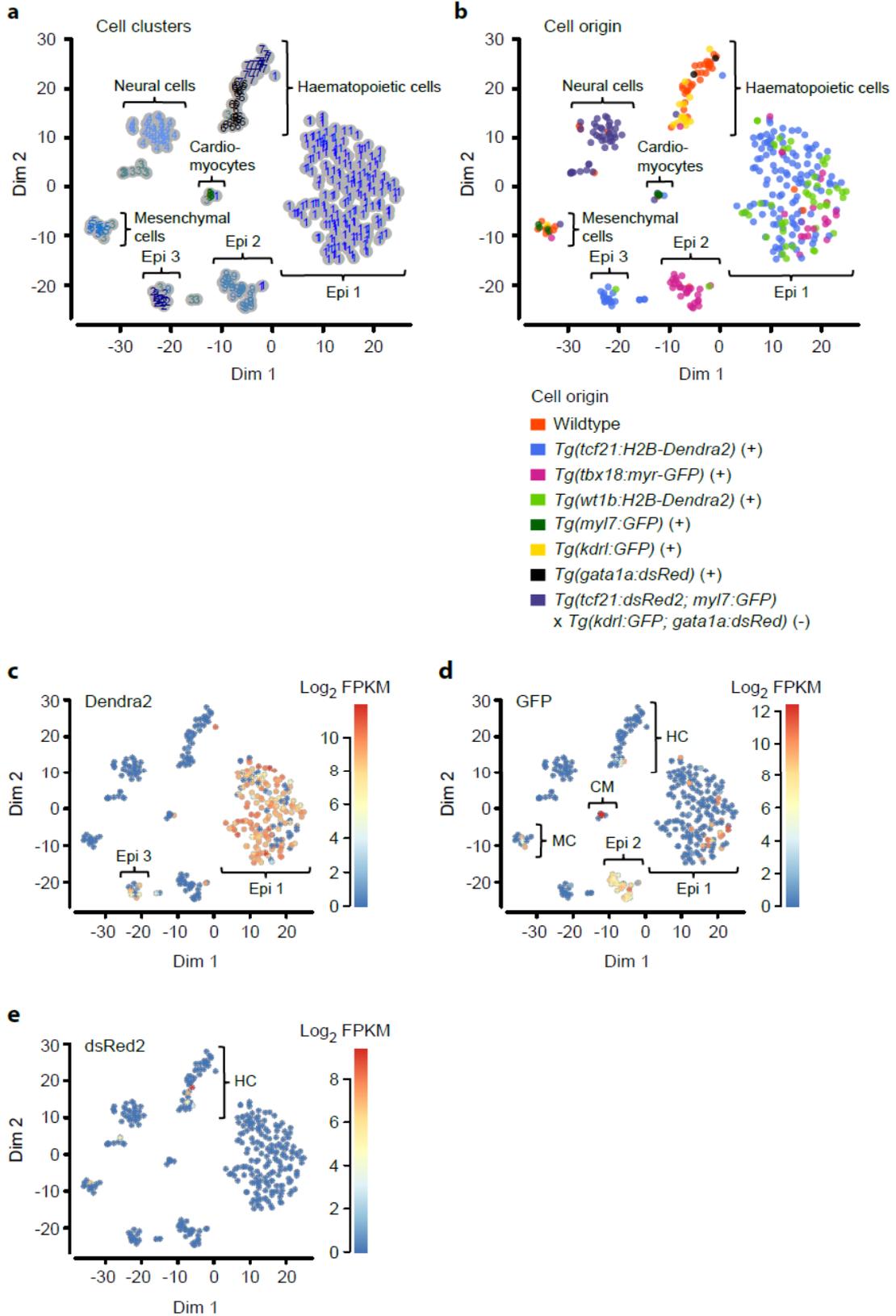
myocardial, haematopoietic, mesenchymal and neural phenotype formed distinct clusters in the t-SNE plot. However, cells belonging to the heterogeneous hierarchical cluster 3 localized to multiple t-SNE clusters. The subsets within cluster 3 identified as myelinating cells and neural progenitors indeed formed a cluster located close to the neural cluster 4. In contrast, cells in cluster 3 that were derived from *Tg(tcf21:H2B-Dendra2)* clustered between the epicardial clusters 2 and 3.

Overlaying fluorophore expression values onto the t-SNE plot showed that the majority of the cells derived from *tcf21:H2B-Dendra2* and *wt1b:H2B-Dendra2* reporters, mostly assigned to epicardial cell clusters 1 and 3, indeed expressed *Dendra2* (figure 4-3c). Similarly, almost all of the *tbx18:myr-GFP* derived cells, mostly located in epicardial cell cluster 1 and 2, expressed GFP (figure 4-3d). Furthermore, all *myl7:GFP* derived cells in the cardiomyocyte cluster expressed high levels of GFP, and some of the *kdrl:GFP* derived cells in the haematopoietic and mesenchymal cell clusters expressed GFP as well. Finally, none of the *tcf21:dsRed2*, *myl7:GFP*, *gata1a:dsRed*, *kdrl:GFP* quadruple negative cells, mostly located in the neural cell cluster, showed expression of GFP or dsRed2 (figure 4-3e). In summary, the transcriptomic fluorophore expression patterns mostly matched the reporter lines that cells in the data set were derived from.

Hierarchical and t-SNE clustering were not completely similar. Two cells assigned to epicardial cell cluster 1 by hierarchical clustering co-localized with epicardial cell cluster 2 on the t-SNE plot and one cell from epicardial cell cluster 2 co-localized with epicardial cell cluster 3. Also, two cells derived from *Tg(tcf21:H2B-Dendra2)* that were assigned to epicardial cell cluster 1 by hierarchical clustering localized to the myocardial and haematopoietic clusters on the t-SNE plot. This suggests that the transcriptomic profile of

some epicardial cells in the data set was ambiguous as to their identity, and that the transcriptomic distance between the epicardial cell clusters was rather small.

Indeed, the three epicardial cell clusters were located close to one another on the t-SNE plot. Similar to the hierarchical clustering, epicardial cell cluster 2 was flanked by the two other epicardial cell clusters. Cardiomyocytes were located in the middle of the plot within equal distance of the other clusters, while mesenchymal cells clustered close to epicardial cell cluster 3 and haematopoietic cells close to epicardial cell cluster 1. The neural cell cluster was the one most distant to the epicardial cell clusters. Taken together, the t-SNE representation mirrors the hierarchical clustering and underlines the presence of three distinct epicardial cell clusters within the data set.



**Figure 4-3: The single cell data set contains three distinct epicardial cell clusters. (a)** Two-dimensional t-SNE clustering of 352 single cell libraries based on 2376 highly variable genes. Numbers indicate assignment to the nine most significant cell clusters within the data set. **(b)** Color-coded cell origins overlaid onto the t-SNE clustering. The legend beneath indicates the reporter line each cell was isolated from and if it was a fluorescent cell (+) or a non-fluorescent cell (-). **(c-e)** FPKM expression values of fluorophore transcripts were overlaid onto the t-SNE clustering. **(c)** Expression of *Dendra2* in epicardial clusters 1 and 3. **(d)** Expression of GFP in epicardial clusters 1 and 2, cardiomyocytes, mesenchymal cells and haematopoietic cells. **(e)** Expression of *dsRed2* in a few haematopoietic cells. Epi = epicardial cell cluster, CM = cardiomyocytes, MC = mesenchymal cells, HC = haematopoietic cells, FPKM = fragments per kilobase of transcript per million transcripts, dim = dimension.

#### **4.11 The presence of *tcf21*, *tbx18* and *wt1b* in the epicardial cell clusters is heterogeneous**

To validate that the three cell clusters consisting of cells derived from *Tg(tcf21:H2B-Dendra2)*, *Tg(tbx18:myr-GFP)* and *Tg(wt1b:H2B-Dendra2)* hearts were indeed epicardial, I analyzed the endogenous expression patterns of *Tcf21*, *Tbx18*, *Wt1b* and other factors that have been used previously as markers of epicardial cells. Overlaying the fragments per kilobase of transcript per million transcripts (FPKM) expression values of transcripts encoding *Tcf21* onto the t-SNE plot showed that *tcf21* was present exclusively in cells within the three cell clusters designated epicardial and that the expression was mostly

restricted to cells in epicardial cell clusters 1 and 3 (figure 4-4a). The expression of *Tbx18* corroborated the epicardial nature of cells in epicardial cell clusters 1 and 2 (figure 4-4b). It is worth noting that the average expression level of *Tbx18* was lower in cells in epicardial cell cluster 2 than in epicardial cell cluster 1. Finally, the expression of *Wt1b* was mostly restricted to cells in epicardial cell cluster 1 (figure 4-4c).

This underlined that most cells in the cell clusters categorized as epicardial according to their origin indeed contained transcripts encoding *Tcf21*, *Tbx18* and *Wt1b*. However, it also showed that these markers were not uniformly expressed across all epicardial cells, but that they rather labelled specific epicardial cell clusters while being largely excluded from others. This might indicate that the presence or the absence of *Tcf21*, *Tbx18* and *Wt1b* is important for the fate and/or function of cells in the individual epicardial cell clusters. For example, only *tbx18* was present in cells in epicardial cell cluster 2 and the protein might regulate the smooth muscle phenotype of these cells. Several components of the smooth muscle contraction machinery were present in cells within epicardial cell cluster 2, which might indicate them being in a late phase of smooth muscle differentiation. A previous study found that *Tbx18* prevents premature trans-differentiation of epicardium-derived cells into smooth muscle cells (Greulich et al., 2012). The fact that the expression level of *Tbx18* was low in cells in epicardial cell cluster 2 might point towards a down-regulation of *Tbx18* to allow differentiation into smooth muscle to be completed. The lack of transcripts encoding *Tcf21* in these cells matches previous findings that *Tcf21* prevents trans-differentiation of epicardium-derived cells into smooth muscle (Braitsch et al., 2012). The expression of *Tcf21* in many cells in epicardial cell cluster 3 suggests that *Tcf21* plays a role in these cells. The expression of *Tcf21*, *Tbx18* and *Wt1b* in overlapping subsets of cells in epicardial cell cluster 1 on the

other hand suggests a role for all three markers here and is further support for the hypothesis that this cell cluster constitutes the main epicardial sheet.

A closer investigation showed that other published epicardial markers, such as the transcription factors *Wt1a* and *Scxa* were expressed in subsets of cells in the epicardial cell clusters (figure 4-4d). The expression pattern of *Wt1a* recapitulated that of its paralogue *Wt1b*, with a strong focus on cells in epicardial cell cluster 1, but also expression in two cells within epicardial cell cluster 3. A previous study identified *Scx* as a marker of a subset of epicardial cells in mouse and chick embryos that gave rise to endothelial cells (Katz et al., 2012). The authors further found the chemo-repellent Semaphorin 3d (*Sema3d*) to co-label the *Scx* positive epicardial cell population. Indeed, *Scxa* and *Sema3d* were both expressed almost exclusively in cells in epicardial cell cluster 1. However, the expression of *Scxa* and *Sema3d* largely overlapped with that of *Tbx18* and *Wt1b*, different to what has been reported (Katz et al., 2012). As mentioned in chapter 4.4, the retinoic acid producing enzyme *Aldh1a2* was also expressed in epicardial cells, however it was mostly confined to cells in epicardial cell cluster 1. The ECM components Fibronectin 1a (*Fn1a*) and *Fn1b* are expressed in the adult zebrafish epicardium and are required for zebrafish heart regeneration (Wang et al., 2013). Fitting these findings, the presence of *fn1a* and *fn1b* in the data set was mostly confined to epicardial cells. The expression of *Fn1a* labelled most cells in epicardial cell cluster 2, while *Fn1b* showed widespread expression in epicardial cell cluster 1. *Fn1a* further was expressed in cells in epicardial cell cluster 1 and in a subset of cells in the mesenchymal cell cluster, while *Fn1b* was expressed in a subset of cells in epicardial cell cluster 2. Epicardial *Igf2* signaling has been found to stimulate cardiomyocyte proliferation in mouse and zebrafish, with *Igf2b* being upregulated following zebrafish heart injury

(Brade et al., 2011; Huang et al., 2013). I found Igf2b indeed to be expressed in cells in all three epicardial cell clusters, and additionally within cells in the mesenchymal cell cluster. In summary, the expression patterns of Wt1a, Scxa, Sema3d, Aldh1a2, Fn1a, Fn1b and Igf2b further validated and refined the previously identified epicardial cell clusters in the data set.

A prerequisite for epicardial trans-differentiation and cellular contribution to other cardiac tissues is epithelial-to-mesenchymal transition (EMT), the conversion of an adherent morphological phenotype into a detached and mobile one. Several factors have been identified as mediators of epicardial EMT, among them the Snail Family Zinc Finger transcription factors Snai1a, Snai1b and Snai2 (Slug). In the mouse, Snai1 and Snai2 are downstream targets of Tbx18 and Wt1 (Martínez-Estrada et al., 2010; Takeichi et al., 2013). Indeed, Snai1a, Snai1b and Snai2 were expressed in subsets of cells in the epicardial cell clusters (figure 4-4e). Snai1a showed a widespread expression across cells in all three epicardial cell clusters and was strongly expressed in cells in epicardial cell cluster 3, while Snai1b was only expressed in a small subset of cells within epicardial cell cluster 1. The expression pattern of Snai2 resembled that of Snai1a, except for a more pronounced expression of Snai2 in cells in epicardial cell cluster 2. The basic helix-loop-helix transcription factor Twist1 is important for neural crest differentiation and for EMT during cardiac valve formation (Shelton and Yutzey, 2008; Soo et al., 2002). Twist1 is expressed in epicardium-derived cells and the encoded protein physically interacts with Snai1 and Snai2 (Lander et al., 2013; Zhou et al., 2010). The zebrafish orthologues Twist1a and Twist1b were indeed expressed in cells in all three epicardial cell clusters, with a strong expression of Twist1a in cells in epicardial cell cluster 2. Nuclear Factor of Activated T-Cells Cytoplasmic 1 (Nfatc1) is a Calcineurin-activated transcription factor

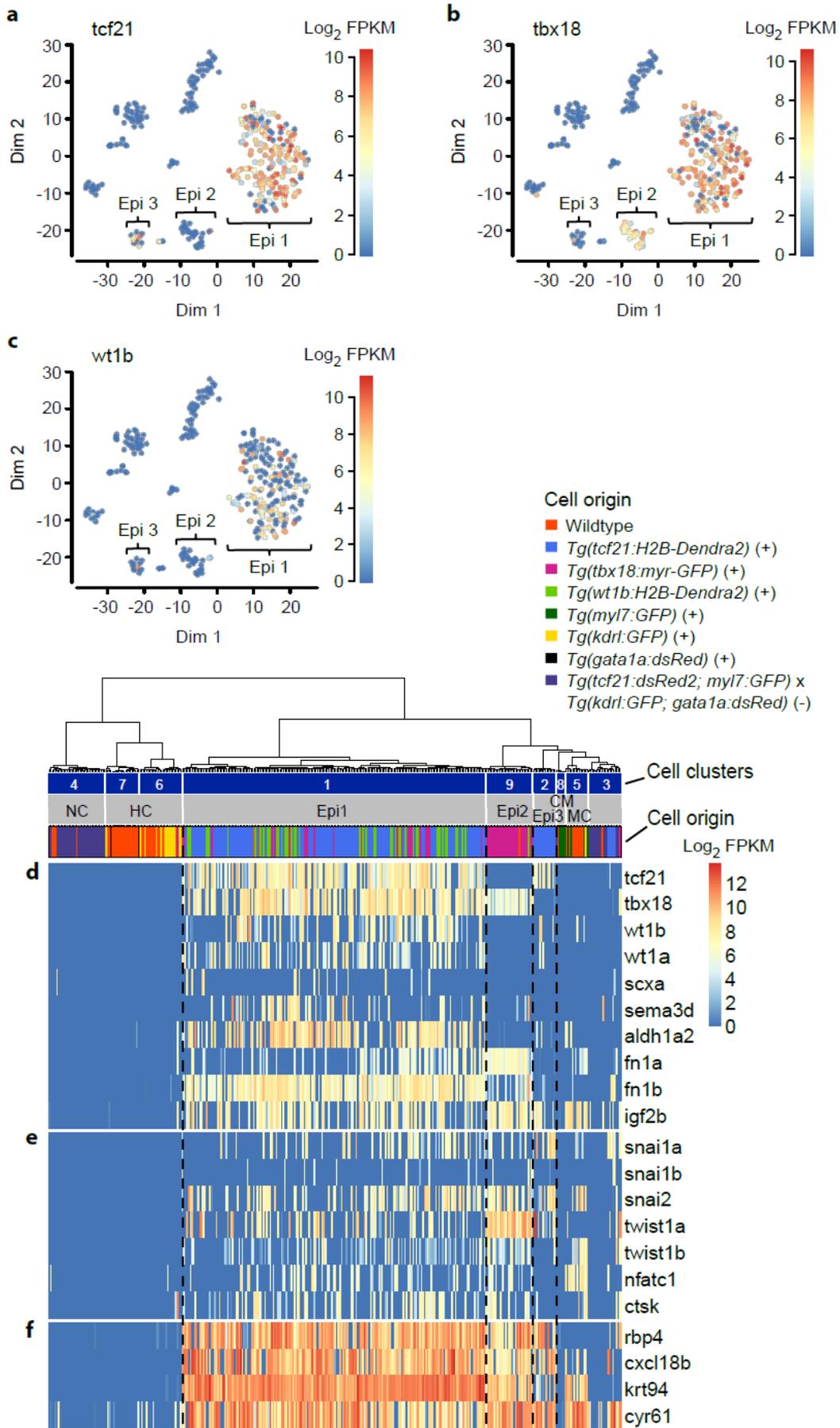
that induces the expression of the ECM remodeling enzyme *Ctsk* and is required for the migration of epicardium derived cells into the myocardium (Combs et al., 2011). Both transcripts encoding *Nfatc1* and *Ctsk* were present in cells in epicardial cell clusters 1 and 2, but not in cells in epicardial cell cluster 3.

The expression of several EMT markers in cells in the epicardial cell clusters indicates that embryonic epicardial cells are still mobile at 5dpf, even though they have formed a coherent epithelial sheet by then. Cells might either migrate within the epicardial sheet, for example to colonize the atrial myocardium, or even leave the epicardial sheet to invade the underlying cardiac tissue. Such mobility is a shared characteristic of the developing embryonic and the regenerating adult epicardium.

As neither *Tcf21*, *Tbx18* nor *Wt1b* was expressed in all epicardial cells within the data set, there might be better pan-epicardial markers. Indeed, a small number of genes showed strong and widespread expression within cells in all three epicardial cell clusters, while being largely absent from cells in the non-epicardial cell clusters (figure 4-4f). The most epicardium-specific of them was *retinol binding protein 4 (rbp4)*, encoding a secreted retinol transporter. In the embryonic zebrafish, *Rbp4* was shown to be expressed in the yolk syncytial layer, the hypochord, the skin and the liver (Li et al., 2007; Sumanas et al., 2005b; Tingaud-Sequeira et al., 2006). *Rbp4* derived from the developing epicardium likely plays a role in providing retinol for retinoic acid production, but might also transport retinol to extra-cardiac tissues like the liver. Furthermore, *Rbp4* regulates the expression of Fibronectin 1 (Li et al., 2007). Other genes with pan-epicardial expression were *cxcl18b*, encoding a neutrophil attracting chemokine (Torraca et al., 2017), *keratin 94 (krt94)* and *cysteine-rich angiogenic inducer 61 (cyr61)*, none of which have been studied in the heart. However, *Cxcl18b*, *Krt94* and *Cyr61* were also expressed in cells in

the mesenchymal cell cluster, and only Rbp4 was almost exclusively expressed in epicardial cells. Therefore, Rbp4 appears to be an exclusively pan-epicardial marker in the developing zebrafish heart. However, this needs to be confirmed by studying the expression of Rbp4 over a range of developmental time points.

In summary, the epicardial cell clusters in the data set indeed contained bona-fide signatures of epicardial gene programs.



**Figure 4-4: Cells within the epicardial cell clusters express epicardial gene signatures.**

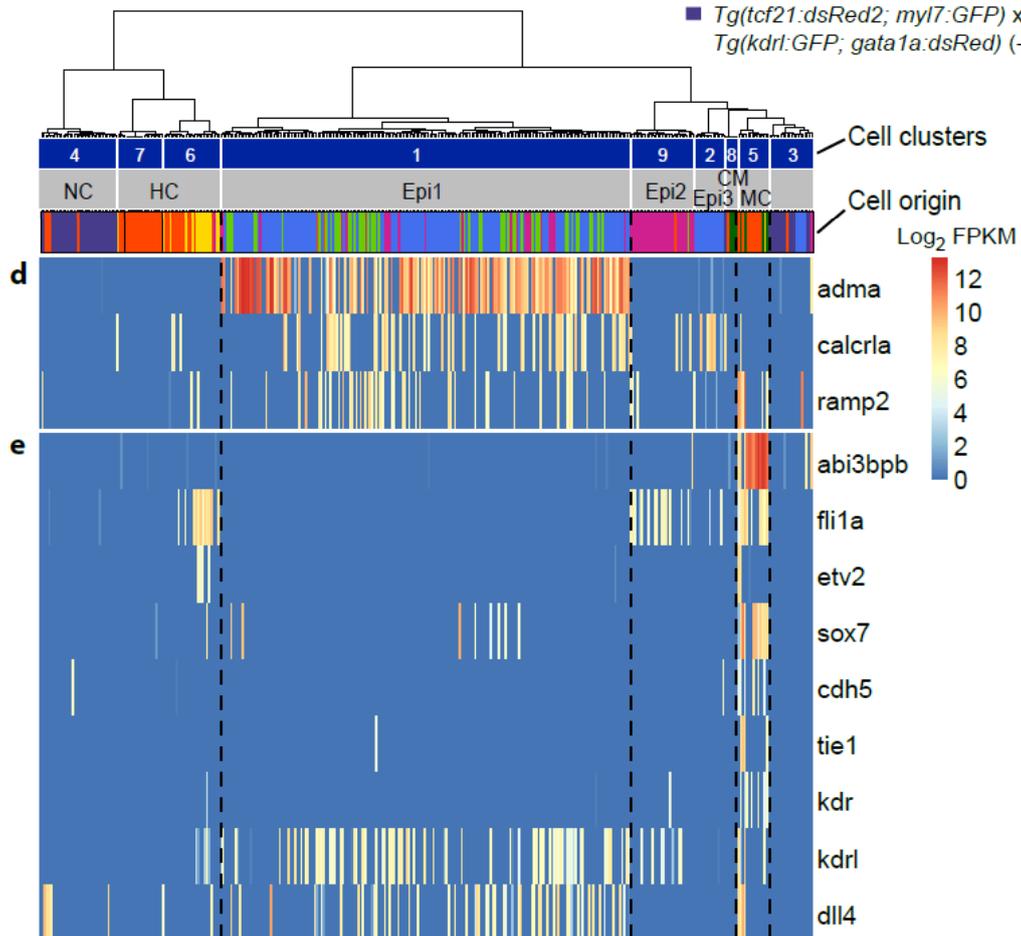
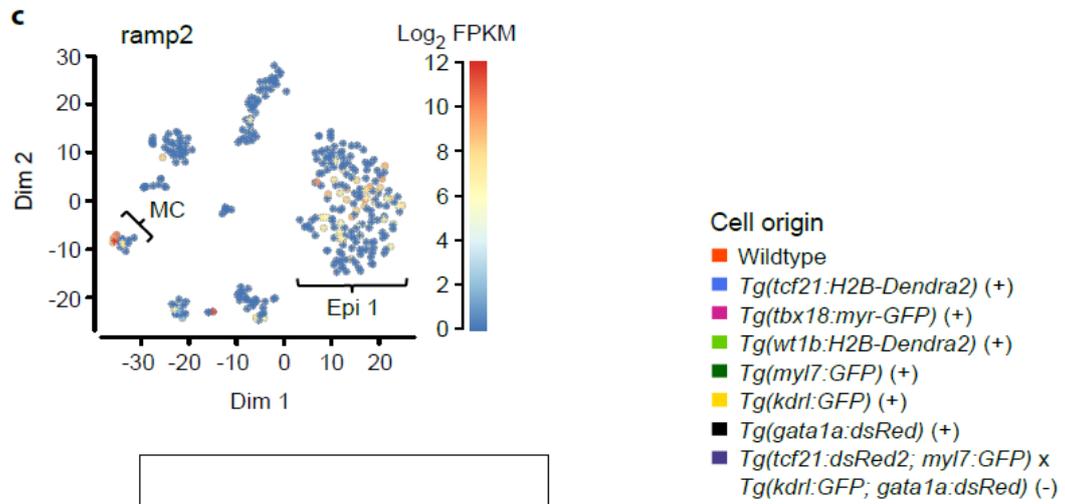
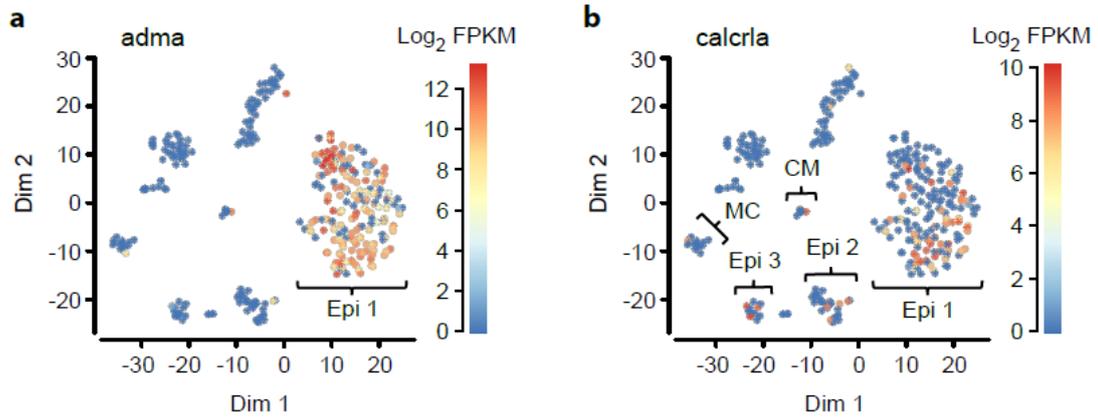
(a-c) FPKM expression values of major epicardial marker genes were overlaid onto the t-SNE clustering. (a) Expression of Tcf21 in epicardial clusters 1 and 3. (b) Expression of Tbx18 in epicardial clusters 1 and 2. (c) Expression of Wt1b, mostly confined to a subset of cells within epicardial cluster 1. (d-f) Heatmap of the clustered single cell data set (columns) showing genes enriched in the epicardial cell clusters (rows). Expression values are FPKM, red color indicates high expression. The phylogenetic tree at the top summarizes the transcriptomic relationships of the single cells within the data set. The significance of each cell cluster is indicated by the height of the respective bifurcation event. White numbers above the heatmap indicate the nine most significant cell clusters. The annotation underneath indicates the cell type each cell cluster contains, as well as the color-coded origin of each single cell. The key to the color code is located at the bottom of the heatmap and indicates the reporter line each cell was isolated from and if it was a fluorescent cell (+) or a non-fluorescent cell (-) (d) Expression of published epicardial marker genes. (e) Expression of epicardial epithelial-to-mesenchymal transition markers. (f) Expression of novel pan-epicardial marker genes. Epi = epicardial cell cluster, NC = neural cells, HC = haematopoietic cells, CM = cardiomyocytes, MC = mesenchymal cells, FPKM = fragments per kilobase of transcript per million transcripts.

#### **4.12 Presence of *adma* in epicardial cell cluster 1 might drive endothelial differentiation of mesenchymal cells in the developing zebrafish heart**

Since *adma* was the most enriched gene in cells in epicardial cell cluster 1 (figure 4-5a), I further investigated which cells the protein might be affecting and what processes it might be regulating. The receptor *Calcrla* was mostly expressed in cells in epicardial cell clusters 1 and 2, but also in a few haematopoietic, myocardial and mesenchymal cells (figure 4-5b). The receptor-associated factor *Ramp2* showed a comparable expression pattern and strongly labelled a subset of mesenchymal cells (figure 4-5c). Since *Calcrla* cannot mediate *Adma*-induced signaling without an associated *Ramp* (McLatchie et al., 1998), I analyzed the overlap between *Calcrla* and *Ramp2* expression and found that co-expression only occurred in some cells within epicardial cell cluster 1 and in cells within the *ramp2* positive subset of mesenchymal cells (figure 4-5d). These cells were labelled by the expression of the mesenchymal cell marker *Abi3bpb*, but also by the expression of the haemangiogenic transcription factor Friend Leukaemia Integration 1a (*Fli1a*), which controls the differentiation of both blood and endothelium (Liu et al., 2008) (figure 4-5e). Some of the *ramp2*, *fli1a* double positive cells also expressed Ets Variant 2 (*Etv2*), another ETS transcription factor required for blood and vessel formation (Sumanas and Lin, 2006). The expression of *Etv2* in the zebrafish embryo is restricted to endothelial cells and their precursors. Another driver of endothelial development is *Sox7*, which is present in endothelial precursor cells and is directly activated by *Etv2* (Behrens et al., 2014). A large subset of mesenchymal cells expressed *Sox7*, including the cluster of cells in which co-expression of *Calcrla*, *Ramp2*, *Fli1a* and *Etv2* occurred. In mouse embryonic stem cells, *Sox7* induces the expression of VE-Cadherin (*Cdh5*), an adhesion protein that establishes

endothelial cell-cell contacts (Costa et al., 2012). Cdh5 and other endothelial cell markers, such as the cell surface receptors Tie1, Kdr and Kdrl, all were expressed in the Fli1a, Etv2, Sox7 expressing mesenchymal cell population. Furthermore, these cells co-expressed the notch ligand Delta-Like 4 (Dll4), which is expressed in endothelial cells under the control of Sox7 and Notch (Sacilotto et al., 2013).

In summary, Adma might play multiple roles in the developing zebrafish heart. First, it might regulate processes within the epicardial sheet itself, potentially including epicardial proliferation, permeability of the epicardial cell layer or a feed-forward loop to reinforce the expression of Adma itself. Second, Adma might regulate mesenchymal cells that show commitment to the endothelial lineage, as underlined by the expression of several transcription factors and structural proteins that are known endothelial markers. Hence, Adma might drive the endothelial differentiation of mesenchymal progenitor cells in the heart, similar to what has been reported in embryonic stem cells (Yurugi-Kobayashi et al., 2006).



**Figure 4-5: Cells within epicardial cell cluster 1 might regulate mesenchymal cells expressing *Calcrla*, *Ramp2* and endothelial markers.** (a-c) FPKM expression values of the transcripts encoding *Adma*, *Calcrla* and *Ramp2* were overlaid onto the t-SNE clustering. (a) Expression of *Adma* in epicardial cluster 1. (b) Expression of *Calcrla* in epicardial clusters 1-3, in mesenchymal cells and in cardiomyocytes. (c) Expression of *Ramp2* in epicardial cluster 1 and in mesenchymal cells. (d-e) Heatmap of the clustered single cell data set (columns) showing genes of interest (rows). Expression values are FPKM, red color indicates high expression. The phylogenetic tree at the top summarizes the transcriptomic relationships of the single cells within the data set. The significance of each cell cluster is indicated by the height of the respective bifurcation event. White numbers above the heatmap indicate the nine most significant cell clusters. The annotation underneath indicates the cell type each cell cluster contains, as well as the color-coded origin of each single cell. The key to the color code is located at the bottom of the heatmap and indicates the reporter line each cell was isolated from and if it was a fluorescent cell (+) or a non-fluorescent cell (-) (d) Direct comparison of the expression patterns of *Adma*, *Calcrla* and *Ramp2*. (e) Expression of endothelial marker genes within the mesenchymal cell cluster. Epi = epicardial cell cluster, MC = mesenchymal cells, NC = neural cells, HC = haematopoietic cells, CM = cardiomyocytes, FPKM = fragments per kilobase of transcript per million transcripts.

### **4.13 Cells in epicardial cell cluster 2 transcriptionally resemble smooth muscle cells and might be regulated cell-autonomously by Wnt11 Related**

When analyzing genes that were enriched in epicardial cell cluster 2, I found that the cells in this cluster were not only labelled by the expression of ECM factors such as *Lox* (figure 4-6a), but also by the expression of the non-canonical Wnt ligand *Wnt11 Related* (*Wnt11r*) which regulates zebrafish heart morphogenesis (Choudhry and Trede, 2013) (figure 4-6b). *Wnt11r* was furthermore expressed in cells in the mesenchymal cell cluster as well as in a few cells within epicardial cell clusters 1 and 3. In *Xenopus*, *Wnt11-R* has been shown to regulate heart development through the receptor *Frizzled-7* (Abu-Elmagd et al., 2017), and *Frizzled 7a* (*Fzd7a*) was expressed in the data set, particularly in cells in epicardial cell cluster 1 (figure 4-6c).

Upon including other enriched genes in the investigation, I found that cells within epicardial cell cluster 2 expressed multiple smooth muscle cell markers such as *Mylka*, *Acta2*, *Myh11a* and *Transgelin* (*Tagln*) (figure 4-6d) as well as the ECM factors *Lox*, *Elnb*, *Acana* and *Fibulin 5* (*Fbln5*) (figure 4-6e). Thus, the transcriptomic profile suggests that the cells within epicardial cell cluster 2 are likely to be smooth muscle cells or myofibroblasts.

Zebrafish and *Xenopus* express *Wnt11r* and *Wnt11-R*, respectively, as a second *Wnt11* related factor, while mammals possess only a single *wnt11* gene (Choudhry and Trede, 2013; Garriock et al., 2005). *Wnt11* is known to activate the planar cell polarity pathway, mediating non-canonical Wnt signaling (Flaherty and Dawn, 2008). Canonical and non-canonical Wnt signaling occurs via *Frizzled* (*Fzd*) receptors (Flaherty and Dawn, 2008). In addition to *Fzd7a*, cells in the data set expressed *Fzd7b* and *Fzd8a*, the latter being known

to mediate the effects of mesodermal Wnt11r during zebrafish endodermal pouch formation (Choe et al., 2013) (figure 4-6f). Cells within epicardial cell cluster 1 furthermore expressed Fzd9a. The expression patterns of the planar cell polarity pathway components Receptor Tyrosine Kinase Like Orphan Receptor 2 (Ror2) (Gao et al., 2011) and Dishevelled-Associated Activator of Morphogenesis 1a (Daam1a) (Habas et al., 2001) resembled those of the Frizzled factors, them for example being present in cells in the epicardial cell clusters 1 and 2. Daam1 is necessary for the activation of the small GTPase Rho, which regulates intracellular actin dynamics (Habas et al., 2001), and Rhoaa expression in the data set overlapped with that of Daam1a. Activated Leukocyte Cell Adhesion Molecule a (Alcama) has been described as a downstream effector of Wnt11r (Choudhry and Trede, 2013) and showed expression in a majority of cells within epicardial cell cluster 2. Thus, multiple components of non-canonical Wnt signaling were expressed within the single cell transcriptomic data set, particularly in epicardial and mesenchymal cells. Wnt11r might be acting through this signaling cascade, potentially regulating several events during zebrafish heart development.

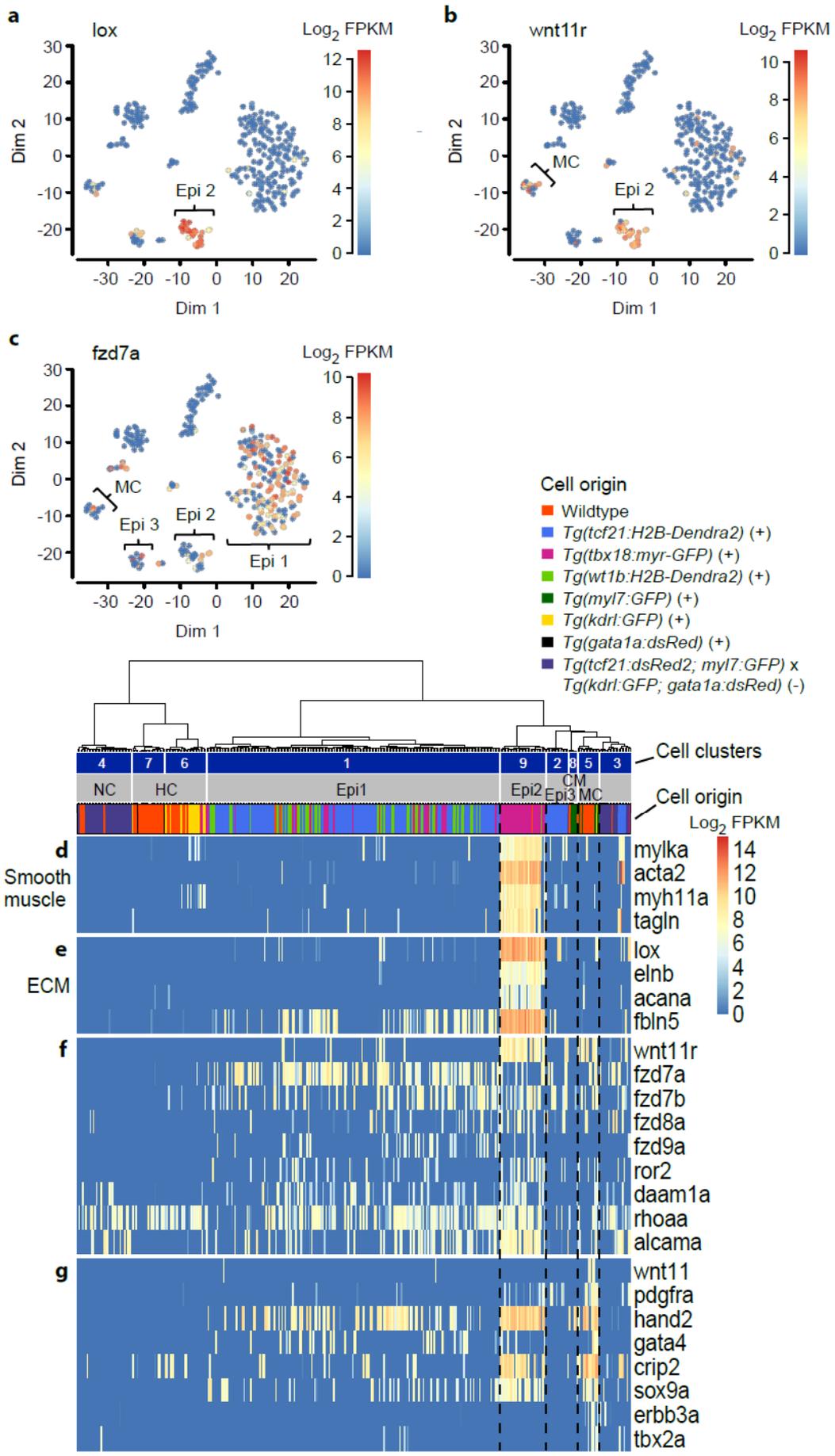
Some of the mesenchymal cells that expressed Wnt11r also expressed Wnt11, which is involved in the specification of cardiac tissue from mesoderm (Eisenberg and Eisenberg, 1999) (figure 4-6g). In mouse and frog, Wnt11 is expressed in the muscle layer of the heart (Garriock et al., 2005; Sinha et al., 2015). The subset of Wnt11r expressing cells within the mesenchymal cell cluster some of which co-expressed Wnt11 furthermore expressed the cardiac progenitor markers *Abi3bbp* (Hodgkinson et al., 2014) and *Pdgfra* (Chong et al., 2013). The additional expression of *Myl7* (figure 4-2) and the GATA Motif Binding Transcription Factor *Gata4* in these cells suggested they might be cardiomyocyte progenitors. The expression of *Myl7* was stronger in the cardiomyocyte cluster than in

the mesenchymal cell cluster, indicating that the latter did not possess a fully matured contractile machinery. In several vertebrate species, Gata4 promotes cardiomyocyte differentiation from progenitor cells (Grépin et al., 1995; Gupta et al., 2013; Haworth et al., 2008; Kuo et al., 1997). Furthermore, *gata4* driven transgenes were found to label de-differentiated cardiomyocytes that restore lost muscle tissue following cardiac injury in the zebrafish (Kikuchi et al., 2010), although this seems to not accurately reflect the endogenous expression of Gata4. Cells in the mesenchymal cell cluster that express Wnt11r and Wnt11 therefore might be differentiating into cardiomyocytes, which was supported by the expression of Hand2, playing a key role in cardiomyocyte formation (Schindler et al., 2014; Tsuchihashi et al., 2011). Hand2 and Gata4 physically interact to promote cardiomyocyte differentiation (Dai et al., 2002b). However, *wnt11r* positive cells in the mesenchymal cell cluster also expressed the transcription factors Crip2 and Sox9a, which are major regulators of cardiac valve formation (Garside et al., 2015; Kim et al., 2014). Sox9a regulates multiple other transcription factors driving heart valve development and induces the Neuregulin receptor Erbb3a to promote proliferation and maturation of mesenchymal cells in the endocardial cushions (Akiyama et al., 2004). Activation of Erbb3a is furthermore required for EMT during the formation of mammalian valve mesenchyme (Camenisch et al., 2002). In addition to Erbb3a, *wnt11r* positive mesenchymal cells expressed Tbx2a, which represses heart chamber-specific genes in the mouse embryo (Christoffels et al., 2004). In mouse and zebrafish embryonic hearts, expression of Tbx2/Tbx2a is restricted to the atrio-ventricular boundary and is required for the formation of the immature valves (Harrelson, 2004; Ribeiro et al., 2007). In conclusion, a subset of mesenchymal cells within the data set expressed markers of cardiomyocyte differentiation, but also markers of cardiac valve development. One

possible scenario is that these mesenchymal progenitors contribute to the cardiac valves, comparable to the formation of valve cushion mesenchyme in other species (De Lange et al., 2004). However, previous results argue against the presence of cushion mesenchyme in the developing zebrafish heart valves (Pestel et al., 2016; Scherz et al., 2008). The subset of mesenchymal cells described here might contribute to parts of the myocardium, such as the trabeculae, which are formed by cardiomyocytes delaminating from the ventricular wall (Staudt et al., 2014). Indeed, Wnt11 was shown to stimulate the differentiation of murine mesenchymal stem cells into cardiomyocytes (He et al., 2011). The action of Wnt11r and Wnt11 in mesenchymal cells might be cell-autonomous, as these cells co-expressed Wnt receptors such as Fzd7b.

Similarly, the prominent expression of Wnt11r in cells in epicardial cell cluster 2 might cell-autonomously regulate these cells via non-canonical Wnt signaling. Cells within epicardial cell cluster 2 expressed several intracellular components of non-canonical Wnt signaling as well as the Wnt11r downstream effector Alcama. In addition, mesenchymal Wnt11 might regulate cells in epicardial cell cluster 2, as mammalian Wnt11 is able to stimulate the expression of Acta2 in airway smooth muscle cells (Kumawat K et al., 2016). In conclusion, Wnt11r and Wnt11 might be involved both in mesenchymal cell differentiation and in smooth muscle formation from epicardial cells in the embryonic zebrafish heart.

Furthermore, Wnt11 over-expression in mesenchymal stem cells was shown to protect cardiomyocytes from hypoxia-induced apoptosis (Zuo et al., 2012). Therefore, Wnt11 expressing mesenchymal cells might play a beneficial role following cardiac injury in the adult zebrafish.



**Figure 4-6: Wnt11r might regulate cells in epicardial cell cluster 2 and cells within the mesenchymal cell cluster.** (a-c) FPKM expression values of transcripts encoding Lox, Wnt11r and Fzd7b were overlaid onto the t-SNE clustering. (a) Expression of Lox in epicardial cluster 2. (b) Expression of Wnt11r in epicardial cluster 2 and in mesenchymal cells. (c) Expression of Fzd7b in epicardial clusters 1-3 and in mesenchymal cells. (d-g) Heatmap of the clustered single cell data set (columns) showing genes of interest (rows). Expression values are FPKM, red color indicates high expression. The phylogenetic tree at the top summarizes the transcriptomic relationships of the single cells within the data set. The significance of each cell cluster is indicated by the height of the respective bifurcation event. White numbers above the heatmap indicate the nine most significant cell clusters. The annotation underneath indicates the cell type each cell cluster contains, as well as the color-coded origin of each single cell. The key to the color code is located at the bottom of the heatmap and indicates the reporter line each cell was isolated from and if it was a fluorescent cell (+) or a non-fluorescent cell (-) (d) Expression of smooth muscle cell markers. (e) Expression of ECM markers. (f) Expression of Wnt11r, the Frizzled receptors Fzd7b and Fzd8a, components of the non-canonical Wnt signaling pathway and the Wnt11r target Alcama. (g) Expression of Wnt11, cardiomyocyte and cardiac valve differentiation marker genes within the mesenchymal cell cluster. Epi = epicardial cell cluster, MC = mesenchymal cells, NC = neural cells, HC = haematopoietic cells, CM = cardiomyocytes, FPKM = fragments per kilobase of transcript per million transcripts, ECM = extracellular matrix.

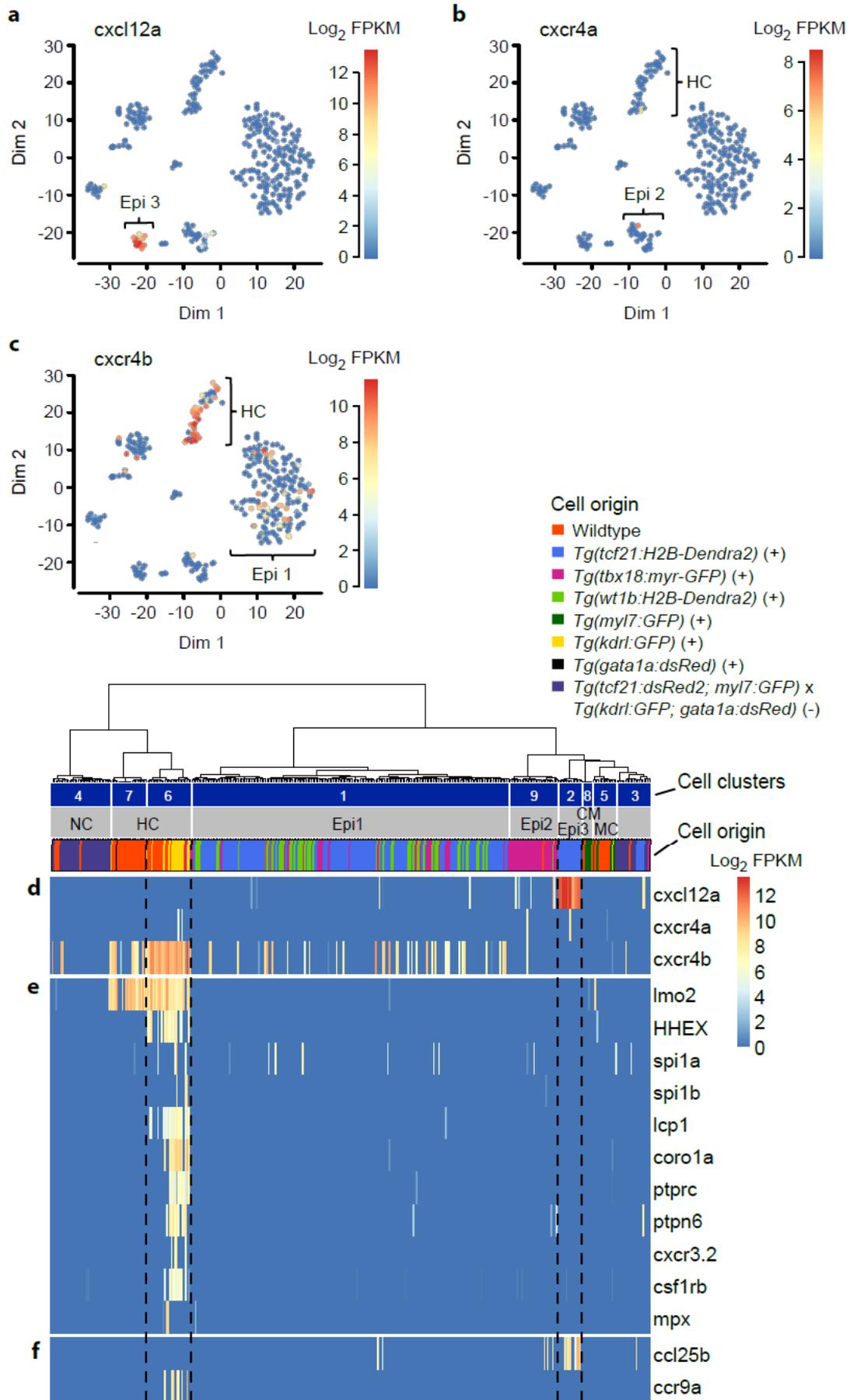
#### **4.14 Presence of *cxcl12a* in epicardial cell cluster 3 might guide myeloid haematopoietic cells into the developing zebrafish heart**

The strong enrichment of transcripts encoding Cxcl12a cells in epicardial cell cluster 3 (figure 4-7a) prompted me to more closely analyze chemokine receptor expression in the data set. The Cxcl12a receptors Cxcr4a and Cxcr4b were both expressed. While Cxcr4a was only expressed in a few haematopoietic and epicardial cells (figure 4-7b), the paralogue Cxcr4b had a much more widespread expression in the same cell clusters and was also present in a few neural cells (figure 4-7c). Closer investigation showed that Cxcr4b was most prominently expressed in the haematopoietic cell cluster 6 (figure 4-7d).

These cells co-expressed the LIM finger protein gene *Lmo2* and, to a lesser extent, the Haematopoietically Expressed Homeobox transcription factor *Hhex*, both of which are essential regulators of haematopoietic development (Paz et al., 2010; Yamada et al., 1998) (figure 4-7e). Furthermore, a subset of cells within cell cluster 6 expressed the E26-Transformation-Specific transcription factors *Spi1a* and *Spi1b*, the zebrafish orthologues of the mammalian *Spi1* (also named *Pu.1*) (Klemsz et al., 1990). Both *Spi1a* and *Spi1b* are essential regulators of leukocyte differentiation (Bukrinsky et al., 2009; Rhodes et al., 2005), similar to *Spi1* (DeKoter, 2000). Several factors involved in myeloid and lymphoid cell differentiation are regulated by *Spi1a* and *Spib* and were expressed in cell cluster 6: The actin bundling factor Lymphocyte Cytosolic Protein 1 / L-Plastin (*Lcp1*) (Herbomel et al., 1999), the actin-organizing factor Coronin 1a (*Coro1a*) (Rhodes et al., 2005), the tyrosine phosphatases CD45 (*Ptprc*) (Anderson et al., 2001) and *Ptpn6* (also named *Shp1*) (Wlodarski et al., 2007), as well as the zebrafish chemokine receptor *Cxcr3.2* (Zakrzewska

et al., 2010). In the latter study, *Cxcr3.2* was found to be myeloid-specific and to regulate macrophage migration. Supporting a macrophage fate, the same subset of cells expressed the colony-stimulating factor receptor *Csf1rb* (also named C-fms). *Csf1r*, the mammalian orthologue of *Csf1rb*, also is a downstream target of *Spi1* (Krysinska et al., 2007) and encodes an essential regulator of macrophage differentiation (Dai et al., 2002a). A subset of *lcp1* positive cells in cluster 6 was *csf1rb* negative and instead expressed Myeloperoxidase (*Mpx*), a marker of differentiated neutrophils (Lieschke et al., 2001), but also a marker of macrophage precursors in the mouse yolk sac (Bertrand et al., 2005). In summary, *Cxcr4b* is strongly expressed within cell cluster 6, which contains myeloid haematopoietic cells that might be a mixture of macrophages and neutrophils. The expression of the ligand *Cxcl12a* in epicardial cell cluster 3 indicates that these cells might be crucial to guide myeloid cell migration within or into the developing zebrafish heart. *Cxcr4b* was furthermore expressed in epicardial cell cluster 1, which raises the possibility that cells within epicardial cell cluster 3 guide cellular rearrangements within the main epicardial sheet. However, *Cxcr4b* was not expressed in cardiomyocytes, unlike what has been reported during adult zebrafish heart regeneration (Itou et al., 2012b). Transcripts encoding another chemokine ligand, *Ccl25b*, were enriched in cells within epicardial cell cluster 3 as well, while its unique receptor *Ccr9a* was expressed in *cxcr4b* positive myeloid cells (figure 4-7f). The mammalian homologue *Ccr9* is mainly expressed in leukocytes (Vicari et al., 1997) and regulates T cell development (Carramolino et al., 2001), but also the migration of inflammatory macrophages (Nakamoto et al., 2012). In the zebrafish embryo, endothelial *Ccl25b* regulates the homing and expansion of haematopoietic stem and progenitor cells in the caudal haematopoietic tissue (Xue et al., 2017). In the developing zebrafish heart, signaling through *Ccl25b-Ccr9a* might thus

reinforce the effect of Cxcl12a-Cxcr4b in guiding the migration of myeloid haematopoietic cells. Furthermore, Ccl25-Ccr9 signaling was found to mediate the infiltration of immune cells into the mouse heart following myocardial infarction (Huang et al., 2016). Therefore, the apparent guidance of immune cell migration by specialized epicardial cells in the developing zebrafish heart might be recapitulated in the adult heart following injury.



**Figure 4-7: Cells within epicardial cell cluster 3 might guide myeloid haematopoietic cells via Cxcl12a and Ccl25b mediated chemokine signaling.** (a-c) FPKM expression values of transcripts encoding Cxcl12a, Cxcr4a and Cxcr4b were overlaid onto the t-SNE clustering. (a) Expression of Cxcl12a in epicardial cluster 3. (b) Expression of Cxcr4a in epicardial cluster 2 and in haematopoietic cells. (c) Expression of Cxcr4b in epicardial cluster 1 and in haematopoietic cells. (d-f) Heatmap of the clustered single cell data set (columns) showing genes of interest (rows). Expression values are FPKM, red color indicates high expression. The phylogenetic tree at the top summarizes the transcriptomic relationships of the single cells within the data set. The significance of each cell cluster is indicated by the height of the respective bifurcation event. White numbers above the heatmap indicate the nine most significant cell clusters. The annotation underneath indicates the cell type each cell cluster contains, as well as the color-coded origin of each single cell. The key to the color code is located at the bottom of the heatmap and indicates the reporter line each cell was isolated from and if it was a fluorescent cell (+) or a non-fluorescent cell (-) (d) Direct comparison of the expression patterns of Cxcl12a, Cxcr4a and Cxcr4b. (e) Expression of myeloid cell differentiation markers in haematopoietic cell cluster 6. (f) Expression of Ccl25b in epicardial cell cluster 3 and Ccr9a in haematopoietic cell cluster 6. Epi = epicardial cell cluster, HC = haematopoietic cells, MC = mesenchymal cells, NC = neural cells, CM = cardiomyocytes, FPKM = fragments per kilobase of transcript per million transcripts.

## 4.15 Summary

I performed single cell RNA sequencing to uncover subpopulations of epicardial cells in the developing zebrafish heart at 5dpf. I found that the epicardium at this developmental time point indeed consisted of multiple transcriptomically distinct cell populations. The largest subpopulation, termed epicardial cell cluster 1, included cells derived from hearts of *tcf21*, *tbx18* and *wt1b* driven reporter embryos and these cells expressed several canonical epicardial markers, such as *Tcf21*, *Tbx18*, *Wt1b*, *Wt1a*, *Scxa*, *Sema3d* and *Aldh1a2*. Cells in epicardial cell cluster 1 was unique in the expression of the cell adhesion proteins *Podxl* and *Jam2b*, as well as in the expression of the signaling peptide *Adma*. By including non-epicardial cell types in the analysis, I identified a population of mesenchymal cells that expressed receptors for *Adma*. This cell population appeared to also belong to the endothelial lineage as it expressed several endothelial markers, such as *Fli1a*, *Etv2*, *Sox7*, *Cdh5*, *Tie1* and *Kdrl*. Therefore, epicardial-derived *adma* might regulate the development of a cell population in the developing zebrafish heart which either differentiates from mesenchyme towards endothelium or de-differentiates from endothelium towards a mesenchymal state.

A second epicardial subpopulation, epicardial cell cluster 2, almost exclusively contained cells derived from *tbx18* driven reporter hearts which expressed *Tbx18* endogenously, but lacked expression of *Tcf21*, *Wt1b*, *Wt1a*, *Scxa* and *Aldh1a2*, with sparse expression of *Sema3d*. Instead, epicardial cell cluster 2 was highly enriched both in components of the smooth muscle contractile machinery, such as *Mylka*, *Acta2* and *Myh11a*, and in ECM factors, such as *Lox*, *Elnb* and *Fbln5*. Additionally, cells in epicardial cell cluster 2 expressed transcripts encoding the non-canonical Wnt ligand *Wnt11r*, the Wnt receptors

Fzd7b and Fzd8a and the Wnt11r downstream effector Alcam. Another Wnt ligand, Wnt11, was expressed by a subset of mesenchymal cells that was different from the Adma responsive subset. These mesenchymal cells expressed Gata4, Hand2 and Myl7, markers of differentiating cardiomyocytes, but they also expressed drivers of cardiac valve formation, such as Crip2, Sox9a and Tbx2a. In conclusion, epicardial cell cluster 2 most likely comprises a population of ECM producing smooth muscle cells that might be regulated by Wnt11r and by Wnt11 expressed by un- or de-differentiated cardiomyocytes potentially contributing to the cardiac valves.

A third epicardial subpopulation, epicardial cell cluster 3, contained cells derived from *tcf21* driven reporter hearts some of which were labelled by the expression of Tcf21 and Wtb1, but did not express Tbx18. Cells in epicardial cell cluster 3 were enriched in transcripts encoding retinoid acid receptors, such as Nr1h5, the tight junction protein Cldn11a and the chemokine ligand Cxcl12a. The gene encoding the Cxcl12a receptor Cxcr4b was strongly expressed in a cluster of myeloid haematopoietic cells expressing Lmo2, Spi1a, Spi1b, Lcp1, Ptprc and Csf1rb. Additionally, the chemokine ligand Ccl25b was expressed in cells in epicardial cell cluster 3, with the receptor Ccr9a being expressed exclusively in *cxcr4b* positive myeloid cells. Therefore, epicardial cell cluster 3 appears to comprise cells that are likely responsive to retinoic acid produced in epicardial cell cluster 1 and that might guide the migration of myeloid haematopoietic cells in or into the developing zebrafish heart.

In conclusion, the epicardial marker gene expression suggests that epicardial cell cluster 1 represents the cells forming the epicardial cell sheet, while cells in epicardial cell clusters 2 and 3 appear to have a distinct phenotype and function. Whereas Tcf21, Tbx18 and Wt1b were not expressed in all epicardial cell clusters, transcripts encoding Rbp4

were exclusive to all epicardial cells and might therefore be a valid pan-epicardial marker gene. Furthermore, Rbp4 potentially has a function that is required by all epicardial cells. All three epicardial subpopulations expressed transcripts encoding markers of EMT, with *twist1a* prominent in epicardial cell cluster 2 and *snai1a*, *snai1b* and *snai2* prominent in epicardial cell cluster 3. This argues for a migratory phenotype of developing epicardial cells.

Additionally, single cell transcriptomics identified clusters of cells that might be neurons and neuronal support cells present within or attached to the zebrafish heart at 5dpf, which has not been described before.

## Chapter 5

### Validation of the cell populations identified via single cell transcriptomics

#### 5.1 Introduction

Cell populations identified through single cell transcriptomics might be a product of technical variability instead of being a product of true biological variation (Grün et al., 2014). To confirm the validity of the epicardial cell clusters identified through single cell transcriptomics, I sequenced bulk epicardial cDNA and compared the bulk sequencing data to the single cell transcriptomic data. Furthermore, I used HCR (Choi et al., 2014a) and immunofluorescence to visualize the expression pattern of the epicardial subpopulation specific markers *Adma*, *Elnb*, *Mylka* and *Cldn11a* in whole-mount zebrafish embryos, focused on the heart region. Finally, I performed ATAC sequencing (Buenrostro et al., 2013) of epicardial and control bulk samples to identify putative enhancer elements that might regulate the expression of marker genes in the epicardial cell populations.

The transcriptomic data obtained from epicardial bulk samples was indeed very similar to that obtained from the epicardial single cell cDNA samples. Marker genes enriched in the different epicardial subpopulations, such as *adma*, *lox* and *cldn11a*, were also enriched in the epicardial bulk samples, indicating the presence of the single cell populations in the bulk samples. HCR showed that *adma* and *cldn11a* transcripts were expressed in the epicardial layer of the developing zebrafish heart, while *elnb* transcripts were expressed in the bulbus arteriosus (BA). Additionally, immunofluorescence staining showed that the expression of *tbx18:myr-Citrine* and *Mylka* overlapped in cells within the smooth muscle

layer of the BA. Analyzing the generated ATAC data, I identified three putative enhancer elements that showed an increased accessibility in epicardial samples versus control samples and had a genomic location close to the loci of *podxl*, a marker of epicardial cell cluster 1, as well as *lox* and *elnb*, markers of epicardial cell cluster 2. I used a reporter assay to visualize the *in vivo* activity of these putative enhancer elements. Corroborating the single cell transcriptomic data, the regulatory element located close to *podxl* was active in the epicardial cell layer, while the regulatory elements located close to *lox* and *elnb* were active in cells within the BA. These findings, together with HCR and immunofluorescence studies, validate the presence of the epicardial subpopulations identified through single cell transcriptomics and confirm the spatial location of epicardial cell clusters 1 and 3 in the epicardial cell layer, as well as the spatial location of epicardial cell cluster 2 within the smooth muscle layer of the BA.

## **5.2 Transcriptomic analysis of epicardial bulk samples derived from the developing zebrafish heart**

To generate transcriptomic data from epicardial cell bulks, I FACS-purified triplicate bulks of up to 1,000 fluorescent cells from *Tg(tcf21:H2B-Dendra2)*, *Tg(tbx18:H2B-Dendra2)*, *Tg(tbx18:myr-GFP)* and *Tg(wt1b:H2B-Dendra2)* hearts (workflow in figure 5-1a). I used *Tg(tbx18:H2B-Dendra2)* hearts for the first two bulk replicates, but found that the relative numbers of fluorescent cells purified were lower than what I expected following the microscopic results described in chapter 3. Therefore, I switched to *Tg(tbx18:myr-GFP)* to process the third bulk replicate and found that, using the same FACS gate, the relative number of fluorescent cells was indeed considerably higher than with

*Tg(tbx18:H2B-Dendra2)*. As a control, I also purified 1,000 non-fluorescent cardiac cells with every fluorescent cell bulk I collected.

As a quality control step following sequencing, I quantified the library size and the number of expressed genes in each bulk sample (figure 5-1b). All bulks, both the ones derived from fluorescent cells ([+]) and the control bulks ([-]), were clustering relatively close together and ranged between 5,000,000 and 9,000,000 read counts in size. Only the third bulk replicate of cells derived from *Tg(wt1b:H2B-Dendra2)* hearts had a larger size of almost 12,000,000 read counts, possibly due to an elevated concentration of this library in the sequencing pool. All bulk samples expressed between 23,000 and 30,000 different genes with no obvious differences between fluorescent cell derived bulks and control bulks. Generally, the epicardial bulk samples showed a higher variability in library size and complexity than the control samples, possibly because I processed varying cell numbers in the individual epicardial samples. However, there was no bias in library size or number of expressed genes clearly separating different conditions in the bulk data set. Next, I analyzed the general similarity between the different bulk samples. Computing principal components (PCs) that described the variability within the bulk data set (comparable to the t-SNE clustering used for single cell samples), I found that the fluorescent cell derived bulks and the control bulks formed clearly separated clusters (figure 5-1c). This separation could be attributed to PC1, which accounted for 79% of the variability within the data set. PC2, the next-significant component, accounted for 7% of the variability and separated the *tbx18* reporter derived fluorescent bulks from the other epicardial bulk samples, indicating they might be derived from cells with a slightly different transcriptomic profile than that of the cells the other epicardial bulk samples were derived from. The fluorescent cell derived bulks purified from *Tg(tbx18:H2B-*

*Dendra2*) hearts showed a particularly strong separation from all other epicardial bulks, including the one purified from *Tg(tbx18:myr-GFP)* hearts, suggesting that the *tbx18* reporter line used for cell purification affected the transcriptomic profile of the resulting bulk data.

To identify genes that were enriched or depleted in the epicardial bulk samples, I performed differential gene expression analysis. Comparing bulks derived from *tcf21:H2B-Dendra2* expressing cells to the respective control bulks, I found that 2079 genes were significantly enriched in *tcf21:H2B-Dendra2* expressing cells and 3374 genes were depleted (figure 5-1d). *adma* and other marker genes of cells in epicardial cell cluster 1, such as *jam2b* and *Irrn4* (*CU984579.1*), (figure 4-2) were among the ten most enriched genes in *tcf21:H2B-Dendra2* expressing cells. This suggests that most of the *tcf21:H2B-Dendra2* expressing cells across the three bulk replicates had a transcriptomic profile similar to that of the cells within epicardial cell cluster 1, as identified through single cell transcriptomics. In contrast, marker genes of the *tcf21* expressing cells in epicardial cell cluster 3, such as *cldn11a* and *cxcl12a* (figure 4-2), were not among the most significantly enriched genes within *tcf21:H2B-Dendra2* expressing cell bulks, possibly because this subpopulation was too small to be detected in the bulk analysis.

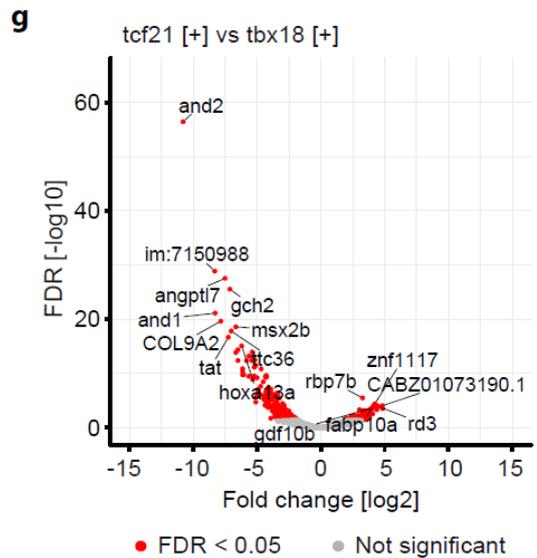
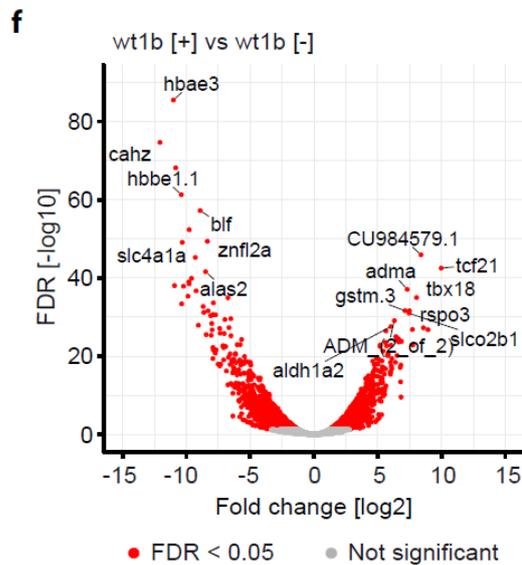
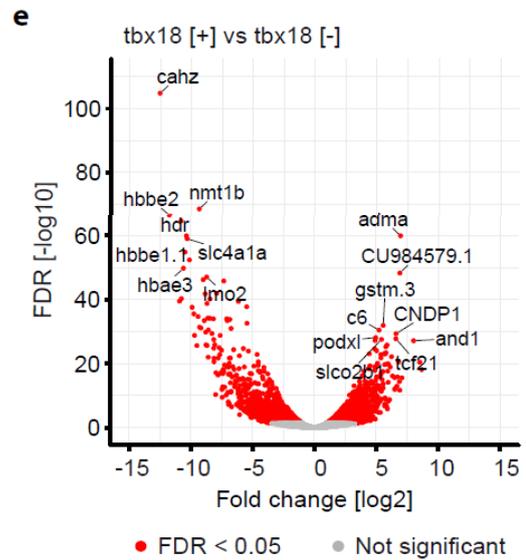
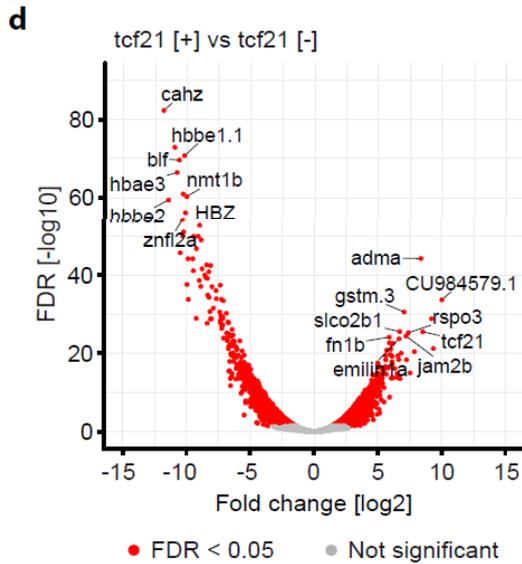
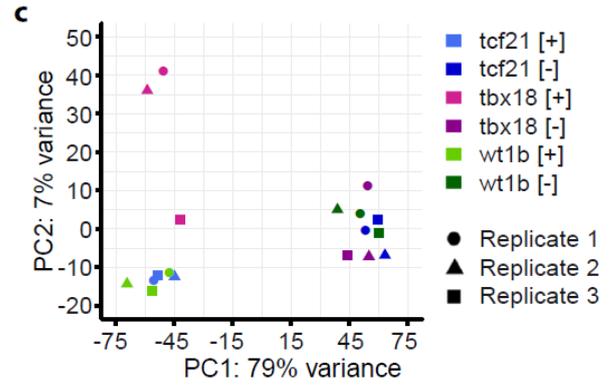
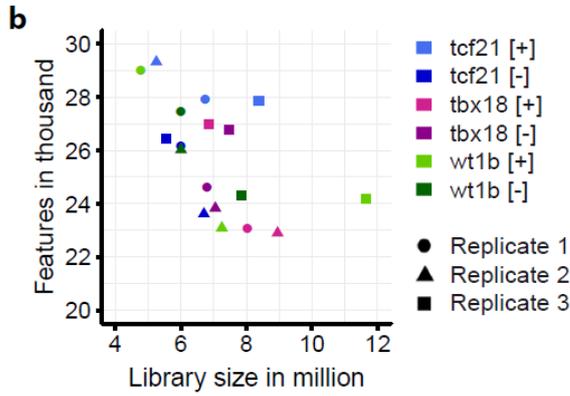
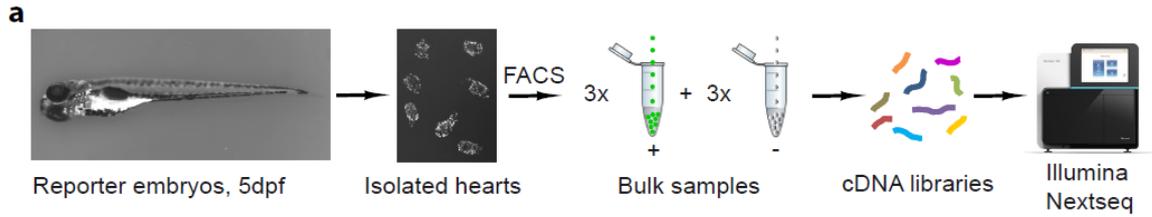
In bulk samples derived from *tbx18* reporter expressing cells, 2660 genes were enriched versus control and 2620 were depleted (figure 5-1e). Again, *adma* was the most significantly enriched gene, fitting the fact that the majority of cells within epicardial cell cluster 1 expressed *Tbx18* (figure 4-4). Other markers of cells in epicardial cell cluster 1, such as *podxl* and *Irrn4*, were also strongly enriched, while genes encoding markers of the *Tbx18* expressing cells in epicardial cell cluster 2, such as *Lox* and *Mylka* (figure 4-2), were not among the most significantly enriched. As with the *tcf21* condition, this suggests that

most of the *tbx18* reporter expressing cells across the three bulk replicates had a transcriptomic profile similar to that of the cells within epicardial cell cluster 1.

In bulk samples derived from *wt1b*:H2B-Dendra2 expressing cells, 1941 genes were enriched versus control and 2339 were depleted (figure 5-1f). As before, genes enriched in epicardial cell cluster 1, such as *adma* and *aldh1a2*, were strongly enriched in the bulk samples as well. This fits the fact that *Wt1b* was almost exclusively expressed in cells within epicardial cell cluster 1 (figure 4-4). In summary, marker genes of cells in epicardial cell cluster 1, particularly *adma*, were among the most significantly enriched genes in all epicardial bulk samples. This supports the validity of epicardial cell cluster 1. Erythroid haematopoietic cell markers, such as *cahz*, *blf* and several *globin* genes, were strongly depleted in all epicardial bulk samples. This suggests that most of the cardiac control cells were circulating red blood cells.

To test whether marker genes of cells in epicardial cell cluster 3 were enriched in bulk samples derived from *tcf21*:H2B-Dendra2 expressing cells and whether marker genes of epicardial cell cluster 2 were enriched in bulk samples derived from *tbx18* reporter expressing cells, I directly compared bulk samples of the two conditions (figure 5-1g). I found that 142 genes were significantly enriched in bulk samples derived from *tcf21*:H2B-Dendra2 expressing cells, while 458 genes were depleted. *fabp10a*, a marker of cells within epicardial cell cluster 3 (figure 4-2), was among the most significantly enriched genes, suggesting that a cell population similar to epicardial cell cluster 3 might indeed be contained within *Tg(tcf21:H2B-Dendra2)* cell bulks. In contrast, genes highly enriched in epicardial cell cluster 2 were not among the most significantly depleted genes when comparing *Tg(tcf21:H2B-Dendra2)* derived to *tbx18* reporter derived cell bulks. However, *angptl7*, a gene expressed in cells in epicardial cell cluster 2 and in the mesenchymal cell

cluster (figure 4-2), was strongly depleted in *tcf21* cell bulks, meaning that it was significantly enriched in the *tbx18* reporter derived cells. *col9a2*, another gene expressed in cells in the mesenchymal cell cluster (data not shown), was also strongly depleted in *tcf21* cell bulks. There is hence the possibility that *col9a2* was present in cells in epicardial cell cluster 2, albeit at a level that was too low to be detected in the single cell samples. In conclusion, several components of the transcriptomic profile characterizing cells in epicardial cell cluster 1, identified through single cell sequencing, were highly enriched within all epicardial bulk samples. Furthermore, single components of the transcriptomic profile characterizing cells in epicardial cell clusters 2 and 3 were expressed within epicardial cell bulks in a manner that fits the expression of *Tcf21* and *Tbx18* in the single cell data set.



**Figure 5-1: Epicardial bulk RNA sequencing to confirm single cell transcriptomics in the embryonic zebrafish heart.** (a) Workflow to generate bulk cDNA libraries from FACS-purified epicardial (+) and control (-) cells, derived from isolated epicardial reporter hearts at 5dpf. 3 replicates (3x) were processed per condition. (b) Library sizes and complexities of epicardial and control bulk samples. The legend indicates the genotype of the cells contained within each bulk sample. (c) Principal component (PC) analysis plot describing epicardial and control bulk samples. The legend indicates the genotype of the cells contained within each bulk. (d) Volcano plot showing depletion and enrichment of genes in *Tg(tcf21:H2B-Dendra2)* derived fluorescent bulk samples, as compared to non-fluorescent control samples. (e) Volcano plot showing depletion and enrichment of genes in *Tg(tbx18:H2B-Dendra2)/Tg(tbx18:myr-GFP)* derived fluorescent bulk samples, as compared to non-fluorescent control samples. (f) Volcano plot showing depletion and enrichment of genes in *Tg(wt1b:H2B-Dendra2)* derived fluorescent bulk samples, as compared to non-fluorescent control samples. (g) Volcano plot showing depletion and enrichment of genes in *Tg(tcf21:H2B-Dendra2)* derived fluorescent bulk samples, as compared to *Tg(tbx18:H2B-Dendra2)/Tg(tbx18:myr-GFP)* derived fluorescent bulk samples. Significantly depleted or enriched genes (FDR<0.05) are colored red. tcf21 indicates samples derived from *Tg(tcf21:H2B-Dendra2)*, tbx18 indicates samples derived from either *Tg(tbx18:H2B-Dendra2)* (bulks 1 and 2) or *Tg(tbx18:myr-GFP)* (bulk 3) and wt1b indicates samples derived from *Tg(wt1b:H2B-Dendra2)*. [+] indicates samples comprising fluorescent cells, [-] indicates control samples comprising non-fluorescent cells. Symbols indicate the individual replicates of each condition. PC = principal component, FDR = false discovery rate.

### **5.3 Epicardial bulk transcriptomic data support the single cell RNA sequencing results**

To further investigate if the bulk derived transcriptomic data provided support for the epicardial subpopulations identified through single cell transcriptomics, I directly compared both data sets. Hence, I summed the read counts of all cells contained within each individual single cell cluster to generate representative pseudo-bulks. I then combined bulk samples and pseudo-bulks and performed PC analysis (figure 5-2a). As with bulk samples only, the most significant PC separated epicardial bulk samples from control bulk samples, accounting for 49% of the variability in the data set. Importantly, all pseudo-bulks representing epicardial single cell clusters were very similar to the epicardial bulk samples. Particularly the pseudo-bulk representing epicardial cell cluster 1 highly matched the epicardial bulk samples, reflecting the findings described in figure 5-1d-f. PC2 separated the other pseudo-bulks from the bulk samples, accounting for 20% of the variability in the data set. The pseudo-bulks representing neural cells (cluster 4) and cardiomyocytes (cluster 8) clustered particularly distinct regarding PC2. Regarding PC1, these pseudo-bulks and the one representing haematopoietic cells (clusters 7 and 6) were very similar to the control bulk samples. The pseudo bulk representing haematopoietic cells clustered closest to the control bulk samples, again suggesting that a large number of cells in the control bulk samples were circulating blood cells. Interestingly, the pseudo bulk that represented mesenchymal cells (cluster 5) clustered closer to the epicardial bulk samples than to the control bulk samples, supporting the transcriptional similarity of epicardial and mesenchymal cells that was apparent in the t-SNE clustering (figure 4-3). In summary, the general transcriptional profiles of the

epicardial bulk samples were very similar to those of pseudo-bulks representing the epicardial cell clusters identified through single cell transcriptomics, confirming the latter as being specific to epicardial cells. In contrast, the transcriptional profiles of pseudo-bulks representing non-epicardial cardiac cell types, such as cardiomyocytes, shared similarities with those of control bulk samples.

To test the validity of the transcriptional profiles characterizing each of the three epicardial cell clusters identified through single cell transcriptomics, I confirmed the expression of marker genes in the epicardial bulk samples. All epicardial bulk samples were enriched in canonical epicardial marker genes, such as *tcf21*, *tbx18* and *wt1b* (figure 5-2b).

In the single cell data set, these canonical epicardial marker genes were enriched within epicardial cell cluster 1. Moreover, markers that were specifically expressed in cells within epicardial cell cluster 1, such as *Adma*, were highly expressed in all epicardial bulk samples as well (figure 5-2c). However, the epicardial bulk samples derived from *tbx18* reporter hearts expressed these marker genes at a slightly lower level than the bulk samples derived from *tcf21* and *wt1b* reporter hearts. This was most evident in bulk replicate 3, which was derived from *Tg(tbx18:myr-GFP)*. The strong enrichment of marker genes such as *adma* within all epicardial bulk samples supports the validity of the transcriptional profile characterizing cells in epicardial cell cluster 1.

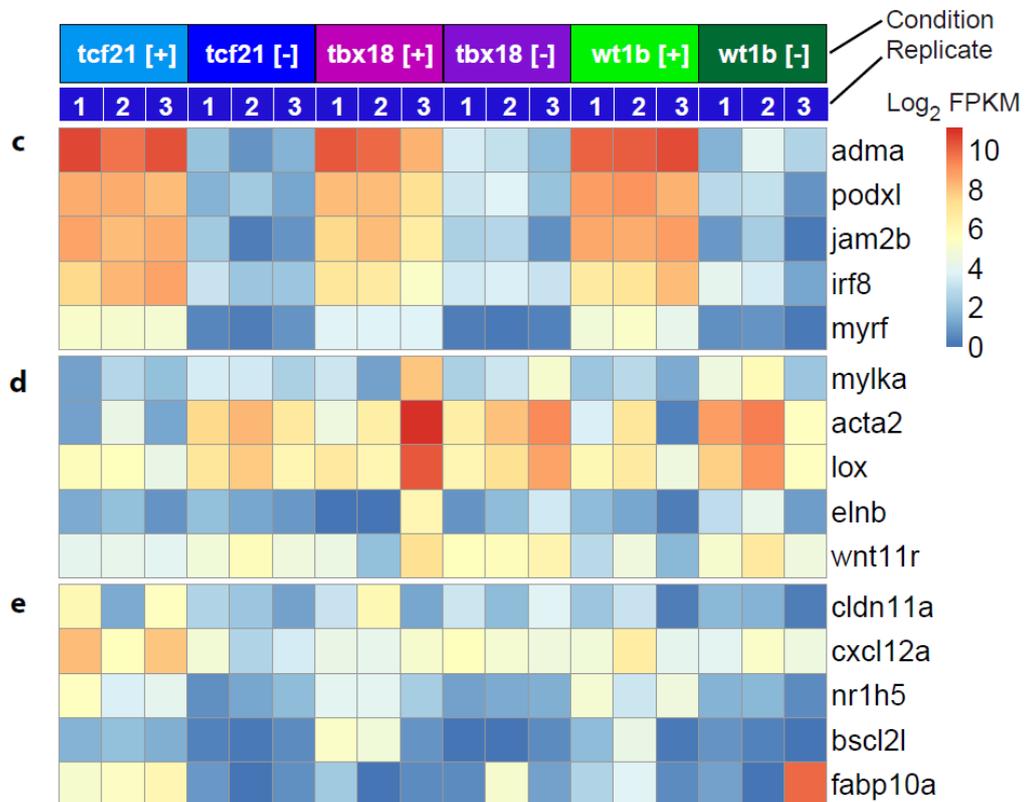
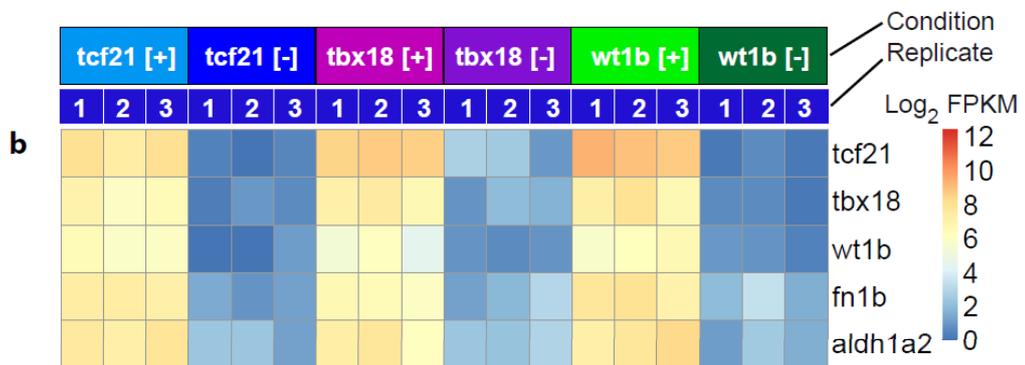
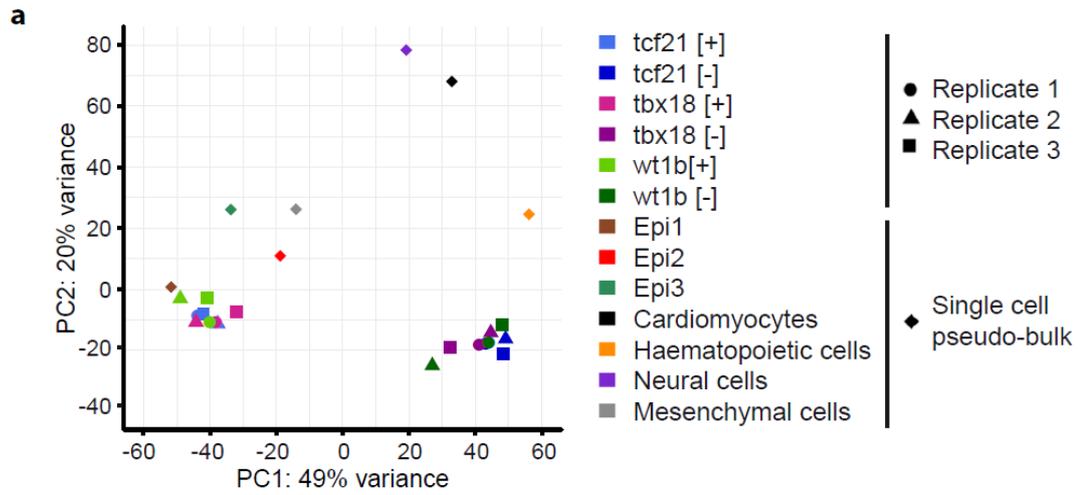
Marker genes of epicardial cell cluster 2, in which most cells expressed *Tbx18* but not *Tcf21* and *Wt1b*, were only weakly expressed in epicardial bulk samples derived from *tcf21* and *wt1b* reporter hearts (figure 5-2d). Even within the epicardial bulk samples derived from *tbx18* reporter hearts, marker genes such as *lox* were only enriched in bulk replicate 3. This discrepancy might be connected to the fact that the first two bulk

replicates were derived from *Tg(tbx18:H2B-Dendra2)* hearts and were generally different from all other epicardial bulk samples (figure 5-1c). As the relative number of fluorescent cells purified from *Tg(tbx18:H2B-Dendra2)* hearts was lower than that purified from *Tg(tbx18:myr-GFP)* hearts, it might be that cells with the transcriptomic profile of cells in epicardial cell cluster 2 were not FACS-purified in bulk replicates 1 and 2, due to low Dendra2 reporter expression or fluorescence intensity. In any case, the strong enrichment of several marker genes, among them *mylka* and *lox*, in the epicardial bulk sample that was derived from *Tg(tbx18:myr-GFP)* hearts supports the validity of the transcriptomic profile characterizing cells in epicardial cell cluster 2, which also was derived from *Tg(tbx18:myr-GFP)*. Also, several genes not present in cells within epicardial cell cluster 2, such as *adma*, showed a decreased expression in the *Tg(tbx18:myr-GFP)* derived bulk sample. In conclusion, the epicardial bulk sample derived from *Tg(tbx18:myr-GFP)* reporter hearts is likely to contain a mixture of cells similar to those in epicardial cell clusters 1 and 2. The fact that marker genes such as *mylka* and *lox* were expressed in control bulks, even in that obtained from *Tg(tbx18:myr-GFP)* hearts, suggests that there are non-epicardial cells within the developing zebrafish heart which share the transcriptomic profile of cells in epicardial cell cluster 2.

Transcripts encoding markers of cells in epicardial cell cluster 3, such as *Cldn11a* and *Cxcl12a*, were only enriched in epicardial bulk samples derived from *Tg(tcf21:H2B-Dendra2)*, consistent with the fact that many cells within this single cell cluster expressed *Tcf21* but mostly lacked expression of *Tbx18* and *Wt1b*. Most of these markers were only expressed at low levels in the control bulk samples, apart from *Fabp10a* which was highly expressed in bulk replicate 3 of the control samples derived from *Tg(wt1b:H2B-Dendra2)* hearts. In conclusion, the elevated expression levels of markers such as *Cldn11a* and

Cxcl12a in epicardial bulk samples derived from *Tg(tcf21:H2B-Dendra2)* hearts support the validity of the transcriptomic profile characterizing cells in epicardial cell cluster 3.

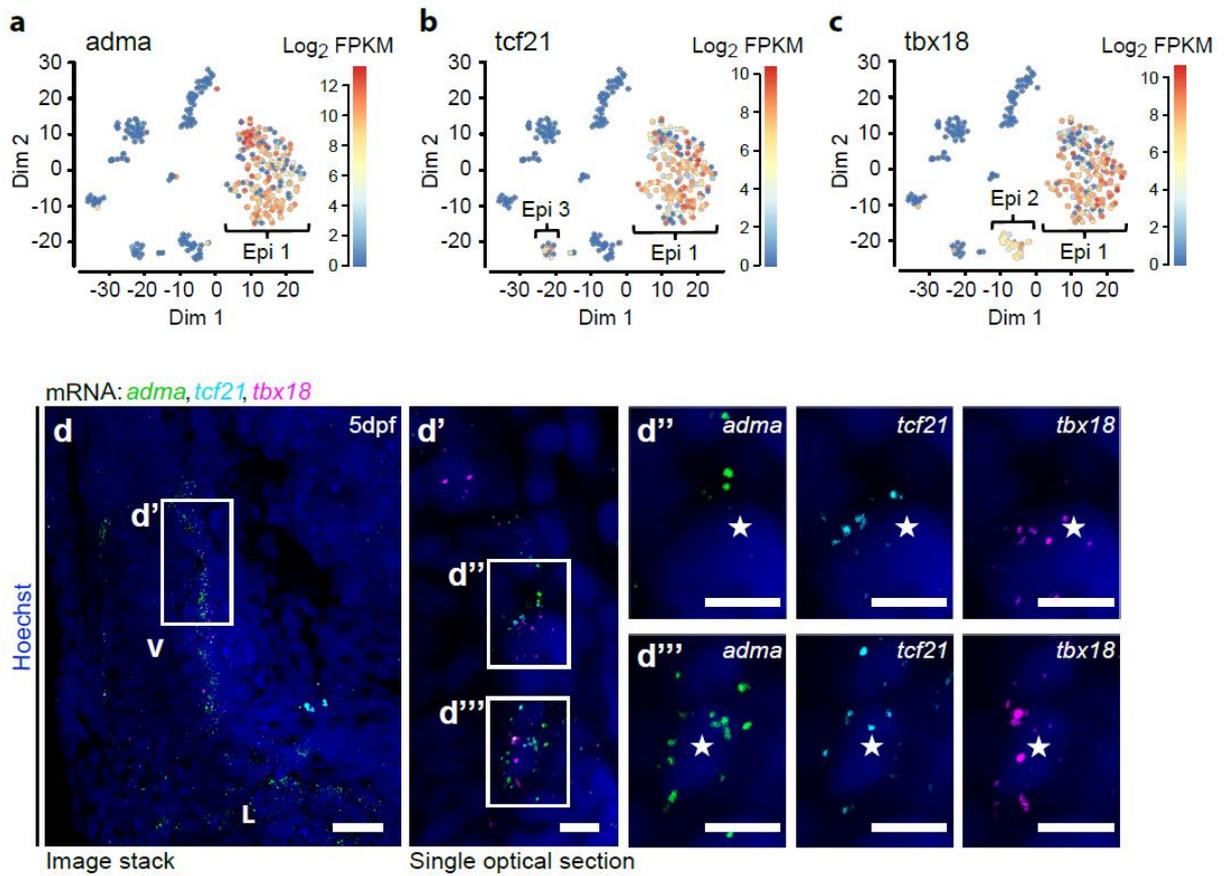
In summary, multiple marker genes identifying cells in the three epicardial subpopulations in the single cell transcriptomic data were also enriched in epicardial bulk samples. Furthermore, the heterogeneous expression of *Tcf21*, *Tbx18* and *Wt1b* in cells within the epicardial single cell clusters correlated with the heterogeneous marker gene expression within the *tcf21*, *tbx18* and *wt1b* reporter derived epicardial bulk samples. This suggests that the transcriptomic profiles of the epicardial subpopulations identified through single cell transcriptomics are valid and specific.



**Figure 5-2: Epicardial bulk transcriptomics support the validity of the epicardial subpopulations identified through single cell transcriptomics.** (a) PC analysis plot showing epicardial and control bulk samples as well as pseudo-bulks generated by combining cells within clusters of the single cell data set. The legend indicates the genotype of the cells contained within each bulk or pseudo-bulk sample. Epi1 = epicardial cell cluster 1, Epi2 = epicardial cell cluster 2, Epi3 = epicardial cell cluster 3. Symbols indicate the individual replicates of each condition. (b) Heatmap showing the expression of canonical epicardial markers in epicardial and control bulk samples. Colored bars indicate the condition each bulk belongs to, white numbers indicate replicate numbers. (c) Heatmap showing the expression of marker genes enriched in epicardial cell cluster 1. (d) Heatmap showing the expression of marker genes enriched in epicardial cell cluster 2. (e) Heatmap showing the expression of marker genes enriched in epicardial cell cluster 3. Red color indicates high expression. *tcf21* indicates samples derived from *Tg(tcf21:H2B-Dendra2)*, *tbx18* indicates samples derived from either *Tg(tbx18:H2B-Dendra2)* (bulks 1 and 2) or *Tg(tbx18:myr-GFP)* (bulk 3) and *wt1b* indicates samples derived from *Tg(wt1b:H2B-Dendra2)*. [+] indicates samples comprising fluorescent cells, [-] indicates control samples comprising non-fluorescent cells. PC = principal component, FPKM = fragments per kilobase of transcript per million transcripts.

#### **5.4 Epicardial cell cluster 1 represents most of the epicardial cell layer that envelops the developing zebrafish heart**

To further validate the identity of epicardial cell cluster 1 and to investigate its location and distribution in the developing zebrafish heart, I performed HCR on key genes expressed in this subpopulation. The expression patterns of *adma* (figure 5-3a), *tcf21* (figure 5-3b) and *tbx18* (figure 5-3c) showed that there was a large number of cells within epicardial cell cluster 1 co-expressing these factors. Indeed, transcripts encoding Adma were present in a large portion of the epicardial cell layer surrounding the ventricle at 5dpf (figure 5-3d). Transcripts encoding Adma were not present in other parts of the heart, but were present in the liver. Upon closer analysis, *adma* could be detected in close proximity to *tcf21* and *tbx18* (figure 5-3d'), raising the possibility that Adma, Tcf21 and Tbx18 might be co-expressed within cardiac cells. Supporting this possibility, several individual cell nuclei were surrounded by all three transcript species (figure 5-3d'' and figure 5-3d'''). These findings suggest that the majority of the epicardial cell layer enveloping the developing zebrafish heart is formed by cells that express Adma, and that these cells might express Tcf21 and Tbx18 as well. This supports the validity of epicardial cell cluster 1. However, the lack of a membrane staining precluded a definitive conclusion as to whether Adma, Tcf21 and Tbx18 were indeed co-expressed in individual cells.



**Figure 5-3: Transcripts encoding the epicardial cell cluster 1 marker *Adma* label the epicardial cell layer.** (a-c) FPKM expression values of transcripts encoding *Adma*, *Tcf21* and *Tbx18* were overlaid onto the single cell t-SNE clustering. (a) Expression of *Adma* in epicardial cluster 1. (b) Expression of *Tcf21* in epicardial clusters 1 and 3. (c) Expression of *Tbx18* in epicardial clusters 1 and 2. (d) Microscopic expression analysis of *adma* in wildtype embryos at 5dpf using hybridization chain reaction. **d'** shows the atrio-ventricular boundary in a higher magnification single optical section. Enlarged images in **d''** and **d'''** visualize the close proximity of *adma*, *tcf21* and *tbx18* transcripts surrounding an individual cell nucleus (asterisk). Scale bar in **d** is 20µm, scale bars in **d' - d'''** are 5µm. Epi = epicardial cell cluster, FPKM = fragments per kilobase of transcript per million transcripts, V = ventricle, L = liver.

## 5.5 Epicardial cell cluster 2 resides within the bulbus arteriosus of the developing zebrafish heart

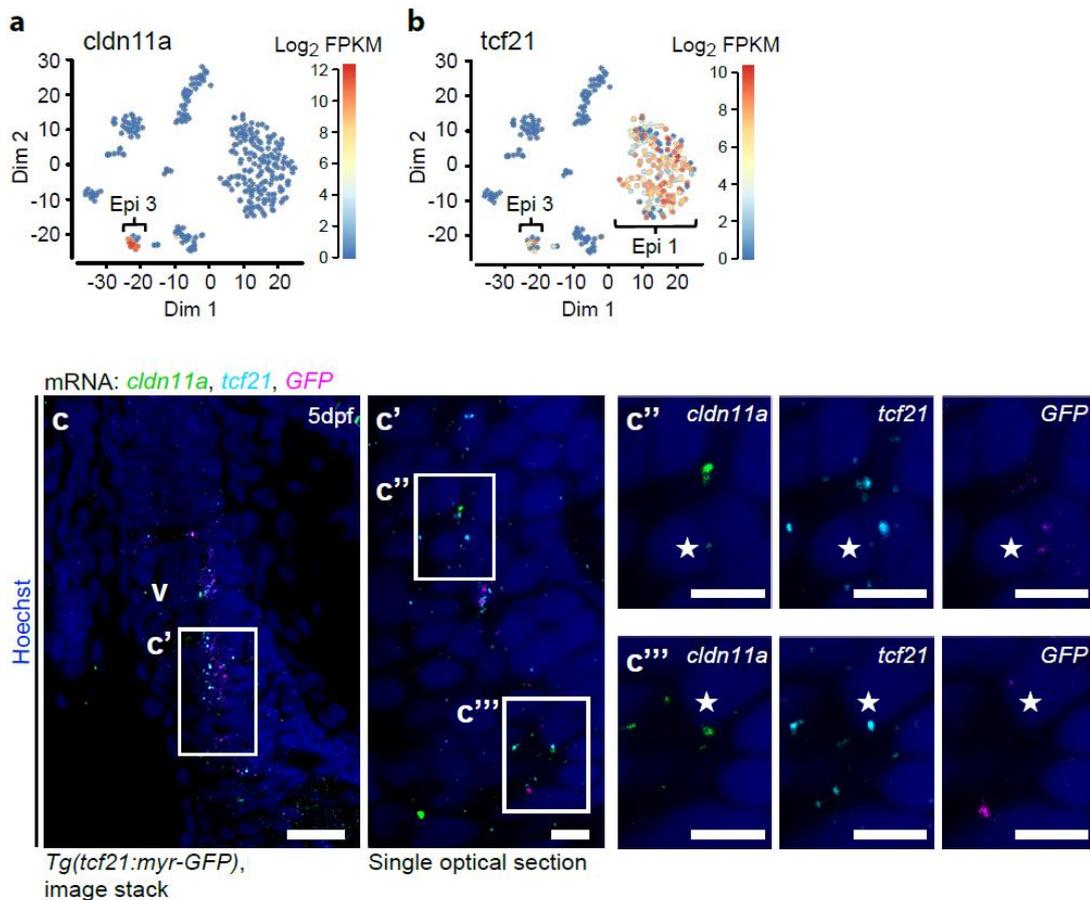
I investigated the presence and location of epicardial cell cluster 2 using HCR and immunofluorescence. Cells within epicardial cell cluster 2 co-expressed ECM molecules such as *Elnb* (figure 5-4a), smooth muscle markers such as *Mylka* (figure 5-4b) and *Tbx18* (figure 5-4c). Visualizing *elnb* transcripts (figure 5-4d) and *Mylka* protein (figure 5-4e) matched published findings (Moriyama et al., 2016) in that both genes were expressed exclusively within the BA, and not within other parts of the developing zebrafish heart. *tbx18* transcripts were mostly located within the epicardial cell layer, but a few transcripts were detected in cells within the BA that were also labelled by *elnb* (figure 5-4d' and figure 5-4d''). Furthermore, *tbx18* driven myr-Citrine labelled the plasma membrane of multiple *Mylka* positive cells within the BA (figure 5-4e' and figure 5-4e''). These findings suggest that cells expressing *Tbx18* are present in the developing zebrafish bulbus arteriosus and that these cells might express *Mylka* and *Elnb* as well. The small number of *tbx18* transcripts and the low fluorescence level of *tbx18* driven myr-Citrine in the BA correlate with a decreased abundance of transcripts encoding *Tbx18* in epicardial cell cluster 2, as compared to the transcript abundance in epicardial cell cluster 1 (figure 5-4c). The second heart field is known to give rise to cardiac outflow tissue and zebrafish second heart field cells express *Ltbp3* and emerge from *Nkx2.5* expressing progenitors (Guner-Ataman et al., 2013; Zhou et al., 2011b). Many cells in epicardial cell cluster 2 indeed possessed transcripts encoding *Ltbp3* (figure 5-4f), but they lacked the expression of *Nkx2.5* (figure 5-4g). This makes it possible that epicardial cell cluster 2 might represent a second heart field derived cell population in the BA, however there is the



**Figure 5-4: The epicardial cluster 2 markers *elnb* and *Mylka* label cells within the bulbus arteriosus.** (a-c) FPKM expression values of transcripts encoding *Elnb*, *Mylka* and *Tbx18* were overlaid onto the single cell t-SNE clustering. (a) Expression of *Elnb* in epicardial cluster 2. (b) Expression of *Mylka* in epicardial clusters 2. (c) Expression of *Tbx18* in epicardial clusters 1 and 2. (d) Microscopic expression analysis of *elnb* in wildtype embryos at 5dpf using hybridization chain reaction. **d'** shows the bulbus arteriosus in a higher magnification single optical section. The enlarged images in **d''** visualize the overlap of *elnb* and *tbx18* transcripts surrounding an individual cell nucleus (asterisk). (e) Microscopic expression analysis of *Mylka* in *Tg(tbx18:myr-Citrine)* embryos at 5dpf using immunofluorescence. **e'** shows the bulbus arteriosus in a higher magnification single optical section. The enlarged images in **e''** visualize an individual cell (asterisk) that is co-labelled by *Mylka* and *tbx18:myr-Citrine*. (f) t-SNE clustering showing the expression of the second heart field marker *Ltbp3* in cells within the epicardial cell clusters 1 and 2 as well as in the mesenchymal cell cluster. (g) t-SNE clustering showing the expression of *Nkx2.5* in cardiomyocytes. Scale bars in d and e are 20µm, scale bars in e' and e'' are 10µm, scale bars in d' and d'' are 5µm. Epi = epicardial cell cluster, MC = mesenchymal cells, CM = cardiomyocytes, FPKM = fragments per kilobase of transcript per million transcripts, BA = bulbus arteriosus, V = ventricle.

## 5.6 Epicardial cell cluster 3 is located within the epicardial cell layer that envelops the developing zebrafish heart

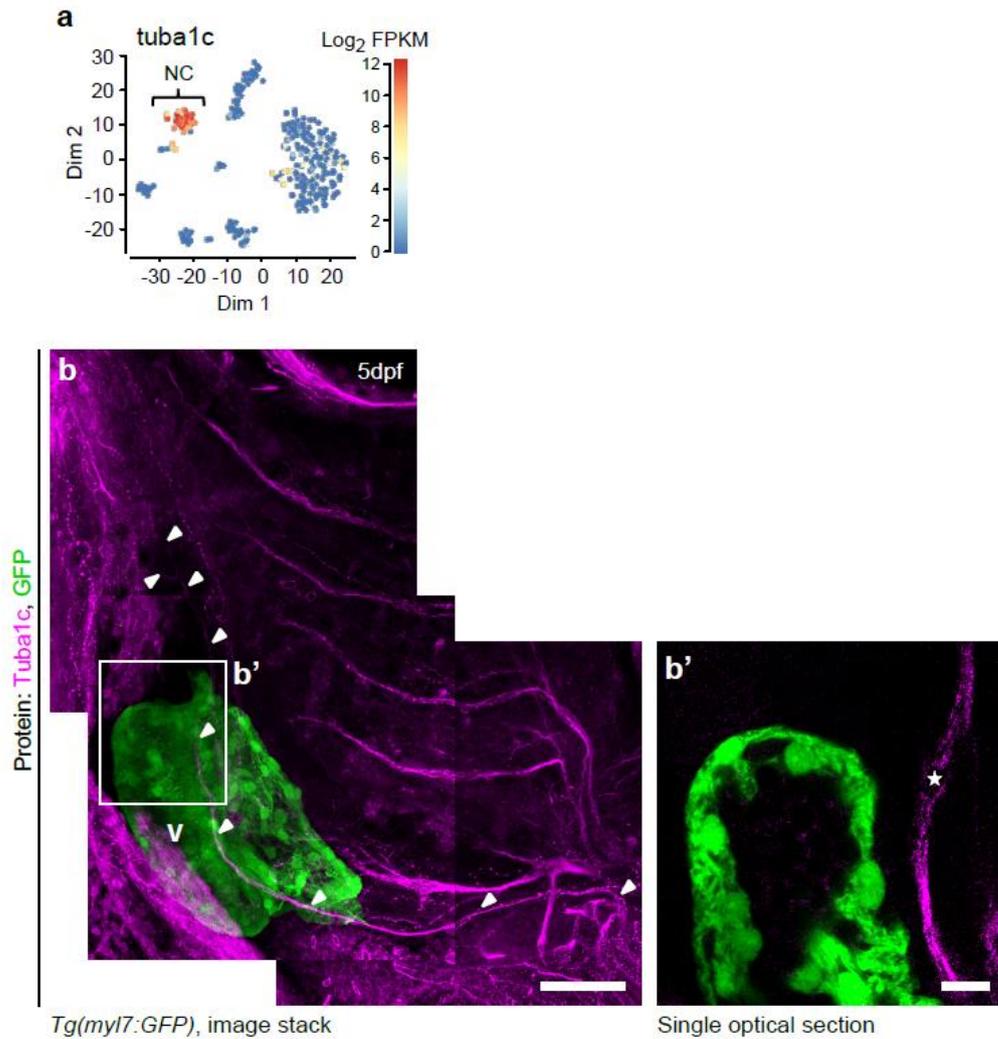
Subsequently, I analyzed the distribution and location of epicardial cell cluster 3 within the developing zebrafish heart. In the single cell transcriptomic data, many cells contained in epicardial cell cluster 3 expressed *Cldn11a* (figure 5-5a), some of which also expressed *Tcf21* (figure 5-5b). Using HCR, I found that there were *cldn11a* transcripts located in the region of the epicardial cell layer, but not in other parts of the heart at 5dpf (figure 5-5c). However, these transcripts were less abundant than *adma* transcripts and they were mostly located at the atrio-ventricular boundary (figure 5-5c'). Closer analysis revealed that *cldn11a* transcripts were located in close proximity to *tcf21* transcripts and transcripts encoding *tcf21* driven GFP (figure 5-5c'' and figure 5-5c'''). All of the three mRNA species were located close to individual cell nuclei, making it possible that they might be expressed in the same cell. Hence, *Cldn11a* expressing cells appear to be present in the epicardial cell layer of the developing zebrafish heart at 5dpf and might also express *Tcf21*. The small number of *cldn11a* transcripts matches the small fraction of *tcf21* reporter derived single cells contained in epicardial cell cluster 3, as compared to those in epicardial cell cluster 1.



**Figure 5-5: Transcripts encoding the epicardial cluster 3 marker Cldn11a are present in the epicardial cell layer. (a-b)** FPKM expression values of transcripts encoding Cldn11a and Tcf21 were overlaid onto the single cell t-SNE clustering. **(a)** Expression of Cldn11a in epicardial cluster 3. **(b)** Expression of Tcf21 in epicardial clusters 1 and 3. **(c)** Microscopic expression analysis of *cldn11a* in *Tg(tcf21:myr-GFP)* embryos at 5dpf using hybridization chain reaction. **c'** shows the atrio-ventricular boundary in a higher magnification single optical section. Enlarged images in **c''** and **c'''** visualize the close proximity of *cldn11a*, *tcf21* and GFP transcripts surrounding an individual cell nucleus (asterisk). Scale bar in c is 20µm, scale bars in c' - c''' are 5µm. Epi = epicardial cell cluster, FPKM = fragments per kilobase of transcript per million transcripts, V = ventricle.

## **5.7 The developing zebrafish heart is in close proximity to Tuba1c positive neurons**

To verify whether the zebrafish heart at 5dpf is innervated by neuronal cells, as suggested by the presence of cluster 4 in the single cell transcriptomic data (figure 4-2), I performed immunofluorescence staining of Tubulin a1c (Tuba1c). This cytoskeleton component is a pan-neuronal marker (Gulati-Leekha and Goldman, 2006) with prominent expression in the neural cell cluster (figure 5-6a). Immunofluorescence staining in *Tg(myh7:GFP)* embryos revealed an extensive Tuba1c positive neural network in the ventral head region close to the heart (figure 5-6b). An extended axonal structure projected from the deeper head regions along the atrio-ventricular boundary and past the BA was observed at 5dpf. This structure appeared to have a partner structure projecting similarly along the opposite side of the heart (data not shown). Closer investigation showed that this axon was located in close proximity to the myocardium and possibly in direct contact with the epicardium (figure 5-6b'). In conclusion, the developing zebrafish heart at 5dpf seems to be closely associated with neuronal structures, as suggested by the presence of neural cells within the single cell transcriptomic data set. Further investigation would be necessary to elucidate whether the identified neuronal structures innervate the developing zebrafish heart at this early stage of development.

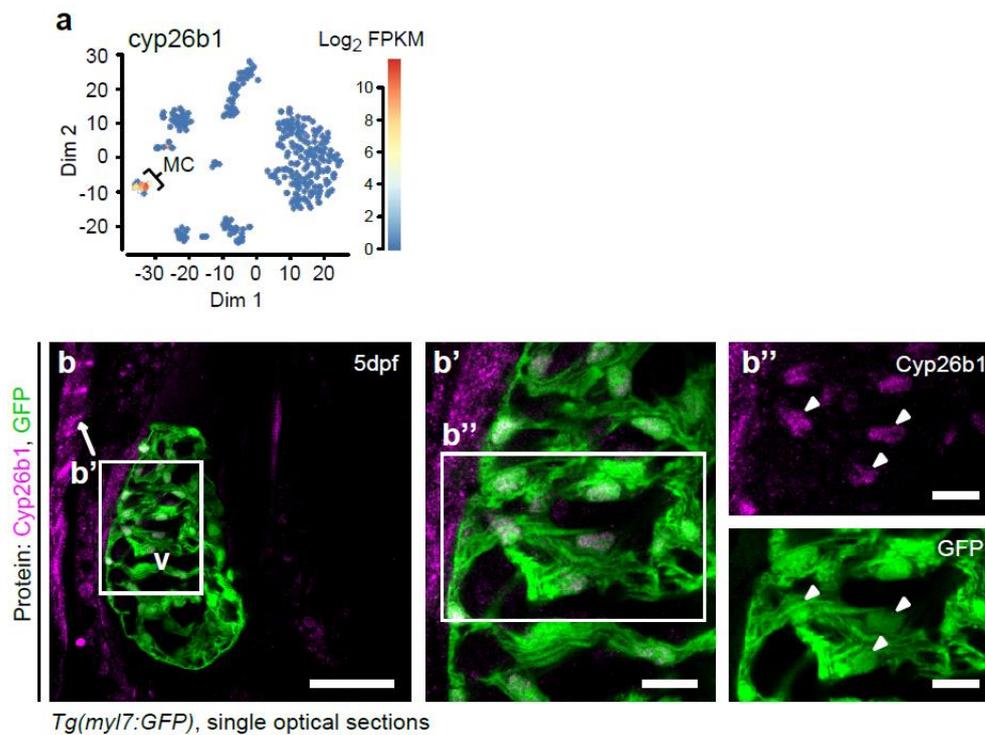


**Figure 5-6: The embryonic zebrafish heart is located adjacent to Tuba1c positive neuronal structures.** (a) FPKM expression values of transcripts encoding Tuba1c were overlaid onto the single cell t-SNE clustering and visualize the enrichment of *tuba1c* in the neural cell cluster 4. (b) Microscopic expression analysis of Tuba1c in *Tg(myf7:GFP)* embryos at 5dpf using immune fluorescence. A Tuba1c positive axon (arrowheads) projected along the atrio-ventricular boundary of the developing heart. The higher magnification single optical section in **b'** shows the close proximity between myocardium and Tuba1c positive neuron (cell body marked by asterisk). Scale bar in b is 50 $\mu$ m, scale bar **b'** is 10 $\mu$ m. NC = neural cells, FPKM = fragments per kilobase of transcript per million transcripts, V = ventricle.

## 5.8 The mesenchymal cell marker Cyp26b1 labels a large portion of the developing zebrafish myocardium

To elucidate the distribution and location of the mesenchymal cell cluster identified in the single cell transcriptomic data, I performed immunofluorescence staining of Cyp26b1, a factor that was strongly enriched in these cells (figure 5-7a). Cyp26b1 was present in extracardiac muscle and in *myl7*:GFP expressing cells within the heart (figure 5-7b). Indeed, Cyp26b1 was present in the cell nuclei of a large fraction of the *myl7*:GFP expressing cells (figure 5-7b' and figure 5-7b''). Therefore, many cells in the developing zebrafish heart that express *myl7*:GFP might possess a transcriptomic profile resembling that of the cells contained within the mesenchymal cell cluster. At least some cells in the mesenchymal cell cluster might thus contribute to the cardiomyocyte lineage. This is supported by the presence of *Tg(myI7:GFP)* derived cells within the mesenchymal cell cluster, as well as by the presence of *myl7* transcripts within these cells (figure 4-2). Potentially, zebrafish cardiomyocytes are still in the process of differentiation at 5dpf and therefore express genes that are markers of a mesenchymal cardiac progenitor population. Not all the Cyp26b1 expressing mesenchymal cells within the single cell transcriptomic data set co-expressed *MyI7*. However, immunofluorescence did not indicate cardiac expression of Cyp26b1 outside the *myl7*:GFP expressing cell population. Performing Cyp26b1 immunostaining in other cardiovascular reporter backgrounds, such as *Tg(flk1:GFP)*, would give further insight into the location of cells with a mesenchymal gene expression profile in the developing zebrafish heart. The subcellular localization of Cyp26b1 indicated that it might restrict retinoic acid directly in the nucleus. Retinoic acid

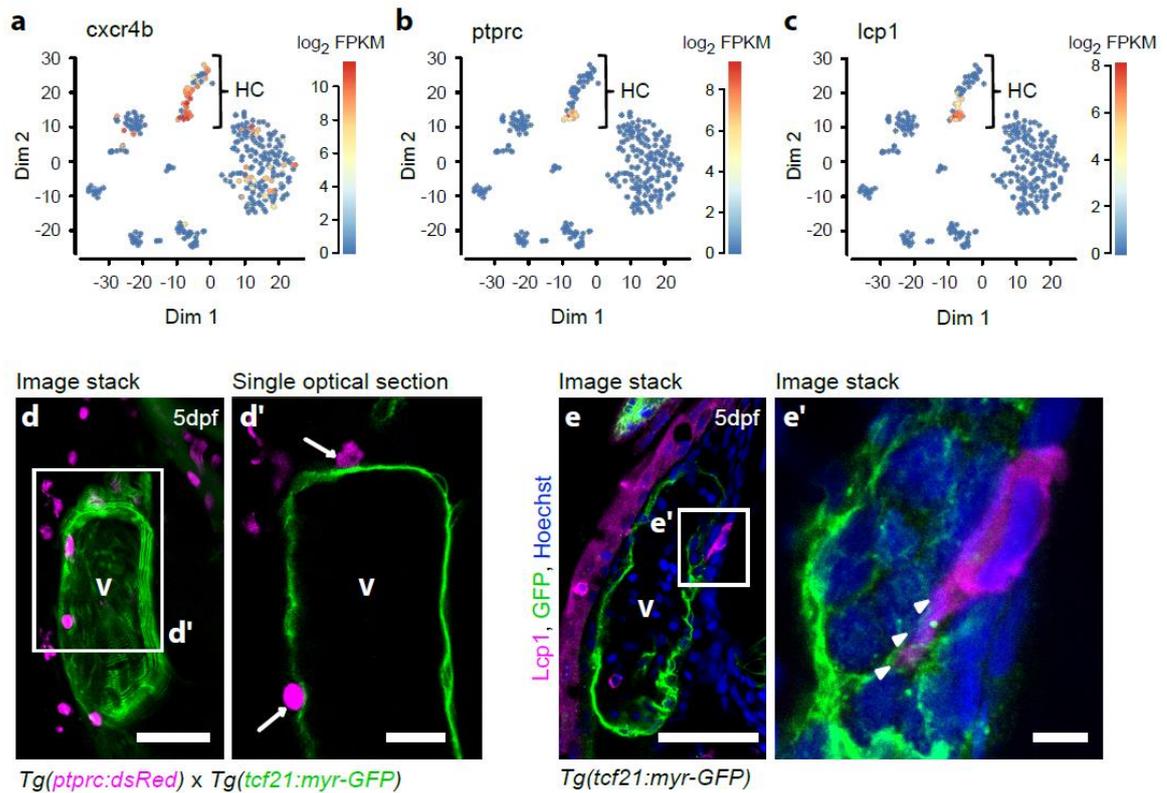
signaling might thus inhibit the formation of cardiac mesenchymal cells, resembling its function during endocardial cushion formation (Bouman et al., 1998).



**Figure 5-7: The embryonic zebrafish heart contains *myf7:GFP* expressing myocardial cells that co-express Cyp26b1.** (a) FPKM expression values of transcripts encoding Cyp26b1 were overlaid onto the single cell t-SNE clustering and visualize the enrichment of *cyp26b1* in the mesenchymal cell cluster 5. (b) Immunofluorescence analysis of Cyp26b1 in *Tg(myf7:GFP)* embryos at 5dpf, showing expression in extracardiac muscle (arrow). The higher magnification image in **b'** shows *myf7:GFP* expressing cell nuclei that are co-labelled by Cyp26b1 (arrowheads). Scale bar in **b** is 50 $\mu\text{m}$ , scale bars in **b'** and **b''** are 10 $\mu\text{m}$ . MC = mesenchymal cells, FPKM = fragments per kilobase of transcript per million transcripts, V = ventricle.

## 5.9 Myeloid haematopoietic cells are in direct contact with the developing zebrafish epicardium

Analysis of the single cell transcriptomic data identified a population of myeloid haematopoietic cells that might be guided into the developing zebrafish heart by Cxcl12a expressing cells within epicardial cell cluster 3 (figure 4-7). Some of these myeloid cells co-expressed transcripts encoding Cxcr4b (figure 5-8a), Ptprc (CD45) (figure 5-8b) and Lcp1 (figure 5-8c). To further investigate, I crossed the existing line *Tg(ptprc:dsRed)* (Bertrand et al., 2008) to the newly generated *Tg(tcf21:myrGFP)* (figure 3-3) and analyzed double positive reporter embryos to reveal that *ptprc:dsRed* expressing cells indeed resided in the tissues surrounding the developing heart (figure 5-8d). Some *ptprc:dsRed* expressing cells directly contacted the epicardial layer from the outside of the heart and appeared to be attached (figure 5-8d'). Immunofluorescence staining of Lcp1 in *Tg(tcf21:myr-GFP)* embryos yielded similar results (figure 5-8e). An Lcp1 expressing cell contacted the epicardium on the outside of the atrio-ventricular boundary and appeared to extend a protrusion into the epicardial cell layer (figure 5-8e'). The fact that *ptprc:dsRed* and Lcp1 expressing cells were in intimate contact with the epicardial cell layer supports a model in which epicardial Cxcl12a acts as a chemo-attractant for these cells. All myeloid cells contacting the epicardium did so on the pericardial side and very few myeloid cells were detected within the heart. This suggests that myeloid cells contact the epicardium, but do not trans-migrate into the heart, or that they do so at a later developmental time point than 5dpf. Time lapse microscopy of *Tg(ptprc:dsRed) x Tg(tcf21:myr-GFP)* embryos could elucidate whether trans-migration occurs following attachment of a myeloid cell to the epicardium.



**Figure 5-8: Myeloid haematopoietic cells are in direct contact with the developing zebrafish epicardium.** (a-c) FPKM expression values of transcripts encoding *Cxcr4b*, *Ptprc* and *Lcp1* were overlaid onto the single cell t-SNE clustering. (a) Expression of *Cxcr4b* in epicardial cell cluster 1 and in the haematopoietic cell clusters 7 and 6. (b) Expression of *Ptprc* in the myeloid haematopoietic cell cluster 6. (c) Expression of *Lcp1* in the myeloid haematopoietic cell cluster 6. (d) Microscopic expression analysis of *Tg(ptprc:dsRed)* x *Tg(tcf21:GFP)* double positive embryos at 5dpf. The higher magnification single optical section in **d'** shows *ptprc:dsRed* expressing cells (arrows) in direct contact with the epicardium. (e) Microscopic expression analysis of *Lcp1* in *Tg(tcf21:GFP)* embryos at 5dpf using immunofluorescence. The higher magnification image stack in **e'** shows an *Lcp1* expressing cell protruding into the epicardial cell layer (arrowheads). Scale bars in **d** and **e** are 50µm, scale bar in **d'** is 10µm, scale bar in **e'** is 5µm. Epi = epicardial cell cluster, HC =

haematopoietic cells, FPKM = fragments per kilobase of transcript per million transcripts, V = ventricle.

## **5.10 Putative enhancers in genomic proximity to markers of epicardial cell cluster 1 and 2 show epicardial subpopulation-specific *in vivo* activity**

Cell type-specific gene expression is regulated by short genomic regions called enhancers that provide a binding platform for transcription factors and produce regulatory enhancer-associated RNAs (Heinz et al., 2015). Enhancers may be located within a gene or be intergenic and the recruited transcription factors are able to control gene promoter activity over long genomic distances. There are multiple ways to identify active enhancer regions: Physical interactions between a promoter and associated enhancers can be queried via chromosome conformation capture-on-chip (4C) (Smyk et al., 2013) and genomic binding of transcription factors as well as the presence of enhancer-indicating chromatin marks, such as H3K27Ac and H3K4me, can be queried via CHIP sequencing (Schmidl et al., 2015). Another method that aides in the identification of active enhancers is the assay for transposase-accessible chromatin (ATAC), which relies on Tn5 transposase-mediated integration of sequencing adapters into regions of accessible chromatin, amplification of the adapter-labelled regions and high throughput sequencing (Buenrostro et al., 2013). However, only a subset of accessible chromatin regions represents enhancers, making it important to test the regulatory activity of the ATAC-identified putative enhancer elements.

To identify enhancer elements that might regulate the differentiation or function of the epicardial subpopulations identified through single cell RNA sequencing, I generated

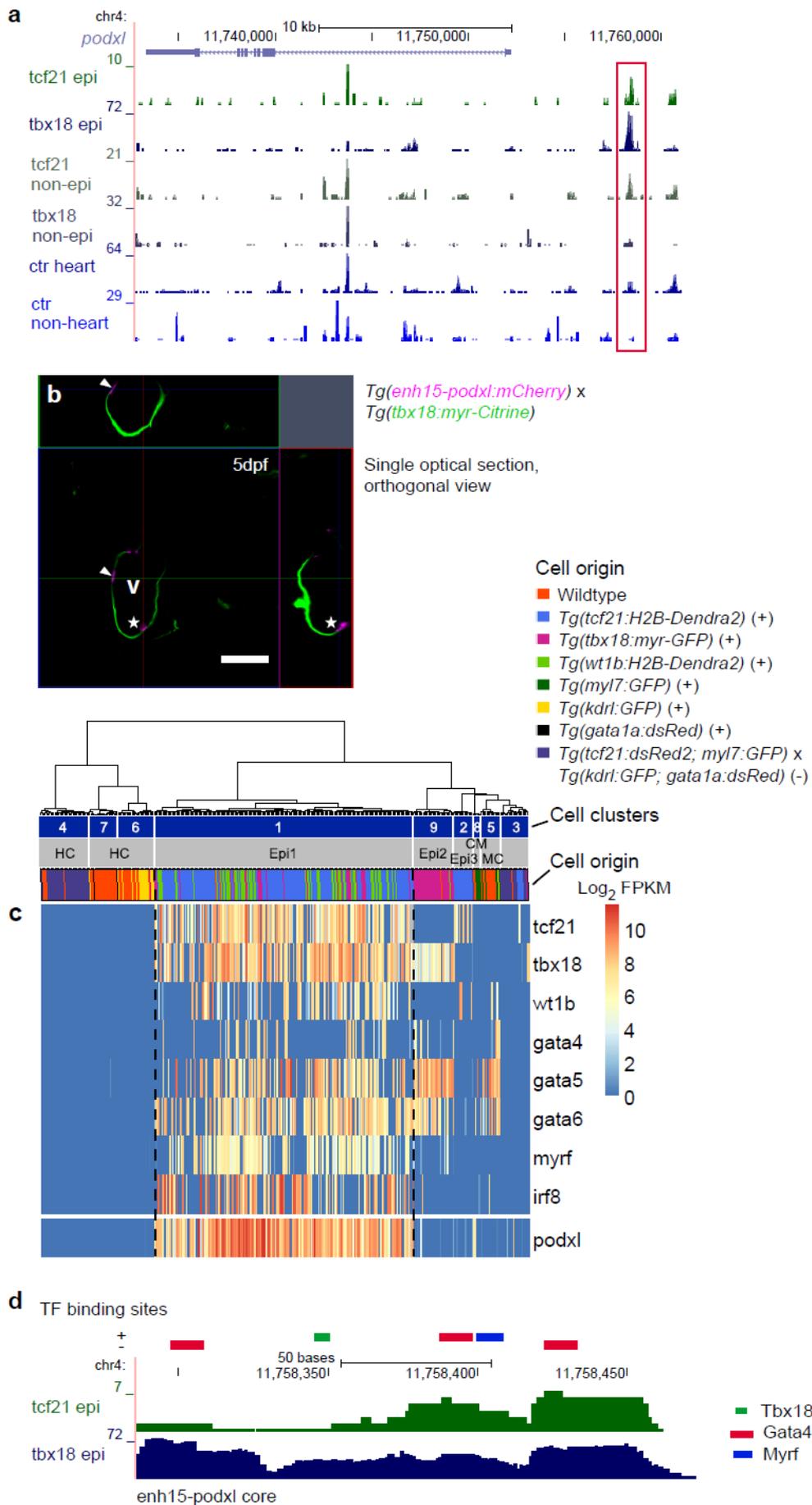
ATAC sequencing data from bulks of *tcf21*:H2B-Dendra2 expressing cardiac cells (*tcf21* epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine expressing cardiac cells (*tbx18* epi), *tcf21*:H2B-Dendra2 expressing non-cardiac cells (*tcf21* non-epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine expressing non-cardiac cells (*tbx18* non-epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine negative cardiac cells (ctr heart) and *tbx18*:myr-GFP/*tbx18*:myr-Citrine negative non-cardiac cells (ctr non-heart) at 5dpf. Since single cell transcriptomics showed *wt1b* positive epicardial cells to mostly express *tcf21* and/or *tbx18* as well, I did not process *wt1b* reporter embryos for ATAC sequencing. Within the ATAC sequencing data, I subsequently searched for regions that were located in proximity to marker genes of the epicardial cell clusters identified through single cell transcriptomics and were accessible in epicardial cells, but mostly inaccessible in non-epicardial control cells.

### **5.10.1 A putative enhancer element located upstream of *podxl* is active in the developing zebrafish epicardial cell layer**

I identified a region located approximately 6.5kb upstream of the epicardial cell cluster 1 marker gene *podxl* that fulfilled the criteria described above (figure 5-9a). I tested the activity of this putative enhancer element (enh15-*podxl*) by inserting the genomic sequence (869 base pairs) into a minimal promoter mCherry reporter vector, analyzing mCherry fluorescence *in vivo*. The reporter vector was based on the AcDs transposition system (Emelyanov et al., 2006), enabling efficient integration into the genome. Importantly, injecting a vector construct that lacked an enhancer element did not result in detectable fluorescence (data not shown). Reporter fluorescence driven by enh15-*podxl* was detectable in epicardial cells labelled by *tbx18*:myr-Citrine (figure 5-9b).

Additionally, cells in the sub-epicardium were labelled by mCherry fluorescence. No other region of the heart or the rest of the embryo was labelled by enh15-podxl driven fluorescence. In conclusion, the activity of enh15-podxl recapitulated the expression pattern of its nearest gene, *podxl*, in the epicardium. However, enh15-podxl activity appeared to be more widespread than the expression of transcripts encoding Podxl and labelled a subset of cells in the sub-epicardial region as well. Nevertheless, enh15-podxl might regulate the expression of Podxl and thereby affect cells within epicardial cell cluster 1. Enhancer knock out would potentially show whether loss of enh15-podxl affects the expression of Podxl and/or epicardial development. Using the Jaspar database (Tan and Lenhard, 2016), I searched the core sequence of enh15-podxl for binding sites of transcription factors that are known regulators of cardiac development and were expressed within epicardial cell cluster 1. In addition to Tcf21, Tbx18 and Wt1b, the cardiac GATA factors Gata4, Gata5 and Gata6, which are necessary for normal heart development (Laforest and Nemer, 2011; Lepore et al., 2006; Pu et al., 2004), showed a fitting expression pattern (figure 5-9c). The most enriched transcription factors within epicardial cell cluster 1 were Irf8 and Myrf. Querying transcription factor binding motifs, I found that the core region of enh15-podxl harbored binding sites of Tbx18, Gata4 and Myrf. I found a single Tbx18 binding motif with the sequence GTGTGA, as well as three Gata4 binding motifs with the sequences CTTTATCAGAA, ATTTATCTCGG and TTTTCTCTGGG and a single Myrf binding motif with the sequence GCCTGGTAC. Tbx18, Gata4 and Myrf might thus be recruited to enh15-podxl and subsequently regulate the expression of Podxl in epicardial cell cluster 1. Gata4 might furthermore be involved in driving the activity of enh15-podxl in the sub-epicardium. CHIP sequencing would test

binding of these factors to enh15-podxl and knock out studies of these transcription factors could verify a regulatory network between Tbx18, Gata4, Myrf and Podxl.



**Figure 5-9: An enhancer element located upstream of *podxl* is active in the developing zebrafish epicardium.** (a) ATAC sequencing data obtained from *tcf21*:H2B-Dendra2 positive epicardial cells (*tcf21* epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine positive epicardial cells (*tbx18* epi), *tcf21*:H2B-Dendra2 positive non-epicardial cells (*tcf21* non-epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine positive non-epicardial cells (*tbx18* non-epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine negative cardiac cells (ctr heart) and *tbx18*:myr-GFP/*tbx18*:myr-Citrine negative non-cardiac cells (ctr non-heart). Shown is the genomic locus of *podxl*. The red frame indicates the putative regulatory element *enh15-podxl*. (b) Microscopic analysis of an *in vivo* reporter assay showing activity of *enh15-podxl* in the epicardium (arrowhead) and the sub-epicardial myocardium (asterisk). (c) Heatmap visualizing the expression of transcription factors in epicardial cell cluster 1. Red color indicates high expression. (d) Transcription factor binding sites within the core of *enh15-podxl*. + and - indicate the DNA strand a binding site is located on. Scale bar in b is 50µm. V = ventricle, Epi = epicardial cell cluster, FPKM = fragments per kilobase of transcript per million transcripts, TF = transcription factor.

### **5.10.2 A putative enhancer element located in intron 3 of *lox* is active in the bulbus arteriosus of the developing zebrafish heart**

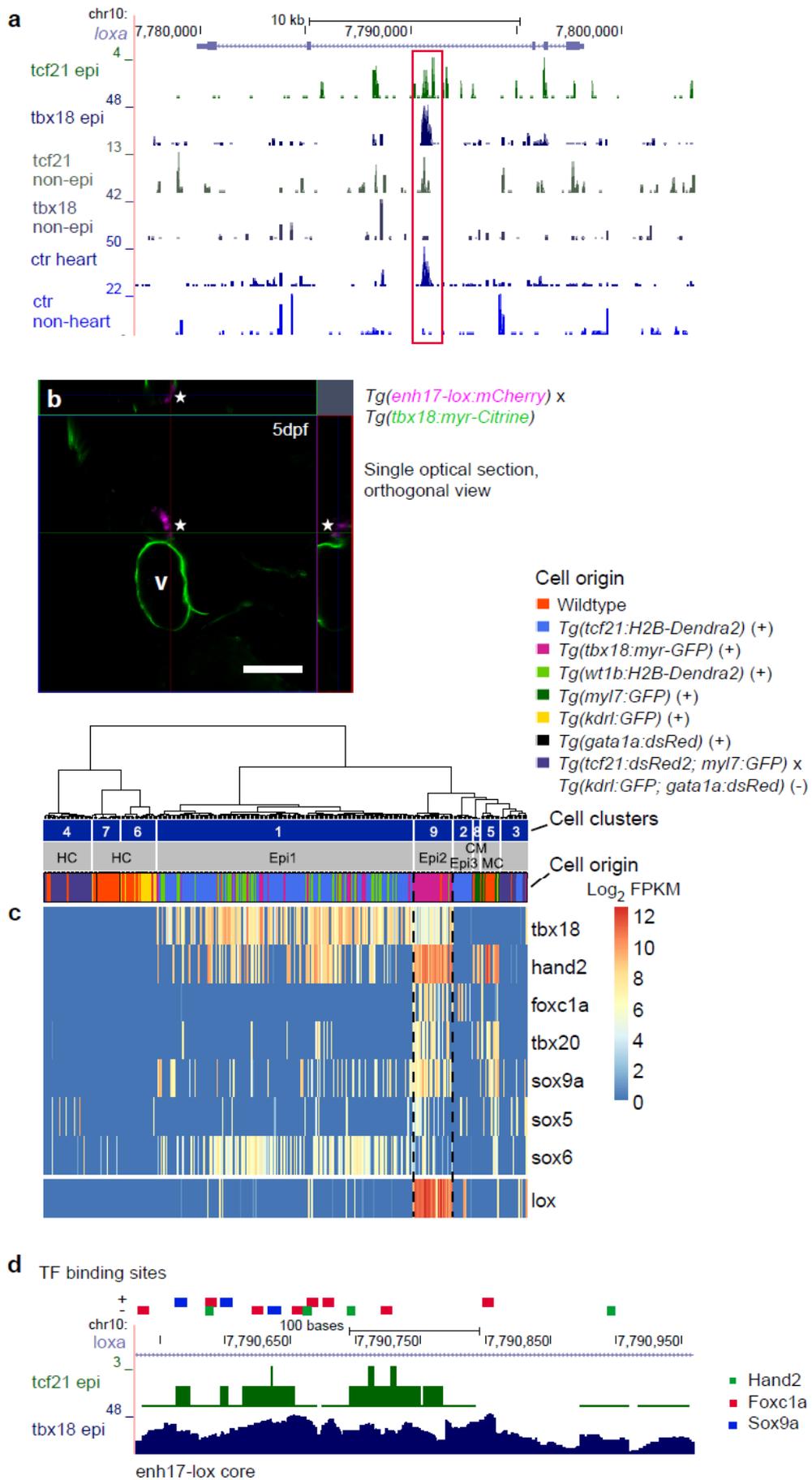
Several accessible chromatin regions were located in close proximity to genes enriched in epicardial cell cluster 2. One of these accessible regions (796 base pairs) resided within intron 3 of the most enriched gene, *lox* (figure 5-10a). Chromatin accessibility in this region was prominent in *tbx18*:myr-GFP/*tbx18*:myr-Citrine expressing epicardial cells and

in *tbx18:myr-GFP/tbx18:myr-Citrine* negative cardiac cells, but far less pronounced in *tcf21:H2B-Dendra2* expressing epicardial cells and in non-cardiac cells. Matching these findings, the identified putative enhancer element (enh17-lox) drove reporter fluorescence exclusively in cells located within the BA (figure 5-10b). These cells were not located within the epicardial cell layer, but further inward within the BA. In conclusion, the landscape of chromatin accessibility and reporter driven enhancer activity shows enh17-lox as able to drive the expression of *Lox* within cells in epicardial cell cluster 2. However, enh17-lox could also be active in cells within the BA that do not belong to epicardial cell cluster 2, as its accessibility was prominent in *Tg(tbx18:myr-GFP)/Tg(tbx18:myr-Citrine)* derived non-fluorescent cardiac cells as well. Enhancer knock out would potentially show whether loss of enh17-lox affects the expression of *Lox*, although there could be redundancy in the regulation of *lox*.

Transcription factors that were expressed within epicardial cell cluster 2 alongside *Tbx18* included *Hand2* and the Forkhead Box transcription factor *Foxc1a* (figure 5-10c). In the mouse, the orthologue *Foxc1* drives the differentiation of embryonic stem cells into cardiomyocytes and regulates cells within the second heart field that subsequently contribute to the cardiac outflow tract (Lambers et al., 2016; Seo and Kume, 2006). The T-Box transcription factor *Tbx20* is another important regulator of heart development (Just et al., 2016; Shelton and Yutzey, 2007) and was enriched in epicardial cell cluster 2. *Tbx20* is expressed in the developing outflow tract and valves in mouse and chick hearts (Chakraborty et al., 2008; Lincoln et al., 2004) and is a downstream target of *Twist1* (Shelton and Yutzey, 2008), which was expressed in epicardial cell cluster 2 as well (figure 4-4e). Additionally, cells within epicardial cell cluster 2 expressed the cartilage marker *Sox9a*, which is essential for cardiac valve development (Akiyama et al., 2004). The

mammalian orthologue Sox9 plays a role in ECM organization during cartilage as well as cardiac valve development and drives the expression of Sox5 and Sox6, further transcription factors expressed in epicardial cell cluster 2 (Akiyama et al., 2002; Lincoln et al., 2007). Binding of Sox9 is promoted by Sox5 and Sox6 to stimulate the expression of Aggrecan, another marker gene of cells in epicardial cell cluster 2 (figure 4-6) (Han and Lefebvre, 2008). Taken together, the transcription factors enriched in epicardial cell cluster 2 are known to form regulatory networks driving cardiac valve and cartilage development and might form a similar network to regulate the contribution of cells in epicardial cell cluster 2 to the smooth muscle layer of the BA.

I identified binding sites for Hand2, Foxc1a and Sox9a in the core region of enh17-lox (figure 5-10d). The four binding sites for Hand2 were dispersed over the core region and featured the binding motifs CAAATG, CAACTG, CACTTT and CATTTT. The eight binding sites for Foxc1a featured the binding motifs CAAATGTA, CTGAAGTA, CAGCAGTA, CACAAGTA, GGGGCGTA, TTGAAGTA, ACTCAGTA and GGTCTGTA. The three binding sites for Sox9a featured the binding motifs TCATTGTAT, CCATAGTTG and GCATAGTTT. The presence of these binding sites argues for a recruitment of Hand2, Foxc1a and Sox9a to enh17-lox, where they might act to drive the expression of Lox and possibly the differentiation of cells within epicardial cell cluster 2 into smooth muscle cells.



**Figure 5-10: An enhancer element located in intron 3 of *lox* is active in the bulbus arteriosus of the developing zebrafish heart.** (a) ATAC sequencing data obtained from *tcf21*:H2B-Dendra2 positive epicardial cells (*tcf21* epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine positive epicardial cells (*tbx18* epi), *tcf21*:H2B-Dendra2 positive non-epicardial cells (*tcf21* non-epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine positive non-epicardial cells (*tbx18* non-epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine negative cardiac cells (ctr heart) and *tbx18*:myr-GFP/*tbx18*:myr-Citrine negative non-cardiac cells (ctr non-heart). Shown is the genomic locus of *lox*. The red frame indicates the putative regulatory element *enh17-lox*. (b) Microscopic analysis of an *in vivo* reporter assay showing activity of *enh17-lox* within the bulbus arteriosus (asterisk). (c) Heatmap visualizing the expression of transcription factors in epicardial cell cluster 2. Red color indicates high expression. (d) Transcription factor binding sites within the core of *enh17-lox*. + and - indicate the DNA strand a binding site is located on. Scale bar in b is 50µm. V = ventricle, Epi = epicardial cell cluster, FPKM = fragments per kilobase of transcript per million transcripts, TF = transcription factor.

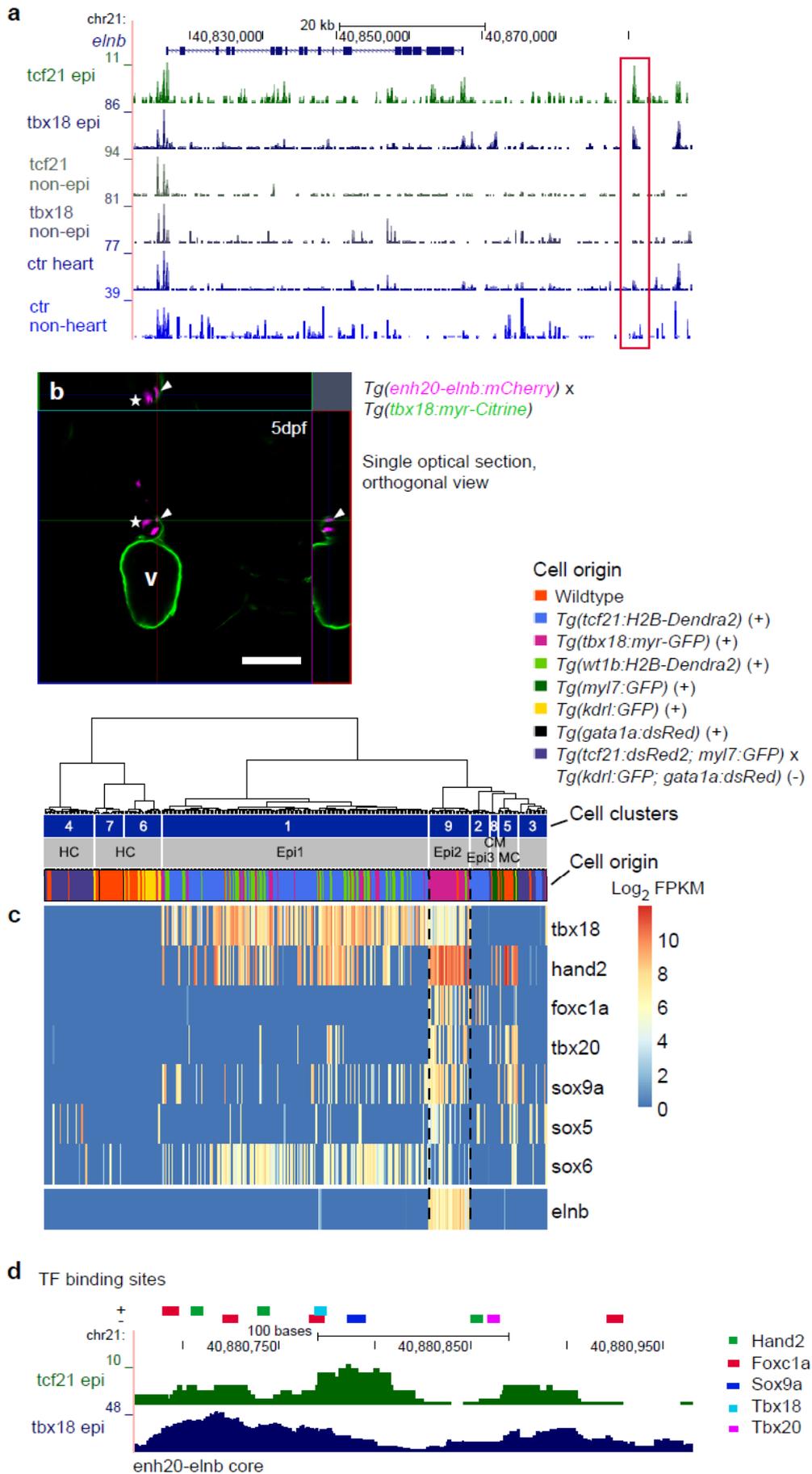
### **5.10.3 A putative enhancer element located upstream of *elnb* is active in the bulbus arteriosus of the developing zebrafish heart**

I additionally identified a region (701 base pairs) approximately 25kb upstream of *elnb*, another gene highly enriched within epicardial cell cluster 2, that was accessible in the epicardial ATAC samples and largely inaccessible in the control conditions (figure 5-11a). Similar to enh17-lox, this putative enhancer element (enh20-elnb) showed activity in cells that were located within the BA, as shown by mCherry reporter activity (figure 5-11b). In addition to cells that appeared to be located within the smooth muscle layer of the BA, enhancer driven fluorescence labelled a single epicardial cell, which co-expressed *tbx18:myr-Citrine*. These findings suggest that enh20-elnb might drive the expression of *Elnb* within the BA and that this enhancer element might be active in cells within epicardial cell cluster 2. Enhancer knock out would be able to show whether loss of enh20-elnb affects the expression of *Elnb*, although there could be redundancy in the regulation of *Elnb*.

Most of the transcription factors expressed in epicardial cell cluster 2 (figure 5-11c) might be recruited to enh20-elnb, as a variety of binding sites was located within the core region of the enhancer element (figure 5-11d). Similar to enh17-lox, enh20-elnb harbored binding sites for Hand2, Foxc1a and Sox9a. The three Hand2 binding sites featured the sequence motifs CATGTG, CAAATT and CAGCTT, while the four Foxc1a binding sites featured the sequence motifs TCGCTGTA, GGCATGTA, ACACTGTA and TCTATGTA. The Sox9a binding motif present in the core region of enh20-elnb had the sequence GCATTGTTT. In addition, the core region of enh20-elnb harbored binding sites for Tbx18 and Tbx20, with the sequence GTGTGA and TGCTGACA, respectively. The

variety of transcription factor binding sites within enh20-elnb suggests that several transcription factors might be involved in the regulation of the target gene. As Elnb was shown to drive the differentiation of smooth muscle cells in the BA (Moriyama et al., 2016), transcription factor binding within enh20-elnb might be essential to drive BA development, potentially making this enhancer a valuable tool for further studies.

The genomic loci of marker genes expressed within epicardial cell cluster 3, such as *Cldn11a* and *Cxcl12a*, did not harbor regions with strong chromatin accessibility. I analyzed four of these regions regarding their *in vivo* activity, but did not detect any reporter fluorescence between 0dpf and 5dpf (data not shown). Potentially, the regulatory regions driving gene expression in epicardial cell cluster 3 only work in cooperation and do not show activity when isolated.



**Figure 5-11: An enhancer element located upstream of *elnb* is active in the bulbus arteriosus of the developing zebrafish heart.** (a) ATAC sequencing data obtained from *tcf21*:H2B-Dendra2 positive epicardial cells (*tcf21* epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine positive epicardial cells (*tbx18* epi), *tcf21*:H2B-Dendra2 positive non-epicardial cells (*tcf21* non-epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine positive non-epicardial cells (*tbx18* non-epi), *tbx18*:myr-GFP/*tbx18*:myr-Citrine negative cardiac cells (ctr heart) and *tbx18*:myr-GFP/*tbx18*:myr-Citrine negative non-cardiac cells (ctr non-heart). Shown is the genomic locus of *elnb*. The red frame indicates the putative regulatory element enh20-elnb. (b) Microscopic analysis of an *in vivo* reporter assay showing activity of enh20-elnb within the bulbus arteriosus (asterisk) and in an epicardial cell covering the bulbus arteriosus (arrowhead). (c) Heatmap visualizing the expression of transcription factors in epicardial cell cluster 2. Red color indicates high expression. (d) Transcription factor binding sites within the core of enh20-elnb. + and - indicate the DNA strand a binding site is located on. Scale bar in b is 50µm. V = ventricle, Epi = epicardial cell cluster, FPKM = fragments per kilobase of transcript per million transcripts, TF = transcription factor.

## 5.11 Summary

I collected multiple lines of evidence to support the validity of the epicardial subpopulations identified through single cell transcriptomics. Analysis of epicardial bulk transcriptomes revealed a general similarity between these and pseudo-bulks generated from single cell epicardial cell clusters. Furthermore, epicardial bulk samples were highly enriched in marker genes expressed in epicardial cell cluster 1, such as *adma* and *podxl*. Marker genes expressed in epicardial cell clusters 2 and 3, such as *lox* and *cldn11a*, were

less enriched in epicardial bulk samples, but showed a heterogeneous enrichment that fits the *tbx18* positive, *tcf21* and *wt1b* negative expression profile of epicardial cell cluster 2 as well as the *tcf21* positive, *tbx18* and *wt1b* negative expression profile of epicardial cell cluster 3. These findings further demonstrate the advantages of single cell RNA sequencing over bulk transcriptomics, as only single cell transcriptomics clearly delineated the different epicardial subpopulations, their transcriptomic profiles and their interactions with other cardiac cell populations.

Visualization of marker gene transcript and protein localization suggested that epicardial cell cluster 1 represents the majority of the epicardial cell layer, while epicardial cell cluster 3 appears to constitute a smaller cell population within the epicardial cell layer and epicardial cell cluster 2 is likely to reside outside the epicardial cell layer and in the BA. Reporter based *in vivo* activity assays of putative enhancer elements located close to the loci of *podxl*, *lox* and *elnb* revealed that the activity of these regulatory elements overlapped with the regions in the developing zebrafish heart where the epicardial subpopulation expressing the respective gene resided. Furthermore, the expression of several transcription factors within the epicardial subpopulations that are able to bind within the identified enhancer elements and are known to form regulatory networks in other settings, such as cardiac valve development, provides a starting point to elucidate the genetic programs driving development and differentiation of the identified epicardial subpopulations.

As a next step, I am planning to utilize the identified active enhancers to drive Cas9 (Cong et al., 2013) expression specifically in cells with enhancer activity. This system will allow for an epicardium specific knock out of epicardial markers, such as *Adma*, *Acta2* and *Cxcl12a*, to study the fate and function of the different epicardial cell clusters.

## Chapter 6

### General discussion, final conclusions and future work

#### 6.1 Investigating the developing zebrafish epicardium using novel *tcf21*, *tbx18* and *wt1b* reporter lines

In this study, I generated novel zebrafish reporter lines that allowed for comparative expression analysis of Tcf21, Tbx18 and Wt1b simultaneously. I subsequently used these lines both to image epicardial reporter fluorescence in the wholemount embryo and to obtain FACS-purified epicardial cells for single cell transcriptomics. I also used the newly generated reporter lines to validate the findings obtained using single cell transcriptomics, both to FACS-purify epicardial cells for bulk transcriptomics and ATAC sequencing and to perform microscopy based validation of subpopulation specific gene signatures.

During imaging based experiments, I compared the newly generated reporter lines to pre-existing epicardial reporter lines that allowed for comparative expression analysis of Tcf21 and Wt1b or Tbx18 and Wt1b in the same embryo. While both existing and newly generated reporter lines indicated that Tcf21, Tbx18 and Wt1b differed in their epicardial expression patterns and that none of them labelled the entire developing epicardium, there were marked differences between some of the newly generated lines and their existing counterparts.

In *tcf21* reporter lines, there were only minor differences between newly generated and pre-existing reporters. The existing reporter line *Tg(tcf21:dsRed2)* and the newly generated reporter line *Tg(tcf21:myr-tdTomato)* labelled comparable numbers of

epicardial cells (figure 3-2 and figure 3-3). Only at 3dpf, the number of *tcf21:myr-tdTomato* expressing epicardial cells was considerably lower than that of *tcf21:dsRed2* expressing epicardial cells. As discussed in chapter 3, this might be due to a prolonged maturation time of the myr-tagged version of tdTomato. The maturation time of native tdTomato (1 hour) is much shorter than that of dsRed2 (10 hours) (Shaner et al., 2004), however the N-terminal presence of the myr tag and the spatial localization close to the plasma membrane might prolong the maturation of myr-tdTomato. Alternatively, transient transcriptional interference by regulatory elements at the genomic site of BAC integration in *Tg(tcf21:myr-tdTomato)* might underlie the observed differences between *Tg(tcf21:myr-tdTomato)* and *Tg(tcf21:dsRed2)* at 3dpf. This would fit the fact that the numbers of epicardial cells labelled in *Tg(tcf21:myr-tdTomato)* and *Tg(tcf21:dsRed2)* were comparable at 5dpf and 7dpf. Both *Tg(tcf21:dsRed2)* and the new reporter lines, *Tg(tcf21:myr-tdTomato)*, *Tg(tcf21:H2B-Dendra2)* and *Tg(tcf21:myr-GFP)*, were generated from the BAC DKEYP 79F12, which appeared to contain all the major regulatory elements controlling transcription of *tcf21*, since most *Tg(tcf21:H2B-Dendra2)* derived single cells possessed endogenous transcripts encoding Tcf21 (figure 4-4). Furthermore, the endogenous DNA sequence featured on DKEYP 79F12 contained all major accessible chromatin regions surrounding the *tcf21* locus, as identified by ATAC sequencing (data not shown). Finally, transcripts encoding GFP in *Tg(tcf21:myr-GFP)* co-localized with endogenous transcripts encoding Tcf21 (figure 5-5).

The number of epicardial cells labelled in the pre-existing line *Tg(tbx18:dsRed2)* was lower than that labelled in the newly generated line *Tg(tbx18:myr-GFP)* (figure 3-2 and figure 3-3), however almost all the *Tg(tbx18:myr-GFP)* derived sequenced single cells expressed endogenous Tbx18 (figure 4-4). This argues for the validity of *Tg(tbx18:myr-*

*GFP*), simultaneously suggesting that *Tg(tbx18:dsRed2)* might not label all Tbx18 expressing epicardial cells in the zebrafish embryo. Although *Tg(tbx18:dsRed2)* is BAC-based, this BAC (CH211-197L9) might lack regulatory elements contained in the BAC DKEYP 117G5, which the new *tbx18* reporter lines are based on. Indeed, DKEYP 117G5 includes a DNA sequence extending from 45kb to 100kb downstream of *tbx18* that is not contained in CH211-197L9 and the endogenous counterpart of this downstream sequence contained multiple regions with high chromatin accessibility in epicardial cells, as identified by ATAC sequencing (data not shown). The epicardial expression of another newly generated *tbx18* reporter line, *Tg(tbx18:myr-Citrine)*, was very similar to that of *Tg(tbx18:myr-GFP)* (figure 3-1 and figure 3-3). However, *Tg(tbx18:H2B-Dendra2)*, which was based on DKEYP 117G5 as well, resembled *Tg(tbx18:dsRed2)* in that it labelled a lower number of epicardial cells than *Tg(tbx18:myr-GFP)*, likely not capturing all of the Tbx18 expressing epicardial subpopulations (figure 5-2d). This suggests that positional effects arising from random integration of the different *tbx18* BAC constructs into the genome, the fluorophore sequence integrated into the *tbx18* BACs or different fluorophore maturation times affected the final expression patterns.

The most striking differences between pre-existing and newly generated reporter lines were present in the *wt1b* setting. The number of epicardial cells labelled in the pre-existing line *Tg(wt1b:GFP)* was much lower than that labelled in the newly generated line *Tg(wt1b:H2B-Dendra2)*. In this line, the number of labelled epicardial cells was as high as the ones found in *tcf21* and *tbx18* reporter lines (figure 3-2 and figure 3-3). The discrepancy could in part be attributed to a population of *wt1b:H2B-Dendra2* expressing cells covering the BA that was not present in *Tg(wt1b:GFP)* hearts, but there were pronounced differences in the ventricular epicardium as well. *Tg(wt1b:H2B-Dendra2)* is a

BAC based reporter line and therefore likely to incorporate more of the regulatory elements acting on *wt1b* than *Tg(wt1b:GFP)*, which is a plasmid-based reporter line (Perner et al., 2007). For this reason, *Tg(wt1b:H2B-Dendra2)* should be better in capturing the endogenous expression pattern of *Wt1b* than *Tg(wt1b:GFP)*. During zebrafish heart regeneration however, GFP expression in *Tg(wt1b:GFP)* was observed to match endogenous *Wt1b* expression (Gonzalez-Rosa et al., 2011). I observed the endogenous epicardial expression of *Wt1b* to be less abundant than suggested in *Tg(wt1b:H2B-Dendra2)* and much sparser than that of *Tcf21* and *Tbx18* at 5dpf (figure 3-2j). Additionally, a significant fraction of sequenced single cells derived from *Tg(wt1b:H2B-Dendra2)* hearts did not endogenously express *Wt1b* and the expression of *Wt1b* in *Tg(wt1b:H2B-Dendra2)* derived cells was less widespread than that of *Tcf21* and *Tbx18* (figure 4-4). In summary, *wt1b* driven H2B-Dendra2 labelled several epicardial cells that did not possess endogenous transcripts encoding *Wt1b*. It is possible that there were fluctuations in the endogenous *wt1b* transcripts that were not accurately reflected by reporter fluorescence, as Dendra2 protein may be stable for longer than *wt1b* endogenous transcripts. Immune fluorescence staining of endogenous *Wt1b* protein levels would be necessary to elucidate if such transcriptional fluctuations manifest on the protein level, or if *wt1b* driven H2B-Dendra2 reports more accurately on the presence of *Wt1b* protein than it reports on the presence of *wt1b* transcripts. At this point, it is hard to say whether *Tg(wt1b:H2B-Dendra2)* reports more accurately on endogenous epicardial *Wt1b* expression than *Tg(wt1b:GFP)*. Endogenous inhibitory regulatory elements not even included on the *wt1b* BAC (CH73 157N22) as well as positional effects that positively affected BAC derived transcription might also be responsible for the high number of *wt1b:H2B-Dendra2* expressing epicardial cells. Indeed, ATAC data in *tcf21* and

*tbx18* reporter expressing epicardial cells showed the presence of accessible chromatin regions in the vicinity of the *wt1b* locus that were not contained in CH73 157N22 (data not shown). Some false-positive epicardial cells might have also been responsible for the discrepancy observed in cell numbers, as cells in the pericardial sac, located adjacent to the epicardium and also labelled with *wt1b* reporters, might have been included in the counting. Furthermore, it was not always readily detectable if an epicardial cell expressed *tbx18:myr-GFP* (green membrane) only or *wt1b:H2B-Dendra2* and *tbx18:myr-GFP* (green nucleus and green membrane).

The reliability of the *wt1b* reporter lines furthermore impacts on the findings obtained following embryonic heart injury (figure 3-4). Here, *wt1b:GFP* expression was not activated in epicardial cells following laser mediated ablation of the embryonic ventricle. However, since *wt1b:H2B-Dendra2* expression is BAC based, it might include additional regulatory elements that might be responsive to embryonic heart injury. It is thus necessary to perform embryonic heart injury in *Tg(wt1b:H2B-Dendra2)* embryos to further elucidate if the embryonic epicardium responds to injury with an activation comparable to that of the adult epicardium.

In summary, the newly generated transgenic lines *Tg(tcf21:myr-tdTomato)*, *Tg(tcf21:H2B-Dendra2)*, *Tg(tcf21:myr-GFP)*, *Tg(tbx18:myr-GFP)* and *Tg(tbx18:myr-Citrine)* appeared to accurately report on the endogenous epicardial expression of Tcf21 and Tbx18, respectively. In contrast, *Tg(tbx18:H2B-Dendra2)* did not capture the full epicardial expression pattern of Tbx18 and *Tg(wt1b:H2B-Dendra2)* ectopically labelled epicardial cells that lacked endogenous *Wt1b* transcripts. Lastly, the fluorescence in *Tg(wt1b:myr-tdTomato)* was very faint, interfering with a proper analysis of epicardial reporter expression.

## 6.2 Epicardial subpopulations in the developing zebrafish heart

### 6.2.1 *wt1b*:H2B-Dendra2 single positive cells cover the bulbus arteriosus

Microscopic analysis of the newly generated epicardial reporter lines found the epicardial expression of *Tcf21*, *Tbx18* and *Wt1b* in the developing zebrafish embryo to be heterogeneous. It also identified an epicardial cell population formed by *wt1b*:H2B-Dendra2 positive, *tcf21*:myr-tdTomato and *tbx18*:myr-GFP negative cells covering the BA (figure 3-3, figure 6-1a). This population might represent the part of the epicardium that is not derived from the PEO but from the pericardium (Peralta et al., 2014). This would explain the expression of *wt1b*:H2B-Dendra2 in these cells, as the pericardial sac was strongly fluorescent as well.

However, single cell RNA sequencing showed that *Tg(wt1b:H2B-Dendra2)* derived transcriptomes did not cluster separately from other epicardial cell transcriptomes and that the epicardial cell clusters which expressed *Wt1b* also expressed *Tcf21* and *Tbx18* (figure 4-3 and figure 4-4). Potentially, the *wt1b*:H2B-Dendra2 single positive cell population on the BA was represented as a subcluster within epicardial cell cluster 1, the single cell cluster that contained all *Tg(wt1b:H2B-Dendra2)* derived single cell transcriptomes. However, epicardial cell cluster 1 contained very few cells that were neither *tcf21* or *tbx18* positive. False positive reporter expression in *Tg(wt1b:H2B-Dendra2)*, as discussed above, might explain the absence of a distinct single cell cluster matching the cell population identified through microscopy. Another possibility is that *Tg(wt1b:H2B-Dendra2)* derived single cells from the BA were underrepresented in the single cell data set, as these cells might have remained attached to the extra-cardiac pericardium during heart purification and thus would have been lost before FACS-

purification. Microscopic analysis of isolated *Tg(wt1b:H2B-Dendra2)* derived hearts would show whether this is the case.

## **6.2.2 Adma, Podxl, Tcf21, Tbx18, Wt1b expressing cells form the epicardial sheet and might regulate mesenchymal cells in the developing zebrafish heart**

Unlike microscopic analysis based on transgenic reporter fluorescence, single cell RNA sequencing identified epicardial subpopulations based on endogenous gene expression patterns. Most of these subpopulations would have remained obscured during bulk population transcriptomics, and only single cell RNA sequencing enabled their identification. Single cell transcriptomics complemented microscopic analysis, providing an entire gene expression profile of each cell population identified. This allowed for an in-depth characterization of the three epicardial subpopulations identified through single cell transcriptomics and of their interactions with non-epicardial cell populations in the developing zebrafish heart.

Epicardial cell cluster 1 was the largest cluster of epicardial cells identified and many cells contained in it were labelled by the expression of *Tcf21*, *Tbx18* and *Wt1b* (figure 4-4, figure 6-1a). Some cells in epicardial cell cluster 1 expressed other genes known to label epicardial cells, such as *Scxa*, *Sema3d* and *Aldh1a2*. Additionally, cells in epicardial cell cluster 1 expressed several genes that to date have not been studied in the epicardium and which were exclusively expressed in this cell cluster. Among them was the signaling peptide *Adma*, which appeared to be expressed in the epicardial cell layer, co-localizing

with the expression of *Tcf21* and *Tbx18* (figure 5-3). It is thus likely that most of the epicardial sheet is formed by cells resembling those contained in epicardial cell cluster 1, suggesting that this cluster constitutes the part of the developing zebrafish epicardium that has previously been studied. Corroborating evidence comes from the fact that most epicardial cells in *Tg(tcf21:myr-tdTomato; tbx18:myr-GFP; wt1b:H2B-Dendra2)* embryos were triple fluorescent (figure 3-3) and that an enhancer element upstream of *podxl*, another marker gene expressed in epicardial cell cluster 1, was active in the epicardial cell layer (figure 5-9).

The enhancer *enh15-podxl* contained binding sites for *Tbx18*, *Gata4* and *Myrf* (figure 5-9). *Tbx18* has been found dispensable for epicardial development (Greulich et al., 2012) and *Gata4* and *Myrf* might be more important for the activity of *enh15-podxl* than *Tbx18*. In the mouse, *Gata4* is essential for the formation of the proepicardium (Watt et al., 2004). Single cell transcriptomics showed that *Gata4* was expressed in epicardial and mesenchymal cell clusters at 5dpf, but not in cardiomyocytes (figure 5-9). In conclusion, a non-myocardial function of *Gata4* might be to act through *enh15-podxl*. *Myrf* activates the transcription of Myelin genes and is not known to play a role during cardiogenesis, however the prominent enrichment in epicardial cell cluster 1 suggests that *Myrf* regulates epicardial development. Due to the presence of a *Myrf* binding site within *enh15-podxl* and the high correlation between the expression of *Myrf* and *Podxl* (figure 5-9), it is possible that *Myrf* drives *Podxl* expression through *enh15-podxl* during epicardial development.

Some of the genes enriched in epicardial cell cluster 1 have been found to interact during epicardial development (Summarized in figure 6-1b). In the mouse, the expression of the retinoic acid producing enzyme *Raldh2* is directly activated by *Wt1* (Guadix et al., 2011)

and *Aldh1a2* might be regulated similarly by *Wt1a* and *Wt1b* in epicardial cell cluster 1. *Wt1* was also shown to activate the expression of  $\alpha 4$ Integrin (*Itga4*) (Kirschner et al., 2006). Transcripts encoding *Wt1a*, *Wt1b* and *Itga4* (data not shown) were enriched in epicardial cell cluster 1, making it possible that *Itga4* is induced by *Wt1a* or *Wt1b* to enhance cell adhesion in the epicardial sheet. Furthermore, factors expressed in epicardial cell cluster 1 are likely to regulate non-epicardial cells in the developing zebrafish heart. The strong enrichment of transcripts encoding *Adma* suggests that it is important in this context. The *Adma* receptors *Calcrla* and *Ramp2* were prominently expressed in epicardial cell cluster 1 itself, but also in a cluster of cells termed mesenchymal due to the enrichment of *abi3bpb* and *cyp26b1* in most of these cells (figure 4-5). The co-expression of multiple transcription factors driving endothelial cell formation as well as other endothelial cell markers in some of the mesenchymal cells suggested that these cells were in the process of endothelial differentiation, potentially resembling mesenchymal-to-endothelial transition observed following cardiac injury in the mouse (Ubil et al., 2014). Epicardial *Adma* might be one of the factors driving this process. Since the coronary vasculature in the zebrafish heart develops only at the juvenile stage (Harrison et al., 2015), endothelial cells derived from the mesenchymal cell cluster might instead contribute to the endocardium and the cardiac valves. Lineage tracing revealed that cardiac valves in the developing mouse heart are mostly formed by endocardial cells that have undergone endothelial-to-mesenchymal transition (De Lange et al., 2004). Microscopic analysis of developing zebrafish heart valves does not support the formation of mesenchymal cushions (Pestel et al., 2016; Scherz et al., 2008). However, zebrafish valve endocardial cells might still activate a mesenchymal transcriptional program to enable the complex cellular rearrangements taking place

within the endocardial layer. Thus, the *Adma* responsive subset of cells in the mesenchymal cell cluster might alternatively be mobilized endothelial cells contributing to the valves. These cells expressed *Nfatc1*, one of the factors driving cardiac valve formation. Other factors of valve formation however, such as *Sox9a* and *Scxa*, were not expressed in endothelial-like mesenchymal cells. Further imaging and lineage tracing experiments will be necessary to clarify the location and tissue contribution of *Adma* responsive mesenchymal cells, since immunofluorescence of *Cyp26b1* exclusively labelled a different, *myl7:GFP* positive subset of cells in the mesenchymal cell cluster (figure 5-7). Additionally, the expression patterns of *Adma* and its receptors suggest it might have a role within the epicardium itself, making *Adma* likely to play multiple roles in the developing zebrafish heart.

### **6.2.3 *Lox*, *Elnb*, *Mylka*, *Acta2*, *Tbx18* expressing smooth muscle cells in the bulbus arteriosus are potentially regulated by *Wnt11r* signaling**

There was more to the developing zebrafish epicardium than epicardial cell cluster 1. Single cell transcriptomics additionally identified an epicardial cell population, termed epicardial cell cluster 2, that expressed *Tbx18* and lacked the expression of *Tcf21* and *Wt1b* (figure 4-2, 4-3, 4-4 and 6-1a). The transcriptomic profile of these cells suggested they were ECM producing smooth muscle cells, as many of them expressed *Lox* and *Elnb*, as well as *Mylka*, *Acta2* and *Myh11a* (figure 4-6). HCR and immunofluorescence supported the presence of epicardial cell cluster 2 and revealed it to be spatially distributed within the BA (figure 5-4). Furthermore, enhancer elements that were accessible in *tbx18:myr-GFP/tbx18:myr-Citrine* expressing epicardial cells and located in

intron 3 of *lox*, as well as upstream of *elnb*, showed exclusive activity within the BA (figure 5-10 and figure 5-11). Transcription factors that might bind within the identified enhancers and were expressed within epicardial cell cluster 2 include *Tbx20* and *Sox9a*, which are crucial regulators of cardiac valve formation at the atrioventricular boundary and in the outflow tract/BA (Chakraborty et al., 2010). Additionally, cells in epicardial cell cluster 2 strongly expressed *Hand2*, which positively regulates BA growth (Schindler et al., 2014). Consistent with these findings, a previous publication found *Elnb* and *Mykx* to specifically label smooth muscle in the developing zebrafish BA (Moriyama et al., 2016). Also, many of the genes enriched in epicardial cell cluster 2 are expressed in the adult zebrafish BA (Singh et al., 2016). Taken together, there is strong support for a *Tbx18* expressing cell population that contributes to the BA of the developing zebrafish heart. Since these cells were part of the smooth muscle layer of the BA, they were different from the *wt1b:H2B-Dendra2* expressing cell population that formed the epicardial cell layer covering the BA.

Epicardial cell cluster 2 might originate from the epicardium, as suggested by the expression of *Tbx18*, or from other cardiac lineages contributing to the BA. A subset of neural crest cells has been shown to contribute to the developing zebrafish BA (Cavanaugh et al., 2015). However, cells in epicardial cell cluster 2 did not express the neural crest marker *Sox10* (figure 4-2). Cells from the second heart field give rise to BA tissue and parts of the zebrafish ventricle (Guner-Ataman et al., 2013; Zhou et al., 2011b). These cells were identified by the expression of *Ltbp3* and *Nkx2.5*. Indeed, cells in epicardial cell cluster 2 expressed *Ltbp3*, raising the possibility they were derived from the second heart field (figure 5-4). However, *Nkx2.5* was not expressed within epicardial cell cluster 2 (figure 5-4) and *Ltbp3* might be expressed in cells that are not second heart

field derived. While there thus is the possibility of a second heart field origin of epicardial cell cluster 2, strong support for an epicardial origin of epicardial cell cluster 2 comes from zebrafish lineage tracing studies using a *tcf21* driven Cre (Kikuchi et al., 2011a). Here, Tcf21 expressing embryonic cells were found to contribute to the adult BA and to express Mylka. The authors concluded that these cells originated from the embryonic epicardium, since they did not detect *tcf21* reporter expression in the smooth muscle layer of the embryonic BA. Smooth muscle cells covering the coronary vasculature were not labelled with this approach, suggesting a different origin for coronary smooth muscle. In summary, it is likely that epicardial cell cluster 2 is derived from the epicardium, however further lineage tracing studies, for example using *wt1b* driven Cre, might provide further corroborating evidence, since Wt1b, like Tcf21, is not expressed in BA smooth muscle (figure 3-3).

Tbx18 was shown to prevent premature differentiation of epicardium derived cells into smooth muscle (Greulich et al., 2012), which might explain the decreased level of Tbx18 expression in smooth muscle marker expressing cells in epicardial cell cluster 2, as compared to *tbx18* positive cells in epicardial cell cluster 1 (figure 4-4). The low level of Tbx18 expression in epicardial cell cluster 2 might explain why this cell population was not detectable during microscopic analysis of *tbx18* driven myr-GFP fluorescence (figure 3-3). Indeed, only the bright fluorescence of Citrine enabled the detection of *tbx18* positive cells in the BA of *tbx18* reporter embryos (figure 5-4). The lack of Tcf21 expression in cells in epicardial cell cluster 2 might be explained by the fact that Tcf21 was shown to prevent epicardial trans-differentiation into smooth muscle (Braitsch et al., 2012). Furthermore, Tbx18 promotes epicardial EMT *in vitro* (Takeichi et al., 2013) and many cells in epicardial cell cluster 2 expressed several markers of EMT, such as Snai2

and Twist1a (figure 4-4), supporting the hypothesis that they are epicardium derived cells.

The expression of Wnt11r as well as that of other Wnt signaling components provided an insight into the regulation of epicardial cell cluster 2 (figure 4-6, summarized in figure 6-1c). The co-expression of Wnt11r, several Frizzled receptors and downstream Wnt effectors in cells in epicardial cell cluster 2 indicates these cells might be subject to cell-autonomous Wnt regulation. Additionally, myocardial Wnt11 was shown to regulate outflow tract development in the mouse (van Vliet et al., 2017). In the single cell data set, Wnt11 was expressed by some cells in epicardial cell cluster 1, but also by a subset of cells in the mesenchymal cell cluster. The transcriptomic profile of these mesenchymal cells resembled that of cardiac valve progenitor cells, many expressing markers such as Crip2, Sox9a, Nfatc1 and Tbx2a. Their widespread presence in the ventricular myocardium (figure 5-7) additionally suggested they might be cells contributing to the myocardium. Potentially, Wnt11r and Wnt11 in these cells synchronize the development of the myocardium with that of the BA, perhaps in response to intra-cardiac fluid forces. Recruitment of epicardium derived cells might thus be an important mechanism to fortify the BA as the zebrafish embryo grows in size and the heart generates increasing force to maintain the circulation.

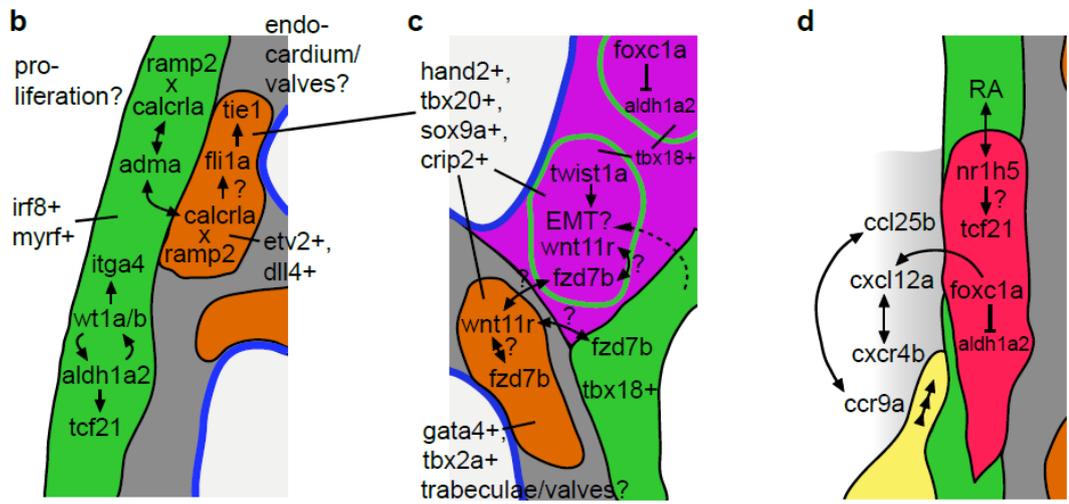
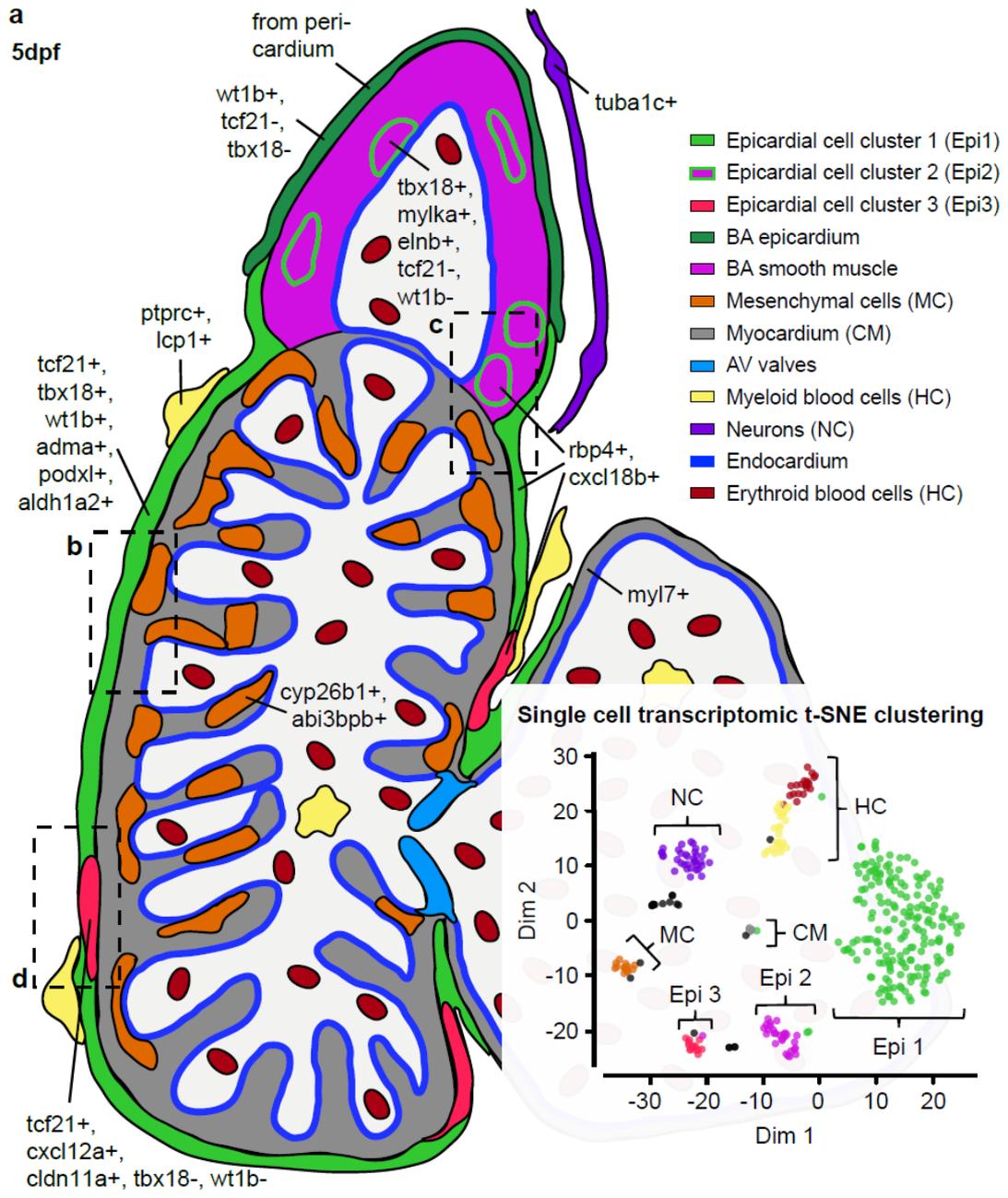
#### **6.2.4 Cldn11a, Cxcl12a, Nr1h5, Tcf21 expressing cells in the epicardial sheet might attract myeloid haematopoietic cells into the developing zebrafish heart**

The third epicardial cell population identified through single cell transcriptomics was termed epicardial cell cluster 3. Many of these cells expressed Tcf21, completely lacked transcripts encoding Tbx18 and mostly lacked transcripts encoding Wt1b (figure 4-2, 4-4 and 6-1a). A diverse set of genes was enriched in epicardial cell cluster 3, including the tight junction marker Cldn11a and the chemokine Cxcl12a. HCR connected the molecular profile with the position within the heart and suggested Cldn11a was expressed in some regions of the developing epicardial cell layer (figure 5-5), making it possible that cells belonging to epicardial cell cluster 3 were interspersed with cells belonging to epicardial cell cluster 1. This was supported by the fact that microscopic analysis identified a small population of *tcf21:myr-tdTomato* single positive epicardial cells (figure 3-3). Marker genes of epicardial cell cluster 3 showed few accessible chromatin regions in their vicinity.

One characteristic of cells in epicardial cell cluster 3 was that they expressed several retinoic acid receptors, such as Nr1h5 (figure 4-2, summarized in figure 6-1d). Epicardial retinoic acid signaling plays an important role both during heart development and regeneration (Kikuchi et al., 2011b; Stuckmann et al., 2003) and stimulates cardiomyocyte proliferation, but also regulates the expression of Tcf21 and Wt1 in the epicardium itself (Braitsch et al., 2012). Retinoic acid produced by Aldh1a2 in epicardial cell cluster 1 might have a similar function in cells contained in epicardial cell cluster 3. The transcription factor Foxc1a was enriched in epicardial cell clusters 2 and 3 and might

regulate the expression of key genes in epicardial cell cluster 2, such as *Lox* and *Elnb* (figure 5-10 and figure 5-11). Additionally, *Foxc1a* has been found to directly inhibit the expression of *Aldh1a2* during somitogenesis (Li et al., 2015). *Aldh1a2* was not expressed in cells within epicardial cell clusters 2 and 3 and *Foxc1a* might thus be crucial to regulate intra-epicardial retinoic acid signaling in the developing zebrafish embryo. Taken together, retinoic acid signaling seemed to be a factor connecting the epicardial cell populations, which was further underlined by the enrichment of the retinol transporter *Rbp4* across the three epicardial cell clusters (figure 4-4).

*Foxc1* activates the expression of *Cxcl12* in mammalian bone marrow cells (Omatsu et al., 2014) and *Foxc1a* might therefore stimulate the expression of *Cxcl12a* in epicardial cell cluster 3. However, *Foxc1a* would need to co-operate with other factors in epicardial cell cluster 2, since *Cxcl12a* was not expressed there. Alternatively, *Foxc1a* mediated transcriptional activation of *Cxcl12a* might be repressed by factors expressed in epicardial cell cluster 2. Epicardial *Cxcl12a* has been studied as a guidance cue for cardiomyocytes following cardiac injury (Itou et al., 2012b). The expression of *Cxcl12a* in epicardial cell cluster 3 appeared to play a role in guiding haematopoietic cells in the developing zebrafish heart (figure 4-7). These cells had a transcriptomic profile indicating they were myeloid in nature. Indeed, report lines and immunofluorescence revealed that both *ptprc:dsRed* and *Lcp1* expressing extracardiac cells were in close contact with the developing zebrafish epicardium (figure 5-8). Hence, epicardial *Cxcl12a* may act to attract myeloid haematopoietic cells to the developing zebrafish heart, where they might fulfill immune surveillance functions.



**Figure 6-1: Subpopulations in the developing zebrafish heart, as identified in this study, and their possible regulation.** (a) The zebrafish heart at 5dpf contains several cell populations from different lineages. Subpopulations expressing epicardial markers are likely located in the epicardial layer (epicardial cell cluster 1, green, expresses Tcf21, Tbx18 and Wt1b; and epicardial cell cluster 3, red, expresses Tcf21, lacks Tbx18 and Wt1b) and within the smooth muscle layer of the bulbus arteriosus (BA) (magenta) (epicardial cell cluster 2, magenta with green outline, expresses Tbx18, lacks Tcf21 and Wt1b). Additionally, a pericardium derived *wt1b* positive epicardial cell population (dark green) covers the BA. A population of mesenchymal cells (orange) might be located in the myocardium (grey) and multiple extra-cardiac myeloid haematopoietic cells (yellow) are in direct contact with the epicardial layer. Neurons (purple) form an axonal structure in proximity to the heart. Text boxes give genes that are enriched (+) or depleted (-) within the individual cell populations. The t-SNE plot shows cell clusters determined by single cell RNA sequencing in colors corresponding to those in the graphic. (b-d) Possible regulatory interactions between cardiac cell populations, as suggested by single cell transcriptomics. Arrows indicate transcriptional activation, blunted arrows indicate transcriptional repression. If a regulatory interaction between two factors has not been demonstrated in the literature, it is labelled with a question mark. Double headed arrows indicate ligand-receptor interactions. A dashed arrow in c indicates a possible derivation of epicardial cell cluster 2 from the epicardial layer, possibly involving Twist1a mediated epithelial-to-mesenchymal transition (EMT). Epi = epicardial cell cluster, MC = mesenchymal cells, CM = cardiomyocytes, NC = neural cells, HC = haematopoietic cells, AV = atrio-ventricular, RA = retinoic acid.

### 6.3 Non-epicardial cell populations in the developing zebrafish heart

In addition to the epicardial subpopulations, single cell transcriptomics identified multiple other cardiac cell populations that might be of interest for future studies. However, there was no endocardial cell population present in the single cell data set, even though one of the reporter constructs used for cell purification, *kdrl:GFP*, was shown to label the zebrafish endocardium (Cavanaugh et al., 2015). Additionally, I could clearly see the *kdrl:GFP* labelled endocardium in *Tg(kdrl:GFP; gata1a:dsRed)* embryos at 5dpf under the microscope (data not shown).

Instead, *kdrl:GFP* purified cells contributed to the myeloid haematopoietic cell cluster (figure 4-7). Potentially, endocardial cells did not respond well to the heart dissociation protocol used before FACS. *kdrl:GFP* expressing haematopoietic cells might be more robust towards the dissociation protocol used, explaining the bias in the single cell population. In conclusion, processing a higher number of cells derived from *Tg(kdrl:GFP)* and adjusting the dissociation protocol (Patra et al., 2017) could enable the generation of single cell transcriptomic data from an endocardial cell population.

#### 6.3.1 Mesenchymal cells

One of the non-epicardial cell populations detected in the developing zebrafish heart through single cell transcriptomics appeared to consist of mesenchymal progenitor cells, as these expressed the mesenchymal cell markers *Abi3bbp* and *Cyp26b1* (figure 4-2). The mesenchymal cell cluster contained at least two subsets of cells with distinct transcriptomic profiles. Cells in the first subset expressed several markers of endothelial

cells (figure 4-5), cells in the second subset expressed markers of cardiomyocytes, such as *MyI7* (figure 4-2). Immunofluorescence of *Cyp26b1* validated the second subset of mesenchymal cells, since *Cyp26b1* was detected within *myI7:GFP* expressing cells (figure 5-7).

Many cells within the mesenchymal cell cluster expressed the cardiac fibroblast marker *Pdgfra* (Moore-Morris et al., 2014; Rudat et al., 2013) and therefore might be cardiac fibroblasts, potentially derived from the epicardium. However, the expression of *Tcf21* was observed to persist during the differentiation of epicardium-derived cardiac fibroblasts (Acharya et al., 2012; Braitsch et al., 2012), and there was no expression of *Tcf21* in the mesenchymal cell cluster (figure 4-4). Cells within the mesenchymal cell cluster might contribute to the developing cardiac valves, as suggested by the expression of several valve markers, such as *Nfatc1* (figure 4-4), *Sox9a* and *Tbx2a* (figure 4-6). Microscopic analysis of valve formation in the zebrafish heart suggested that valves are formed through endocardial invagination directly and not from mesenchymal endocardial cushions as in mammals (Scherz et al., 2008). However, it is unknown how much of a mesenchymal phenotype endocardial cells adopt during invagination. The expression of endothelial markers within the mesenchymal cell cluster makes it possible that these could be endocardial cells that underwent EMT during valve formation. *Thrombospondin 1b* (*Thbs1b*), a factor expressed in the mesenchymal cell cluster (figure 4-2), was found to be significantly enriched in the zebrafish heart between 48hpf and 56hpf (Steed et al., 2016). Potentially, this time point marks the earliest presence of a mesenchymal transcriptomic profile in the embryonic zebrafish heart. Interestingly, this coincides with the onset of atrio-ventricular valve formation in the developing zebrafish heart (Beis, 2005).

The time between 48hpf and 56hpf also marks the onset of myocardial trabeculation, a process during which cardiomyocytes delaminate from the cortical myocardium to form trabeculae that expand into the ventricular lumen (Staudt et al., 2014). This and the expression of *Myl7* within cells in the mesenchymal cell cluster makes it possible that these could be derived from the myocardium, contributing to the forming trabeculae. The widespread presence of *Cyp26b1* in ventricular cardiomyocytes (figure 5-7) as well as the contribution of *Tg(myl7:GFP)* derived cells to the mesenchymal cell cluster (figure 4-2) matched this hypothesis. Delaminated cardiomyocytes lose their polarity (Jiménez-Amilburu et al., 2016) and might adopt other mesenchymal characteristics. Apart from *Erb3a* (figure 4-6), components of Neuregulin signaling were not expressed in the mesenchymal cell cluster, although Neuregulin signaling is crucial for trabeculation (Liu et al., 2010). Another essential driver of trabeculation is blood flow, and most of the cells in the mesenchymal cell cluster expressed the flow-responsive Krüppel-Like Factor 2 zebrafish paralogues *Klf2a* and *Klf2b* (Steed et al., 2016).

There is the alternative possibility that cells contained in the mesenchymal cell cluster were not derived from differentiated cardiac cells, but were mesenchymal cells directly contributing to the developing heart. The second heart field contains both endocardial and myocardial progenitor cells (Milgrom-Hoffman et al., 2011), raising the possibility that the mesenchymal cell cluster was derived from the second heart field. Also, some cells within the mesenchymal cell cluster expressed *Ltb3*. However, the expression of *Ltb3* did not label the entire mesenchymal cell cluster and cells in this cluster did not express *Nkx2.5*. *Cyp26a1* and *Cyp26c1* are required for the addition of second heart field progenitor cells to the developing zebrafish heart by restricting retinoic acid (Rydeen and Waxman, 2016), however *Cyp26a1* and *Cyp26c1* were not expressed in the single cell

data set. *Cyp26b1* might fulfill a similar function and restrict retinoic acid signaling in the mesenchymal cell cluster to facilitate their addition to the developing heart. Potentially, *Cyp26b1* also acts to attenuate the maturation of these cells, like it does during skeletal development in the zebrafish (Laue et al., 2008). *Abi3bp* on the other hand was found to promote cardiac progenitor differentiation (Hodgkinson et al., 2014). The concerted action of *Cyp26b1* and *Abi3bp* might thus set the pace for the maturation of the cells in the mesenchymal cell cluster.

In conclusion, my data raise the following possibilities regarding the origin and function of the cells in the mesenchymal cell cluster. Their origin might be in the endocardium, in the myocardium and/or in the second heart field and their contribution might be to the cardiac valves and/or to the myocardium, potentially to the trabeculae. Microscopic analysis of the BAC reporter line *Tg(cyp26b1:YFP)* (Spoorendonk et al., 2008) and immunofluorescence of *Cyp26b1* in *ltbp3* lineage traced cells (Zhou et al., 2011b), will be necessary to elucidate the origin and the exact contribution of cells in the mesenchymal cell cluster identified in this study. The mesenchymal cell cluster is of considerable interest for studies focused on the epicardium, as it might be regulated by epicardial *Adma* signaling and itself might regulate epicardial contribution to the BA via *Wnt11* signaling. Furthermore, cardiomyocytes were found to differentiate from *hand2* and *tbx20* positive progenitor cells during heart regeneration (Lepilina et al., 2006). These adult progenitor cells might be similar to the cells in the mesenchymal cell cluster, as these expressed both *Hand2* and *Tbx20* (figure 5-10).

### 6.3.2 Neural cells

Another non-epicardial cell population identified in the developing zebrafish heart consisted of neuronal cells (figure 4-2). Neuronal support cells were present as well in the single cell RNA sequencing data set. Immunofluorescence confirmed the presence of Tuba1c positive neuronal structures adjacent to the developing heart (figure 5-6). The development of nerves in the zebrafish heart has not been studied so far, although nerves are important promoters of cardiac regeneration (Mahmoud et al., 2015). There are possible roles for the epicardium in this context. At 5dpf, neuronal structures appeared fully formed and neuronal single cells did not express chemokine receptors like *Cxcr4b*. However, the epicardium might guide nerves at an earlier time point during development, for example through *Cxcl12a*. Additionally, the epicardium might play a role during the neural innervation of the developing heart, as innervating axons are likely to be in direct contact with epicardial cells.

### 6.4 Future studies

One important focus of future studies should be to elucidate the functions of the epicardial subpopulations identified in this study. Single cell transcriptomics revealed several possible cell fates and interactions between epicardial and non-epicardial cardiac cell populations and loss of function experiments would prove useful to further investigate these. Knock out of *adma* in an epicardium-specific manner, connected to microscopic analysis of *Tg(tcf21:H2B-Dendra2)* or *Tg(tbx18:myr-Citrine)* and immunofluorescence studies of markers such as *Cyp26b1*, would elucidate the

importance of *Adma* during development, either functioning cell-autonomously within the epicardium itself and/or in cells of the mesenchymal cell cluster that express endothelial genes. Similarly, epicardium-specific knock out of genes encoding markers such as *Elnb*, *Acta2* or *Wnt11r*, connected to analysis of *Tg(myl7:GFP)* and expression of *Mylka*, would be suited to further investigate the importance of epicardial cell cluster 2 in fortifying the developing BA and the impact of *Wnt11r* signaling on this process. General knock out of *elnb* was shown to induce the loss of smooth muscle in the BA and its replacement by myocardium (Moriyama et al., 2016). It would thus be interesting to see if the epicardium specific loss of *Elnb*, *Acta2* or *Wnt11r* recapitulates this phenotype or if it refines these results. Finally, epicardium specific knock out of *cxcl12a* plus monitoring of the migration of *ptprc:dsRed* and/or *Lcp1* positive cells in contact with the epicardium, relative to the wildtype situation, would elucidate if epicardial cells might be involved in guiding myeloid haematopoietic cells in the developing heart. To this end, I am generating transgenic lines in which Cas9 (Cong et al., 2013) is either driven by the epicardial enhancer elements identified in this study or by a 3.4kb *tcf21* promoter region. With this I aim to express Cas9 during the earliest steps of epicardial development, in order to generate a specific knock out by CRISPR/Cas9 technology in every cell contributing to any of the three epicardial subpopulations. Vectors that encode single guide RNAs targeting genes of interest, such as *adma*, *acta2* and *cxcl12a*, will be injected into Cas9 expressing transgenic embryos to ensure the continued presence of these guide RNAs until the onset of epicardial development. The impact of each knock out condition on epicardial development and cell fate decisions will then be analyzed using reporter lines and immunofluorescence, as described above.

*Tg(wt1b:Cre)*, a lineage tracing line recently generated in the lab, crossed to a ubiquitous *Tg(ubb:LOXP-AmCyan-LOXP-ZsYellow)* labelling line (Ghayé et al., 2015) could verify whether *Tbx18* and *Mylka* expressing cells in the BA originate from the epicardium.

Finally, closer investigation of the identified enhancers *enh15-podxl*, *enh17-lox* and *enh20-elnb* could provide valuable insights into the transcriptional regulation of key genes expressed in epicardial cell clusters 1 and 2. Specific perturbation of transcription factor binding sites in the core of the active enhancers will enable the investigation of transcription factors that drive the activity of *enh15-podxl*, *enh17-lox* and *enh20-elnb*. Knocking out the identified transcription factors and CRISPR-targeting their binding sites on the endogenous enhancer sequences will further allow to study the importance of such regulatory sequences and the respective transcription factors during the formation and/or function of epicardial cell clusters 1 and 2. This will refine the current knowledge on gene regulatory networks driving epicardial development (Braitsch and Yutzey, 2013). Ultimately, it will be important to further explore the findings obtained in the zebrafish embryonic epicardium in the adult zebrafish and mammalian models of heart injury. Some of the factors potentially regulating the identified embryonic epicardial cell clusters and their interactions with non-epicardial cell populations are known to be involved in heart regeneration. These include EMT markers (Kim et al., 2010), genes expressed in epicardial cell cluster 1, such as *aldh1a2* (Kikuchi et al., 2011b) and *fn1b* (Wang et al., 2013), genes expressed in epicardial cell cluster 2, such as *hand2* (Schindler et al., 2014), as well as genes expressed in epicardial cell cluster 3, such as *cxcl12a* (Itou et al., 2012b; Lien et al., 2006). This underlines that embryonic and adult regenerating epicardial cells share similarities in their genetic programs. It would therefore be interesting to analyze if counterparts of the embryonic epicardial subpopulations are present in the adult

zebrafish and necessary during heart regeneration. For example, epicardium mediated recruitment of myeloid haematopoietic cells to the injured zebrafish heart could serve to promote the regenerative process (Aurora et al., 2014; Lai et al., 2017). Adult cells resembling those in epicardial cell cluster 2 might furthermore provide the regenerating coronary vasculature with smooth muscle and thereby aid heart repair (Lepilina et al., 2006; Marín-Juez et al., 2016) or guide regeneration of the epicardium itself (Wang et al., 2015). Finally, Gata4 expressing cells resembling those in the mesenchymal cell cluster (figure 4-6) might be crucial for myocardial repair, as *gata4* reporter positive cardiomyocytes are major contributors to myocardial regeneration (Kikuchi et al., 2010). This would be particularly interesting if cells in the mesenchymal cell cluster turned out to be derived from the developing myocardium, as discussed above, since myocardial regeneration in the zebrafish is thought to be mediated by de-differentiated pre-existing cardiomyocytes (Jopling et al., 2010).

## **6.5 Final remarks**

The epicardium is heterogeneous in nature and gives rise to multiple cell types during cardiogenesis. Focusing on the genomic profiling of transcriptional and epigenetic cell states, I identified multiple epicardial cell populations in the developing zebrafish embryo that are likely to have distinct and specific fates and functions. I furthermore propose new associations between cardiac cell populations during heart development. My studies will build into the known regulatory networks driving cardiac development and will enable a more accurate functional analysis of epicardial development and the epicardial response to cardiac injury.

## References

- Abu-Elmagd, M., Mulvaney, J., and Wheeler, G.N. (2017). Frizzled-7 is required for *Xenopus* heart development. *Biol. Open* 6.
- Acharya, A., Baek, S.T., Huang, G., Eskiocak, B., Goetsch, S., Sung, C.Y., Banfi, S., Sauer, M.F., Olsen, G.S., Duffield, J.S., et al. (2012). The bHLH transcription factor Tcf21 is required for lineage-specific EMT of cardiac fibroblast progenitors. *Development* 139, 2139–2149.
- Aggarwal, S., Yurlova, L., Snaidero, N., Reetz, C., Frey, S., Zimmermann, J., Pähler, G., Janshoff, A., Friedrichs, J., Müller, D.J., et al. (2011). A size barrier limits protein diffusion at the cell surface to generate lipid-rich myelin-membrane sheets. *Dev. Cell* 21, 445–456.
- Airik, R., Bussen, M., Singh, M.K., Petry, M., and Kispert, A. (2006). Tbx18 regulates the development of the ureteral mesenchyme. *J. Clin. Invest.* 116, 663–674.
- Akiyama, H., Chaboissier, M.C., Martin, J.F., Schedl, A., and De Crombrughe, B. (2002). The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* 16, 2813–2828.
- Akiyama, H., Chaboissier, M.-C., Behringer, R.R., Rowitch, D.H., Schedl, A., Epstein, J.A., and de Crombrughe, B. (2004). Essential role of Sox9 in the pathway that controls formation of cardiac valves and septa. *Proc. Natl. Acad. Sci.* 101, 6502–6507.
- Alexander, J.M., and Bruneau, B.G. (2010). Lessons for cardiac regeneration and repair through development. *Trends Mol. Med.* 16, 426–434.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106.
- Anderson, K.L., Nelson, S.L., Perkin, H.B., Smith, K.A., Klemsz, M.J., and Torbett, B.E. (2001). PU.1 is a Lineage-specific Regulator of Tyrosine Phosphatase CD45. *J. Biol. Chem.* 276, 7637–7642.
- Arrate, M.P., Rodriguez, J.M., Tran, T.M., Brock, T.A., and Cunningham, S.A. (2001). Cloning of Human Junctional Adhesion Molecule 3 (JAM3) and Its Identification as the JAM2 Counter-receptor. *J. Biol. Chem.* 276, 45826–45832.
- Arrenberg, A.B., Stainier, D.Y.R., Baier, H., and Huisken, J. (2010). Optogenetic control of cardiac function. *Science* 330, 971–974.
- Aurora, A.B., Porrello, E.R., Tan, W., Mahmoud, A.I., Hill, J.A., Bassel-Duby, R., Sadek, H.A., and Olson, E.N. (2014). Macrophages are required for neonatal heart regeneration. *J. Clin. Invest.* 124, 1382–1392.

Bagley, R.G., Honma, N., Weber, W., Boutin, P., Rouleau, C., Shankara, S., Kataoka, S., Ishida, I., Roberts, B.L., and Teicher, B.A. (2008). Endosialin/TEM 1/CD248 is a pericyte marker of embryonic and tumor neovascularization. *Microvasc. Res.* 76, 180–188.

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

Balmer, G.M., Bollini, S., Dubé, K.N., Martinez-Barbera, J.P., Williams, O., and Riley, P.R. (2014). Dynamic haematopoietic cell contribution to the developing and adult epicardium. *Nat. Commun.* 5, 4054.

Behrens, A.N., Zierold, C., Shi, X., Ren, Y., Koyano-Nakagawa, N., Garry, D.J., and Martin, C.M. (2014). Sox7 Is Regulated by ETV2 During Cardiovascular Development. *Stem Cells Dev.* 23, 34–36.

Beis, D., Bartman T, Jin SW, Scott IC, D'Amico LA, Ober EA, Verkade H, Frantsve J, Field HA, Wehman A, Baier H, Tallafuss A, Bally-Cuif L, Chen JN, Stainier DY, Jungblut B (2005). Genetic and cellular analyses of zebrafish atrioventricular cushion and valve development. *Development* 132, 4193–4204.

Belmadani, A., Tran PB, Ren D, Assimacopoulos S, Grove EA, Miller RJ. (2005). The Chemokine Stromal Cell-Derived Factor-1 Regulates the Migration of Sensory Neuron Progenitors. *J. Neurosci.* 25, 3995–4003.

Bertrand, J.Y., Jalil, A., Klaine, M., Jung, S., Cumano, A., and Godin, I. (2005). Three pathways to mature macrophages in the early mouse yolk sac. *Blood* 106, 3004–3011.

Bertrand, J.Y., Kim, A.D., Teng, S., and Traver, D. (2008). CD41+ cmyb+ precursors colonize the zebrafish pronephros by a novel migration route to initiate adult hematopoiesis. *Development* 135, 1853–1862.

Binder, Z.A., Siu, I.-M., Eberhart, C.G., ap Rhys, C., Bai, R.-Y., Staedtke, V., Zhang, H., Smoll, N.R., Piantadosi, S., Piccirillo, S.G., et al. (2013). Podocalyxin-Like Protein Is Expressed in Glioblastoma Multiforme Stem-Like Cells and Is Associated with Poor Outcome. *PLoS One* 8, e75945.

Bleul, C.C., Fuhlbrigge, R.C., Casasnovas, J.M., Aiuti, a, and Springer, T. a (1996). A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1). *J. Exp. Med.* 184, 1101–1109.

Boldajipour, B., Mahabaleswar, H., Kardash, E., Reichman-Fried, M., Blaser, H., Minina, S., Wilson, D., Xu, Q., and Raz, E. (2008). Control of Chemokine-Guided Cell Migration by Ligand Sequestration. *Cell* 132, 463–473.

Bollini, S., Vieira, J.M.N., Howard, S., Dubè, K.N., Balmer, G.M., Smart, N., and Riley, P.R. (2014). Re-activated adult epicardial progenitor cells are a heterogeneous population molecularly distinct from their embryonic counterparts. *Stem Cells Dev.* 23, 1719–1730.

Bouman, H.G., Broekhuizen, M.L., Baasten, A.M., Gittenberger-de Groot, A.C., and Wenink, A.C. (1998). Diminished growth of atrioventricular cushion tissue in stage 24 retinoic acid-treated chicken embryos. *Dev. Dyn.* *213*, 50–58.

Brade, T., Kumar, S., Cunningham, T.J., Chatzi, C., Zhao, X., Cavallero, S., Li, P., Sucov, H.M., Ruiz-Lozano, P., and Duester, G. (2011). Retinoic acid stimulates myocardial expansion by induction of hepatic erythropoietin which activates epicardial Igf2. *Development* *138*, 139–148.

Braitsch, C., and Yutzey, K. (2013). Transcriptional Control of Cell Lineage Development in Epicardium-Derived Cells. *J. Dev. Biol.* *1*, 92–111.

Braitsch, C.M., Combs, M.D., Quaggin, S.E., and Yutzey, K.E. (2012). Pod1/Tcf21 is regulated by retinoic acid signaling and inhibits differentiation of epicardium-derived cells into smooth muscle in the developing heart. *Dev. Biol.* *368*, 345–357.

Bressan, M., Liu, G., and Mikawa, T. (2013). Early Mesodermal Cues Assign Avian Cardiac Pacemaker Fate Potential in a Tertiary Heart Field. *Science* (80-. ). *340*, 744–748.

Buckingham, M., Meilhac, S., and Zaffran, S. (2005). Building the mammalian heart from two sources of myocardial cells. *Nat. Rev. Genet.* *6*, 826–837.

Buckler, a J., Pelletier, J., Haber, D. a, Glaser, T., and Housman, D.E. (1991). Isolation, characterization, and expression of the murine Wilms' tumor gene (WT1) during kidney development. *Mol. Cell. Biol.* *11*, 1707–1712.

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

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* *523*, 486–490.

Bujalka, H., Koenning, M., Jackson, S., Perreau, V.M., Pope, B., Hay, C.M., Mitew, S., Hill, A.F., Lu, Q.R., Wegner, M., et al. (2013). MYRF Is a Membrane-Associated Transcription Factor That Autoproteolytically Cleaves to Directly Activate Myelin Genes. *PLoS Biol.* *11*.

Bukrinsky, A., Griffin, K.J.P., Zhao, Y., Lin, S., and Banerjee, U. (2009). Essential role of spi-1-like (spi-1l) in zebrafish myeloid cell differentiation. *Blood* *113*, 2038–2046.

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., and Fishman, M.C.

(2005). High-throughput assay for small molecules that modulate zebrafish embryonic heart rate. *Nat. Chem. Biol.* *1*, 263–264.

Bussmann, J., and Schulte-Merker, S. (2011). Rapid BAC selection for tol2-mediated transgenesis in zebrafish. *Development* *138*, 4327–4332.

Bussmann, J., Wolfe, S.A., and Siekmann, A.F. (2011). Arterial-venous network formation during brain vascularization involves hemodynamic regulation of chemokine signaling. *Development* *138*, 1717–1726.

Cai, C.-L., Martin, J.C., Sun, Y., Cui, L., Wang, L., Ouyang, K., Yang, L., Bu, L., Liang, X., Zhang, X., et al. (2008). A myocardial lineage derives from Tbx18 epicardial cells. *Nature* *454*, 104–108.

Camenisch, T.D., Schroeder, J. a, Bradley, J., Klewer, S.E., and McDonald, J. a (2002). Heart-valve mesenchyme formation is dependent on hyaluronan-augmented activation of ErbB2-ErbB3 receptors. *Nat. Med.* *8*, 850–855.

Cannon, J.E., Place, E.S., Eve, A.M.J., Bradshaw, C.R., Sesay, A., Morrell, N.W., and Smith, J.C. (2013). Global analysis of the haematopoietic and endothelial transcriptome during zebrafish development. *Mech. Dev.* *130*, 122–131.

Cao, J., Navis, A., Cox, B.D., Dickson, A.L., Gemberling, M., Karra, R., Bagnat, M., and Poss, K.D. (2015). Single epicardial cell transcriptome sequencing identifies Caveolin-1 as an essential factor in zebrafish heart regeneration. *Development* *1*, 232–243.

Caron, K.M., and Smithies, O. (2001). Extreme hydrops fetalis and cardiovascular abnormalities in mice lacking a functional Adrenomedullin gene. *Proc. Natl. Acad. Sci.* *98*, 615–619.

Carramolino, L., Zaballos, A., Kremer, L., Villares, R., Martín, P., Ardavín, C., Martínez-A, C., and Márquez, G. (2001). Expression of CCR9  $\beta$ -chemokine receptor is modulated in thymocyte differentiation and is selectively maintained in CD8<sup>+</sup> T cells from secondary lymphoid organs. *Blood* *97*, 850–857.

Cavallero, S., Shen, H., Yi, C., Lien, C.L., Kumar, S.R., and Sucov, H.M. (2015). CXCL12 Signaling is Essential for Maturation of the Ventricular Coronary Endothelial Plexus and Establishment of Functional Coronary Circulation. *Dev. Cell* *33*, 469–477.

Cavanaugh, A.M., Huang, J., and Chen, J.N. (2015). Two developmentally distinct populations of neural crest cells contribute to the zebrafish heart. *Dev. Biol.* *404*, 103–112.

Chablais, F., and Jazwinska, a. (2012). The regenerative capacity of the zebrafish heart is dependent on TGF signaling. *Development* *139*, 1921–1930.

Chablais, F., Veit, J., Rainer, G., and Jaźwińska, A. (2011). The zebrafish heart regenerates

after cryoinjury-induced myocardial infarction. *BMC Dev. Biol.* *11*, 21.

Chakraborty, S., Cheek, J., Sakthivel, B., Aronow, B.J., and Yutzey, K.E. (2008). Shared gene expression profiles in developing heart valves and osteoblast progenitor cells. *Physiol. Genomics* *35*, 75–85.

Chakraborty, S., Combs, M.D., and Yutzey, K.E. (2010). Transcriptional regulation of heart valve progenitor cells. In *Pediatric Cardiology*, pp. 414–421.

Chehrehasa, F., Meedeniya, A.C.B., Dwyer, P., Abrahamsen, G., and Mackay-Sim, A. (2009). EdU, a new thymidine analogue for labelling proliferating cells in the nervous system. *J. Neurosci. Methods* *177*, 122–130.

Chen, Z., Huang, W., Dahme, T., Rottbauer, W., Ackerman, M.J., and Xu, X. (2008). Depletion of zebrafish essential and regulatory myosin light chains reduces cardiac function through distinct mechanisms. *Cardiovasc. Res.* *79*, 97–108.

Chi, N.C., Shaw, R.M., Jungblut, B., Huisken, J., Ferrer, T., Arnaout, R., Scott, I., Beis, D., Xiao, T., Baier, H., et al. (2008). Genetic and physiologic dissection of the vertebrate cardiac conduction system. *PLoS Biol.* *6*, 1006–1019.

Choe, C.P., Collazo, A., Trinh, L.A., Pan, L., Moens, C.B., and Crump, J.G. (2013). Wnt-Dependent Epithelial Transitions Drive Pharyngeal Pouch Formation. *Dev. Cell* *24*, 296–309.

Choi, H.M.T., Beck, V.A., and Pierce, N.A. (2014a). Next-generation in situ hybridization chain reaction: Higher gain, lower cost, greater durability. *ACS Nano* *8*, 4284–4294.

Choi, T.Y., Ninov, N., Stainier, D.Y.R., and Shin, D. (2014b). Extensive conversion of hepatic biliary epithelial cells to hepatocytes after near total loss of hepatocytes in zebrafish. *Gastroenterology* *146*, 776–788.

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

Chong, J.J.H., Reinecke, H., Iwata, M., Torok-Storb, B., Stempien-Otero, A., and Murry, C.E. (2013). Progenitor Cells Identified by PDGFR-Alpha Expression in the Developing and Diseased Human Heart. *Stem Cells Dev.* *22*, 1932–1943.

Chong, S.W., Emelyanov, A., Gong, Z., and Korzh, V. (2001). Expression pattern of two zebrafish genes, *cxcr4a* and *cxcr4b*. *Mech. Dev.* *109*, 347–354.

Choudhry, P., and Trede, N.S. (2013). DiGeorge Syndrome Gene *tbx1* Functions through *wnt11r* to Regulate Heart Looping and Differentiation. *PLoS One* *8*.

Christoffels, V.M., Hoogaars, W.M.H., Tessari, A., Clout, D.E.W., Moorman, A.F.M., and

Campione, M. (2004). T-Box Transcription Factor Tbx2 Represses Differentiation and Formation of the Cardiac Chambers. *Dev. Dyn.* 229, 763–770.

Christoffels, V.M., Grieskamp, T., Norden, J., Mommersteeg, M.T.M., Rudat, C., and Kispert, A. (2009). Tbx18 and the fate of epicardial progenitors. *Nature* 458, E8–E9.

Combs, M.D., Braitsch, C.M., Lange, A.W., James, J.F., and Yutzey, K.E. (2011). NFATC1 promotes epicardium-derived cell invasion into myocardium. *Development* 138, 1747–1757.

Comes, N., Buie, L.K., and Borrás, T. (2011). Evidence for a role of angiopoietin-like 7 (ANGPTL7) in extracellular matrix formation of the human trabecular meshwork: Implications for glaucoma. *Genes to Cells* 16, 243–259.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823.

Costa, G., Mazan, A., Gandillet, A., Pearson, S., Lacaud, G., and Kouskoff, V. (2012). SOX7 regulates the expression of VE-cadherin in the haemogenic endothelium at the onset of haematopoietic development. *Development* 139, 1587–1598.

Cunningham, S.A., Arrate, M.P., Rodriguez, J.M., Bjercke, R.J., Vanderslice, P., Morris, A.P., and Brock, T.A. (2000). A novel protein with homology to the junctional adhesion molecule. Characterization of leukocyte interactions. *J. Biol. Chem.* 275, 34750–34756.

Cunningham, S.A., Rodriguez, J.M., Pia Arrate, M., Tran, T.M., and Brock, T.A. (2002). JAM2 interacts with  $\alpha 4\beta 1$ . Facilitation by JAM3. *J. Biol. Chem.* 277, 27589–27592.

Dackor, R.T., Fritz-Six, K., Dunworth, W.P., Gibbons, C.L., Smithies, O., and Caron, K.M. (2006). Hydrops Fetalis, Cardiovascular Defects, and Embryonic Lethality in Mice Lacking the Calcitonin Receptor-Like Receptor Gene. *Mol. Cell. Biol.* 26, 2511–2518.

Dai, X.M., Ryan, G.R., Hapel, A.J., Dominguez, M.G., Russell, R.G., Kapp, S., Sylvestre, V., and Stanley, E.R. (2002a). Targeted disruption of the mouse colony-stimulating factor 1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies, and reproductive defects. *Blood* 99, 111–120.

Dai, Y.S., Cserjesi, P., Markham, B.E., and Molkenin, J.D. (2002b). The transcription factors GATA4 and dHAND physically interact to synergistically activate cardiac gene expression through a p300-dependent mechanism. *J. Biol. Chem.* 277, 24390–24398.

Darby, I., Skalli, O., and Gabbiani, G. (1990). Alpha-smooth muscle actin is transiently expressed by myofibroblasts during experimental wound healing. *Lab. Invest.* 63, 21–29.

David, N.B., Sapède, D., Saint-Etienne, L., Thisse, C., Thisse, B., Dambly-Chaudière, C., Rosa, F.M., and Ghysen, A. (2002). Molecular basis of cell migration in the fish lateral line:

role of the chemokine receptor CXCR4 and of its ligand, SDF1. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 16297–16302.

Deák, F., Wagener, R., Kiss, I., and Paulsson, M. (1999). The matrilins: A novel family of oligomeric extracellular matrix proteins. *Matrix Biol.* *18*, 55–64.

Debruin, E.J., Hughes, M.R., Sina, C., Liu, A., Cait, J., Jian, Z., Lopez, M., Lo, B., Abraham, T., and McNagny, K.M. (2014). Podocalyxin regulates murine lung vascular permeability by altering endothelial cell adhesion. *PLoS One* *9*.

DeKoter, R.P. and Singh, H. (2000). Regulation of B Lymphocyte and Macrophage Development by Graded Expression of PU.1. *Science* (80-. ). *288*, 1439–1441.

DeLaughter, D.M., Bick, A.G., Wakimoto, H., McKean, D., Gorham, J.M., Kathiriya, I.S., Hinson, J.T., Homsy, J., Gray, J., Pu, W., et al. (2016). Single-Cell Resolution of Temporal Gene Expression during Heart Development. *Dev. Cell* *39*, 480–490.

Dempsey, W.P., Fraser, S.E., and Pantazis, P. (2012). PhOTO zebrafish: A transgenic resource for in vivo lineage tracing during development and regeneration. *PLoS One* *7*.

Denninger, A.R., Breglio, A., Maheras, K.J., Leduc, G., Cristiglio, V., Demé, B., Gow, A., and Kirschner, D.A. (2015). Claudin-11 Tight Junctions in Myelin Are a Barrier to Diffusion and Lack Strong Adhesive Properties. *Biophys. J.* *109*, 1387–1397.

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

Donà, E., Barry, J.D., Valentin, G., Quirin, C., Khmelinskii, A., Kunze, A., Durdu, S., Newton, L.R., Fernandez-Minan, A., Huber, W., et al. (2013). Directional tissue migration through a self-generated chemokine gradient. *Nature*.

Duband, J.L., Gimona, M., Scatena, M., Sartore, S., and Small, J. V. (1993). Calponin and SM 22 as differentiation markers of smooth muscle: Spatiotemporal distribution during avian embryonic development. *Differentiation* *55*, 1–11.

Dunworth, W.P., Fritz-Six, K.L., and Caron, K.M. (2008). Adrenomedullin stabilizes the lymphatic endothelial barrier in vitro and in vivo. *Peptides* *29*, 2243–2249.

Dutton, K.A., Pauliny, A., Lopes, S.S., Elworthy, S., Carney, T.J., Rauch, J., Geisler, R., Haffter, P., and Kelsh, R.N. (2001). Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. *Development* *128*, 4113–4125.

D'Uva G, Aharonov A, Lauriola M, Kain D, Yahalom-Ronen Y, Carvalho S, Weisinger K, Bassat E, Rajchman D, Yifa O, Lysenko M, Konfino T, Hegesh J, Brenner O, Neeman M, Yarden Y, Leor J, Sarig R, Harvey RP and Tzahor E. (2015). ERBB2 triggers mammalian heart regeneration by promoting cardiomyocyte dedifferentiation and proliferation. *Nat*

Cell Biol. 17(5):627-38.

Eisenberg, C.A., and Eisenberg, L.M. (1999). WNT11 promotes cardiac tissue formation of early mesoderm. *Dev. Dyn.* 216, 45–58.

Emelyanov, A., Gao, Y., Naqvi, N.I., and Parinov, S. (2006). Trans-kingdom transposition of the maize Dissociation element. *Genetics* 174, 1095–1104.

Emery, B., Agalliu, D., Cahoy, J.D., Watkins, T.A., Dugas, J.C., Mulinyawe, S.B., Ibrahim, A., Ligon, K.L., Rowitch, D.H., and Barres, B.A. (2009). Myelin Gene Regulatory Factor Is a Critical Transcriptional Regulator Required for CNS Myelination. *Cell* 138, 172–185.

Eto, T., Kitamura, K., and Kato, J. (1999). Biological and clinical roles of adrenomedullin in circulation control and cardiovascular diseases. *Clin. Exp. Pharmacol. Physiol.* 26, 371–380.

Fan, J., Salathia, N., Liu, R., Kaeser, G.E., Yung, Y.C., Herman, J.L., Kaper, F., Fan, J.-B., Zhang, K., Chun, J., et al. (2016). Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* 13, 241–244.

Field, H.A., Ober, E.A., Roeser, T., and Stainier, D.Y.R. (2003). Formation of the digestive system in zebrafish. I. Liver morphogenesis. *Dev. Biol.* 253, 279–290.

Flaherty, M.P., and Dawn, B. (2008). Noncanonical Wnt11 Signaling and Cardiomyogenic Differentiation. *Trends Cardiovasc. Med.* 18, 260–268.

Forouhar, A.S., Liebling M, Hickerson A, Nasiraei-Moghaddam A, Tsai HJ, Hove JR, Fraser SE, Dickinson ME, Gharib M. (2006). The Embryonic Vertebrate Heart Tube Is a Dynamic Suction Pump. *Science* (80- ). 312, 751–753.

Frischknecht, R., and Seidenbecher, C.I. (2012). Brevican: A key proteoglycan in the perisynaptic extracellular matrix of the brain. *Int. J. Biochem. Cell Biol.* 44, 1051–1054.

Fritz-Six, K.L., Dunworth, W.P., Li, M., and Caron, K.M. (2008). Adrenomedullin signaling is necessary for murine lymphatic vascular development. *J. Clin. Invest.* 118, 40–50.

Gansner, J.M., Mendelsohn, B.A., Hultman, K.A., Johnson, S.L., and Gitlin, J.D. (2007). Essential role of lysyl oxidases in notochord development. *Dev. Biol.* 307, 202–213.

Gao, B., Song, H., Bishop, K., Elliot, G., Garrett, L., English, M.A., Andre, P., Robinson, J., Sood, R., Minami, Y., et al. (2011). Wnt Signaling Gradients Establish Planar Cell Polarity by Inducing Vangl2 Phosphorylation through Ror2. *Dev. Cell* 20, 163–176.

Garriock, R.J., D'Agostino, S.L., Pilcher, K.C., and Krieg, P.A. (2005). Wnt11-R, a protein closely related to mammalian Wnt11, is required for heart morphogenesis in *Xenopus*. *Dev. Biol.* 279, 179–192.

Garside, V.C., Cullum, R., Alder, O., Lu, D.Y., Vander Werff, R., Bilenky, M., Zhao, Y., Jones, S.J.M., Marra, M.A., Underhill, T.M., et al. (2015). SOX9 modulates the expression of key transcription factors required for heart valve development. *Development* 142, 4340–4350.

Gemberling, M., Karra, R., Dickson, A.L., and Poss, K.D. (2015). Nrg1 is an injury-induced cardiomyocyte mitogen for the endogenous heart regeneration program in zebrafish. *Elife* 4.

Ghaye, A.P., Bergemann, D., Tarifeño-Saldivia, E., Flasse, L.C., Von Berg, V., Peers, B., Voz, M.L., and Manfredi, I. (2015). Progenitor potential of nkx6.1-expressing cells throughout zebrafish life and during beta cell regeneration. *BMC Biol.* 13, 70.

von Gise, A., Zhou, B., Honor, L.B., Ma, Q., Petryk, A., and Pu, W.T. (2011). WT1 regulates epicardial epithelial to mesenchymal transition through  $\beta$ -catenin and retinoic acid signaling pathways. *Dev. Biol.* 356, 421–431.

Gonzalez-Quesada, C., Cavalera, M., Biernacka, A., Kong, P., Lee, D.W., Saxena, A., Frunza, O., Dobaczewski, M., Shinde, A., and Frangogiannis, N.G. (2013). Thrombospondin-1 induction in the diabetic myocardium stabilizes the cardiac matrix in addition to promoting vascular rarefaction through angiotensin-2 upregulation. *Circ. Res.* 113, 1331–1344.

Gonzalez-Rosa, J.M., Martin, V., Peralta, M., Torres, M., and Mercader, N. (2011). Extensive scar formation and regression during heart regeneration after cryoinjury in zebrafish. *Development* 138, 1663–1674.

González-Rosa, J.M., Peralta, M., and Mercader, N. (2012). Pan-epicardial lineage tracing reveals that epicardium derived cells give rise to myofibroblasts and perivascular cells during zebrafish heart regeneration. *Dev. Biol.* 370, 173–186.

Gourdie, R.G., Mima, T., Thompson, R.P., and Mikawa, T. (1995). Terminal diversification of the myocyte lineage generates Purkinje fibers of the cardiac conduction system. *Development* 121, 1423–1431.

Gow, A., Southwood, C.M., Li, J.S., Pariali, M., Riordan, G.P., Brodie, S.E., Danias, J., Bronstein, J.M., Kachar, B., and Lazzarini, R.A. (1999). CNS Myelin and Sertoli Cell Tight Junction Strands Are Absent in *Osp*/Claudin-11 Null Mice. *Cell* 99, 649–659.

Grün, D., and Van Oudenaarden, A. (2015). Design and Analysis of Single-Cell Sequencing Experiments. *Cell* 163, 799–810.

Grépin, C., Robitaille, L., Antakly, T., and Nemer, M. (1995). Inhibition of transcription factor GATA-4 expression blocks in vitro cardiac muscle differentiation. *Mol. Cell. Biol.* 15, 4095–4102.

Greulich, F., Farin, H.F., Schuster-Gossler, K., and Kispert, A. (2012). Tbx18 function in

epicardial development. *Cardiovasc. Res.* *96*, 476–483.

Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat. Methods* *11*, 637–640.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* *525*, 251–255.

Guadix, J.A., Ruiz-Villalba, A., Lettice, L., Velecela, V., Muñoz-Chápuli, R., Hastie, N.D., Pérez-Pomares, J.M., and Martínez-Estrada, O.M. (2011). *Wt1* controls retinoic acid signalling in embryonic epicardium through transcriptional activation of *Raldh2*. *Development* *138*, 1093–1097.

Gulati-Leekha, A., and Goldman, D. (2006). A reporter-assisted mutagenesis screen using  $\alpha$ 1-tubulin-GFP transgenic zebrafish uncovers missteps during neuronal development and axonogenesis. *Dev. Biol.* *296*, 29–47.

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

Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-Cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* *23*, 2126–2135.

Gupta, V., Gemberling, M., Karra, R., Rosenfeld, G.E., Evans, T., and Poss, K.D. (2013). An injury-responsive *gata4* program shapes the zebrafish cardiac ventricle. *Curr. Biol.* *23*, 1221–1227.

Gurskaya, N.G., Verkhusha, V. V., Shcheglov, A.S., Staroverov, D.B., Chepurnykh, T. V., Fradkov, A.F., Lukyanov, S., and Lukyanov, K.A. (2006). Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nat. Biotechnol.* *24*, 461–465.

Habas, R., Kato, Y., and He, X. (2001). Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel formin homology protein Daam1. *Cell* *107*, 843–854.

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

Han, Y., and Lefebvre, V. (2008). L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer. *Mol. Cell. Biol.* *28*, 4999–5013.

Harrelson, Z., Kelly RG, Goldin SN, Gibson-Brown JJ, Bollag RJ, Silver LM, Papaioannou VE.

(2004). *Tbx2* is essential for patterning the atrioventricular canal and for morphogenesis of the outflow tract during heart development. *Development* *131*, 5041–5052.

Harrison, M.R.M., Bussmann, J., Huang, Y., Zhao, L., Osorio, A., Burns, C.G., Burns, C.E., Sucov, H.M., Siekmann, A.F., and Lien, C.L. (2015). Chemokine-Guided Angiogenesis Directs Coronary Vasculature Formation in Zebrafish. *Dev. Cell* *33*, 442–454.

Haworth, K.E., Kotecha, S., Mohun, T.J., and Latinkic, B. V (2008). GATA4 and GATA5 are essential for heart and liver development in *Xenopus* embryos. *BMC Dev. Biol.* *8*, 74.

He, Z., Li, H., Zuo, S., Pasha, Z., Wang, Y., Yang, Y., Jiang, W., Ashraf, M., and Xu, M. (2011). Transduction of *Wnt11* Promotes Mesenchymal Stem Cell Transdifferentiation into Cardiac Phenotypes. *Stem Cells Dev.* *20*, 1771–1778.

Heim, R., Prasher, D.C., and Tsien, R.Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci.* *91*, 12501–12504.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* *16*, 144–154.

Hellqvist, M., Mahlapuu, M., Samuelsson, L., Enerbäck, S., and Carlsson, P. (1996). Differential activation of lung-specific genes by two forkhead proteins, FREAC-1 and FREAC-2. *J. Biol. Chem.* *271*, 4482–4490.

Herbomel, P., Thisse, B., and Thisse, C. (1999). Ontogeny and behaviour of early macrophages in the zebrafish embryo. *Development* *126*, 3735–3745.

Hidai, H., Bardales, R., Goodwin, R., Quertermous, T., and Quertermous, E.E. (1998). Cloning of capsulin, a basic helix-loop-helix factor expressed in progenitor cells of the pericardium and the coronary arteries. *Mech. Dev.* *73*, 33–43.

Hirakow, R. (1992). Epicardial formation in staged human embryos. *Kaibogaku Zasshi* *67*, 616–622.

Hodgkinson, C.P., Naidoo, V., Patti, K.G., Gomez, J.A., Schmeckpeper, J., Zhang, Z., Davis, B., Pratt, R.E., Mirotsov, M., and Dzau, V.J. (2013). *Abi3bp* is a multifunctional autocrine/paracrine factor that regulates mesenchymal stem cell biology. *Stem Cells* *31*, 1669–1682.

Hodgkinson, C.P., Gomez, J.A., Payne, A.J., Zhang, L., Wang, X., Dal-Pra, S., Pratt, R.E., and Dzau, V.J. (2014). *Abi3bp* regulates cardiac progenitor cell proliferation and differentiation. *Circ. Res.* *115*, 1007–1016.

Hu, N., Sedmera, D., Yost, H.J., and Clark, E.B. (2000). Structure and function of the developing zebrafish heart. *Anat. Rec.* *260*, 148–157.

Hu, N., Joseph Yost, H., and Clark, E.B. (2001). Cardiac morphology and blood pressure in the adult zebrafish. *Anat. Rec.* *264*, 1–12.

Huang, K.Y., Chen, G. Der, Cheng, C.H., Liao, K.Y., Hung, C.C., Chang, G.D., Hwang, P.P., Lin, S.Y., Tsai, M.C., Khoo, K.H., et al. (2011). Phosphorylation of the zebrafish M6Ab at serine 263 contributes to filopodium formation in pc12 cells and neurite outgrowth in zebrafish embryos. *PLoS One* *6*.

Huang, Y., Harrison, M.R., Osorio, A., Kim, J., Baugh, A., Duan, C., Sucov, H.M., and Lien, C.L. (2013). Igf Signaling is Required for Cardiomyocyte Proliferation during Zebrafish Heart Development and Regeneration. *PLoS One* *8*.

Huang, Y., Wang, D., Wang, X., Zhang, Y., Liu, T., Chen, Y., Tang, Y., Wang, T., Hu, D., and Huang, C. (2016). Abrogation of CC chemokine receptor 9 ameliorates ventricular remodeling in mice after myocardial infarction. *Sci. Rep.* *6*, 32660.

Ichikawa-Shindo, Y., Sakurai, T., Kamiyoshi, A., Kawate, H., Iinuma, N., Yoshizawa, T., Koyama, T., Fukuchi, J., Iimuro, S., Moriyama, N., et al. (2008). The GPCR modulator protein RAMP2 is essential for angiogenesis and vascular integrity. *J. Clin. Invest.* *118*, 29–39.

Iriyama, C., Matsuda, S., Katsumata, R., and Hamaguchi, M. (2001). Cloning and sequencing of a novel human gene which encodes a putative hydroxylase. *J. Hum. Genet.* *46*, 289–292.

Itou, J., Kawakami, H., Burgoyne, T., and Kawakami, Y. (2012a). Life-long preservation of the regenerative capacity in the fin and heart in zebrafish. *Biol. Open* *1*, 739–746.

Itou, J., Oishi, I., Kawakami, H., Glass, T.J., Richter, J., Johnson, A., Lund, T.C., and Kawakami, Y. (2012b). Migration of cardiomyocytes is essential for heart regeneration in zebrafish. *Development* *139*, 4133–4142.

Jensen, B., Boukens, B.J.D., Postma, A. V., Gunst, Q.D., van den Hoff, M.J.B., Moorman, A.F.M., Wang, T., and Christoffels, V.M. (2012). Identifying the Evolutionary Building Blocks of the Cardiac Conduction System. *PLoS One* *7*.

Jiang, D.-S., Wei, X., Zhang, X.-F., Liu, Y., Zhang, Y., Chen, K., Gao, L., Zhou, H., Zhu, X.-H., Liu, P.P., et al. (2014). IRF8 suppresses pathological cardiac remodelling by inhibiting calcineurin signalling. *Nat. Commun.* *5*, 3303.

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* *21*, 1543–1551.

Jiménez-Amilburu, V., Rasouli, S.J., Staudt, D.W., Nakajima, H., Chiba, A., Mochizuki, N., and Stainier, D.Y.R. (2016). In Vivo Visualization of Cardiomyocyte Apicobasal Polarity Reveals Epithelial to Mesenchymal-like Transition during Cardiac Trabeculation. *Cell Rep.*

17, 2687–2699.

Jin, S.-W., Beis D, Mitchell T, Chen JN, Stainier DY. (2005). Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. *Development* 132, 5199–5209.

Jin, D., Harada, K., Ohnishi, S., Yamahara, K., Kangawa, K., and Nagaya, N. (2008). Adrenomedullin induces lymphangiogenesis and ameliorates secondary lymphoedema. *Cardiovasc. Res.* 80, 339–345.

Johnson, P.R.A., Black, J.L., Carlin, S., Ge, Q., and Underwood, P.A. (2000). The production of extracellular matrix proteins by human passively sensitized airway smooth-muscle cells in culture: The effect of beclomethasone. *Am. J. Respir. Crit. Care Med.* 162, 2145–2151.

Jopling, C., Sleep, E., Raya, M., Martí, M., Raya, A., and Belmonte, J.C.I. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature* 464, 606–609.

Just, S., Raphel, L., Berger, I.M., Bühler, A., Keßler, M., and Rottbauer, W. (2016). Tbx20 is an essential regulator of embryonic heart growth in zebrafish. *PLoS One* 11.

Kamm, K.E., and Stull, J.T. (2001). Dedicated Myosin Light Chain Kinases with Diverse Cellular Functions. *J. Biol. Chem.* 276, 4527–4530.

Kanamori-Katayama, M., Kaiho, A., Ishizu, Y., Okamura-Oho, Y., Hino, O., Abe, M., Kishimoto, T., Sekihara, H., Nakamura, Y., Suzuki, H., et al. (2011). LRRN4 and UPK3B are markers of primary mesothelial cells. *PLoS One* 6.

Katz, T.C., Singh, M.K., Degenhardt, K., Rivera-Feliciano, J., Johnson, R.L., Epstein, J.A., and Tabin, C.J. (2012). Distinct Compartments of the Proepicardial Organ Give Rise to Coronary Vascular Endothelial Cells. *Dev. Cell* 22, 639–650.

Keegan, B.R., Feldman JL, Begemann G, Ingham PW, Yelon D. (2005). Retinoic Acid Signaling Restricts the Cardiac Progenitor Pool. *Science* (80-. ). 307, 247–249.

Kerjaschki, D., Sharkey, D.J., and Farquhar, M.G. (1984). Identification and characterization of podocalyxin--the major sialoprotein of the renal glomerular epithelial cell. *J. Cell Biol.* 98, 1591–1596.

Kershaw, D.B., Beck, S.G., Wharram, B.L., Wiggins, J.E., Goyal, M., Thomas, P.E., and Wiggins, R.C. (1997). Molecular cloning and characterization of human podocalyxin-like protein. *J. Biol. Chem.* 272, 15708–15714.

Kershaw, D. B., Beck, S.G., Wharram, B.L.Wiggins, J.E., Goyal, M., Thomas, P.E., Wiggins, R.C. (1997). Molecular Cloning and Characterization of Human Podocalyxin-like Protein. Orthologous relationship to rabbit PCLP1 and rat podocalyxin. *J. Biol. Chem.* 272, 15708–15714.

Kharchenko, P. V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* 11, 740–742.

Kikuchi, K. (2014). Advances in understanding the mechanism of zebrafish heart regeneration. *Stem Cell Res.* 13, 542–555.

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., and Poss, K.D. (2010). Primary contribution to zebrafish heart regeneration by *gata4+* cardiomyocytes. *Nature* 464, 601–605.

Kikuchi, K., Gupta, V., Wang, J., Holdway, J.E., Wills, A.A., Fang, Y., and Poss, K.D. (2011a). *tcf21+* epicardial cells adopt non-myocardial fates during zebrafish heart development and regeneration. *Development* 138, 2895–2902.

Kikuchi, K., Holdway, J.E., Major, R.J., Blum, N., Dahn, R.D., Begemann, G., and Poss, K.D. (2011b). Retinoic Acid Production by Endocardium and Epicardium Is an Injury Response Essential for Zebrafish Heart Regeneration. *Dev. Cell* 20, 397–404.

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., et al. (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* 5, 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., et al. (2010). PDGF signaling is required for epicardial function and blood vessel formation in regenerating zebrafish hearts. *Proc. Natl. Acad. Sci.* 107, 17206–17210.

Kim, J.-D., Kim, H.-J., Koun, S., Ham, H.-J., Kim, M.-J., Rhee, M., and Huh, T.-L. (2014). Zebrafish *Crip2* plays a critical role in atrioventricular valve development by downregulating the expression of ECM genes in the endocardial cushion. *Mol. Cells* 37, 406–411.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. *Dev. Dyn. an Off. Public* 203, 253–310.

Kirschner, K.M., Wagner, N., Wagner, K.D., Wellmann, S., and Scholz, H. (2006). The Wilms tumor suppressor *Wt1* promotes cell adhesion through transcriptional activation of the  $\alpha 4$  integrin gene. *J. Biol. Chem.* 281, 31930–31939.

Klein, R.S., Rubin, J.B., Gibson, H.D., DeHaan, E.N., Alvarez-Hernandez, X., Segal, R. a, and Luster, a D. (2001). SDF-1 alpha induces chemotaxis and enhances Sonic hedgehog-induced proliferation of cerebellar granule cells. *Development* 128, 1971–1981.

Klemsz, M.J., McKercher, S.R., Celada, A., Van Beveren, C., and Maki, R.A. (1990). The macrophage and B cell-specific transcription factor PU.1 is related to the *ets* oncogene. *Cell* 61, 113–124.

Knaut, H., Werz, C., Geisler, R., The Tübingen 2000 Screen Consortium, and Nüsslein-Volhard, C. (2003). A zebrafish homologue of the chemokine receptor Cxcr4 is a germ-cell guidance receptor. *Nature* *421*, 279–282.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The Technology and Biology of Single-Cell RNA Sequencing. *Mol. Cell* *58*, 610–620.

Komiyama, M., Ito, K., and Shimada, Y. (1987). Origin and development of the epicardium in the mouse embryo. *Anat. Embryol. (Berl.)* *176*, 183–189.

Kraus, F., Haenig, B., and Kispert, A. (2001). Cloning and expression analysis of the mouse T-box gene *Tbx18*. *Mech. Dev.* *100*, 83–86.

Kroeger, P.T., and Wingert, R.A. (2014). Using zebrafish to study podocyte genesis during kidney development and regeneration. *Genesis* *52*, 771–792.

Kroehne, V., Freudenreich, D., Hans, S., Kaslin, J., and Brand, M. (2011). Regeneration of the adult zebrafish brain from neurogenic radial glia-type progenitors. *Development* *138*, 4831–4841.

Krysinska, H., Hoogenkamp, M., Ingram, R., Wilson, N., Tagoh, H., Laslo, P., Singh, H., and Bonifer, C. (2007). A two-step, PU.1-dependent mechanism for developmentally regulated chromatin remodeling and transcription of the *c-fms* gene. *Mol. Cell. Biol.* *27*, 878–887.

Kumawat K, Koopmans Tim, Menzen MH, Prins A, Smit M, Halayko AJ, and Gosens R (2016). Cooperative signaling by TGF-1 and WNT-11 drives sm-actin expression in smooth muscle via Rho kinase-actin-MRTF-A signaling. *Am J Physiol Lung Cell Mol Physiol* *311*, L529-37.

Kuo, C.T., Morrissey, E.E., Anandappa, R., Sigrist, K., Lu, M.M., Parmacek, M.S., Soudais, C., and Leiden, J.M. (1997). GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev.* *11*, 1048–1060.

Laforest, B., and Nemer, M. (2011). GATA5 interacts with GATA4 and GATA6 in outflow tract development. *Dev. Biol.* *358*, 368–378.

Lai, S.L., Marín-Juez, R., Moura, P.L., Kuenne, C., Lai, J.K.H., Tsedeke, A.T., Guenther, S., Looso, M., and Stainier, D.Y.R. (2017). Reciprocal analyses in zebrafish and medaka reveal that harnessing the immune response promotes cardiac regeneration. *Elife* *6*.

Lambers, E., Arnone, B., Fatima, A., Qin, G., Andrew Wasserstrom, J., and Kume, T. (2016). *Foxc1* Regulates Early Cardiomyogenesis and Functional Properties of Embryonic Stem Cell Derived Cardiomyocytes. *Stem Cells* *34*, 1487–1500.

Lander, R., Nasr, T., Ochoa, S.D., Nordin, K., Prasad, M.S., and Labonne, C. (2013).

Interactions between Twist and other core epithelial-mesenchymal transition factors are controlled by GSK3-mediated phosphorylation. *Nat. Commun.* **4**, 1542.

Lange, A.W., and Yutzey, K.E. (2006). NFATc1 expression in the developing heart valves is responsive to the RANKL pathway and is required for endocardial expression of cathepsin K. *Dev. Biol.* **292**, 407–417.

De Lange, F.J., Moorman, A.F.M., Anderson, R.H., Männer, J., Soufan, A.T., De Gier-De Vries, C., Schneider, M.D., Webb, S., Van Den Hoff, M.J.B., and Christoffels, V.M. (2004). Lineage and morphogenetic analysis of the cardiac valves. *Circ. Res.* **95**, 645–654.

Lataillade, J.J., Clay, D., Bourin, P., Hérodin, F., Dupuy, C., Jasmin, C., and Le Bousse-Kerdilès, M.C. (2002). Stromal cell-derived factor 1 regulates primitive hematopoiesis by suppressing apoptosis and by promoting G0/G1 transition in CD34+ cells: Evidence for an autocrine/paracrine mechanism. *Blood* **99**, 1117–1129.

Laue, K., Jänicke, M., Plaster, N., Sonntag, C., and Hammerschmidt, M. (2008). Restriction of retinoic acid activity by Cyp26b1 is required for proper timing and patterning of osteogenesis during zebrafish development. *Development* **135**, 3775–3787.

Laugwitz, K.-L., Moretti, A., Lam, J., Gruber, P., Chen, Y., Woodard, S., Lin, L.-Z., Cai, C.-L., Lu, M.M., Reth, M., et al. (2005). Postnatal isl1+ cardioblasts enter fully differentiated cardiomyocyte lineages. *Nature* **433**, 647–653.

Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326.

Lawson, N.D., and Wolfe, S.A. (2011). Forward and Reverse Genetic Approaches for the Analysis of Vertebrate Development in the Zebrafish. *Dev. Cell* **21**, 48–64.

Lee, M.P., and Yutzey, K.E. (2011). Twist1 directly regulates genes that promote cell proliferation and migration in developing heart valves. *PLoS One* **6**.

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

Lepore, J.J., Mericko, P.A., Cheng, L., Lu, M.M., Morrisey, E.E., and Parmacek, M.S. (2006). GATA-6 regulates semaphorin 3C and is required in cardiac neural crest for cardiovascular morphogenesis. *J. Clin. Invest.* **116**, 929–939.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.

Li, G., Xu, A., Sim, S., Priest, J.R., Tian, X., Khan, T., Quertermous, T., Zhou, B., Tsao, P.S., Quake, S.R., et al. (2016). Transcriptomic Profiling Maps Anatomically Patterned

Subpopulations among Single Embryonic Cardiac Cells. *Dev. Cell* 39, 491–507.

Li, J., Yue, Y., Dong, X., Jia, W., Li, K., Liang, D., Dong, Z., Wang, X., Nan, X., Zhang, Q., et al. (2015). Zebrafish *foxc1a* plays a crucial role in early somitogenesis by restricting the expression of *aldh1a2* directly. *J. Biol. Chem.* 290, 10216–10228.

Li, L., Jin, H., Xu, J., Shi, Y., and Wen, Z. (2011a). *Irf8* regulates macrophage versus neutrophil fate during zebrafish primitive myelopoiesis. *Blood* 117, 1359–1369.

Li, P., Pashmforoush, M., and Sucov, H.M. (2010). Retinoic acid regulates differentiation of the secondary heart field and TGFbeta-mediated outflow tract septation. *Dev. Cell* 18, 480–485.

Li, P., Cavallero, S., Gu, Y., Chen, T.H.P., Hughes, J., Hassan, a B., Brüning, J.C., Pashmforoush, M., and Sucov, H.M. (2011b). IGF signaling directs ventricular cardiomyocyte proliferation during embryonic heart development. *Development* 138, 1795–1805.

Li, Y.X., Zdanowicz, M., Young, L., Kumiski, D., Leatherbury, L., and Kirby, M.L. (2003). Cardiac neural crest in zebrafish embryos contributes to myocardial cell lineage and early heart function. *Dev. Dyn.* 226, 540–550.

Li, Z., Korzh, V., and Gong, Z. (2007). Localized *rbp4* expression in the yolk syncytial layer plays a role in yolk cell extension and early liver development. *BMC Dev. Biol.* 7, 117.

Liao, W., Bisgrove, B.W., Sawyer, H., Hug, B., Bell, B., Peters, K., Grunwald, D.J., and Stainier, D.Y. (1997). The zebrafish gene *cloche* acts upstream of a *flk-1* homologue to regulate endothelial cell differentiation. *Development* 124, 381–389.

Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.

Liebling, M., Forouhar, A.S., Gharib, M., Fraser, S.E., and Dickinson, M.E. (2005). Four-dimensional cardiac imaging in living embryos via postacquisition synchronization of nongated slice sequences. *J. Biomed. Opt.* 10, 54001.

Liebling, M., Forouhar, A.S., Wolleschensky, R., Zimmermann, B., Ankerhold, R., Fraser, S.E., Gharib, M., and Dickinson, M.E. (2006). Rapid three-dimensional imaging and analysis of the beating embryonic heart reveals functional changes during development. *Dev. Dyn.* 235, 2940–2948.

Lien, C.L., Schebesta, M., Makino, S., Weber, G.J., and Keating, M.T. (2006). Gene expression analysis of zebrafish heart regeneration. *PLoS Biol.* 4, 1386–1396.

Lieschke, G.J., Oates, A.C., Crowhurst, M.O., Ward, A.C., and Layton, J.E. (2001). Morphologic and functional characterization of granulocytes and macrophages in embryonic and adult zebrafish. *Blood* 98, 3087–3096.

- Lin, C.W., Sun, M.S., Liao, M.Y., Chung, C.H., Chi, Y.H., Chiou, L.T., Yu, J., Lou, K.L., and Wu, H.C. (2014). Podocalyxin-like 1 promotes invadopodia formation and metastasis through activation of Rac1/Cdc42/cortactin signaling in breast cancer cells. *Carcinogenesis* *35*, 2425–2435.
- Lincoln, J., Alfieri, C.M., and Yutzey, K.E. (2004). Development of heart valve leaflets and supporting apparatus in chicken and mouse embryos. *Dev. Dyn.* *230*, 239–250.
- Lincoln, J., Kist, R., Scherer, G., and Yutzey, K.E. (2007). Sox9 is required for precursor cell expansion and extracellular matrix organization during mouse heart valve development. *Dev. Biol.* *305*, 120–132.
- Liu, J., and Stainier, D.Y.R. (2010). Tbx5 and Bmp signaling are essential for proepicardium specification in zebrafish. *Circ. Res.* *106*, 1818–1828.
- Liu, S., and Trapnell, C. (2016). Single-cell transcriptome sequencing: recent advances and remaining challenges. *F1000Research*.
- Liu, F., Walmsley, M., Rodaway, A., and Patient, R. (2008). Fli1 Acts at the Top of the Transcriptional Network Driving Blood and Endothelial Development. *Curr. Biol.* *18*, 1234–1240.
- Liu, J., Bressan, M., Hassel, D., Huisken, J., Staudt, D., Kikuchi, K., Poss, K.D., Mikawa, T., and Stainier, D.Y.R. (2010). A dual role for ErbB2 signaling in cardiac trabeculation. *Development* *137*, 3867–3875.
- Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* *167*, 1170–1187.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Lugus, J.J., Park, C., Ma, Y.D., and Choi, K. (2009). Both primitive and definitive blood cells are derived from Flk-1 + mesoderm. *Blood* *113*, 563–566.
- Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* *5*, 2122.
- Lyons, M.S., Bell, B., Stainier, D., and Peters, K.G. (1998). Isolation of the zebrafish homologues for the tie-1 and tie-2 endothelium-specific receptor tyrosine kinases. *Dev. Dyn.* *212*, 133–140.
- van der Maaten, L. (2014). Accelerating t-SNE using Tree-Based Algorithms. *J. Mach. Learn. Res.* *15*, 3221–3245.
- Mably, J.D., Mohideen, M.A.P.K., Burns, C.G., Chen, J.N., and Fishman, M.C. (2003). Heart

of glass Regulates the Concentric Growth of the Heart in Zebrafish. *Curr. Biol.* *13*, 2138–2147.

Maclean, G., Dollé, P., and Petkovich, M. (2009). Genetic disruption of CYP26B1 severely affects development of neural crest derived head structures, but does not compromise hindbrain patterning. *Dev. Dyn.* *238*, 732–745.

MacNicol, A.M., Hardy, L.L., Spencer, H.J., and MacNicol, M.C. (2015). Neural stem and progenitor cell fate transition requires regulation of Musashi1 function. *BMC Dev. Biol.* *15*, 15.

Mahmoud, A.I., O’Meara, C.C., Gemberling, M., Zhao, L., Bryant, D.M., Zheng, R., Gannon, J.B., Cai, L., Choi, W.Y., Egnaczyk, G.F., et al. (2015). Nerves Regulate Cardiomyocyte Proliferation and Heart Regeneration. *Dev. Cell* *34*, 387–399.

Mall, M., Kareta, M.S., Chanda, S., Ahlenius, H., Perotti, N., Zhou, B., Grieder, S.D., Ge, X., Drake, S., Euong Ang, C., et al. (2017). Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature* *544*, 245–249.

Mann, K.D., Hoyt, C., Feldman, S., Blunt, L., Raymond, A., and Page-McCaw, P.S. (2010). Cardiac response to startle stimuli in larval zebrafish: sympathetic and parasympathetic components. *Am. J. Physiol. Regul. Integrative Comp. Physiol.* *298*, R1288–R1297.

Marín-Juez, R., Marass, M., Gauthier, S., Rossi, A., Lai, S.-L., Materna, S.C., Black, B.L., and Stainier, D.Y.R. (2016). Fast revascularization of the injured area is essential to support zebrafish heart regeneration. *Proc. Natl. Acad. Sci.* *113*, 11237–11242.

Martínez-Estrada, O.M., Lettice, L.A., Essafi, A., Guadix, J.A., Slight, J., Velecela, V., Hall, E., Reichmann, J., Devenney, P.S., Hohenstein, P., et al. (2010). Wt1 is required for cardiovascular progenitor cell formation through transcriptional control of Snail and E-cadherin. *Nat. Genet.* *42*, 89–93.

Masters, M., and Riley, P.R. (2014). The epicardium signals the way towards heart regeneration. *Stem Cell Res.* *13*, 683–692.

Matrone, G., Taylor, J.M., Wilson, K.S., Baily, J., Love, G.D., Girkin, J.M., Mullins, J.J., Tucker, C.S., and Denvir, M.A. (2013). Laser-targeted ablation of the zebrafish embryonic ventricle: A novel model of cardiac injury and repair. *Int. J. Cardiol.* *168*, 3913–3919.

McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2016). scater: pre-processing, quality control, normalisation and visualisation of single-cell RNA-seq data in R. *bioRxiv* 69633.

McLatchie, L.M., Fraser, N.J., Main, M.J., Wise, A., Brown, J., Thompson, N., Solari, R., Lee, M.G., and Foord, S.M. (1998). RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor. *Nature* *393*, 333–339.

Meng, X., Ezzati, P., and Wilkins, J.A. (2011). Requirement of podocalyxin in TGF-beta induced epithelial mesenchymal transition. *PLoS One* 6.

Mercer, S.E., Odelberg, S.J., and Simon, H.G. (2013). A dynamic spatiotemporal extracellular matrix facilitates epicardial-mediated vertebrate heart regeneration. *Dev. Biol.* 382, 457–469.

Merki, E., Zamora, M., Raya, A., Kawakami, Y., Wang, J., Zhang, X., Burch, J., Kubalak, S.W., Kaliman, P., Izpisua Belmonte, J.C., et al. (2005). Epicardial retinoid X receptor alpha is required for myocardial growth and coronary artery formation. *Proc. Natl. Acad. Sci. U. S. A.* 102, 18455–18460.

Miao, M., Bruce, A.E.E., Bhanji, T., Davis, E.C., and Keeley, F.W. (2007). Differential expression of two tropoelastin genes in zebrafish. *Matrix Biol.* 26, 115–124.

Mickoleit, M., Schmid, B., Weber, M., Fahrbach, F.O., Hombach, S., Reischauer, S., and Huisken, J. (2014). High-resolution reconstruction of the beating zebrafish heart. *Nat. Methods* 11, 919–922.

Milan, D.J., Giokas, A.C., Serluca, F.C., Peterson, R.T., and MacRae, C.A. (2006). Notch1b and neuregulin are required for specification of central cardiac conduction tissue. *Development* 133, 1125–1132.

Milgrom-Hoffman, M., Harrelson, Z., Ferrara, N., Zelzer, E., Evans, S.M., and Tzahor, E. (2011). The heart endocardium is derived from vascular endothelial progenitors. *Development* 138, 4777–4787.

Molnar, J., Fong, K.S.K., He, Q.P., Hayashi, K., Kim, Y., Fong, S.F.T., Fogelgren, B., Szauter, K.M., Mink, M., and Csiszar, K. (2003). Structural and functional diversity of lysyl oxidase and the LOX-like proteins. In *Biochimica et Biophysica Acta - Proteins and Proteomics*, pp. 220–224.

Del Monte, G., Casanova, J.C., Guadix, J.A., MacGrogan, D., Burch, J.B.E., Pérez-Pomares, J.M., and De La Pompa, J.L. (2011). Differential notch signaling in the epicardium is required for cardiac inflow development and coronary vessel morphogenesis. *Circ. Res.* 108, 824–836.

Moore, A.W., McInnes, L., Kreidberg, J., Hastie, N.D., and Schedl, A. (1999). YAC complementation shows a requirement for Wt1 in the development of epicardium, adrenal gland and throughout nephrogenesis. *Development* 126, 1845–1857.

Moore-Morris, T., Guimarães-Camboa, N., Banerjee, I., Zambon, A.C., Kisseleva, T., Velayoudon, A., Stallcup, W.B., Gu, Y., Dalton, N.D., Cedenilla, M., et al. (2014). Resident fibroblast lineages mediate pressure overload-induced cardiac fibrosis. *J. Clin. Invest.* 124, 2921–2934.

Morita, K., Sasaki, H., Fujimoto, K., Furuse, M., and Tsukita, S. (1999). Claudin-11/OSP-

based tight junctions of myelin sheaths in brain and Sertoli cells in testis. *J. Cell Biol.* **145**, 579–588.

Moriyama, Y., Ito, F., Takeda, H., Yano, T., Okabe, M., Kuraku, S., Keeley, F.W., and Koshiba-Takeuchi, K. (2016). Evolution of the fish heart by sub/neofunctionalization of an elastin gene. *Nat. Commun.* **7**, 10397.

Münzel, E.J., Schaefer, K., Obirei, B., Kremmer, E., Burton, E.A., Kuscha, V., Becker, C.G., Brösamle, C., Williams, A., and Becker, T. (2012). Claudin k is specifically expressed in cells that form myelin during development of the nervous system and regeneration of the optic nerve in adult zebrafish. *Glia* **60**, 253–270.

Nakamoto, N., Ebinuma, H., Kanai, T., Chu, P., Ono, Y., Mikami, Y., Ojiro, K., Lipp, M., Love, P.E., Saito, H., et al. (2012). CCR9 + macrophages are required for acute liver inflammation in mouse models of hepatitis. *Gastroenterology* **142**, 366–376.

Naumann, U., Cameroni, E., Pruenster, M., Mahabaleshwar, H., Raz, E., Zerwes, H.G., Rot, A., and Thelen, M. (2010). CXCR7 functions as a scavenger for CXCL12 and CXCL11. *PLoS One* **5**.

Naylor, A.J., McGettrick, H.M., Maynard, W.D., May, P., Barone, F., Croft, A.P., Egginton, S., and Buckley, C.D. (2014). A differential role for CD248 (Endosialin) in PDGF-mediated skeletal muscle angiogenesis. *PLoS One* **9**.

Nichols, J.T., Pan, L., Moens, C.B., and Kimmel, C.B. (2013). *barx1* represses joints and promotes cartilage in the craniofacial skeleton. *Development* **140**, 2765–2775.

Nicoli, S., Tobia, C., Gualandi, L., De Sena, G., and Presta, M. (2008). Calcitonin receptor-like receptor guides arterial differentiation in zebrafish. *Blood* **111**, 4965–4972.

Niederreither, K., Subbarayan, V., Dolle, P., and Chambon, P. (1999). Embryonic retinoic acid synthesis is essential for heart morphogenesis in the mouse. *Nat. Genet.* **21**, 444–448.

Nosedá, M., Harada, M., McSweeney, S., Leja, T., Belian, E., Stuckey, D.J., Abreu Paiva, M.S., Habib, J., Macaulay, I., de Smith, A.J., et al. (2015). PDGFR $\alpha$  demarcates the cardiogenic clonogenic Sca1<sup>+</sup> stem/progenitor cell in adult murine myocardium. *Nat. Commun.* **6**, 6930.

Nunomura, W., Takakuwa, Y., Cherr, G.N., and Murata, K. (2007). Characterization of protein 4.1R in erythrocytes of zebrafish (*Danio rerio*): Unique binding properties with transmembrane proteins and calmodulin. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **148**, 124–138.

Okuda, Y., Ogura, E., Kondoh, H., and Kamachi, Y. (2010). B1 SOX coordinate cell specification with patterning and morphogenesis in the early zebrafish embryo. *PLoS Genet.* **6**, 36.

- Olson, E.N. (2006). Gene regulatory networks in the evolution and development of the heart. *Science* 313, 1922–1927.
- Omatsu, Y., Seike, M., Sugiyama, T., Kume, T., and Nagasawa, T. (2014). Foxc1 is a critical regulator of haematopoietic stem/progenitor cell niche formation. *Nature* 508, 536–540.
- Ormestad, M., Astorga J, Landgren H, Wang T, Johansson BR, Miura N, Carlsson P (2006). Foxf1 and Foxf2 control murine gut development by limiting mesenchymal Wnt signaling and promoting extracellular matrix production. *Development* 133, 833–843.
- Otte, K., Kranz, H., Kober, I., Thompson, P., Hoefer, M., Haubold, B., Rimmel, B., Voss, H., Kaiser, C., Albers, M., et al. (2003). Identification of farnesoid X receptor beta as a novel mammalian nuclear receptor sensing lanosterol. *Mol. Cell. Biol.* 23, 864–872.
- de Pater, E., Clijsters, L., Marques, S.R., Lin, Y.-F., Garavito-Aguilar, Z. V., Yelon, D., and Bakkers, J. (2009). Distinct phases of cardiomyocyte differentiation regulate growth of the zebrafish heart. *Development* 136, 1633–1641.
- Patra, C., Kontarakis, Z., Kaur, H., Rayrikar, A., Mukherjee, D., and Stainier, D.Y.R. (2017). The zebrafish ventricle: A hub of cardiac endothelial cells for in vitro cell behavior studies. *Sci. Rep.* 7, 2687.
- Paw, B.H., Davidson, A.J., Zhou, Y., Li, R., Pratt, S.J., Lee, C., Trede, N.S., Brownlie, A., Donovan, A., Liao, E.C., et al. (2003). Cell-specific mitotic defect and dyserythropoiesis associated with erythroid band 3 deficiency. *Nat. Genet.* 34, 59–64.
- Payne, V.A., Grimsey, N., Tuthill, A., Virtue, S., Gray, S.L., Nora, E.D., Semple, R.K., O’Rahilly, S., and Rochford, J.J. (2008). The human lipodystrophy gene BSCL2/Seipin may be essential for normal adipocyte differentiation. *Diabetes* 57, 2055–2060.
- Paz, H., Lynch, M.R., Bogue, C.W., and Gasson, J.C. (2010). The homeobox gene Hhex regulates the earliest stages of definitive hematopoiesis. *Blood* 116, 1254–1262.
- Peralta, M., Steed, E., Harlepp, S., González-Rosa, J.M., Monduc, F., Ariza-Cosano, A., Cortés, A., Rayón, T., Gómez-Skarmeta, J.L., Zapata, A., et al. (2013). Heartbeat-driven pericardiac fluid forces contribute to epicardium morphogenesis. *Curr. Biol.* 23, 1726–1735.
- Peralta, M., González-Rosa, J.M., Marques, I.J., and Mercader, N. (2014). The Epicardium in the Embryonic and Adult Zebrafish. *J. Dev. Biol.* 2, 101–116.
- Pérez-Pomares, J.M., Phelps, A., Sedmerova, M., and Wessels, A. (2003). Epicardial-like cells on the distal arterial end of the cardiac outflow tract do not derive from the proepicardium but are derivatives of the cephalic pericardium. *Dev. Dyn.* 227, 56–68.
- Perner, B., Englert, C., and Bollig, F. (2007). The Wilms tumor genes wt1a and wt1b

control different steps during formation of the zebrafish pronephros. *Dev. Biol.* 309, 87–96.

Peshkovsky, C., Totong, R., and Yelon, D. (2011). Dependence of cardiac trabeculation on neuregulin signaling and blood flow in zebrafish. *Dev. Dyn.* 240, 446–456.

Pestel, J., Ramadass, R., Gauvrit, S., Helker, C., Herzog, W., and Stainier, D.Y.R. (2016). Real-time 3D visualization of cellular rearrangements during cardiac valve formation. *Development* 143, 2217–2227.

Peterkin, T., Gibson, A., and Patient, R. (2003). GATA-6 maintains BMP-4 and Nkx2 expression during cardiomyocyte precursor maturation. *EMBO J.* 22, 4260–4273.

Peterkin, T., Gibson, A., and Patient, R. (2007). Redundancy and evolution of GATA factor requirements in development of the myocardium. *Dev. Biol.* 311, 623–635.

Peterkin, T., Gibson, A., and Patient, R. (2009). Common genetic control of haemangioblast and cardiac development in zebrafish. *Development* 136, 1465–1474.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.

Plavicki, J.S., Hofsteen, P., Yue, M.S., Lanham, K.A., Peterson, R.E., and Heideman, W. (2014). Multiple modes of proepicardial cell migration require heartbeat. *BMC Dev. Biol.* 14, 18.

Pombal, M.A., Carmona, R., Megías, M., Ruiz, A., Pérez-Pomares, J.M., and Muñoz-Chápuli, R. (2008). Epicardial development in lamprey supports an evolutionary origin of the vertebrate epicardium from an ancestral pronephric external glomerulus. *Evol. Dev.* 10, 210–216.

Porrello, E.R., and Olson, E.N. (2014). A neonatal blueprint for cardiac regeneration. *Stem Cell Res.* 13, 556–570.

Porrello, E.R., Mahmoud, A.I., Simpson, E., Hill, J.A., Richardson, J.A., Olson, E.N., and Sadek, H.A. (2011). Transient regenerative potential of the neonatal mouse heart. *Science* 331, 1078–1080.

Poss, K.D., Wilson LG, Keating MT. (2002). Heart Regeneration in Zebrafish. *Science* (80-). 298, 2188–2190.

Powell, G.T., and Wright, G.J. (2011). Jamb and jamc are essential for vertebrate myocyte fusion. *PLoS Biol.* 9.

Powell, G.T., and Wright, G.J. (2012). Genomic organisation, embryonic expression and biochemical interactions of the zebrafish junctional adhesion molecule family of receptors. *PLoS One* 7.

- Prall, O.W.J., Menon, M.K., Solloway, M.J., Watanabe, Y., Zaffran, S., Bajolle, F., Biben, C., McBride, J.J., Robertson, B.R., Chaulet, H., et al. (2007). An Nkx2-5/Bmp2/Smad1 Negative Feedback Loop Controls Heart Progenitor Specification and Proliferation. *Cell* 128, 947–959.
- Pu, W.T., Ishiwata, T., Juraszek, A.L., Ma, Q., and Izumo, S. (2004). GATA4 is a dosage-sensitive regulator of cardiac morphogenesis. *Dev. Biol.* 275, 235–244.
- Qian, F., Zhen, F., Ong, C., Jin, S.W., Soo, H.M., Stainier, D.Y.R., Lin, S., Peng, J., and Wen, Z. (2005). Microarray analysis of zebrafish cloche mutant using amplified cDNA and identification of potential downstream target genes. *Dev. Dyn.* 233, 1163–1172.
- Qiu, M., Hua, S., Agrawal, M., Li, G., Cai, J., Chan, E., Zhou, H., Luo, Y., and Liu, M. (1998). Molecular Cloning and Expression of Human Grap-2, a Novel Leukocyte-Specific SH2- and SH3-Containing Adaptor-like Protein That Binds to Gab-1. *Biochem. Biophys. Res. Commun.* 253, 443–447.
- Quaggin, S.E., Vanden Heuvel, G.B., and Igarashi, P. (1998). Pod-1, a mesoderm-specific basic-helix-loop-helix protein expressed in mesenchymal and glomerular epithelial cells in the developing kidney. *Mech. Dev.* 71, 37–48.
- Quarles, R.H. (2007). Myelin-associated glycoprotein (MAG): Past, present and beyond. *J. Neurochem.* 100, 1431–1448.
- Raya, A., Koth, C.M., Büscher, D., Kawakami, Y., Itoh, T., Raya, R.M., Sternik, G., Tsai, H.-J.J., Rodríguez-Esteban, C., and Izpisua-Belmonte, J.C. (2003). Activation of Notch signaling pathway precedes heart regeneration in zebrafish. *Proc. Natl. Acad. Sci. U. S. A.* 100, 11889–11895.
- Reijntjes, S., Rodaway, A., and Maden, M. (2007). The retinoic acid metabolising gene, CYP26B1, patterns the cartilaginous cranial neural crest in zebrafish. *Int. J. Dev. Biol.* 51, 351–360.
- 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.
- Reiter, J.F., Verkade, H., and Stainier, D.Y.R. (2001). Bmp2b and Oep promote early myocardial differentiation through their regulation of gata5. *Dev. Biol.* 234, 330–338.
- Reynaud, C., Baas, D., Gleyzal, C., Le Guellec, D., and Sommer, P. (2008). Morpholino knockdown of lysyl oxidase impairs zebrafish development, and reflects some aspects of copper metabolism disorders. *Matrix Biol.* 27, 547–560.
- Rhodes, J., Hagen, A., Hsu, K., Deng, M., Liu, T.X., Look, A.T., and Kanki, J.P. (2005). Interplay of pu.1 and Gata1 determines myelo-erythroid progenitor cell fate in zebrafish.

Dev. Cell 8, 97–108.

Ribeiro, I., Kawakami, Y., Büscher, D., Raya, Á., Rodríguez-León, J., Morita, M., Rodríguez Esteban, C., and Izpisua Belmonte, J.C. (2007). Tbx2 and Tbx3 regulate the dynamics of cell proliferation during heart remodeling. *PLoS One* 2.

Risebro, C.A., Vieira, J.M., Klotz, L., and Riley, P.R. (2015). Characterisation of the human embryonic and foetal epicardium during heart development. *Development* 142, 3630–3636.

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

Robb, L., Mifsud, L., Hartley, L., Biben, C., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., and Harvey, R.P. (1998). epicardin: A novel basic helix-loop-helix transcription factor gene expressed in epicardium, branchial arch myoblasts, and mesenchyme of developing lung, gut, kidney, and gonads. *Dev. Dyn.* 213, 105–113.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.

Rotem, A., Ram, O., Shores, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* 33, 1165–1172.

Rudat, C., and Kispert, A. (2012). Wt1 and epicardial fate mapping. *Circ. Res.* 111, 165–169.

Rudat, C., Norden, J., Taketo, M.M., and Kispert, A. (2013). Epicardial function of canonical Wnt-, Hedgehog-, Fgfr1/2-, and Pdgfra-signalling. *Cardiovasc. Res.* 100, 411–421.

Rydeen, A.B., and Waxman, J.S. (2016). Cyp26 Enzymes Facilitate Second Heart Field Progenitor Addition and Maintenance of Ventricular Integrity. *PLoS Biol.* 14.

Sacilotto, N., Monteiro, R., Fritzsche, M., Becker, P.W., Sanchez-del-Campo, L., Liu, K., Pinheiro, P., Ratnayaka, I., Davies, B., Goding, C.R., et al. (2013). Analysis of Dll4 regulation reveals a combinatorial role for Sox and Notch in arterial development. *Proc. Natl. Acad. Sci.* 110, 11893–11898.

Sander, V., Suñe, G., Jopling, C., Morera, C., and Belmonte, J.C.I. (2013). Isolation and in vitro culture of primary cardiomyocytes from adult zebrafish hearts. *Nat. Protoc.* 8, 800–809.

Sassetti, C., Tangemann, K., Singer, M.S., Kershaw, D.B., and Rosen, S.D. (1998). Identification of podocalyxin-like protein as a high endothelial venule ligand for L-

selectin: parallels to CD34. *J. Exp. Med.* *187*, 1965–1975.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* *33*, 495–502.

Scherz, P.J., Huisken, J., Sahai-Hernandez, P., and Stainier, D.Y.R. (2008). High-speed imaging of developing heart valves reveals interplay of morphogenesis and function. *Development* *135*, 1179–1187.

Schindler, Y.L., Garske, K.M., Wang, J., Firulli, B.A., Firulli, A.B., Poss, K.D., and Yelon, D. (2014). Hand2 elevates cardiomyocyte production during zebrafish heart development and regeneration. *Development* *141*, 3112–3122.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* *12*, 963–965.

Schnabel, K., Wu, C.C., Kurth, T., and Weidinger, G. (2011). Regeneration of cryoinjury induced necrotic heart lesions in zebrafish is associated with epicardial activation and cardiomyocyte proliferation. *PLoS One* *6*.

Schoenebeck, J.J., Keegan, B.R., and Yelon, D. (2007). Vessel and Blood Specification Override Cardiac Potential in Anterior Mesoderm. *Dev. Cell* *13*, 254–267.

Scholey, J.M., Taylor, K.A., and Kendrick-Jones, J. (1980). Regulation of non-muscle myosin assembly by calmodulin-dependent light chain kinase. *Nature* *287*, 233–235.

Schwartz, G.A., Henion, T.R., Nugent, J.D., Caplan, B., and Tobet, S. (2006). Stromal Cell-Derived Factor-1 (Chemokine C-X-C Motif Ligand 12) and Chemokine C-X-C Motif Receptor 4 Are Required for Migration of Gonadotropin-Releasing Hormone Neurons to the Forebrain. *J. Neurosci.* *26*, 6834–6840.

Sedmera, D., Pexieder, T., Vuillemin, M., Thompson, R.P., and Anderson, R.H. (2000). Developmental patterning of the myocardium. *Anat. Rec.* *258*, 319–337.

Seo, S., and Kume, T. (2006). Forkhead transcription factors, Foxc1 and Foxc2, are required for the morphogenesis of the cardiac outflow tract. *Dev. Biol.* *296*, 421–436.

Serluca, F.C. (2008). Development of the proepicardial organ in the zebrafish. *Dev. Biol.* *315*, 18–27.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E., and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* *22*, 1567–1572.

Sharf, R., Azriel, A., Lejbkowitz, F., Winograd, S.S., Ehrlich, R., and Levi, B.Z. (1995). Functional domain analysis of interferon consensus sequence binding protein (ICSBP) and

its association with interferon regulatory factors. *J. Biol. Chem.* 270, 13063–13069.

Shelton, E.L., and Yutzey, K.E. (2007). Tbx20 regulation of endocardial cushion cell proliferation and extracellular matrix gene expression. *Dev. Biol.* 302, 376–388.

Shelton, E.L., and Yutzey, K.E. (2008). Twist1 function in endocardial cushion cell proliferation, migration, and differentiation during heart valve development. *Dev. Biol.* 317, 282–295.

Shi, G., Harrison, K., Wilson, G.L., Moratz, C., and Kehrl, J.H. (2002). RGS13 regulates germinal center B lymphocytes responsiveness to CXC chemokine ligand (CXCL12 and CXCL13). *J. Immunol.* 169, 2507–2515.

Simoës, F.C., Peterkin, T., and Patient, R. (2011). Fgf differentially controls cross-antagonism between cardiac and haemangioblast regulators. *Development* 138, 3235–3245.

Singh, A.R., Sivadas, A., Sabharwal, A., Vellarikal, S.K., Jayarajan, R., Verma, A., Kapoor, S., Joshi, A., Scaria, V., and Sivasubbu, S. (2016). Chamber specific gene expression landscape of the zebrafish heart. *PLoS One* 11.

Singh, R., Hoogaars, W.M., Barnett, P., Grieskamp, T., Sameer Rana, M., Buermans, H., Farin, H.F., Petry, M., Heallen, T., Martin, J.F., et al. (2012). Tbx2 and Tbx3 induce atrioventricular myocardial development and endocardial cushion formation. *Cell. Mol. Life Sci.* 69, 1377–1389.

Sinha, T., Lin, L., Li, D., Davis, J., Evans, S., Wynshaw-Boris, A., and Wang, J. (2015). Mapping the dynamic expression of Wnt11 and the lineage contribution of Wnt11-expressing cells during early mouse development. *Dev. Biol.* 398, 177–192.

Smart, N., Risebro, C.A., Melville, A.A.D., Moses, K., Schwartz, R.J., Chien, K.R., and Riley, P.R. (2007). Thymosin  $\beta$ 4 induces adult epicardial progenitor mobilization and neovascularization. *Nature* 445, 177–182.

Smart, N., Risebro, C.A., Clark, J.E., Ehler, E., Miquerol, L., Rossdeutsch, A., Marber, M.S., and Riley, P.R. (2010). Thymosin  $\beta$ 4 facilitates epicardial neovascularization of the injured adult heart. In *Annals of the New York Academy of Sciences*, pp. 97–104.

Smart, N., Bollini, S., Dubé, K.N., Vieira, J.M., Zhou, B., Davidson, S., Yellon, D., Riegler, J., Price, A.N., Lythgoe, M.F., et al. (2011). De novo cardiomyocytes from within the activated adult heart after injury. *Nature* 474, 640–644.

Smith, C.L., Baek, S.T., Sung, C.Y., and Tallquist, M.D. (2011). Epicardial-derived cell epithelial-to-mesenchymal transition and fate specification require PDGF receptor signaling. *Circ. Res.* 108.

Smyk, M., Szafranski, P., Startek, M., Gambin, A., and Stankiewicz, P. (2013).

Chromosome conformation capture-on-chip analysis of long-range cis-interactions of the SOX9 promoter. *Chromosom. Res.* *21*, 781–788.

Snider, P., Hinton, R.B., Moreno-Rodriguez, R.A., Wang, J., Rogers, R., Lindsley, A., Li, F., Ingram, D.A., Menick, D., Field, L., et al. (2008). Periostin is required for maturation and extracellular matrix stabilization of noncardiomyocyte lineages of the heart. *Circ. Res.* *102*, 752–760.

Soo, K., O'Rourke, M.P., Khoo, P.-L., Steiner, K. a, Wong, N., Behringer, R.R., and Tam, P.P.L. (2002). Twist function is required for the morphogenesis of the cephalic neural tube and the differentiation of the cranial neural crest cells in the mouse embryo. *Dev. Biol.* *247*, 251–270.

Spoorendonk, K.M., Peterson-Maduro, J., Renn, J., Trowe, T., Kranenbarg, S., Winkler, C., and Schulte-Merker, S. (2008). Retinoic acid and *Cyp26b1* are critical regulators of osteogenesis in the axial skeleton. *Development* *135*, 3765–3774.

Stainier, D.Y. (2001). Zebrafish genetics and vertebrate heart formation. *Nat. Rev. Genet.* *2*, 39–48.

Stainier, D.Y., Lee, R.K., and Fishman, M.C. (1993). Cardiovascular development in the zebrafish. I. Myocardial fate map and heart tube formation. *Development* *119*, 31–40.

Staudt, D.W., Liu, J., Thorn, K.S., Stuurman, N., Liebling, M., and Stainier, D.Y.R. (2014). High-resolution imaging of cardiomyocyte behavior reveals two distinct steps in ventricular trabeculation. *Development* *141*, 585–593.

Steed, E., Faggianelli, N., Roth, S., Ramspacher, C., Concordet, J.-P., and Vermot, J. (2016). *klf2a* couples mechanotransduction and zebrafish valve morphogenesis through fibronectin synthesis. *Nat. Commun.* *7*, 11646.

Stefanovic, S., and Christoffels, V.M. (2015). GATA-dependent transcriptional and epigenetic control of cardiac lineage specification and differentiation. *Cell. Mol. Life Sci.* *72*, 3871–3881.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* *16*, 133–145.

Stevens, S.M., von Gise, A., VanDusen, N., Zhou, B., and Pu, W.T. (2016). Epicardium is required for cardiac seeding by yolk sac macrophages, precursors of resident macrophages of the adult heart. *Dev. Biol.* *413*, 153–159.

Stuckmann, I., Evans, S., and Lassar, A.B. (2003). Erythropoietin and retinoic acid, secreted from the epicardium, are required for cardiac myocyte proliferation. *Dev. Biol.* *255*, 334–349.

Sumanas, S., and Lin, S. (2006). *Ets1*-related protein is a key regulator of vasculogenesis in

zebrafish. *PLoS Biol.* 4, 0060–0069.

Sumanas, S., Zhang, B., Dai, R., and Lin, S. (2005a). 15-Zinc finger protein Bloody Fingers is required for zebrafish morphogenetic movements during neurulation. *Dev. Biol.* 283, 85–96.

Sumanas, S., Joraniak, T., and Lin, S. (2005b). Identification of novel vascular endothelial-specific genes by the microarray analysis of the zebrafish cloche mutants. *Blood* 106, 534–541.

Sutton, M.G.S.J., and Sharpe, N. (2000). Left Ventricular Remodeling After Myocardial Infarction : Pathophysiology and Therapy. *Circulation* 101, 2981–2988.

Takeda, T., Go, W.Y., Orlando, R.A., and Farquhar, M.G. (2000). Expression of podocalyxin inhibits cell-cell adhesion and modifies junctional properties in Madin-Darby canine kidney cells. *Mol. Biol. Cell* 11, 3219–3232.

Takeichi, M., Nimura, K., Mori, M., Nakagami, H., and Kaneda, Y. (2013). The Transcription Factors Tbx18 and Wt1 Control the Epicardial Epithelial-Mesenchymal Transition through Bi-Directional Regulation of Slug in Murine Primary Epicardial Cells. *PLoS One* 8.

Tan, G., and Lenhard, B. (2016). TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* 32, 1555–1556.

Tandon, P., Miteva, Y. V, Kuchenbrod, L.M., Cristea, I.M., and Conlon, F.L. (2013). Tcf21 regulates the specification and maturation of proepicardial cells. *Development* 140, 2409–2421.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382.

Terman, B.I., Dougher-Vermazen, M., Carrion, M.E., Dimitrov, D., Armellino, D.C., Gospodarowicz, D., and Böhlen, P. (1992). Identification of the KDR tyrosine kinase as a receptor for vascular endothelial cell growth factor. *Biochem. Biophys. Res. Commun.* 187, 1579–1586.

Tingaud-Sequeira, A., Fogue, J., André, M., and Babin, P.J. (2006). Epidermal transient down-regulation of retinol-binding protein 4 and mirror expression of apolipoprotein Eb and estrogen receptor 2a during zebrafish fin and scale development. *Dev. Dyn.* 235, 3071–3079.

Tissier-Seta, J.P., Mucchielli, M.L., Mark, M., Mattei, M.G., Goridis, C., and Brunet, J.F. (1995). Barx1, a new mouse homeodomain transcription factor expressed in cranio-facial ectomesenchyme and the stomach. *Mech. Dev.* 51, 3–15.

- Tiveron, M.-C., Rossel, M., Moepps, B., Zhang, Y.L., Seidenfaden, R., Favor, J., König, N., and Cremer, H. (2006). Molecular Interaction between Projection Neuron Precursors and Invading Interneurons via Stromal-Derived Factor 1 (CXCL12)/CXCR4 Signaling in the Cortical Subventricular Zone/Intermediate Zone. *J. Neurosci.* *26*, 13273–13278.
- Tomanek, R.J. (2005). Formation of the coronary vasculature during development. *Angiogenesis* *8*, 273–284.
- Tomkowicz, B., Rybinski, K., Sebeck, D., Sass, P., Nicolaidis, N.C., Grasso, L., and Zhou, Y. (2010). Endosialin/TEM-1/CD248 regulates pericyte proliferation through PDGF receptor signaling. *Cancer Biol. Ther.* *9*, 908–915.
- Torraca, V., Otto, N.A., Tavakoli-Tameh, A., and Meijer, A.H. (2017). The inflammatory chemokine Cxcl18b exerts neutrophil-specific chemotaxis via the promiscuous chemokine receptor Cxcr2 in zebrafish. *Dev. Comp. Immunol.* *67*, 57–65.
- Tournoij, E., Weber, G.J., Akkerman, J.W.N., De Groot, P.G., Zon, L.I., Moll, F.L., and Schulte-Merker, S. (2010). Mlck1a is expressed in zebrafish thrombocytes and is an essential component of thrombus formation. *J. Thromb. Haemost.* *8*, 588–595.
- Traver, D., Paw, B.H., Poss, K.D., Penberthy, W.T., Lin, S., and Zon, L.I. (2003). Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat. Immunol.* *4*, 1238–1246.
- Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* *509*, 371–375.
- Trichas, G., Begbie, J., and Srinivas, S. (2008). Use of the viral 2A peptide for bicistronic expression in transgenic mice. *BMC Biol.* *6*, 40.
- Tsuchihashi, T., Maeda, J., Shin, C.H., Ivey, K.N., Black, B.L., Olson, E.N., Yamagishi, H., and Srivastava, D. (2011). Hand2 function in second heart field progenitors is essential for cardiogenesis. *Dev. Biol.* *351*, 62–69.
- Ubil, E., Duan, J., Pillai, I.C.L., Rosa-Garrido, M., Wu, Y., Bargiacchi, F., Lu, Y., Stanbouly, S., Huang, J., Rojas, M., et al. (2014). Mesenchymal–endothelial transition contributes to cardiac neovascularization. *Nature* *514*, 585–590.
- Valentin, G., Haas, P., and Gilmour, D. (2007). The Chemokine SDF1a Coordinates Tissue Migration through the Spatially Restricted Activation of Cxcr7 and Cxcr4b. *Curr. Biol.* *17*, 1026–1031.
- Vicari, A.P., Figueroa, D.J., Hedrick, J.A., Foster, J.S., Singh, K.P., Menon, S., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Bacon, K.B., et al. (1997). TECK: A novel CC chemokine specifically expressed by thymic dendritic cells and potentially involved in T cell development. *Immunity* *7*, 291–301.

- Vieira, J.M., Howard, S., Villa del Campo, C., Bollini, S., Dubé, K.N., Masters, M., Barnette, D.N., Rohling, M., Sun, X., Hankins, L.E., et al. (2017). BRG1-SWI/SNF-dependent regulation of the Wt1 transcriptional landscape mediates epicardial activity during heart development and disease. *Nat. Commun.* *8*, 16034.
- van Vliet, P.P., Lin, L., Boogerd, C.J., Martin, J.F., Andelfinger, G., Grossfeld, P.D., and Evans, S.M. (2017). Tissue specific requirements for WNT11 in developing outflow tract and dorsal mesenchymal protrusion. *Dev. Biol.* *429*, 249–259.
- Wallace, K.N., Akhter, S., Smith, E.M., Lorent, K., and Pack, M. (2005). Intestinal growth and differentiation in zebrafish. *Mech. Dev.* *122*, 157–173.
- Wang, J., and Knaut, H. (2014). Chemokine signaling in development and disease. *Development* *141*, 4199–4205.
- Wang, Y., and Navin, N.E. (2015). Advances and Applications of Single-Cell Sequencing Technologies. *Mol. Cell* *58*, 598–609.
- Wang, J., Panáková, D., Kikuchi, K., Holdway, J.E., Gemberling, M., Burris, J.S., Singh, S.P., Dickson, A.L., Lin, Y.-F., Sabeh, M.K., et al. (2011). The regenerative capacity of zebrafish reverses cardiac failure caused by genetic cardiomyocyte depletion. *Development* *138*, 3421–3430.
- Wang, J., Karra, R., Dickson, A.L., and Poss, K.D. (2013). Fibronectin is deposited by injury-activated epicardial cells and is necessary for zebrafish heart regeneration. *Dev. Biol.* *382*, 427–435.
- Wang, J., Cao, J., Dickson, A.L., and Poss, K.D. (2015). Epicardial regeneration is guided by cardiac outflow tract and Hedgehog signalling. *Nature* *522*, 226–230.
- Watt, A.J., Battle, M.A., Li, J., and Duncan, S.A. (2004). GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 12573–12578.
- Welch, J.E., Schatte, E.C., O'Brien, D. a, and Eddy, E.M. (1992). Expression of a glyceraldehyde 3-phosphate dehydrogenase gene specific to mouse spermatogenic cells. *Biol. Reprod.* *46*, 869–878.
- Whitesell, T.R., Kennedy, R.M., Carter, A.D., Rollins, E.L., Georgijevic, S., Santoro, M.M., and Childs, S.J. (2014). An  $\alpha$ -smooth muscle actin (acta2/ $\alpha$ sma) zebrafish transgenic line marking vascular mural cells and visceral smooth muscle cells. *PLoS One* *9*.
- Wilhelm, A., Aldridge, V., Haldar, D., Naylor, A.J., Weston, C.J., Hedegaard, D., Garg, A., Fear, J., Reynolds, G.M., Croft, A.P., et al. (2015). CD248/endosialin critically regulates hepatic stellate cell proliferation during chronic liver injury via a PDGF-regulated mechanism. *Gut* [gutjnl-2014-308325](https://doi.org/10.1136/gutjnl-2014-308325).

- Wills, Q.F., Livak, K.J., Tipping, A.J., Enver, T., Goldson, A.J., Sexton, D.W., and Holmes, C. (2013). Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nat. Biotechnol.* *31*, 748–752.
- Wlodarski, P., Zhang, Q., Liu, X., Kasprzycka, M., Marzec, M., and Wasik, M.A. (2007). PU.1 activates transcription of SHP-1 gene in hematopoietic cells. *J. Biol. Chem.* *282*, 6316–6323.
- Woodcock-Mitchell, J., Mitchell, J.J., Low, R.B., Kieny, M., Sengel, P., Rubbia, L., Skalli, O., Jackson, B., and Gabbiani, G. (1988). Alpha-smooth muscle actin is transiently expressed in embryonic rat cardiac and skeletal muscles. *Differentiation* *39*, 161–166.
- Woods, I.G., Lyons, D.A., Voas, M.G., Pogoda, H.M., and Talbot, W.S. (2006). *nsf* Is Essential for Organization of Myelinated Axons in Zebrafish. *Curr. Biol.* *16*, 636–648.
- Wu, S.P., Dong, X.R., Regan, J.N., Su, C., and Majesky, M.W. (2013). Tbx18 regulates development of the epicardium and coronary vessels. *Dev. Biol.* *383*, 307–320.
- Xin, M., Olson, E.N., and Bassel-Duby, R. (2013). Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. *Nat. Rev. Mol. Cell Biol.* *14*, 529–541.
- Xu, J., Liu, H., Lan, Y., Aronow, B.J., Kalinichenko, V. V., and Jiang, R. (2016). A Shh-Foxf-Fgf18-Shh Molecular Circuit Regulating Palate Development. *PLoS Genet.* *12*.
- Xue, Y., Lv, J., Zhang, C., Wang, L., Ma, D., and Liu, F. (2017). The Vascular Niche Regulates Hematopoietic Stem and Progenitor Cell Lodgment and Expansion via *klf6a-ccl25b*. *Dev. Cell* *42*, 349–362.e4.
- Yamada, Y., Warren, A.J., Dobson, C., Forster, A., Pannell, R., and Rabbitts, T.H. (1998). The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 3890–3895.
- Yamaguchi, Y., Cavallero, S., Patterson, M., Shen, H., Xu, J., Kumar, S.R., and Sucov, H.M. (2015). Adipogenesis and epicardial adipose tissue: a novel fate of the epicardium induced by mesenchymal transformation and PPAR $\gamma$  activation. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 2070–2075.
- Yurugi-Kobayashi, T., Itoh, H., Schroeder, T., Nakano, A., Narazaki, G., Kita, F., Yanagi, K., Hiraoka-Kanie, M., Inoue, E., Ara, T., et al. (2006). Adrenomedullin/cyclic AMP pathway induces notch activation and differentiation of arterial endothelial cells from vascular progenitors. *Arterioscler. Thromb. Vasc. Biol.* *26*, 1977–1984.
- Zacharias, D.A., Violin JD, Newton AC, Tsien RY. (2002). Partitioning of Lipid-Modified Monomeric GFPs into Membrane Microdomains of Live Cells. *Science* (80-. ). *296*, 913–916.

- Zakrzewska, A., Cui, C., Stockhammer, O.W., Benard, E.L., Spaink, H.P., and Meijer, A.H. (2010). Macrophage-specific gene functions in Spi1-directed innate immunity. *Blood* 116.
- Zeisel, a., Machado, a. B.M., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, a., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* (80-. ). 347, 1138–1142.
- Zhou, B., Gise, A. von, Ma, Q., Rivera-Feliciano, J., and Pu, W.T. (2008a). Nkx2-5- and Isl1-expressing cardiac progenitors contribute to proepicardium. *Biochem. Biophys. Res. Commun.* 375, 450–453.
- Zhou, B., Ma, Q., Rajagopal, S., Wu, S.M., Domian, I., Rivera-Feliciano, J., Jiang, D., von Gise, A., Ikeda, S., Chien, K.R., et al. (2008b). Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart. *Nature* 454, 109–113.
- Zhou, B., von Gise, A., Ma, Q., Hu, Y.W., and Pu, W.T. (2010). Genetic fate mapping demonstrates contribution of epicardium-derived cells to the annulus fibrosus of the mammalian heart. *Dev. Biol.* 338, 251–261.
- Zhou, B., Honor, L.B., He, H., Qing, M., Oh, J.H., Butterfield, C., Lin, R.Z., Melero-Martin, J.M., Dolmatova, E., Duffy, H.S., et al. (2011a). Adult mouse epicardium modulates myocardial injury by secreting paracrine factors. *J. Clin. Invest.* 121, 1894–1904.
- Zhou, B., Honor, L.B., Ma, Q., Oh, J.H., Lin, R.Z., Melero-Martin, J.M., von Gise, A., Zhou, P., Hu, T., He, L., et al. (2012). Thymosin beta 4 treatment after myocardial infarction does not reprogram epicardial cells into cardiomyocytes. *J. Mol. Cell. Cardiol.* 52, 43–47.
- Zhou, Y., Cashman, T.J., Nevis, K.R., Obregon, P., Carney, S.A., Liu, Y., Gu, A., Mosimann, C., Sondalle, S., Peterson, R.E., et al. (2011b). Latent TGF- $\beta$  binding protein 3 identifies a second heart field in zebrafish. *Nature* 474, 645–648.
- Ziaeeian, B., and Fonarow, G.C. (2016). Epidemiology and aetiology of heart failure. *Nat. Rev. Cardiol.* 13, 368–378.
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* 65, 631–643.e4.
- Zuo, S., Jones, W.K., Li, H., He, Z., Pasha, Z., Yang, Y., Wang, Y., Fan, G.-C., Ashraf, M., and Xu, M. (2012). Paracrine Effect of Wnt11-Overexpressing Mesenchymal Stem Cells on Ischemic Injury. *Stem Cells Dev.* 21, 598–608.

## Appendix A - HCR probe sequences

Probe name	Probe sequence 5' -> 3' (amplifier sequences colored)
adma B5 #1	CTCACTCCCAATCTCTATCTACCCTACAAATCCAATaaaaGCTGGGAATGCTGGTGCA GCAGTGATACTCTTCTAATAGTCTGACTCTCCatTTTCACTTCATATCACTCACTCCCAAT CTCTATCTACCC
adma B5 #2	CTCACTCCCAATCTCTATCTACCCTACAAATCCAATaaaaGTCGTCTGGTCTGACAAAC TGCCTGTTTTCTAAAGCCCTCAGTGCAGGAAatTTTCACTTCATATCACTCACTCCCAAT CTCTATCTACCC
adma B5 #3	CTCACTCCCAATCTCTATCTACCCTACAAATCCAATaaaaGTGCATTCCCAATCTTGAG TTTGTGTTGAGGTCATGGAGACGGTGTGCCatTTTCACTTCATATCACTCACTCCCAAT CTCTATCTACCC
adma B5 #4	CTCACTCCCAATCTCTATCTACCCTACAAATCCAATaaaaATTGCTCTTCCAGTCGCTG TGCTTGACAAAGAATGCGCTGGAGAACTCTTatTTTCACTTCATATCACTCACTCCCAAT CTCTATCTACCC
adma B5 #5	CTCACTCCCAATCTCTATCTACCCTACAAATCCAATaaaaCAGGACACAGTGACACTG GTCTAGCACAATAATCCCACGGCTTGATTAatTTTCACTTCATATCACTCACTCCCAAT CTCTATCTACCC
cldn11a B4 #1	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAACCCAGTCCAGCCGAGAAA ACTCATCACA AAAACCCGTGAAGTGAAGCAAGATTTTCACATTTACAgACCTCAACCT ACCTCCAACCTCTCAC
cldn11a B4 #2	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAACTGCGTGAGTGTGATGCA GTGATAAAGCGCAGTTGAGATCACACAGTCTGATTTTCACATTTACAgACCTCAACCT ACCTCCAACCTCTCAC
cldn11a B4 #3	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAAACATGGACATGAGCACCA TCACAACGGCAGGCAATCCAAGAATCGATGCTATTTTCACATTTACAgACCTCAACCT ACCTCCAACCTCTCAC
cldn11a B4 #4	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAAAGACAGAGAGCTGTCCCC ACCCATCCTGAATACAGAGAAAAGCCAAAGGAATTTTCACATTTACAgACCTCAACCT ACCTCCAACCTCTCAC
cldn11a B4 #5	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAACTGTGTGGGCAGGGCCT TGTTTGGAGTAGTAGAATCGGTTGTTTTCAGTATTTTCACATTTACAgACCTCAACCTA CCTCCAACCTCTCAC
cldn11a B4 #6	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAAATGGGAAACCACACAGTG GAGACGATGCCACAGAAAGCTGTCAGCAGTATATTTTCACATTTACAgACCTCAACCT ACCTCCAACCTCTCAC
elnb B4 #1	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAAATACACAGAACCAGCAGGA GAAATCCATGCAAGAGCAGAGGTACCTTCTCATTTTCACATTTACAgACCTCAACCT ACCTCCAACCTCTCAC
elnb B4 #2	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAAATGGGTAGCCTGTTCCAG TACCTCTGGTGAAGTGAATAGCACCAAACATTTTCACATTTACAgACCTCAACCTA CCTCCAACCTCTCAC
elnb B4 #3	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAAGAAGTCCAACACCAGCTG TTCCACCTGGTAATCTCCAGTTCTAAAGTCATTTTCACATTTACAgACCTCAACCTAC CTCCAACCTCTCAC

elnb B4 #4	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAACCTGGAATACCTGCAGCTA ATCCAGTGCCTGGACCTAACCTGTACCTCTAATTTTCACATTTACAgACCTCAACCTA CCTCCAACCTCTCAC
elnb B4 #5	CCTCAACCTACCTCCAACCTCACCATATTCgCTTCTAAAACCTCCTGCACCAAGAAAG CCTTGAAGAGCTGCATATTTGGCGGCTTTGAATTTTCACATTTACAgACCTCAACCTA CCTCCAACCTCTCAC

## Appendix B - Illumina Nextera indexing primer sequences

Primer name	Index sequence	Full primer sequence 5' -> 3'
Ad1_noMX	-----	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCA GATGTG
S517	TCTTACGC	AATGATACGGCGACCACCGAGATCTACAC <u>CGCTAAGAT</u> CGTCGG CAGCGTC
S502	ATAGAGAG	AATGATACGGCGACCACCGAGATCTACAC <u>CTCTCTATT</u> CGTCGGC AGCGTC
S503	AGAGGATA	AATGATACGGCGACCACCGAGATCTACACT <u>TATCCTCTT</u> CGTCGGC AGCGTC
S504	TCTACTCT	AATGATACGGCGACCACCGAGATCTACAC <u>AGAGTAGAT</u> CGTCGG CAGCGTC
S505	CTCCTTAC	AATGATACGGCGACCACCGAGATCTACAC <u>GTAAGGAGT</u> CGTCGG CAGCGTC
S506	TATGCAGT	AATGATACGGCGACCACCGAGATCTACAC <u>ACTGCATAT</u> CGTCGG CAGCGTC
S507	TACTCCTT	AATGATACGGCGACCACCGAGATCTACACA <u>AAGGAGTAT</u> CGTCGG CAGCGTC
S508	AGGCTTAG	AATGATACGGCGACCACCGAGATCTACAC <u>CTAAGCCTT</u> CGTCGG CAGCGTC
N701	TAAGGCGA	CAAGCAGAAGACGGCATAACGAGAT <u>TCGCCTTAGT</u> CTCGTGGGCT CGGAGATGT
N702	CGTACTAG	CAAGCAGAAGACGGCATAACGAGAT <u>CTAGTACGGT</u> CTCGTGGGCT CGGAGATGT
N703	AGGCAGAA	CAAGCAGAAGACGGCATAACGAGAT <u>TTTCTGCCT</u> GTCTCGTGGGCT CGGAGATGT
N704	TCCTGAGC	CAAGCAGAAGACGGCATAACGAGAT <u>GCTCAGGAGT</u> CTCGTGGGC TCGGAGATGT
N705	GGACTCCT	CAAGCAGAAGACGGCATAACGAGAT <u>AGGAGTCCGT</u> CTCGTGGGC TCGGAGATGT
N706	TAGGCATG	CAAGCAGAAGACGGCATAACGAGAT <u>CATGCCTA</u> GTCTCGTGGGCT CGGAGATGT
N707	CTCTCTAC	CAAGCAGAAGACGGCATAACGAGAT <u>GTAGAGAGGT</u> CTCGTGGGC TCGGAGATGT
N708	CAGAGAGG	CAAGCAGAAGACGGCATAACGAGAT <u>CCTCTCTGGT</u> CTCGTGGGCT CGGAGATGT
N709	GCTACGCT	CAAGCAGAAGACGGCATAACGAGAT <u>AGCGTAGCGT</u> CTCGTGGGC TCGGAGATGT
N710	CGAGGCTG	CAAGCAGAAGACGGCATAACGAGAT <u>CAGCCTCGT</u> CTCGTGGGCT CGGAGATGT
N711	AAGAGGCA	CAAGCAGAAGACGGCATAACGAGAT <u>TGCCTCTTGT</u> CTCGTGGGCT CGGAGATGT
N712	GTAGAGGA	CAAGCAGAAGACGGCATAACGAGAT <u>TCCTCTACGT</u> CTCGTGGGCT CGGAGATGT

N714	GCTCATGA	CAAGCAGAAGACGGCATAACGAGATTCATGAGCGTCTCGTGGGCT CGGAGATGT
N715	ATCTCAGG	CAAGCAGAAGACGGCATAACGAGATCCTGAGATGTCTCGTGGGCT CGGAGATGT
N716	ACTCGCTA	CAAGCAGAAGACGGCATAACGAGATTAGCGAGTGTCTCGTGGGCT CGGAGATGT
N718	GGAGCTAC	CAAGCAGAAGACGGCATAACGAGATGTAGCTCCGTCTCGTGGGCT CGGAGATGT
N719	GCGTAGTA	CAAGCAGAAGACGGCATAACGAGATTACTACGCGTCTCGTGGGCT CGGAGATGT
N720	CGGAGCCT	CAAGCAGAAGACGGCATAACGAGATAGGCTCCGGTCTCGTGGGCT CGGAGATGT
N721	TACGCTGC	CAAGCAGAAGACGGCATAACGAGATGCAGCGTAGTCTCGTGGGC TCGGAGATGT
N722	ATGCGCAG	CAAGCAGAAGACGGCATAACGAGATCTGCGCATGTCTCGTGGGCT CGGAGATGT

## Appendix C - R scripts

### Single cell transcriptomic analysis

```
library(scater)
if (!require("scater")) {
  source("http://bioconductor.org/biocLite.R")
  biocLite("scater")
  library(scater)
}
```

```
library(scran)
if (!require("scran")) {
  source("http://bioconductor.org/biocLite.R")
  biocLite("scran")
  library(scran)
}
```

```
library(limSolve)
if (!require("limSolve")) {
  source("http://bioconductor.org/biocLite.R")
  biocLite("limSolve")
  library(limSolve)
}
```

```
library(RUVSeq)
if (!require("RUVSeq")) {
  source("http://bioconductor.org/biocLite.R")
  biocLite("RUVSeq")
}
```

```
library(EDASeq)
if (!require("EDASeq")) {
  source("http://bioconductor.org/biocLite.R")
  biocLite("EDASeq")
}
```

```
library(RColorBrewer)
if (!require("RColorBrewer")) {
  source("http://bioconductor.org/biocLite.R")
  biocLite("RColorBrewer")
  library(RColorBrewer)
}
```

```
library(pheatmap)
if (!require("pheatmap")) {
  install.packages("pheatmap", dependencies=TRUE)
  library(pheatmap)
}
```

```
library(edgeR)
if (!require("edgeR")) {
  source("http://bioconductor.org/biocLite.R")
  biocLite("edgeR")
}
```

```

library(edgeR)
}

library(statmod)
if (!require("statmod")) {
source("http://bioconductor.org/biocLite.R")
biocLite("statmod")
library(statmod)
}

library(scde)
if (!require("scde")) {
require(devtools)
devtools::install_github('hms-dbmi/scde', build_vignettes = FALSE)
library(scde)
}

library(biomaRt)
if (!require("biomaRt")) {
source("http://bioconductor.org/biocLite.R")
biocLite("biomaRt")
library(biomaRt)
}

library(GO.db)
if (!require("GO.db")) {
source("http://bioconductor.org/biocLite.R")
biocLite("GO.db")
library(GO.db)
}

library(org.Dr.eg.db)
if (!require("org.Dr.eg.db")) {
source("http://bioconductor.org/biocLite.R")
biocLite("org.Dr.eg.db")
library(org.Dr.eg.db)
}

library(Rtsne)
if (!require("Rtsne")) {
source("http://bioconductor.org/biocLite.R")
biocLite("Rtsne")
library(Rtsne)
}

library(ggplot2)
if (!require("ggplot2")) {
source("http://bioconductor.org/biocLite.R")
biocLite("ggplot2")
library(ggplot2)
}

# read in data
endo_data <- read.table("Genes_single_cell_zebrafish_heart_5dpf")
dim(endo_data)
spike_data <- read.table("ERCC_single_cell_zebrafish_heart_5dpf")
dim(spike_data)

```

```

mito_data <- read.table("MT_single_cell_zebrafish_heart_5dpf")
dim(mito_data)

# quality control using the scater package
# combine data into a single cell expression set
all_data <- rbind(endo_data, mito_data, spike_data)
rownames(all_data) <- make.names(all_data$gene_name, unique=TRUE)
metadata <- AnnotatedDataFrame(all_data[1:3])
sce <- newSCESet(countData=all_data[,4:ncol(all_data)], featureData=metadata)
dim(sce)

# calculate quality control metrics
nrows <- c(nrow(endo_data), nrow(mito_data), nrow(spike_data))
is.spike <- rep(c(FALSE, FALSE, TRUE), nrows)
is.mito <- rep(c(FALSE, TRUE, FALSE), nrows)
sce <- calculateQCMetrics(sce, feature_controls=list(ERCC=is.spike, Mt=is.mito))
head(colnames(pData(sce)))
isSpike(sce) <- "ERCC"

sum(grepl("TRUE",sce@featureData$is_feature_control_ERCC))
sum(grepl("TRUE",sce@featureData$is_feature_control_Mt))

# perform quality control on cells
# identify cells 3 median absolute deviations (MADs) below median values
libsize.drop <- isOutlier(sce$total_counts, nmads=3, type="lower", log=TRUE)
feature.drop <- isOutlier(sce$total_features, nmads=3, type="lower", log=TRUE)

# identify cells 3 median absolute deviations (MADs) above median values
mito.drop <- isOutlier(sce$pct_counts_feature_controls_Mt, nmads=3, type="higher")
spike.drop <- isOutlier(sce$pct_counts_feature_controls_ERCC, nmads=3, type="higher")

# exclude identified cells from dataset
sce <- sce[!(libsize.drop | feature.drop | spike.drop | mito.drop )]
data.frame(ByLibSize=sum(libsize.drop), ByFeature=sum(feature.drop),
           BySpike=sum(spike.drop), ByMito=sum(mito.drop), Remaining=ncol(sce))

# exclude genes with an average number of counts smaller than the cutoff value
cutoff <- 0.1
ave.counts <- rowMeans(counts(sce))
keep <- ave.counts >= cutoff
sum(keep)
sce <- sce[keep,]

# exclude mitochondrial genes
sce <- sce[!pData(sce)$is_feature_control_Mt,]

# analysis of heterogeneity using the scde package (pagoda pipeline)
# extract counts of endogenous genes from sce dataset (excluding ERCC)
count_data <- data.frame(counts(sce)[1:20862,])
dim(count_data)
count_data <- apply(count_data,2,function(x) {storage.mode(x) <- 'integer'; x})

# generate color code for group factors
x <- sub("^.*_.*_.*[1-9]", "\\1", colnames(count_data)) # 9 groups
l2cols <- c("orangered","black","gold","darkgreen","violetred","royalblue2","chartreuse3",
"slateblue4")[as.integer(factor(x, levels = c("ctr","ref.blood","ref.endo","ref.myo","tbx18","tcf21","wt1b",
"tmkg")))]

```

```

# fit error models (estimating 8 subpopulations)
knn <- knn.error.models(count_data, k = ncol(count_data)/8, n.cores = 1, min.count.threshold = 1,
min.nonfailed = 5, max.model.plots = 10)
# exclude cells that don't show positive correlation with the expected expression magnitudes (very poor fits)
valid.cells <- knn$corr.a > 0
table(valid.cells)
knn <- knn[valid.cells, ]

# normalize variance for expected technical and biological noise
windows()
varinfo <- pagoda.varnorm(knn, counts = count_data, trim = 3/ncol(count_data), max.adj.var = 5, n.cores =
1, plot = TRUE)
# control for sequencing depth
varinfo <- pagoda.subtract.aspect(varinfo, colSums(count_data[, rownames(knn)]>0))

# create a list of GO terms
# get ids for gene names
ids <- unlist(lapply(mget(rownames(count_data), org.Dr.egALIAS2EG, ifnotfound = NA), function(x) x[1]))
rids <- names(ids)
names(rids) <- ids
# get GO terms
gos.interest <- unique(c(ls(org.Dr.egGO2ALLEGS)[1:500]))
go.env <- lapply(mget(gos.interest, org.Dr.egGO2ALLEGS), function(x) as.character(na.omit(rids[x])))
go.env <- clean.gos(go.env) # remove GOs with too few or too many genes
go.env <- list2env(go.env) # convert to an environment
# test
class(go.env)
head(ls(go.env)) # Look at gene set names
head(get(ls(go.env)[1], go.env)) # Look at one gene set

# first round of analysis
# calculate weighted PC magnitude
set.seed(0)
pwpca <- pagoda.pathway.wPCA(varinfo, go.env, n.components = 1, n.cores = 1)
# calculate significance of overdispersion for each GO (Z-score of 1.96 corresponds to pvalue of 0.05)
windows()
df <- pagoda.top.aspects(pwpca, return.table = TRUE, plot = TRUE, z.score = 1.96)
head(df)

# calculate de novo overdispersion
set.seed(0)
clpca <- pagoda.gene.clusters(varinfo, trim = 7.1/ncol(varinfo$mat), n.clusters = 50, n.cores = 1, plot =
TRUE)
windows()
df <- pagoda.top.aspects(pwpca, clpca, return.table = TRUE, plot = TRUE, z.score = 1.96)
head(df)

# get full info on the top aspects
tam <- pagoda.top.aspects(pwpca, clpca, n.cells = NULL, z.score = qnorm(0.01/2, lower.tail = FALSE))
# determine overall cell clustering
hc <- pagoda.cluster.cells(tam, varinfo)
col.cols <- rbind(groups = cutree(hc, 9))

# reduce to unique clusters
windows()
tamr <- pagoda.reduce.loading.redundancy(tam, pwpca, clpca)

```

```

tamr2 <- pagoda.reduce.redundancy(tamr, distance.threshold = 0.9, plot = TRUE, cell.clustering = hc,
labRow = NA, labCol = NA, box = TRUE, margins = c(0.5, 0.5), trim = 0, col.cols = col.cols)
windows()
pagoda.view.aspects(tamr2, cell.clustering = hc, box = TRUE, labCol = NA, margins = c(0.5, 20), col.cols =
rbind(l2cols))

# cell cycle correction
# get cell cycle GO terms and view the top genes
windows()
cc.pattern <-
pagoda.show.pathways(c("GO:0000070", "GO:0000086", "GO:0000278", "GO:0000819", "GO:0000910", "GO:
0000280", "GO:0000793"), varinfo, go.env, show.cell.dendrogram = TRUE, cell.clustering = hc,
showRowLabels = TRUE)
# exclude cell cycle GO terms
varinfo.cc <- pagoda.subtract.aspect(varinfo, cc.pattern)

# second round of analysis (cell cycle corrected)
# calculate weighted PC magnitude
set.seed(0)
pw pca.cc <- pagoda.pathway.wPCA(varinfo.cc, go.env, n.components = 1, n.cores = 1)
# calculate significance of overdispersion for each GO (Z-score of 1.96 corresponds to pvalue of 0.05)
windows()
df.cc <- pagoda.top.aspects(pw pca.cc, return.table = TRUE, plot = TRUE, z.score = 1.96)
head(df.cc)

# calculate de novo overdispersion
set.seed(0)
clpca.cc <- pagoda.gene.clusters(varinfo.cc, trim = 7.1/ncol(varinfo.cc$mat), n.clusters = 50, n.cores = 1,
plot = TRUE)
windows()
df.cc <- pagoda.top.aspects(pw pca.cc, clpca.cc, return.table = TRUE, plot = TRUE, z.score = 1.96)
head(df.cc)

# get full info on the top aspects
tam.cc <- pagoda.top.aspects(pw pca.cc, clpca.cc, n.cells = NULL, z.score = qnorm(0.01/2, lower.tail =
FALSE))
# determine overall cell clustering (9 clusters)
hc.cc <- pagoda.cluster.cells(tam.cc, varinfo.cc)
col.cols <- rbind(groups = cutree(hc.cc, 9))

# reduce to unique clusters
windows()
tamr.cc <- pagoda.reduce.loading.redundancy(tam.cc, pw pca.cc, clpca.cc)
tamr2.cc <- pagoda.reduce.redundancy(tamr.cc, distance.threshold = 0.9, plot = TRUE, cell.clustering =
hc.cc, labCol = NA, box = TRUE, margins = c(0.5, 17.5), trim = 0, col.cols = rbind(l2cols), cexRow=1.0)
windows()
pagoda.view.aspects(tamr2.cc, cell.clustering = hc.cc, box = TRUE, labCol = NA, labRow = NA, margins =
c(0.5, 0.5), col.cols = col.cols)

# compile a browsable app, showing top clusters
app <- make.pagoda.app(tamr2.cc, tam.cc, varinfo.cc, go.env, pw pca.cc, clpca.cc, col.cols = col.cols,
cell.clustering <- hc.cc, title = "Zebrafish main epicardium 5dpf") #, embedding = NULL
app <- readRDS("pagoda_app_cc_352_cells")

# show app in the browser (port 1468)
show.app(app, "pollen", browse = TRUE, port = 1468)

```

```

# perform differential gene expression analysis
# use pagoda clustering to identify cell population to subset for
# subset for epicardial cell cluster 1
subset1 <- c(which(col.cols==1))
# subset for everything else
subset2 <- c(which(col.cols!=1))

# subset for epicardial cell cluster 2
subset3 <- c(which(col.cols==9))
# subset for everything else
subset4 <- c(which(col.cols!=9))

# subset for epicardial cell cluster 3
subset5 <- c(which(col.cols==2))
# subset for everything else
subset6 <- c(which(col.cols!=2))

# subset for mesenchymal cells
subset7 <- c(which(col.cols==5))
# subset for everything else
subset8 <- c(which(col.cols!=5))

# subset for erythroid haematopoietic cells
subset9 <- c(which(col.cols==7))
# subset for everything else
subset10 <- c(which(col.cols!=7))

# subset for myocardial cells
subset11 <- c(which(col.cols==8))
# subset for everything else
subset12 <- c(which(col.cols!=8))

# subset for neural cells
subset13 <- c(which(col.cols==4))
# subset for everything else
subset14 <- c(which(col.cols!=4))

# subset for myeloid haematopoietic cells
subset15 <- c(which(col.cols==6))
# subset for everything else
subset16 <- c(which(col.cols!=6))

# subset for rest
subset17 <- c(which(col.cols==3))
# subset for everything else
subset18 <- c(which(col.cols!=3))

# factor determining populations to compare
sg <- factor(colnames(count_data)%in%colnames(count_data), levels = c("pop1", "pop2"))
sg[subset1] <- "pop1"
sg[subset2] <- "pop2"

# the group factor should be named accordingly
x <- sub("^.*_.*_.*_.*[1-9]", "\\1", colnames(count_data))
names(sg) <- colnames(count_data)
table(sg,x)

```

```

# calculate error models
o.ifm <- scde.error.models(counts = count_data, n.cores = 1, threshold.segmentation = TRUE,
save.crossfit.plots = FALSE, save.model.plots = FALSE, verbose = 1)
head(o.ifm)
# exclude cells that don't show positive correlation with the expected expression magnitudes (very poor fits)
valid.cells <- o.ifm$corr.a > 0
table(valid.cells)
o.ifm <- o.ifm[valid.cells, ]

# estimate gene expression prior
o.prior <- scde.expression.prior(models = o.ifm, counts = count_data, length.out = 400, show.plot = FALSE)

# define two groups of cells
groups <- factor(rownames(o.ifm)%in%rownames(o.ifm), levels = c("pop1", "pop2"))
groups[subset1] <- "pop1"
groups[subset2] <- "pop2"
names(groups) <- row.names(o.ifm)

# run differential expression test on all genes and save as .csv
ediff.batch <- scde.expression.difference(o.ifm, count_data, o.prior, groups = groups, batch = x,
n.randomizations = 100, n.cores = 1, return.posteriors = TRUE, verbose = 1)
dge_results <- cbind(rownames(ediff.batch$results),ediff.batch$results)
colnames(dge_results)[1] <- "gene_id"
rownames(dge_results) <- NULL
dge_results <- dge_results[order(-dge_results$Z),]
write.csv(dge_results, file="Pagoda_dge_epi1_vs_rest_352_cells.csv", row.names=FALSE)

```

## Plots

```

# quality control plots
sce_qcplot <- as.data.frame(sce$total_counts/1e6)
sce_qcplot <- cbind(sce_qcplot, sce$total_features/1e3, sce$pct_counts_feature_controls_Mt,
sce$pct_counts_feature_controls_ERCC)
colnames(sce_qcplot) <- c("Total_counts", "Total_features", "Pct_features_MT", "Pct_features_ERCC")

# plot library sizes
dev.new(width=9, height=9)
ggplot(data=sce_qcplot, aes(sce_qcplot$Total_counts)) +
geom_histogram(binwidth=0.055, col="black", fill="grey70") +
theme_minimal(base_size = 30) +
theme(
  axis.text.x = element_text(family = "Arial", size = 30, colour = "black",face = "plain", angle=0, hjust=0.5),
  axis.text.y = element_text(family = "Arial", size = 30, colour = "black",face = "plain"),
  axis.title.x = element_text(family = "Arial", size = 32, vjust = -2),
  axis.title.y = element_text(family = "Arial", size = 32, vjust = 0),
  axis.line.x = element_line(colour = "black", size = 2),
  axis.line.y = element_line(colour = "black", size = 2),
  axis.ticks = element_line(colour = "black", size = 2),
  legend.position = "none",
  plot.margin = margin(0.5, 0.5, 0.5, 0.5, "cm")) +
scale_y_continuous(name="Number of cells", limits=c(0, 45), breaks=c(0,10,20,30,40)) +
scale_x_continuous(name="Library size in million", limits=c(0,2.5), breaks=c(0,0.5,1,1.5,2,2.5))

# plot library complexities
dev.new(width=9, height=9)

```

```

ggplot(data=sce_qcplot, aes(sce_qcplot$Total_features)) +
geom_histogram(binwidth=0.15, col="black", fill="grey70") +
theme_minimal(base_size = 30) +
theme(
  axis.text.x = element_text(family = "Arial", size = 30, colour = "black",face = "plain", angle=0, hjust=0.5),
  axis.text.y = element_text(family = "Arial", size = 30, colour = "black",face = "plain"),
  axis.title.x = element_text(family = "Arial", size = 32, vjust = -2, hjust=0.8),
  axis.title.y = element_text(family = "Arial", size = 32, vjust = 0),
  axis.line.x = element_line(colour = "black", size = 2),
  axis.line.y = element_line(colour = "black", size = 2),
  axis.ticks = element_line(colour = "black", size = 2),
  legend.position = "none",
  plot.margin = margin(0.5, 1, 0.5, 0.5, "cm")) +
scale_y_continuous(name="Number of cells", limits=c(0, 35), breaks=c(0,10,20,30)) +
scale_x_continuous(name="Expressed features in thousand", limits=c(0,6), breaks=seq(0,6,by=1))

```

```
# plot proportion of mitochondrial reads
```

```

dev.new(width=9, height=9)
ggplot(data=sce_qcplot, aes(sce_qcplot$Pct_features_MT)) +
geom_histogram(binwidth=0.3, col="black", fill="grey70") +
theme_minimal(base_size = 30) +
theme(
  axis.text.x = element_text(family = "Arial", size = 30, colour = "black",face = "plain", angle=0, hjust=0.5),
  axis.text.y = element_text(family = "Arial", size = 30, colour = "black",face = "plain"),
  axis.title.x = element_text(family = "Arial", size = 32, vjust = -2, hjust=0.8),
  axis.title.y = element_text(family = "Arial", size = 32, vjust = 0),
  axis.line.x = element_line(colour = "black", size = 2),
  axis.line.y = element_line(colour = "black", size = 2),
  axis.ticks = element_line(colour = "black", size = 2),
  legend.position = "none",
  plot.margin = margin(0.5, 1, 0.5, 0.5, "cm")) +
scale_y_continuous(name="Number of cells", limits=c(0, 35), breaks=c(0,10,20,30)) +
scale_x_continuous(name="Mitochondrial features [%]", limits=c(0,12), breaks=seq(0,12,by=2))

```

```
# plot proportion of ERCC reads
```

```

dev.new(width=9, height=9)
ggplot(data=sce_qcplot, aes(sce_qcplot$Pct_features_ERCC)) +
geom_histogram(binwidth=1.15, col="black", fill="grey70") +
theme_minimal(base_size = 30) +
theme(
  axis.text.x = element_text(family = "Arial", size = 30, colour = "black",face = "plain", angle=0, hjust=0.5),
  axis.text.y = element_text(family = "Arial", size = 30, colour = "black",face = "plain"),
  axis.title.x = element_text(family = "Arial", size = 32, vjust = -2, hjust=0.8),
  axis.title.y = element_text(family = "Arial", size = 32, vjust = 0),
  axis.line.x = element_line(colour = "black", size = 2),
  axis.line.y = element_line(colour = "black", size = 2),
  axis.ticks = element_line(colour = "black", size = 2),
  legend.position = "none",
  plot.margin = margin(0.5, 1, 0.5, 0.5, "cm")) +
scale_y_continuous(name="Number of cells", limits=c(0, 35), breaks=c(0,10,20,30)) +
scale_x_continuous(name="Spike in features [%]", limits=c(0,50), breaks=seq(0,50,by=10))

```

```
# plot gene expression levels
```

```

sce_qcplot_genes <- as.data.frame(log10(ave.counts))
colnames(sce_qcplot_genes) <- c("Log10_expression")
dev.new(width=9, height=9)
ggplot(data=sce_qcplot_genes, aes(sce_qcplot_genes$Log10_expression)) +

```

```

geom_histogram(binwidth=0.08, col="black", fill="grey70") +
theme_minimal(base_size = 30) +
theme(
  axis.text.x = element_text(family = "Arial", size = 30, colour = "black",face = "plain", angle=0, hjust=0.5),
  axis.text.y = element_text(family = "Arial", size = 30, colour = "black",face = "plain"),
  axis.title.x = element_text(family = "Arial", size = 32, vjust = -2, hjust=0.8),
  axis.title.y = element_text(family = "Arial", size = 32, vjust = 0),
  axis.line.x = element_line(colour = "black", size = 2),
  axis.line.y = element_line(colour = "black", size = 2),
  axis.ticks = element_line(colour = "black", size = 2),
  legend.position = "none",
  plot.margin = margin(0.5, 1, 0.5, 0.5, "cm")) +
scale_y_continuous(name="Number of features", limits=c(0, 850), breaks=seq(0,800,by=200)) +
scale_x_continuous(name="Mean feature count [log10]", limits=c(-3,4.5), breaks=seq(-3,4,by=1)) +
geom_vline(xintercept=log10(cutoff), col=2, lwd=1.5, lty=2)

# convert raw counts to FPKM (using edgeR)
rpkm_meta <-
cbind(as.data.frame(featureData(sce)$gene_name[1:20862]),as.data.frame(featureData(sce)$gene_length[
1:20862]))
colnames(rpkm_meta) <- c("gene_name","gene_length")
y_norm <- DGEList(counts=count_data, genes=rpkm_meta)
RPKM <- rpkm(y_norm, y_norm$genes$gene_length)
RPKM_log <- as.data.frame(log2(RPKM+1))

# draw t-SNE plot with cell origin color coded
# recalculate clustering distance, we'll need to specify return.details=T
cell.clustering.cc <-
pagoda.cluster.cells(tam.cc,varinfo.cc,include.aspects=TRUE,verbose=TRUE,return.details=T)
# clustering cells based on 2376 genes and 99 aspect patterns

# calculate t-SNE, fix the seed to ensure reproducible results
set.seed(0);
tSNE.pagoda.cc <- Rtsne(cell.clustering.cc$distance,is_distance=T,initial_dims=100,perplexity=10)

# plot
dev.new(width=10, height=9)
par(mfrow=c(1,1), mar = c(7.5,7.5,1.0,0.5), mgp = c(6.3,1.0,0.5), cex = 1.1, bty='n');
plot(tSNE.pagoda.cc$Y, col=adjustcolor(l2cols,alpha=.75), cex=1.2, pch=19, xaxt = 'n', yaxt='n',
  xlab="Dim 1", ylab="", cex.lab=2.5)
axis(1, at=seq(-40,30, by=10), cex.axis=2.5, las=2, lwd=5)
axis(2, at=seq(-30,30, by=10), cex.axis=2.5, las=1, lwd=5)
mtext("Dim 2", side=2, line=5.0, cex=2.5)

# draw t-SNE plot of cell clusters
# recalculate clustering distance, we'll need to specify return.details=T
cell.clustering.cc <-
pagoda.cluster.cells(tam.cc,varinfo.cc,include.aspects=TRUE,verbose=TRUE,return.details=T)
# clustering cells based on 2376 genes and 99 aspect patterns

# calculate t-SNE, fix the seed to ensure reproducible results
labels <- c("blue", "darkblue", "cadetblue4", "cornflowerblue", "dodgerblue3", "black", "navy",
"darkgreen", "steelblue")[as.integer(as.data.frame(col.cols))]
set.seed(0);
tSNE.pagoda.cc <- Rtsne(cell.clustering.cc$distance,is_distance=T,initial_dims=100,perplexity=10)

```

```

# plot
dev.new(width=10, height=9)
par(mfrow=c(1,1), mar = c(7.5,7.5,1.0,0.5), mgp = c(6.3,1.0,0.5), cex = 1.1, bty='n');
plot(tSNE.pagoda.cc$Y, col="grey70", cex=2.2, pch=19, xaxt = 'n', yaxt='n',
     xlab="Dim 1", ylab="", cex.lab=2.5)
axis(1, at=seq(-40,30, by=10), cex.axis=2.5, las=2, lwd=5)
axis(2, at=seq(-30,30, by=10), cex.axis=2.5, las=1, lwd=5)
mtext("Dim 2", side=2, line=5.0, cex=2.5)
text(tSNE.pagoda.cc$Y[,1],tSNE.pagoda.cc$Y[,2],col=cols,col=labels,cex=1.05)

# generate legend for cell origin
plot(c(0, 1000), c(0, 1000), type="n", xlab="", ylab="")
legend("center", legend=c("wildtype","Tg(tcf21:H2B-Dendra2) [+]","Tg(tbx18:H2B-Dendra2)
[+]","Tg(wt1b:H2B-Dendra2) [+]","Tg(myl7:EGFP) [+]","Tg(kdrl:EGFP) [+]","Tg(gata1a:dsRed)
[+]","Tg(tcf21:dsRed;myl7:EGFP;kdrl:EGFP;gata1a:dsRed) [-]"),
fill=c("orangered", "royalblue2", "violetred", "chartreuse3", "darkgreen", "gold", "black", "slateblue4"),
border="white", bty="n", cex=1.2
)

# draw t-SNE plot with FPKM expression values
g <- "tcf21" #choose a gene
l <- RPKM_log[g[1],]
mi <- min(l,na.rm=TRUE)
ma <- max(l,na.rm=TRUE)
ColorRamp <- colorRampPalette(rev(brewer.pal(n = 7,name = "RdYlBu")))(100)
ColorLevels <- seq(mi, ma, length=length(ColorRamp))
v <- round((l - mi)/(ma - mi)*99 + 1,0)

# recalculate clustering distance, we'll need to specify return.details=T
cell.clustering.cc <-
pagoda.cluster.cells(tam.cc,varinfo.cc,include.aspects=TRUE,verbose=TRUE,return.details=T)
# calculate t-SNE, fix the seed to ensure reproducible results
set.seed(0);
tSNE.pagoda.cc <- Rtsne(cell.clustering.cc$distance,is_distance=T,initial_dims=100,perplexity=10)

# plot
dev.new(width=11, height=9)
layout(matrix(data=c(1,3,2,4), nrow=2, ncol=2), widths=c(9,1,9,1), heights=c(9,0,1,1))
par(mar = c(7.5,7.5,2.5,3), mgp = c(6.3,1.0,0.5), cex = 1.1, bty='n');
plot(tSNE.pagoda.cc$Y, col="grey70", cex=1.2, pch=19, xaxt = 'n', yaxt='n',
     xlab="Dim 1", ylab="", cex.lab=2.5)
axis(1, at=seq(-40,30, by=10), cex.axis=2.5, las=2, lwd=5)
axis(2, at=seq(-30,30, by=10), cex.axis=2.5, las=1, lwd=5)
mtext("Dim 2", side=2, line=5.0, cex=2.5)
points(tSNE.pagoda.cc$Y,col=adjustcolor(ColorRamp[as.integer(v)],alpha=.75),pch=8,cex=.9)
par(mar = c(7.5,1.0,3.0,1.5))
image(1, ColorLevels,
matrix(data=ColorLevels, ncol=length(ColorLevels),nrow=1), col=ColorRamp,
     xlab="", ylab="", main="", cex.main=.8, adj=.2, xaxt="n", yaxt='n')
axis(2, cex.axis=2.0, las=1, lwd=4)
mtext(expression(Log[2]~"FPKM"), side=3, line=0.7, adj=.8, cex=2.0)
layout(1)

# draw heatmap using FPKM of top differentially expressed genes in each cluster
RPKM_log <- RPKM_log[,hc.cc$order]

```

```

dev.new(width=8, height=10)
pheatmap(RPKM_log[c("nmnat2", "gpm6ab", "tbr1b", "nsfa", "myt1la",
"cahz", "slc4a1a", "dmtn", "epb41b", "alas2",
"grap2a", "rgs13", "CABZ01070258.1", "blf", "BX004828.2",
"adma", "podxl", "tspan12", "sema3e", "cemip", "ppfibp1a", "ADAMTSL5", "aldh1a2", "irf8", "ADM_2_of_2.", "r
spo3", "myrf", "jam2b", "gstm.3", "CU984579.1",
"lox", "myh11a", "htra1a", "mylka", "acta2", "acana", "fbln5", "cfd", "lox1", "pi15b", "elnb", "emid1", "si.dkey.164
f24.2", "mylka.1", "LTBP4", "cd248a",
"cldn11a", "bscl2l", "cxcl12a", "farp10a", "CABZ01073190.1", "bgnb", "cntfr", "artnb", "ANPEP_1_of_5.", "nr1h
5", "scube1", "si.dkey.12l12.1", "si.ch211.251b21.1", "rgs5b", "prss35",
"myl7", "cmlc1", "gapdh", "zgc.101560", "pdkmb",
"abi3bbp", "cyp26b1", "postna", "angptl7", "thbs1b", "matn4", "pdgfra", "sox7", "tie1", "kdr",
"mbpa", "cldnk", "sox10", "msi1", "sox2", "sox3", "sox19a", "foxf1", "barx1", "foxf2a"
)],),
cluster_rows=FALSE, show_rownames=TRUE, cluster_cols=FALSE, show_colnames=FALSE, cex=.7)

# draw heatmap with FPKM
# epicardial markers
RPKM_log <- RPKM_log[,hc.cc$order]
dev.new(width=14, height=10)
pheatmap(RPKM_log[c("tcf21", "tbx18", "wt1b", "wt1a", "scxa", "sema3d", "aldh1a2", "fn1a", "fn1b",
"igf2b",
"snai1a", "snai1b", "snai2", "twist1a", "twist1b", "nfatc1", "ctsk",
"rbp4", "cxcl18b", "krt94", "cyr61"
)],),
cluster_rows=FALSE, show_rownames=TRUE, cluster_cols=FALSE, show_colnames=FALSE,
gaps_row = c(10,17), fontsize = 20)

# epicardial cell cluster 1
RPKM_log <- RPKM_log[,hc.cc$order]
dev.new(width=14, height=10)
pheatmap(RPKM_log[c("adma", "calcr1a", "ramp2",
"abi3bbp", "fli1a", "etv2", "sox7", "cdh5", "tie1", "kdr", "kdr1", "dll4"
)],),
cluster_rows=FALSE, show_rownames=TRUE, cluster_cols=FALSE, show_colnames=FALSE,
gaps_row = c(3), fontsize = 20)

# epicardial cluster 2
RPKM_log <- RPKM_log[,hc.cc$order]
dev.new(width=14, height=10)
pheatmap(RPKM_log[c("mylka", "acta2", "myh11a", "tagln",
"lox", "elnb", "acana", "fbln5",
"wnt11r", "fzd7a", "fzd7b", "fzd8a", "fzd9a", "ror2", "daam1a", "rhoaa", "alcama",
"wnt11", "pdgfra", "hand2", "gata4", "crip2", "sox9a", "erbb3a", "tbx2a"
)],),
cluster_rows=FALSE, show_rownames=TRUE, cluster_cols=FALSE, show_colnames=FALSE,
gaps_row = c(4,8,17), fontsize = 20)

# epicardial cell cluster 3
RPKM_log <- RPKM_log[,hc.cc$order]
dev.new(width=14, height=10)
pheatmap(RPKM_log[c("cxcl12a", "cxcr4a", "cxcr4b",
"lmo2", "HHEX", "spi1a", "spi1b", "lcp1", "coro1a", "ptprc", "ptpn6", "cxcr3.2", "csf1rb", "mpx",
"ccl25b", "ccr9a"
)],),
cluster_rows=FALSE, show_rownames=TRUE, cluster_cols=FALSE, show_colnames=FALSE,
gaps_row = c(3,14), fontsize = 20)

```

```

# transcription factors
RPKM_log <- RPKM_log[,hc.cc$order]
dev.new(width=7, height=4)
pheatmap(RPKM_log[c("tcf21", "tbx18", "wt1b", "gata4", "gata5", "gata6", "myrf", "irf8", "podxl"),],
cluster_rows=FALSE, show_rownames=TRUE, cluster_cols=FALSE, show_colnames=FALSE,
gaps_row = 8, fontsize = 20)

dev.new(width=7, height=3.5)
pheatmap(RPKM_log[c("tbx18", "hand2", "foxc1a", "tbx20", "sox9a", "sox5", "sox6", "lox"),],
cluster_rows=FALSE, show_rownames=TRUE, cluster_cols=FALSE, show_colnames=FALSE,
gaps_row = 7, fontsize = 20)

dev.new(width=7, height=3.5)
pheatmap(RPKM_log[c("tbx18", "hand2", "foxc1a", "tbx20", "sox9a", "sox5", "sox6", "elnb"),],
cluster_rows=FALSE, show_rownames=TRUE, cluster_cols=FALSE, show_colnames=FALSE,
gaps_row = 7, fontsize = 20)

```

### **Bulk transcriptomic analysis**

```

library(DESeq2)
if (!require("DESeq2")) {
source("http://bioconductor.org/biocLite.R")
biocLite("DESeq2")
library(DESeq2)
}

library(biomaRt)
if (!require("biomaRt")) {
install.packages("biomaRt", dependencies=TRUE)
library(biomaRt)
}

library(pheatmap)
if (!require("pheatmap")) {
install.packages("pheatmap", dependencies=TRUE)
library(pheatmap)
}

library(RColorBrewer)
if (!require("RColorBrewer")) {
install.packages("RColorBrewer", dependencies=TRUE)
library(RColorBrewer)
}

library(ggplot2)
if (!require("ggplot2")) {
install.packages("ggplot2", dependencies=TRUE)
library(ggplot2)
}

library(ggrepel)
if (!require("ggrepel")) {
install.packages("ggrepel", dependencies=TRUE)
library(ggrepel)
}

```

```

}

library(calibrate)
if (!require("calibrate")) {
install.packages("calibrate", dependencies=TRUE)
library(calibrate)
}

library(edgeR)
if (!require("edgeR ")) {
source("http://bioconductor.org/biocLite.R")
biocLite("edgeR ")
library(edgeR)
}

library(statmod)
if (!require("statmod ")) {
source("http://bioconductor.org/biocLite.R")
biocLite("statmod ")
library(statmod)
}

library(scater)
if (!require("scater")) {
source("http://bioconductor.org/biocLite.R")
biocLite("scater")
library(scater)
}

library(scrn)
if (!require("scrn")) {
source("http://bioconductor.org/biocLite.R")
biocLite("scrn")
library(scran)
}

# read in bulk data
all_data <- read.delim("D:/Oxford/RNA seq/Seq data processing/RNA bulk tcf21 tbx18 wt1b
5dpf/featureCounts.final.txt", header=TRUE, sep="\t", stringsAsFactors=FALSE, check.names=FALSE)
rownames(all_data) <- all_data$gene_id
all_data <- all_data[grepl("ENS",rownames(all_data)),]
head(all_data)
# subset for counts to compare
count_data <- data.frame(c(all_data[,4:6],all_data[,10:12]))
rownames(count_data) <- all_data[,1]
head(count_data)

# generate DESeq object
sample_info <- data.frame(c(rep("tcf21",3),rep("ctr",3)))
batch <- factor(c(1,2,3,1,2,3))
sample_info <- cbind(sample_info,batch)
rownames(sample_info) <- colnames(count_data)
colnames(sample_info) <- c("condition","batch")
sample_info

dds <- DESeqDataSetFromMatrix(countData=count_data, colData=sample_info, design= ~ batch +
condition)

```

```

dds
mcols(dds) <- DataFrame(mcols(dds), all_data[,2])
colnames(mcols(dds)) <- c("gene_name")
head(mcols(dds))

# filter out rows with only 0 or 1 counts
dds <- dds[rowSums(counts(dds))>1, ]

# ensure correct comparison
dds$condition <- relevel(dds$condition,ref="ctr")

# run DE analysis
dds <- DESeq(dds)
res <- results(dds, alpha=0.05)
res
summary(res)

# save results as .csv
resSig <- subset(res, padj<0.05)
row_resSig <- data.frame(rownames(resSig))
id_name <- merge(all_data, row_resSig, by.x="gene_id", by.y="rownames.resSig.")
resSig2 <- cbind(id_name,resSig)
# get gene descriptions
bm <- useMart("ensembl")
bm <- useDataset("drerio_gene_ensembl", mart=bm)
attributes <- listAttributes(bm)
eg2ds <- getBM(mart=bm, attributes=c('ensembl_gene_id','description'), filters='ensembl_gene_id',
values=row_resSig)
head(eg2ds)
name_count_resSig <- merge(eg2ds,resSig2, by.x="ensembl_gene_id", by.y="gene_id")
name_count_resSig <- name_count_resSig[order(name_count_resSig$padj),]
head(name_count_resSig)
write.csv(name_count_resSig, file="DESeq2_DEG_new_tcf21_vs_wt1b_5dpf.csv", row.names=FALSE)

```

## Plots

```

# calculate quality metrics using the scater package
metadata <- AnnotatedDataFrame(all_data[1:3])
sce_batch <- newSCESet(countData=all_data[,4:ncol(all_data)], featureData=metadata)
dim(sce_batch)

sce_batch <- calculateQCMetrics(sce_batch)
sce_batch_qcplot <- as.data.frame(sce_batch$total_counts/1e6)
sce_batch_qcplot <- cbind(sce_batch_qcplot, sce_batch$total_features/1e3)
colnames(sce_batch_qcplot) <- c("Total_counts", "Total_features")
rownames(sce_batch_qcplot) <- colnames(sce_batch)
sce_batch_qcplot$batch <- rep(c(16,17,15),6)
sce_batch_qcplot$condition <- c(rep("tcf21",3),rep("tcf21_ctr",3),rep("tbx18",3),rep("tbx_ctr",3),
rep("wt1b",3),rep("wt1b_ctr",3))
sce_batch_qcplot$color <- c(rep("royalblue2",3),rep("blue3",3),rep("violetred",3),rep("magenta4",3),
rep("chartreuse3",3),rep("darkgreen",3))

# plot library sizes vs complexities
dev.new(width=9, height=9)

```

```

ggplot(data=sce_batch_qcplot, aes(x=sce_batch_qcplot$Total_counts,
y=sce_batch_qcplot$Total_features)) +
geom_point(size=5, col=sce_batch_qcplot$color, shape=sce_batch_qcplot$batch) +
theme_minimal(base_size = 30) +
theme(
  axis.text.x = element_text(size = 30, colour = "black",face = "plain", angle=0, hjust=0.5),
  axis.text.y = element_text(size = 30, colour = "black",face = "plain"),
  axis.title.x = element_text(size = 32, vjust = -2),
  axis.title.y = element_text(size = 32, vjust = 0),
  axis.line.x = element_line(colour = "black", size = 2),
  axis.line.y = element_line(colour = "black", size = 2),
  axis.ticks = element_line(colour = "black", size = 2),
  legend.position = "none",
  plot.margin = margin(0.5, 0.5, 0.5, 0.5, "cm")) +
scale_y_continuous(name="Features in thousand", limits=c(20, 30), breaks=seq(20,30,by=2)) +
scale_x_continuous(name="Library size in million", limits=c(4,12), breaks=seq(4,12,by=2))

# generate legend for cell origin
plot(c(0, 1000), c(0, 1000), type="n", xlab="", ylab="")
legend("center", legend=c("Tg(tcf21:H2B-Dendra2) [+]","Tg(tcf21:H2B-Dendra2) [-]","Tg(tbx18:H2B-
Dendra2) [+]","Tg(tbx18:H2B-Dendra2) [-]","Tg(wt1b:H2B-Dendra2) [+]","Tg(wt1b:H2B-Dendra2) [-]"),
fill=c("royalblue2", "blue3", "violetred", "magenta4", "chartreuse3", "darkgreen"),
border="white", bty="n", cex=1.2
)

# normalize count data with variance stabilizing transformation
vsd <- varianceStabilizingTransformation(dds, blind=FALSE)
head(assay(vsd),3)

# generate PCA plot
data <- plotPCA(vsd, intgroup=c("condition","batch"), returnData=TRUE)
data$batch <- rep(c(16,17,15),6)
data$condition <- c(rep("tcf21",3),rep("tcf21_ctr",3),rep("tbx18",3),rep("tbx_ctr",3),
rep("wt1b",3),rep("wt1b_ctr",3))
data$color <- c(rep("royalblue2",3),rep("blue3",3),rep("violetred",3),rep("magenta4",3),
rep("chartreuse3",3),rep("darkgreen",3))
percentVar <- round(100 * attr(data,"percentVar"))

dev.new(width=10, height=9)
ggplot(data=data, aes(x=data$PC1, y=data$PC2)) +
geom_point(size=5, col=data$color, shape=data$batch) +
theme_minimal(base_size = 30) +
theme(
  axis.text.x = element_text(size = 30, colour = "black",face = "plain", angle=0, hjust=0.5),
  axis.text.y = element_text(size = 30, colour = "black",face = "plain"),
  axis.title.x = element_text(size = 32, vjust = -2, hjust=0.8),
  axis.title.y = element_text(size = 32, vjust = 0),
  axis.line.x = element_line(colour = "black", size = 2),
  axis.line.y = element_line(colour = "black", size = 2),
  axis.ticks = element_line(colour = "black", size = 2),
  legend.position = "none",
  plot.margin = margin(0.5, 1, 0.5, 0.5, "cm")) +
scale_y_continuous(name=paste0("PC2: ", percentVar[2], "% variance"), limits=c(-20,50),
breaks=seq(0,50,by=10)) +

```

```
scale_x_continuous(name=paste0("PC1: ", percentVar[1], "% variance"), limits=c(-75,75), breaks=seq(-75,75,by=30))
```

```
# produce heatmap of enriched genes from single cell clusters
# convert raw counts to FPKM (using edgeR)
rpkm_meta <- cbind(as.data.frame(all_data$gene_name),as.data.frame(all_data$gene_length))
colnames(rpkm_meta) <- c("gene_name","gene_length")
y_norm <- DGEList(counts=all_data[,4:ncol(all_data)], genes=rpkm_meta)
RPKM <- rpkm(y_norm, y_norm$genes$gene_length)
RPKM_log <- as.data.frame(log2(RPKM+1))
rownames(RPKM_log) <- make.names(all_data$gene_name, unique=TRUE)
```

```
# epicardial markers
dev.new(width=9.5, height=3.5)
pheatmap(RPKM_log[c("tcf21","tbx18","wt1b","fn1b","aldh1a2",
"and1","and2"),],
cluster_rows=FALSE, show_rownames=TRUE, cluster_cols=FALSE, show_colnames=FALSE,
gaps_row = c(5), fontsize = 20)
```

```
# markers expressed in epicardial cell clusters 1-3
dev.new(width=14, height=7)
pheatmap(RPKM_log[c("adma","podxl","jam2b","irf8","myrf",
"mylka","acta2","lox","elnb","wnt11r",
"cldn11a","cxcl12a","nr1h5","bscl2l","fabp10a"
)],,
cluster_rows=FALSE, show_rownames=TRUE, cluster_cols=FALSE, show_colnames=FALSE,
gaps_row = c(5,10), fontsize = 20)
```

```
# volcano plot of pvalue vs log2FC
resName <- data.frame(cbind(mcols(dds), res))
resName$up_down <- resName$log2FoldChange / abs(resName$log2FoldChange)
resName <- resName[order(resName$up_down,resName$padj,decreasing=FALSE),]
resName$Significant <- ifelse(resName$padj < 0.05, "FDR < 0.05", "Not Sig")
dim(resName)
```

```
up_start <- length(which(resName$up_down == -1)) + 1
label_data <- rbind(resName[1:10,],resName[up_start:(up_start + 9),])
label_exclude <- grep("-", label_data$gene_name)
# exclude due to font Helvetica "-" not displaying properly in PDF
label_data <- label_data[-label_exclude,]
```

```
dev.new(width=10, height=9)
ggplot(resName, aes(x = log2FoldChange, y = -log10(padj))) +
  geom_point(aes(color = Significant)) +
  scale_color_manual(values = c("red", "grey")) +
  theme_minimal(base_size = 23) +
  theme(
  axis.text.x = element_text(size = 23, colour = "black",face = "plain", angle=0, hjust=0.5),
  axis.text.y = element_text(size = 23, colour = "black",face = "plain"),
  axis.title.x = element_text(size = 24, vjust = -2, hjust=0.5),
  axis.title.y = element_text(size = 24, vjust = 0),
  axis.line.x = element_line(colour = "black", size = 1.5),
  axis.line.y = element_line(colour = "black", size = 1.5),
  axis.ticks = element_line(colour = "black", size = 1.5),
  legend.position = "none",
```

```

    plot.margin = margin(0.5, 1, 0.5, 0.5, "cm")) +
  scale_y_continuous(name="FDR [-log10]", limits=c(0,65), breaks=seq(0,65,by=20)) +
  scale_x_continuous(name="Fold change [log2]", limits=c(-15,15), breaks=seq(-15,15,by=5)) +
  geom_text_repel(
    data = label_data,
    aes(label = gene_name),
    size = 7,
    box.padding = unit(0.35, "lines"),
    point.padding = unit(0.3, "lines")
  )

```

## Single and bulk transcriptomics

```

# run scater.scrit first to generate sce single cell dataset
# extract counts of endogenous genes from sce dataset (excluding ERCC)
sce <- readRDS("scater_sce_dataset_QC_filter")
count_data_sc <- data.frame(counts(sce)[1:20862,])
dim(count_data_sc)

# use pagoda clustering to identify cell population to subset for
col.cols <- readRDS("pagoda_col_cols_352_cells")

# subset for epicardial cell cluster 1
adma_epicardium <- c(which(col.cols==1))
counts_adma_positive_epicardium <- count_data_sc[adma_epicardium]
counts_adma_positive_epicardium_merged <-
as.data.frame(rowSums(counts_adma_positive_epicardium))
colnames(counts_adma_positive_epicardium_merged) <- "adma_positive_epicardium"

# subset for epicardial cell cluster 2
acta2_epicardium <- c(which(col.cols==9))
counts_acta2_positive_epicardium <- count_data_sc[acta2_epicardium]
counts_acta2_positive_epicardium_merged <-
as.data.frame(rowSums(counts_acta2_positive_epicardium))
colnames(counts_acta2_positive_epicardium_merged) <- "acta2_positive_epicardium"

# subset for epicardial cell cluster 3
cldn11a_epicardium <- c(which(col.cols==2))
counts_cldn11a_positive_epicardium <- count_data_sc[cldn11a_epicardium]
counts_cldn11a_positive_epicardium_merged <-
as.data.frame(rowSums(counts_cldn11a_positive_epicardium))
colnames(counts_cldn11a_positive_epicardium_merged) <- "cldn11a_positive_epicardium"

# subset for myocardial cells
myocardium <- c(which(col.cols==8))
counts_myocardium <- count_data_sc[myocardium]
counts_myocardium_merged <- as.data.frame(rowSums(counts_myocardium))
colnames(counts_myocardium_merged) <- "myocardium"

# subset for haematopoietic cells
haematopoietic_cells = c(which(col.cols==7), which(col.cols==6))
counts_haematopoietic_cells <- count_data_sc[haematopoietic_cells]
counts_haematopoietic_cells_merged <- as.data.frame(rowSums(counts_haematopoietic_cells))
colnames(counts_haematopoietic_cells_merged) <- "haematopoietic_cells"

```

```

# subset for neural cells
neural_cells = c(which(col.cols==4))
counts_neural_cells <- count_data_sc[,neural_cells]
counts_neural_cells_merged <- as.data.frame(rowSums(counts_neural_cells))
colnames(counts_neural_cells_merged) <- "neural_cells"

# subset for mesenchymal cells
mesenchymal_cells = c(which(col.cols==5))
counts_mesenchymal_cells <- count_data_sc[,mesenchymal_cells]
counts_mesenchymal_cells_merged <- as.data.frame(rowSums(counts_mesenchymal_cells))
colnames(counts_mesenchymal_cells_merged) <- "mesenchymal_cells"

# join merged counts
counts_sc_subsets <- cbind(counts_adma_positive_epicardium_merged,
counts_acta2_positive_epicardium_merged,
counts_cldn11a_positive_epicardium_merged, counts_myocardium_merged,
counts_haematopoietic_cells_merged, counts_neural_cells_merged,
counts_mesenchymal_cells_merged)

counts_sc_subsets$genes <- rownames(counts_sc_subsets)
head(counts_sc_subsets)

# read in bulk data and generate DESeq object
# generate bulk count dataframe
all_data <- read.delim("D:/Oxford/RNA seq/Seq data processing/RNA bulk tcf21 tbx18 wt1b
5dpf/featureCounts.final.txt", header=TRUE, sep="\t", stringsAsFactors=FALSE, check.names=FALSE)
rownames(all_data) <- all_data$gene_id
all_data <- all_data[grepl("ENS",rownames(all_data)),]
head(all_data)
count_data_batch <- all_data[,4:ncol(all_data)]
rownames(count_data_batch) <- make.names(all_data$gene_name, unique=TRUE)
count_data_batch$genes <- rownames(count_data_batch)
head(count_data_batch)

# join bulk data and merged single cell data
count_data_batch_sc_subsets <- merge(count_data_batch, counts_sc_subsets, by="genes")
rownames(count_data_batch_sc_subsets) <- count_data_batch_sc_subsets$genes
count_data_batch_sc_subsets$genes <- NULL
head(count_data_batch_sc_subsets)

sample_info <-
data.frame(c(rep("tcf21_batch",3),rep("ctr_tcf21_batch",3),rep("tbx18_batch",3),rep("ctr_tbx18_batch",3)
,rep("wt1b_batch",3),rep("ctr_wt1b_batch",3),
"Epi 1", "Epi 2", "Epi 3", "Cardiomyocytes", "Haematopoietic cells", "Neural cells", "Mesenchymal cells"))

batch <- factor(c(rep(1:3,6),rep(1,7)))
sample_info <- cbind(sample_info,batch)
rownames(sample_info) <- colnames(count_data_batch_epi_subsets)
colnames(sample_info) <- c("condition","batch")
sample_info

dds_batch_sc <- DESeqDataSetFromMatrix(countData=count_data_batch_sc_subsets,
colData=sample_info, design= ~ batch + condition)
dds_batch_sc
mcols(dds_batch_sc) <- DataFrame(mcols(dds_batch_sc), rownames(count_data_batch_sc_subsets))
colnames(mcols(dds_batch_sc)) <- c("gene_name")
head(mcols(dds_batch_sc))

```

```

# filter out rows with only 0 or 1 counts
dds_batch_sc <- dds_batch_sc[rowSums(counts(dds_batch_sc))>1, ]

# normalize count data with variance stabilizing transformation
vsd_batch_sc <- varianceStabilizingTransformation(dds_batch_sc, blind=FALSE)
head(assay(vsd_batch_sc),3)

# generate PCA plot
data <- plotPCA(vsd_batch_sc, intgroup=c("condition","batch"), returnData=TRUE)
data$batch <- c(rep(c(16,17,15),6),rep(18,7))
data$color <- c(rep("royalblue2",3),rep("blue3",3),rep("violetred",3),rep("magenta4",3),
rep("chartreuse3",3),rep("darkgreen",3),"sienna4","red1","seagreen4","black","darkorange","purple3","snow4"))
percentVar <- round(100 * attr(data,"percentVar"))

dev.new(width=10, height=9)
ggplot(data=data, aes(x=data$PC1, y=data$PC2)) +
geom_point(size=5, col=data$color, shape=data$batch) +
theme_minimal(base_size = 23) +
theme(
  axis.text.x = element_text(size = 23, colour = "black",face = "plain", angle=0, hjust=0.5),
  axis.text.y = element_text(size = 23, colour = "black",face = "plain"),
  axis.title.x = element_text(size = 24, vjust = -2, hjust=0.5),
  axis.title.y = element_text(size = 24, vjust = 0),
  axis.line.x = element_line(colour = "black", size = 1.5),
  axis.line.y = element_line(colour = "black", size = 1.5),
  axis.ticks = element_line(colour = "black", size = 1.5),
  legend.position = "none",
  plot.margin = margin(0.5, 1, 0.5, 0.5, "cm")) +
scale_y_continuous(name=paste0("PC2: ", percentVar[2], "% variance"), limits=c(-40,80),
breaks=seq(0,80,by=20)) +
scale_x_continuous(name=paste0("PC1: ", percentVar[1], "% variance"), limits=c(-60,60), breaks=seq(-
60,60,by=20))

# generate legend for cell origin
plot(c(0, 1000), c(0, 1000), type="n", xlab="", ylab="")
legend("center", legend=c("Tg(tcf21:H2B-Dendra2) [+]","Tg(tcf21:H2B-Dendra2) [-]","Tg(tbx18:H2B-
Dendra2) [+]","Tg(tbx18:H2B-Dendra2) [-]","Tg(wt1b:H2B-Dendra2) [+]","Tg(wt1b:H2B-Dendra2) [-]","
epi1", "epi2", "epi3", "CM", "HC", "NC", "MC"),
fill=c("royalblue2", "blue3", "violetred", "magenta4", "chartreuse3", "darkgreen",
"sienna4", "red1", "seagreen4", "black", "darkorange", "purple3","snow4"),
border="white", bty="n", cex=1.2
)

```

## Microscopy

```

library(ggplot2)
if (!require("ggplot2")) {
source("http://bioconductor.org/biocLite.R")
biocLite("ggplot2")
library(ggplot2)
}

library(Hmisc)

```

```

if (!require("Hmisc")) {
source("http://bioconductor.org/biocLite.R")
biocLite("Hmisc")
library(Hmisc)
}

ANALYSIS NEW TRIPLE TRANSGENIC REPORTER LINE
### 3dpf
samples <- c("2016-11-12_1", "2016-11-14_2",
"2016-11-21_2", "2016-11-21_3",
"2016-12-11_3")

A <- c(37,25,19,28,23)
B <- c(47,39,33,49,43)
C <- c(33,20,15,27,21)
D <- c(31,44,43,44,30)
E <- c(17,16,15,22,13)
F <- c(1,1,0,1,2)
G <- c(7,4,4,11,9)
H <- c(4,9,11,11,4)
I <- c(3,6,7,8,3)
J <- c(16,4,0,5,8)
K <- c(3,4,4,0,0)
L <- c(7,15,14,11,13)

no_samples <- length(A)
df1_3dpf <- as.data.frame(cbind(c(A,B,C,D,E,F,G,H,I,J,K,L),
c(rep("tcf21 total",no_samples),rep("tbx18 total",no_samples),rep("tcf21 tbx18 double total",no_samples),
rep("wt1b total",no_samples),rep("triple pos",no_samples),rep("tcf21 single pos",no_samples),rep("tbx18
single pos",no_samples),
rep("wt1b single pos",no_samples),rep("wt1b single pos BA",no_samples),rep("tcf21 tbx18 double
pos",no_samples),
rep("tcf21 wt1b double pos",no_samples),rep("tbx18 wt1b double pos",no_samples))))))

df1_3dpf <- cbind(df1_3dpf,rep("3dpf",dim(df1_3dpf)[1]))
rownames(df1_3dpf) <- make.names(rep(samples,dim(df1_3dpf)[1]/length(samples)), unique=TRUE)
colnames(df1_3dpf) <- c("count","cell_type","time_point")
df1_3dpf$count <- as.numeric(as.character(df1_3dpf$count))

### 5dpf
samples <- c("2016-11-14_1", "2016-11-14_2", "2016-11-14_3", "2016-11-14_4",
"2016-11-23_1", "2016-11-23_2", "2016-11-23_3",
"2016-12-13_1", "2016-12-13_2", "2016-12-13_3")

A <- c(54,48,46,42,37,30,37,43,41,44)
B <- c(56,50,47,50,41,37,45,51,51,56)
C <- c(49,42,38,41,35,28,34,41,38,42)
D <- c(66,51,56,50,46,51,57,48,52,53)
E <- c(43,33,30,28,30,27,31,27,27,30)
F <- c(2,5,7,0,1,0,0,0,1,1)
G <- c(0,1,2,0,1,0,0,3,0,4)
H <- c(13,10,18,12,10,13,12,12,10,12)
I <- c(13,10,17,12,10,11,11,12,9,11)
J <- c(6,9,8,13,5,1,3,14,11,12)
K <- c(3,1,1,1,1,2,3,2,2,1)
L <- c(7,7,7,9,5,9,11,7,13,10)

```

```

no_samples <- length(A)
df1_5dpf <- as.data.frame(cbind(c(A,B,C,D,E,F,G,H,I,J,K,L),
c(rep("tcf21 total",no_samples),rep("tbx18 total",no_samples),rep("tcf21 tbx18 double total",no_samples),
rep("wt1b total",no_samples),rep("triple pos",no_samples),rep("tcf21 single pos",no_samples),rep("tbx18
single pos",no_samples),
rep("wt1b single pos",no_samples),rep("wt1b single pos BA",no_samples),rep("tcf21 tbx18 double
pos",no_samples),
rep("tcf21 wt1b double pos",no_samples),rep("tbx18 wt1b double pos",no_samples))))

df1_5dpf <- cbind(df1_5dpf,rep("5dpf",dim(df1_5dpf)[1]))
rownames(df1_5dpf) <- make.names(rep(samples,dim(df1_5dpf)[1]/length(samples)), unique=TRUE)
colnames(df1_5dpf) <- c("count","cell_type","time_point")
df1_5dpf$count <- as.numeric(as.character(df1_5dpf$count))

### 7dpf
samples <- c("2016-11-16_1", "2016-11-16_2", "2016-11-16_3",
"2016-11-25_3",
"2016-12-15_1", "2016-12-15_2")

A <- c(51,50,65,51,46,53)
B <- c(49,57,59,48,49,55)
C <- c(43,47,48,42,42,47)
D <- c(67,61,62,64,63,72)
E <- c(39,39,37,35,41,47)
F <- c(5,1,14,5,3,0)
G <- c(0,0,0,0,0,0)
H <- c(19,10,11,19,14,11)
I <- c(18,9,11,16,12,11)
J <- c(4,8,11,7,1,0)
K <- c(3,2,3,4,1,6)
L <- c(6,10,11,6,7,8)

no_samples <- length(A)
df1_7dpf <- as.data.frame(cbind(c(A,B,C,D,E,F,G,H,I,J,K,L),
c(rep("tcf21 total",no_samples),rep("tbx18 total",no_samples),rep("tcf21 tbx18 double total",no_samples),
rep("wt1b total",no_samples),rep("triple pos",no_samples),rep("tcf21 single pos",no_samples),rep("tbx18
single pos",no_samples),
rep("wt1b single pos",no_samples),rep("wt1b single pos BA",no_samples),rep("tcf21 tbx18 double
pos",no_samples),
rep("tcf21 wt1b double pos",no_samples),rep("tbx18 wt1b double pos",no_samples))))

df1_7dpf <- cbind(df1_7dpf,rep("7dpf",dim(df1_7dpf)[1]))
rownames(df1_7dpf) <- make.names(rep(samples,dim(df1_7dpf)[1]/length(samples)), unique=TRUE)
colnames(df1_7dpf) <- c("count","cell_type","time_point")
df1_7dpf$count <- as.numeric(as.character(df1_7dpf$count))

df1 <- rbind(df1_3dpf, df1_5dpf, df1_7dpf)

### plot data
data_df1 <- df1[c(grep("tcf21 total",df1$cell_type), grep("tbx18 total",df1$cell_type),grep("wt1b
total",df1$cell_type),grep("triple pos",df1$cell_type)),]

## plot data for one time point
data2_df1 <- data_df1[grep("7dpf", data_df1$time_point),]
dev.new(width=10, height=8)

```

```

ggplot(data2_df1, aes(x=cell_type, y=count, fill=cell_type)) +
geom_dotplot(binaxis="y", stackdir="center", stackratio=1.5, dotsize=40, binwidth=.1) +
theme_gray(base_size = 30) +
theme(
  axis.text.x = element_text(family = "Arial", size = 32, colour = "black",face = "plain", angle=25, hjust=1),
  axis.text.y = element_text(family = "Arial", size = 30, colour = "black",face = "plain"),
  axis.title.y = element_text(family = "Arial", size = 32, vjust = 5),
  axis.line.x = element_line(colour = "black", size = 2),
  axis.line.y = element_line(colour = "black", size = 2),
  axis.ticks = element_line(colour = "black", size = 2),
  legend.position = "none",
  plot.margin = margin(0, 0, 0, 4, "cm")) +
scale_y_continuous(name="Epicardial cell count", limits=c(0, 80), breaks=c(0,20,40,60,80)) +
scale_x_discrete(name="", labels=c("tcf21:myr-tdTomato", "tbx18:myr-GFP", "wt1b:H2B-Dendra2",
"Triple"),
  limits=c("tcf21 total", "tbx18 total", "wt1b total", "triple pos")) +
scale_fill_manual(values=c("limegreen", "magenta3", "white", "green4")) +
#stat_summary(fun.data=mean_sdl, fun.args = list(mult=1),geom="crossbar",
color="black",alpha=0,width=0.5) +
labs(title="") +
geom_boxplot(fill="white", alpha=0, lwd=1, fatten=2, width=0.6)

## plot complete data set for all timepoints
data3_df1 <- df1[c(grep("tcf21 single pos",df1$cell_type), grep("tbx18 single pos",df1$cell_type),
grep("wt1b single pos",df1$cell_type),
grep("tcf21 tbx18 double pos",df1$cell_type), grep("tcf21 wt1b double pos",df1$cell_type),
grep("tbx18 wt1b double pos",df1$cell_type), grep("triple pos",df1$cell_type)),]

dev.new(width=30, height=15)
ggplot(data3_df1, aes(x=cell_type, y=count, fill=cell_type)) +
#geom_dotplot(binaxis="y", stackdir="center", stackratio=1.2, dotsize=11, binwidth=.1,
position=position_dodge(0.7)) +
theme_gray(base_size = 30) +
theme(
  axis.text.x = element_text(family = "Arial", size = 28, colour = "black",face = "plain", angle=25, hjust=1),
  axis.text.y = element_text(family = "Arial", size = 28, colour = "black",face = "plain"),
  axis.title.y = element_text(family = "Arial", size = 28, vjust = 5),
  axis.line.x = element_line(colour = "black", size = 2),
  axis.line.y = element_line(colour = "black", size = 2),
  axis.ticks = element_line(colour = "black", size = 2),
  legend.position = "none",
  plot.margin = margin(0, 0, 0, 1, "cm"),
  strip.text.x = element_text(family = "Arial", size = 28, face="plain"),
  strip.background = element_blank()) +
scale_y_continuous(name="Epicardial cell count", limits=c(0, 50), breaks=c(0,10,20,30,40,50)) +
scale_x_discrete(name="", labels=c(1:8),
  limits=c("tcf21 single pos", "tbx18 single pos", "wt1b single pos", "wt1b single pos BA",
"tcf21 tbx18 double pos", "tcf21 wt1b double pos", "tbx18 wt1b double pos", "triple pos")) +
scale_fill_manual(values=rep("black",8)) +
#stat_summary(fun.data=mean_sdl, fun.args = list(mult=1),geom="crossbar",
color="black",alpha=0,width=0.6, position=position_dodge(0.7)) +
labs(title="") +
geom_boxplot(fill=c(rep(c("magenta3","limegreen","green4","green4","darkgray","gray","lightgray","white"),3)), alpha=1, lwd=1, fatten=2, width=0.8) +
facet_grid(. ~ time_point)

### test significance

```

```
# cutoff 4 samples: 0.0083, 0.0017, 0.00017
t.test(df1_3dpf$count[1:5],y=df1_3dpf$count[21:25],alternative = "greater", paired=FALSE)
t.test(df1_3dpf$count[6:10],y=df1_3dpf$count[21:25],alternative = "greater", paired=FALSE)
t.test(df1_3dpf$count[16:20],y=df1_3dpf$count[21:25],alternative = "greater", paired=FALSE)

t.test(df1_5dpf$count[1:10],y=df1_5dpf$count[41:50],alternative = "greater", paired=FALSE)
t.test(df1_5dpf$count[11:20],y=df1_5dpf$count[41:50],alternative = "greater", paired=FALSE)
t.test(df1_5dpf$count[31:40],y=df1_5dpf$count[41:50],alternative = "greater", paired=FALSE)

t.test(df1_7dpf$count[1:6],y=df1_7dpf$count[25:30],alternative = "greater", paired=FALSE)
t.test(df1_7dpf$count[7:12],y=df1_7dpf$count[25:30],alternative = "greater", paired=FALSE)
t.test(df1_7dpf$count[19:24],y=df1_7dpf$count[25:30],alternative = "greater", paired=FALSE)
```

#### ANALYSIS EXISTING DOUBLE TRANSGENIC REPORTER LINES

```
TG(TCF21:DSRED2;WT1B:GFP)
### 3dpf
samples <- c("2015-05-29_1", "2015-05-29_2", "2015-05-29_3", "2015-05-29_4", "2015-05-29_5", "2015-05-29_6")

A <- c(44,47,38,44,45,50)
B <- c(17,19,14,17,14,16)
C <- c(13,15,12,14,11,12)
D <- c(31,32,26,30,34,38)
E <- c(4,4,2,3,3,4)

no_samples <- length(A)
df2_3dpf <- as.data.frame(cbind(c(A,B,C,D,E),
c(rep("tcf21 total",no_samples),rep("wt1b total",no_samples),rep("tcf21 wt1b double pos",no_samples),
rep("tcf21 single pos",no_samples),rep("wt1b single pos",no_samples))))

df2_3dpf <- cbind(df2_3dpf,rep("3dpf",dim(df2_3dpf)[1]))
rownames(df2_3dpf) <- make.names(rep(samples,dim(df2_3dpf)[1]/length(samples)), unique=TRUE)
colnames(df2_3dpf) <- c("count","cell_type","time_point")
df2_3dpf$count <- as.numeric(as.character(df2_3dpf$count))

### 5dpf
samples <- c("2014-05-29_1", "2014-05-29_2", "2014-10-17_1", "2014-10-17_2", "2014-10-17_3", "2014-10-17_4",
"2014-11-01_1", "2014-11-01_2", "2014-11-01_3")

A <- c(46,39,43,53,47,53,61,54,55)
B <- c(6,7,5,4,5,6,9,8,10)
C <- c(3,3,3,3,3,4,5,7,8)
D <- c(43,36,40,50,44,49,56,47,47)
E <- c(3,4,2,1,2,2,4,1,2)

no_samples <- length(A)
df2_5dpf <- as.data.frame(cbind(c(A,B,C,D,E),
c(rep("tcf21 total",no_samples),rep("wt1b total",no_samples),rep("tcf21 wt1b double pos",no_samples),
rep("tcf21 single pos",no_samples),rep("wt1b single pos",no_samples))))

df2_5dpf <- cbind(df2_5dpf,rep("5dpf",dim(df2_5dpf)[1]))
rownames(df2_5dpf) <- make.names(rep(samples,dim(df2_5dpf)[1]/length(samples)), unique=TRUE)
colnames(df2_5dpf) <- c("count","cell_type","time_point")
```

```

df2_5dpf$count <- as.numeric(as.character(df2_5dpf$count))

### 7dpf
samples <- c("2015-09-11_1", "2015-09-11_2", "2015-09-11_4")

A <- c(63,55,57)
B <- c(24,24,18)
C <- c(12,14,8)
D <- c(51,41,49)
E <- c(12,10,10)

no_samples <- length(A)
df2_7dpf <- as.data.frame(cbind(c(A,B,C,D,E),
c(rep("tcf21 total",no_samples),rep("wt1b total",no_samples),rep("tcf21 wt1b double pos",no_samples),
rep("tcf21 single pos",no_samples),rep("wt1b single pos",no_samples))))

df2_7dpf <- cbind(df2_7dpf,rep("7dpf",dim(df2_7dpf)[1]))
rownames(df2_7dpf) <- make.names(rep(samples,dim(df2_7dpf)[1]/length(samples)), unique=TRUE)
colnames(df2_7dpf) <- c("count","cell_type","time_point")
df2_7dpf$count <- as.numeric(as.character(df2_7dpf$count))

df2 <- rbind(df2_3dpf, df2_5dpf, df2_7dpf)

### plot data
data_df2 <- df2[c(grep("tcf21 total", df2$cell_type),grep("wt1b total", df2$cell_type),grep("tcf21 wt1b
double pos", df2$cell_type)),]

## plot data for one time point
data_df2 <- data_df2[grep("3dpf", data_df2$time_point),]
windows()
ggplot(data_df2, aes(x=cell_type, y=count, fill=cell_type)) +
geom_dotplot(binaxis="y", stackdir="center", stackratio=1.2, dotsize=23, binwidth=.1) +
theme_minimal(base_size = 18) +
theme(
  axis.text.x = element_text(size = 16, colour = "black",face = "bold", angle=45, hjust=1),
  axis.text.y = element_text(size = 16, colour = "black",face = "bold"),
  axis.line.x = element_line(colour = "black", size = 1),
  axis.line.y = element_line(colour = "black", size = 1),
  axis.ticks = element_line(colour = "black", size = 1.2),
  legend.position = "none") +
scale_y_continuous(name="Counts", limits=c(0, 80), breaks=c(0,10,20,30,40,50,60,70,80)) +
scale_x_discrete(name="", labels=c("tcf21:dsRed2", "wt1b:GFP", "double"),
  limits=c("tcf21 total", "wt1b total", "tcf21 wt1b double pos")) +
scale_fill_manual(values=c("magenta3", "gray", "green4")) +
stat_summary(fun.data=mean_sdl, fun.args = list(mult=1),geom="crossbar",
color="black",alpha=0,width=0.5) +
labs(title="")

### test significance
# cutoff for 5 samples 0.005, 0.001, 0.0001
t.test(df2_3dpf$count[1:6],y=df2_3dpf$count[13:18],alternative = "greater", paired=FALSE) #tcf21
t.test(df2_3dpf$count[7:12],y=df2_3dpf$count[13:18],alternative = "greater", paired=FALSE) #wt1b

t.test(df2_5dpf$count[1:9],y=df2_5dpf$count[19:27],alternative = "greater", paired=FALSE)
t.test(df2_5dpf$count[10:18],y=df2_5dpf$count[19:27],alternative = "greater", paired=FALSE)

```

```
t.test(df2_7dpf$count[1:3],y=df2_7dpf$count[7:9],alternative = "greater", paired=FALSE)
t.test(df2_7dpf$count[4:6],y=df2_7dpf$count[7:9],alternative = "greater", paired=FALSE)
```

```
TG(TBX18:DSRED2;WT1B:GFP)
```

```
### 3dpf
```

```
samples <- c("2015-05-29_1", "2015-05-29_2", "2015-05-29_3", "2015-05-29_4", "2015-05-29_5", "2015-05-29_6")
```

```
A <- c(20,14,12,14,12,13)
```

```
B <- c(17,14,12,13,11,12)
```

```
C <- c(8,3,4,3,3,4)
```

```
D <- c(12,11,8,11,9,9)
```

```
E <- c(9,11,8,10,8,8)
```

```
no_samples <- length(A)
```

```
df3_3dpf <- as.data.frame(cbind(c(A,B,C,D,E),
```

```
c(rep("tbx18 total",no_samples),rep("wt1b total",no_samples),rep("tbx18 wt1b double pos",no_samples),
```

```
rep("tbx18 single pos",no_samples),rep("wt1b single pos",no_samples))))
```

```
df3_3dpf <- cbind(df3_3dpf,rep("3dpf",dim(df3_3dpf)[1]))
```

```
rownames(df3_3dpf) <- make.names(rep(samples,dim(df3_3dpf)[1]/length(samples)), unique=TRUE)
```

```
colnames(df3_3dpf) <- c("count","cell_type","time_point")
```

```
df3_3dpf$count <- as.numeric(as.character(df3_3dpf$count))
```

```
### 5dpf
```

```
samples <- c("2014-05-14_1", "2014-05-14_2", "2014-05-20_1", "2014-10-17_1", "2014-10-17_2", "2014-10-17_3", "2014-10-17_4", "2014-11-01_1", "2014-11-01_2", "2014-11-01_3")
```

```
A <- c(22,20,23,25,20,22,20,20,23,20)
```

```
B <- c(4,4,6,7,8,7,8,3,4,7)
```

```
C <- c(1,1,1,4,3,2,3,0,2,3)
```

```
D <- c(21,19,22,21,17,20,17,20,21,17)
```

```
E <- c(3,3,5,3,5,5,5,3,2,4)
```

```
no_samples <- length(A)
```

```
df3_5dpf <- as.data.frame(cbind(c(A,B,C,D,E),
```

```
c(rep("tbx18 total",no_samples),rep("wt1b total",no_samples),rep("tbx18 wt1b double pos",no_samples),
```

```
rep("tbx18 single pos",no_samples),rep("wt1b single pos",no_samples))))
```

```
df3_5dpf <- cbind(df3_5dpf,rep("5dpf",dim(df3_5dpf)[1]))
```

```
rownames(df3_5dpf) <- make.names(rep(samples,dim(df3_5dpf)[1]/length(samples)), unique=TRUE)
```

```
colnames(df3_5dpf) <- c("count","cell_type","time_point")
```

```
df3_5dpf$count <- as.numeric(as.character(df3_5dpf$count))
```

```
### 7dpf
```

```
samples <- c("2015-09-11_1", "2015-09-11_2", "2015-09-11_3", "2015-09-11_4", "2015-09-11_5", "2015-09-11_6")
```

```
A <- c(36,41,27,38,35,29)
```

```
B <- c(38,24,27,24,22,27)
```

```
C <- c(18,9,5,10,7,10)
```

```
D <- c(18,32,22,28,28,19)
```

```
E <- c(20,15,22,14,15,17)
```

```

no_samples <- length(A)
df3_7dpf <- as.data.frame(cbind(c(A,B,C,D,E),
c(rep("tbx18 total",no_samples),rep("wt1b total",no_samples),rep("tbx18 wt1b double pos",no_samples),
rep("tbx18 single pos",no_samples),rep("wt1b single pos",no_samples))))

df3_7dpf <- cbind(df3_7dpf,rep("7dpf",dim(df3_7dpf)[1]))
rownames(df3_7dpf) <- make.names(rep(samples,dim(df3_7dpf)[1]/length(samples)), unique=TRUE)
colnames(df3_7dpf) <- c("count","cell_type","time_point")
df3_7dpf$count <- as.numeric(as.character(df3_7dpf$count))

df3 <- rbind(df3_3dpf, df3_5dpf, df3_7dpf)

### plot data
data_df3 <- df3[c(grep("tbx18 total", df3$cell_type),grep("wt1b total", df3$cell_type),grep("tbx18 wt1b
double pos", df3$cell_type)),]

## plot data for one time point
data_df3 <- data_df3[grep("3dpf", data_df3$time_point),]
windows()
ggplot(data_df3, aes(x=cell_type, y=count, fill=cell_type)) +
geom_dotplot(binaxis="y", stackdir="center", stackratio=1.2, dotsize=25, binwidth=.1) +
theme_minimal(base_size = 18) +
theme(
  axis.text.x = element_text(size = 16, colour = "black",face = "bold", angle=45, hjust=1),
  axis.text.y = element_text(size = 16, colour = "black",face = "bold"),
  axis.line.x = element_line(colour = "black", size = 1),
  axis.line.y = element_line(colour = "black", size = 1),
  axis.ticks = element_line(colour = "black", size = 1.2),
  legend.position = "none") +
scale_y_continuous(name="Counts", limits=c(0, 80), breaks=c(0,10,20,30,40,50,60,70,80)) +
scale_x_discrete(name="", labels=c("tbx18:dsRed2", "wt1b:GFP", "double"),
  limits=c("tbx18 total", "wt1b total", "tbx18 wt1b double pos")) +
scale_fill_manual(values=c("magenta3", "gray", "green4")) +
stat_summary(fun.data=mean_sdl, fun.args = list(mult=1),geom="crossbar",
color="black",alpha=0,width=0.5) +
labs(title="")

### test significance
# cutoff for 5 samples 0.005, 0.001, 0.0001
t.test(df3_3dpf$count[1:6],y=df3_3dpf$count[13:18],alternative = "greater", paired=FALSE)
t.test(df3_3dpf$count[7:12],y=df3_3dpf$count[13:18],alternative = "greater", paired=FALSE)

t.test(df3_5dpf$count[1:10],y=df3_5dpf$count[21:30],alternative = "greater", paired=FALSE)
t.test(df3_5dpf$count[11:20],y=df3_5dpf$count[21:30],alternative = "greater", paired=FALSE)

t.test(df3_7dpf$count[1:6],y=df3_7dpf$count[13:18],alternative = "greater", paired=FALSE)
t.test(df3_7dpf$count[7:12],y=df3_7dpf$count[13:18],alternative = "greater", paired=FALSE)

### Plot both Tg(tcf21:dsRed2;wt1b:GFP) and Tg(tbx18:dsRed2;wt1b:GFP)
df4 <- rbind(df2, df3)
data_df4 <- df4[c(grep("tcf21 total", df4$cell_type),grep("tbx18 total", df4$cell_type),grep("wt1b total",
df4$cell_type),

```

```

grep("tcf21 wt1b double pos", df4$cell_type),grep("tbx18 wt1b double pos", df4$cell_type)),]

## plot data for one time point
data2_df4 <- data_df4[grep("7dpf", data_df4$time_point),]
dev.new(width=9, height=8)
ggplot(data2_df4, aes(x=cell_type, y=count, fill=cell_type)) +
geom_dotplot(binaxis="y", stackdir="center", stackratio=1.2, dotsize=40, binwidth=.1) +
theme_gray(base_size = 30) +
theme(
  axis.text.x = element_text(family = "Arial", size = 32, colour = "black",face = "plain", angle=30, hjust=1),
  axis.text.y = element_text(family = "Arial", size = 30, colour = "black",face = "plain"),
  axis.title.y = element_text(family = "Arial", size = 32, vjust = 5),
  axis.line.x = element_line(colour = "black", size = 2),
  axis.line.y = element_line(colour = "black", size = 2),
  axis.ticks = element_line(colour = "black", size = 2),
  legend.position = "none",
  plot.margin = margin(0, 0, 0, 1, "cm")) +
scale_y_continuous(name="Epicardial cell count", limits=c(0, 80), breaks=c(0,20,40,60,80)) +
scale_x_discrete(name="", labels=c("tcf21:dsRed2", "tbx18:dsRed2", "wt1b:GFP", "tcf21/wt1b double",
"tbx18/wt1b double"),
  limits=c("tcf21 total", "tbx18 total", "wt1b total", "tcf21 wt1b double pos", "tbx18 wt1b double pos")) +
scale_fill_manual(values=c("magenta3", "gray", "green4", "limegreen", "black")) +
#stat_summary(fun.data=mean_sdl, fun.args = list(mult=1),geom="crossbar",
color="black",alpha=0,width=0.5) +
labs(title="") +
geom_boxplot(fill="white", alpha=0, lwd=1, fatten=2, width=0.9)

```

LARVAL HEART INJURY TG(TCF21:DSRED2;WT1B:GFP)

### 0hpi

```

samples <- c("2015-05-20_1", "2015-05-20_2", "2015-05-20_3", "2015-05-20_4", "2015-05-20_5",
"2015-05-20_6", "2015-05-20_7", "2015-05-20_8")

```

```
A <- c(52,53,57,59)
```

```
B <- c(48,42,50,49)
```

```
C <- c(6,5,4,4)
```

```
D <- c(7,8,7,6)
```

```
E <- c(5,5,3,4)
```

```
F <- c(6,7,7,6)
```

```
no_samples <- length(A)
```

```
df5_0hpi <- as.data.frame(cbind(c(A,B,C,D,E,F),
```

```
c(rep("tcf21 total",2*no_samples),rep("wt1b total",2*no_samples),rep("tcf21 wt1b double
pos",2*no_samples))))
```

```
df5_0hpi <-
```

```
cbind(df5_0hpi,rep("0hpi",dim(df5_0hpi)[1]),rep(c(rep("ctr",no_samples),rep("inj",no_samples)),3))
```

```
rownames(df5_0hpi) <- make.names(rep(samples,dim(df5_0hpi)[1]/length(samples)), unique=TRUE)
```

```
colnames(df5_0hpi) <- c("count","cell_type","time_point","condition")
```

```
df5_0hpi$count <- as.numeric(as.character(df5_0hpi$count))
```

### 15hpi

```

samples <- c("2015-05-20_9", "2015-05-20_10", "2015-05-20_11", "2015-05-20_12", "2015-05-20_13",
"2015-05-20_14", "2015-05-20_15", "2015-05-20_16", "2015-05-20_17")

```

```
A <- c(62,60,60,56)
```

```

B <- c(61,57,62,61,59)
C <- c(8,8,5,6)
D <- c(16,11,8,8,4)
E <- c(7,6,5,4)
F <- c(13,9,7,7,3)

no_samples <- length(A)
df5_15hpi <- as.data.frame(cbind(c(A,B,C,D,E,F),
c(rep("tcf21 total",(2*no_samples)+1),rep("wt1b total",(2*no_samples)+1),rep("tcf21 wt1b double
pos",(2*no_samples)+1))))

df5_15hpi <-
cbind(df5_15hpi,rep("15hpi",dim(df5_15hpi)[1]),rep(c(rep("ctr",no_samples),rep("inj",no_samples+1)),3))
rownames(df5_15hpi) <- make.names(rep(samples,dim(df5_15hpi)[1]/length(samples)), unique=TRUE)
colnames(df5_15hpi) <- c("count","cell_type","time_point","condition")
df5_15hpi$count <- as.numeric(as.character(df5_15hpi$count))

### 24hpi
samples <- c("2015-05-20_18", "2015-05-20_19", "2015-05-20_20", "2015-05-20_21", "2015-05-20_22",
"2015-05-20_23", "2015-05-20_24", "2015-05-20_25")

A <- c(67,61,61,58)
B <- c(80,86,76,75)
C <- c(5,8,4,6)
D <- c(10,9,11,7)
E <- c(4,8,2,4)
F <- c(8,8,9,5)

no_samples <- length(A)
df5_24hpi <- as.data.frame(cbind(c(A,B,C,D,E,F),
c(rep("tcf21 total",2*no_samples),rep("wt1b total",2*no_samples),rep("tcf21 wt1b double
pos",2*no_samples))))

df5_24hpi <-
cbind(df5_24hpi,rep("24hpi",dim(df5_24hpi)[1]),rep(c(rep("ctr",no_samples),rep("inj",no_samples)),3))
rownames(df5_24hpi) <- make.names(rep(samples,dim(df5_24hpi)[1]/length(samples)), unique=TRUE)
colnames(df5_24hpi) <- c("count","cell_type","time_point","condition")
df5_24hpi$count <- as.numeric(as.character(df5_24hpi$count))

df5 <- rbind(df5_0hpi, df5_15hpi, df5_24hpi)

## plot data for one time point
data_df5 <- df5[grep("24hpi", df5$time_point),]
dev.new(width=9, height=7)
ggplot(data_df5, aes(x=cell_type, y=count, fill=condition)) +
geom_dotplot(binaxis="y", stackdir="center", stackratio=1.5, dotsize=40, binwidth=.1,
position=position_dodge(0.8)) +
theme_gray(base_size = 30) +
theme(
axis.text.x = element_text(family = "Arial", size = 32, colour = "black",face = "plain", angle=25, hjust=1),
axis.text.y = element_text(family = "Arial", size = 30, colour = "black",face = "plain"),
axis.title.y = element_text(family = "Arial", size = 32, vjust = 5),
axis.line.x = element_line(colour = "black", size = 2),
axis.line.y = element_line(colour = "black", size = 2),
axis.ticks = element_line(colour = "black", size = 2),

```

```

legend.position = "right",
plot.margin = margin(0, 0, 0, 1, "cm")) +
scale_y_continuous(name="Epicardial cell count", limits=c(0, 100), breaks=c(0,20,40,60,80,100)) +
scale_x_discrete(name="", labels=c("tcf21:dsRed2", "wt1b:GFP", "Double"),
limits=c("tcf21 total", "wt1b total", "tcf21 wt1b double pos")) +
scale_fill_manual(values=c("green4", "magenta3")) +
#stat_summary(fun.data=mean_sdl, fun.args = list(mult=1),geom="crossbar",
color="black",alpha=0,width=0.5) +
labs(title="") +
geom_boxplot(alpha=0, lwd=1, fatten=2, width=0.7, position=position_dodge(0.8))

```

```
### test significance
```

```
# cutoff for 6 samples 0.0033, 0.0007, 0.00007
```

```
t.test(df5_0hpi$count[1:4],y=df5_0hpi$count[5:8], paired=FALSE) #tcf21
```

```
t.test(df5_0hpi$count[9:12],y=df5_0hpi$count[13:16], paired=FALSE) #wt1b
```

```
t.test(df5_0hpi$count[17:20],y=df5_0hpi$count[21:24], paired=FALSE) #double
```

```
t.test(df5_15hpi$count[1:4],y=df5_15hpi$count[5:9], paired=FALSE) #tcf21
```

```
t.test(df5_15hpi$count[10:13],y=df5_15hpi$count[14:18], paired=FALSE) #wt1b
```

```
t.test(df5_15hpi$count[19:22],y=df5_15hpi$count[23:27], paired=FALSE) #double
```

```
t.test(df5_24hpi$count[1:4],y=df5_24hpi$count[5:8], paired=FALSE) #tcf21
```

```
t.test(df5_24hpi$count[9:12],y=df5_24hpi$count[13:16], paired=FALSE) #wt1b
```

```
t.test(df5_24hpi$count[17:20],y=df5_24hpi$count[21:24], paired=FALSE) #double
```

## Session info

### Single cell transcriptomics

R version 3.3.2 (2016-10-31)

Platform: x86\_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 15063)

locale:

[1] LC\_COLLATE=English\_United Kingdom.1252

[2] LC\_CTYPE=English\_United Kingdom.1252

[3] LC\_MONETARY=English\_United Kingdom.1252

[4] LC\_NUMERIC=C

[5] LC\_TIME=English\_United Kingdom.1252

attached base packages:

[1] stats4 parallel stats graphics grDevices utils datasets

[8] methods base

other attached packages:

[1] devtools\_1.12.0 limSolve\_1.5.5.1

[3] scran\_1.2.0 scater\_1.2.0

[5] ggplot2\_2.2.0 Rtsne\_0.11

[7] org.Dr.e.g.db\_3.4.0 GO.db\_3.4.0

[9] AnnotationDbi\_1.36.0 biomaRt\_2.30.0

[11] scde\_1.99.4 flexmix\_2.3-13

[13] lattice\_0.20-34 statmod\_1.4.26

[15] pheatmap\_1.0.8 RColorBrewer\_1.1-2

[17] RUVSeq\_1.8.0 edgeR\_3.16.2

[19] limma\_3.30.3 EDASeq\_2.8.0

[21] ShortRead\_1.32.0 GenomicAlignments\_1.10.0

[23] SummarizedExperiment\_1.4.0 Rsamtools\_1.26.1

[25] GenomicRanges\_1.26.1 GenomeInfoDb\_1.10.1

[27] Biostrings\_2.42.0 XVector\_0.14.0

[29] IRanges\_2.8.1 S4Vectors\_0.12.0  
[31] BiocParallel\_1.8.1 Biobase\_2.34.0  
[33] BiocGenerics\_0.20.0

loaded via a namespace (and not attached):

[1] ggbeeswarm\_0.5.0 minqa\_1.2.4  
[3] colorspace\_1.3-0 rjson\_0.2.15  
[5] hwriter\_1.3.2 modeltools\_0.2-21  
[7] dynamicTreeCut\_1.63-1 RcppArmadillo\_0.7.500.0.0  
[9] MatrixModels\_0.4-1 splines\_3.3.2  
[11] R.methodsS3\_1.7.1 extRemes\_2.0-8  
[13] Lmoments\_1.2-3 tximport\_1.2.0  
[15] DESeq\_1.26.0 geneplotter\_1.52.0  
[17] nloptr\_1.0.4 Cairo\_1.5-9  
[19] pbkrtest\_0.4-6 annotate\_1.52.0  
[21] R.oo\_1.21.0 shinydashboard\_0.5.3  
[23] shiny\_0.14.2 assertthat\_0.1  
[25] Matrix\_1.2-7.1 lazyeval\_0.2.0  
[27] htmltools\_0.3.5 quantreg\_5.29  
[29] tools\_3.3.2 gtable\_0.2.0  
[31] reshape2\_1.4.2 dplyr\_0.5.0  
[33] Rcpp\_0.12.12 nlme\_3.1-128  
[35] rtracklayer\_1.34.1 stringr\_1.1.0  
[37] lme4\_1.1-12 mime\_0.5  
[39] lpSolve\_5.6.13 XML\_3.98-1.5  
[41] distillery\_1.0-2 zlibbioc\_1.20.0  
[43] MASS\_7.3-45 zoo\_1.7-13  
[45] scales\_0.4.1 aroma.light\_3.4.0  
[47] pcaMethods\_1.66.0 rhdf5\_2.18.0  
[49] SparseM\_1.74 memoise\_1.0.0  
[51] gridExtra\_2.2.1 latticeExtra\_0.6-28  
[53] stringi\_1.1.2 RSQLite\_1.0.0  
[55] genefilter\_1.56.0 Rook\_1.1-1  
[57] GenomicFeatures\_1.26.0 chron\_2.3-47  
[59] matrixStats\_0.51.0 bitops\_1.0-6  
[61] RMTstat\_0.3 plyr\_1.8.4  
[63] magrittr\_1.5 R6\_2.2.0  
[65] DBI\_0.5-1 withr\_1.0.2  
[67] mgcv\_1.8-15 survival\_2.40-1  
[69] RCurl\_1.95-4.8 nnet\_7.3-12  
[71] tibble\_1.2 car\_2.1-3  
[73] viridis\_0.3.4 locfit\_1.5-9.1  
[75] grid\_3.3.2 data.table\_1.9.6  
[77] digest\_0.6.10 xtable\_1.8-2  
[79] httpuv\_1.3.3 brew\_1.0-6  
[81] R.utils\_2.5.0 munsell\_0.4.3  
[83] beeswarm\_0.2.3 vipor\_0.4.4  
[85] quadprog\_1.5-5

### Bulk transcriptomics

R version 3.3.2 (2016-10-31)

Platform: x86\_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 15063)

locale:

[1] LC\_COLLATE=English\_United Kingdom.1252

[2] LC\_CTYPE=English\_United Kingdom.1252

[3] LC\_MONETARY=English\_United Kingdom.1252  
[4] LC\_NUMERIC=C  
[5] LC\_TIME=English\_United Kingdom.1252

attached base packages:

[1] parallel stats4 stats graphics grDevices utils datasets  
[8] methods base

other attached packages:

[1] devtools\_1.12.0      scran\_1.2.0  
[3] BiocParallel\_1.8.1    scater\_1.2.0  
[5] statmod\_1.4.26       edgeR\_3.16.2  
[7] limma\_3.30.3         calibrate\_1.7.2  
[9] MASS\_7.3-45          ggrepel\_0.6.5  
[11] ggplot2\_2.2.0        RColorBrewer\_1.1-2  
[13] pheatmap\_1.0.8       biomaRt\_2.30.0  
[15] DESeq2\_1.14.0        SummarizedExperiment\_1.4.0  
[17] Biobase\_2.34.0       GenomicRanges\_1.26.1  
[19] GenomeInfoDb\_1.10.1   IRanges\_2.8.1  
[21] S4Vectors\_0.12.0     BiocGenerics\_0.20.0

loaded via a namespace (and not attached):

[1] dynamicTreeCut\_1.63-1 viridis\_0.3.4    splines\_3.3.2  
[4] Formula\_1.2-1        shiny\_0.14.2    assertthat\_0.1  
[7] latticeExtra\_0.6-28   vipor\_0.4.4     RSQLite\_1.0.0  
[10] lattice\_0.20-34      chron\_2.3-47    digest\_0.6.10  
[13] XVector\_0.14.0       colorspace\_1.3-0   htmltools\_0.3.5  
[16] httpuv\_1.3.3         Matrix\_1.2-7.1   plyr\_1.8.4  
[19] XML\_3.98-1.5         genefilter\_1.56.0   zlibbioc\_1.20.0  
[22] xtable\_1.8-2         scales\_0.4.1     htmlTable\_1.7  
[25] tibble\_1.2           annotate\_1.52.0   withr\_1.0.2  
[28] nnet\_7.3-12          lazyeval\_0.2.0   survival\_2.40-1  
[31] magrittr\_1.5         mime\_0.5         memoise\_1.0.0  
[34] foreign\_0.8-67       beeswarm\_0.2.3   shinydashboard\_0.5.3  
[37] tools\_3.3.2          data.table\_1.9.6   matrixStats\_0.51.0  
[40] stringr\_1.1.0        munsell\_0.4.3     locfit\_1.5-9.1  
[43] cluster\_2.0.5        AnnotationDbi\_1.36.0   rhdf5\_2.18.0  
[46] grid\_3.3.2           RCurl\_1.95-4.8    tximport\_1.2.0  
[49] rjson\_0.2.15         bitops\_1.0-6     gtable\_0.2.0  
[52] DBI\_0.5-1            reshape2\_1.4.2    R6\_2.2.0  
[55] zoo\_1.7-13          gridExtra\_2.2.1   knitr\_1.15  
[58] dplyr\_0.5.0         Hmisc\_4.0-0      stringi\_1.1.2  
[61] ggbeeswarm\_0.5.0     Rcpp\_0.12.12     geneplotter\_1.52.0  
[64] rpart\_4.1-10         acepack\_1.4.1