mgr inż. Eugeniusz Tralle

Cell lineage tracing in zebrafish heart development

Rozprawa na stopień doktora nauk medycznych i nauk o zdrowiu
w dyscyplinie nauki medyczne

Promotor: dr hab. Cecilia L. Winata

Laboratorium Genomiki Rozwoju Danio Pręgowanego
Międzynarodowy Instytut Biologii Molekularnej i Komórkowej w Warszawie

Obrona rozprawy doktorskiej przed Radą Dyscypliny Nauk Medycznych
Warszawskiego Uniwersytetu Medycznego

Warszawa 2023 r.

Słowa kluczowe

danio pręgowany, rozwój serca, regulacja ekspresji genów, transkryptomika pojedynczych komórek, śledzenie lineażu komórek;

Keywords

zebrafish, heart development, regulation of gene expression, single cell transcriptomics, lineage tracing;

The scholarship and work of Eugeniusz Tralle was funded by:

1. Postgraduate School of Molecular Medicine and Warsaw Medical University through the "Program Interdyscyplinarnych Studiów Doktoranckich wykorzystujących Sekwencjonowanie nowej generacji (NGS) w medycynie spersonalizowanej" scholarship program POWR.03.02.00-00-I041/16-00
2. „Reconstructing cardiovascular cell lineage evolution, one cell at a time" OPUS grant, National Science Center 2019/35/B/NZ2/02548. Laureate: C. L. Winata

Acknowledgments

I would like to express my deepest gratitude to my mentor and supervisor dr hab. Cecilia L. Winata for giving me the opportunity to be a part of your team. Your continued guidance, unparalleled support and understanding have made my budding academic career not only possible but also fulfilling and rewarding.

It's true that science is a collaborative effort, and it certainly seems that 'my' science is doubly so. Therefore, I'm extremely grateful to all the past and present members of the ZDG lab for their friendship, scientific discussions and help throughout my time there. In particular, I would like to extend my heartfelt thanks to dr Agata Sulej, dr Kasia Nieścierowicz and Maciek Łapiński whose expertise, kindness and support have helped me grow beyond my greatest expectations both as a scientist and as a person.

Special thanks to all the current and former Zebrafish Core Facility and IIMCB Core Facility members, especially Magda Góra, Magda Gral, dr Kasia Misztal and dr Kamil Jastrzębski – your knowledge and patience while I was still learning the ropes were extremely valuable to me (and this project).

Many thanks to dr Zuzanna Nowak-Życzyńska and prof. dr hab. Witold Rant for guiding me at the beginning of my academic journey and being the first supporters of my work and development in this field.

Special thanks to prof. dr hab. Bożena Kamińska-Kaczmarek and dr Natalia Ochocka-Lewicka who have kindly shared their expertise needed to complete crucial parts of this project.

I cannot begin to express my gratitude to my wife Olga Tralle, whose continuous support, unwavering confidence in me and ability to empower others have given true meaning to any endeavor I undertake. Your presence in my life is of immeasurable importance to me.

I am also grateful to my parents and brothers, Aleksy Tralle, Irena Morocka-Tralle, Wojciech Tralle and Leon Morocki, for creating and fostering an environment that allowed me to develop and prosper from the earliest possible moment.

Heartfelt thanks to all my friends, especially Julita Młyńska, Pietro Girardini, Patrycja Taradejna, Bartek Kulik and Natalia Drobniak-Kulik, Liliana Przyborowska and Łukasz Przyborowski. The positive influence your friendship has exerted on my life cannot be overestimated. You provided me with support and understanding at every opportunity and continue to do so despite my occasional failures. I'm also extremely grateful to all the friends and colleagues I met during my studies, especially the NGSchool team: Kasia Kędzierska, Maja Kuzman, German Demidov, Anamaria Elek and Andrey Prjibelski. Meeting you all has sealed my decision to pursue an academic career and your brilliance, drive and passion have been a constant source of inspiration and comfort for me.

Table of contents

List of figures and tables

List of abbreviations

hpf - hours post fertilization

dpf - days post fertilization

CP - cardiac progenitor

FHF - first heart field

SHF - second heart field

ALPM - anterior lateral plate mesoderm

CM - cardiomyocyte

WT - wild-type

TF - transcription factor

TFBS - transcription factor binding site

ZFIN - The Zebrafish Information Network

UMI - unique molecular identifier

MO - morpholino

KD - knock-down

LT/ST - low throughput/standard throughput

Streszczenie

Serce kręgowców jest jednym z pierwszych narządów powstających podczas rozwoju embrionalnego. Proces rozpoczyna się podczas gastrulacji, od utworzenia się dwóch populacji komórek macierzystych. Zlokalizowane są one po obu bokach wzdłuż pionowej osi ciała i podzielone na pierwsze i drugie pole sercowe. W ostatnich latach opublikowane zostały doniesienia o wcześniej nieopisanym zróżnicowaniu wśród komórek progenitorowych znajdujących się w polach serca. W tej pracy wykorzystałem danio pręgowane, organizm modelowy często używany w badaniach nad organogenezą, aby zbadać, kiedy i jak powstaje to zróżnicowanie. W pierwszej części rozprawy opisuję obecny stan wiedzy na temat rozwoju serca i konserwacji mechanizmów rozwoju serca wśród kręgowców, koncentrując się na ostatnich doniesieniach o heterogeniczności komórek tworzących serce. W drugim rozdziale opisuję badanie potencjalnych sekwencji wzmacniających ekspresję (enhancerów) w pobliżu *locus isl1a,* kluczowego dla rozwoju serca czynnika transkrypcyjnego, w celu zidentyfikowania sekwencji zdolnej do regulacji ekspresji tego genu w sercu. W trzecim rozdziale opisuję analizę transkryptomiczną pojedynczych komórek eksprymujących *nkx2.5,* inny kluczowy dla rozwoju serca czynnik transkrypcyjny, w określonych etapach powstawania serca danio pręgowanego. Połączyłem ten eksperyment z wyciszeniem ekspresji genów *nkx2.5*/*nkx2.7* przy użyciu morpholino w celu dalszego zbadania udziału genu *nkx2.5* w tworzeniu heterogeniczności progenitorów serca. Ponadto ustanowiłem w naszym laboratorium wcześniej opublikowaną metodę do śledzenia lineażu komórek kompatybilną z sekwencjonowaniem pojedynczych komórek, opartą na edycji sekwencji przez system CRISPR/Cas9. Wyniki przedstawione w tej pracy stanowią część większego projektu, z wciąż trwającą analizą danych transkryptomicznych z pojedynczych komórek. Stanowią one fundament dla przyszłych projektów, skupiających się na zgłębianiu pytań ujawnionych przez eksperymenty opisane w tej pracy.

Abstract

The vertebrate heart is one of the first formed organs during embryonic development. It begins during gastrulation with a formation of bilateral progenitor cell populations organized along the anteroposterior axis into the first and second heart fields. Recently, previously undescribed cardiac progenitor heterogeneity within these heart fields has been reported. I utilize the zebrafish, a potent model organism to study organogenesis, to investigate the questions of when and how this heterogeneity is established. In the first part of this thesis, I describe the current state of knowledge on heart development and its conservation across vertebrates, focusing on the recent reports of heterogeneity in the cells making up the heart. In the second chapter, I describe the first avenue of research I undertook while pursuing the answers to questions posed in this project. I investigate the putative enhancers proximal to the *isl1a*, a key cardiac transcription factor, with the aim of identifying an enhancer sequence capable of driving gene expression in a tissue-specific manner. In the third chapter, I describe a time course single-cell transcriptomic analysis on cells expressing *nkx2.5*, another key cardiac transcription factor, across the timeline of heart development in the zebrafish embryo. I coupled this experiment with a *nkx2.5/nkx2.7* loss of function approach in order to further investigate the role of *nkx2.5* in establishing the heterogeneity of cardiac progenitors. Finally, I establish in our lab a previously published single cell sequencing-compatible lineage tracing system based on CRISPR/Cas9 barcode editing. The results presented in this thesis are part of a larger, ongoing project with single cell data analysis ongoing. Nevertheless, it lays the foundation for future projects focusing on pursuing the questions uncovered by this work.

1. Introduction

The zebrafish (*Danio rerio*) are small freshwater fish which originally live in shallow, slow-moving streams or in periodically appearing still pools of southeast Asia and the Indian subcontinent. Their habitats were found as far east as Bangladesh and their northernmost populations were identified in Pakistan and Nepal (Parichy, 2015). The zebrafish are omnivorous, with their diet consisting of small insects, zooplankton as well as algae and plant material (Parichy, 2015). Zebrafish tend to live in shoals of six or more, which possibly provides protection from predators or facilitates mating (Parichy, 2015). Worldwide, they are widely recognized as popular aquarium pets and the transgenic, fluorescent zebrafish were among the first genetically engineered animals made available for purchase to the public in the US in 2003 (https://www.glofish.com).

The zebrafish belong to the bony fishes (*Teleostei*) class of ray-finned fishes (*Actinopterygii*) according to the traditional morphology-based classification described in *Fishes of the World* (Nelson, 2006). Accordingly, it falls under the *Cypriniformes* order, which comprises the largest, highly diverse group of freshwater teleosts. The family *Cyprinidae* encompasses some of the most widespread relatives of the zebrafish including commercially important Eurasian carp (*Cyprinus carpio*) and other popular aquarium fishes such as the blackline rasbora (*Rasbora borapetensis*) or the goldfish (*Carassius auratus*) (Nelson, 2006).

1.1. Zebrafish as a model organism

Over the last few decades, the zebrafish has been gaining prominence as a model organism to study a wide range of biological processes, from vertebrate embryonic development, through regeneration to aging (Veldman et al., 2008, Chia et al., 2022). It has also been utilized for drug discovery and toxicological assays (MacRae et al., 2015). The zebrafish is also widely used in medical research to model various human diseases, such as cancer, obesity, liver injury and cardiovascular disease (Bakkers, 2011, Zhang et al., 2018, Astell et al., 2020, Katoch et al., 2021).

The utility of the zebrafish in research stems in large part from key beneficial features of the species. It is characterized by high fecundity, with hundreds of embryos produced from a single spawning pair, coupled with external fertilization. This leads to ease of collection of embryonic material for studies. The zebrafish have a relatively short generation time of three

months, with a typical lifespan of 3 years. By 12 hpf (hours post fertilization), a typical vertebrate body plan has been established, and by 5 dpf (days post fertilization), most organs are already in place (Westerfield et al., 1988; Kimmel et al., 1995). As a result, observations of important developmental processes consume less time in comparison to mammalian models. Furthermore, the zebrafish embryo develops externally and is translucent, allowing for ease of microscopic observations of early developmental processes (Westerfield, 2000). Until 120 hpf a zebrafish embryo is not independently feeding and therefore not considered a protected animal by European law (Directive 2010/63/EU), making research focusing on zebrafish embryos before that stage easier from the ethical and 3R (Replacement, Reduction and Refinement) standpoint. Lastly, zebrafish husbandry protocols and detailed characterization of its development are well established, together with a database of gene expression and mutants, which makes it even more convenient as a subject for research (Kimmel et al., 1995; Beier, 1998).

These qualities were leveraged in the first large-scale forward genetic screen of zebrafish, culminating in 37 research papers on over 1500 developmental mutations, published in a special issue of Development in 1996 (Mullins, Acedo, Priya, Solnica-Krezel, Wilson, 2021). Since then, the number of zebrafish-focused research published has increased nearly every year (Fig. 1). Moreover, the collection of zebrafish genetic mutants has been a valuable resource, thanks to which the molecular mechanisms underlying various developmental processes have been elucidated. On the other hand, reverse genetics in the zebrafish model organism is facilitated by the ease of embryonic microinjection, enabling the introduction of foreign molecules for the purpose of imaging, such as fluorescent dye or reporter constructs, as well as genetic manipulation for knockdown or overexpression studies (Nasevicius and Ekker, 2000). The well-established transposition system also enabled rapid generation of fluorescent transgenic lines marking different populations of cell types (Parinov et al., 2004, Kawakami, 2007).

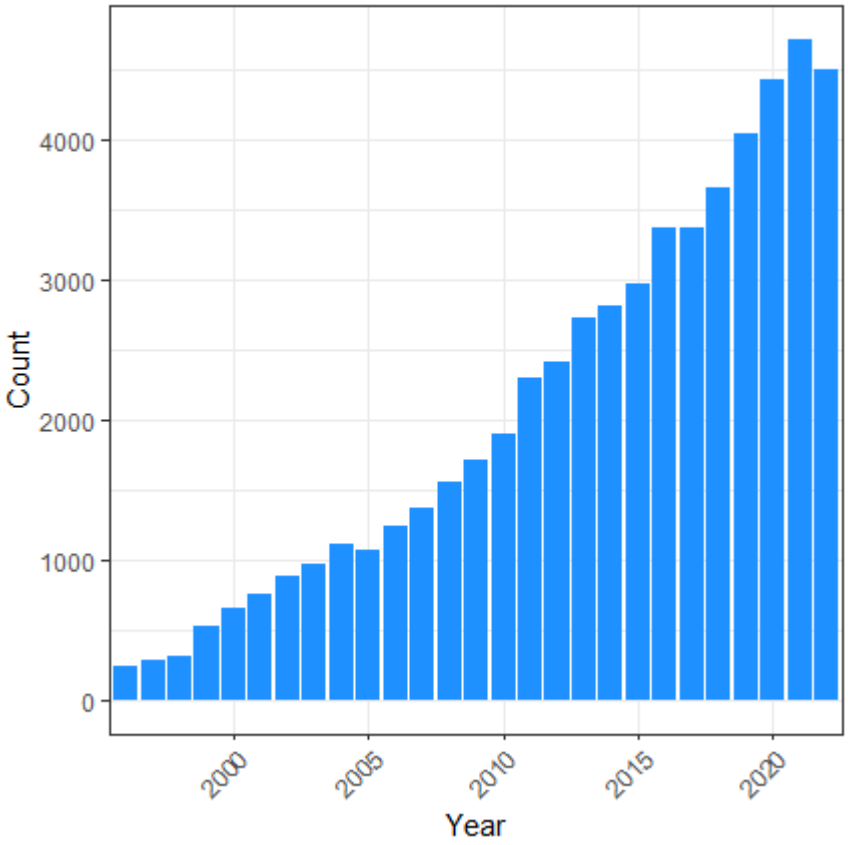


Fig. 1. Number of research papers containing the keyword "zebrafish" available in the PubMed database published between 1996 and 2022. Data source: PubMed.

The zebrafish genome project was completed in 2013 with the publication of the first zebrafish genome assembly, Zv9 (Howe et al., 2013). Since then, continuous efforts have been undertaken to refine it further, with the most recent complete release, GRCz11, published in 2017. These efforts enabled conducting more focused research using zebrafish, including targeted mutation screens. They also revealed the conservation of many molecular pathways across vertebrate taxa (Garcia et al., 2017, Paone et al., 2018), cementing zebrafish as a suitable model organism to study human disease and vertebrate embryonic development. Currently, the zebrafish genome size is estimated at 1.4 Gb with ~26 000 protein-coding genes arranged on 25 chromosomes (GRCz11, 2017). Compared to the human genome assembly, approximately 70% of human protein coding genes have at least one orthologue in the zebrafish genome (Howe et al., 2013). This comparison indicates the immense potential the zebrafish presents for modeling human genetic diseases as many diseases with genetic components can be modeled in zebrafish thanks to this extent of conservation (Gore et al., 2018, Astell et al., 2020).

The zebrafish is a particularly good model organism for the study of vertebrate heart development. Aside from its general characteristics described above, its larvae initially do not rely on a functioning heart and vasculature for oxygen delivery. This facilitates loss of function studies, as zebrafish mutants with impacted cardiac function - or lack thereof - can develop for several days, allowing for comprehensive phenotype characterization (Liu et al., 2012). Although the zebrafish heart is two-chambered and lacking some of the complexity of the mammalian heart, the molecular mechanisms leading to heart formation are conserved throughout development (Liu et al., 2012). Furthermore, zebrafish has been successfully used to model congenital heart disease which results from either structural or functional defects (Tu et al, 2012).

1.2. Vertebrate heart development

The vertebrate heart comes in various architectures across taxa, from the two-chambered fish heart to the four-chambered mammalian heart. Even within mammals, the differences in heart anatomy can be extensive. While the human heart weighs 250 - 300 g and is shaped to accommodate the bipedal nature of humans, the murine heart only weighs ~0,2 g, does not rest on the diaphragm and has an ellipsoidal shape, accounting for the position it takes within the context of a quadrupedal animal's thoracic cavity (Wessels et al., 2003). Furthermore, due to differences in the gestational period across vertebrates, even the timelines of heart development are variable, with processes occurring after birth in a murine heart being completed before birth in humans (Wessels et al., 2003). Despite these differences, the underlying molecular processes leading to heart formation are broadly conserved across vertebrates, making observations from one taxonomic group applicable in the broader context (Evans et al., 2010, Staudt and Stainier, 2012).

Vertebrate heart development begins at gastrulation. During that time, two *Mesp1*-expressing cell populations can be identified within the mesoderm, positioned on the sides of the primitive streak (Saga et al., 1999, Bondue et al., 2011, Chan et al., 2013). Mesp1 has been proposed as an early cardiac progenitor (CP) marker alongside Flk1 and Pdgfra (Bondue et al., 2011). Studies performed in mice have also shown that the lineage segregation of cardiac progenitors occurs during early gastrulation, prior to the onset of *Mesp1* expression (Devine et al. 2014; Lescroart et al. 2014).

As embryonic development progresses, these CPs quickly migrate through the primitive streak to join at the midline and form the cardiac crescent in mice, whereas in human and avian embryos the bilateral CP pools remain separated until later in development (Kirby, 2007, Evans et al., 2010). In the zebrafish, the earliest CPs appear as a bilateral population of cells in the lateral plate mesoderm (Bakkers, 2011). These bilateral populations of cells subsequently migrate towards the midline, fusing to form the early heart tube (Bakkers, 2011). The initial heterogeneity in the cardiogenic cells is established prior to their migration towards the midline, such that the cells in the fused cardiogenic mesoderm can already be identified as two distinct populations: the first and second heart fields (Kelly et al. 2001, Meilhac et al. 2004). The first heart field cells reside in the anterior part of the cardiogenic mesoderm, located in proximity to the developing head, while the second heart field cells are positioned in the posterior area (Kirby, 2007).

In both mammals and fishes, these two distinct populations of cardiac precursors are established and maintained through the interaction of signaling pathways including bone morphogenic protein (BMP), Wnt and Nodal (Marques et al., 2009, Staudt and Stainier, 2012), which originate from the neighboring structures. On the one hand, pharyngeal endoderm secretes BMPs and Wnt inhibitors, which promote cardiac progenitor fate. At the same time, CP population is limited by Noggin and Chordin signaling from the notochord, which inhibits BMP signaling in the center of the embryo. Another key limiting factor is Wnt signaling from the neural tube, which inhibits cardiac but promotes vascular fate. In this way, the CP population arises only in the regions of the embryo where BMPs and Wnt antagonist expression intersects, allowing for BMP-dependent activation of a key early cardiac fate driver - Nkx2-5 (Gilbert and Barresi, 2016).

The first heart field provides cells that initially structure the heart. During mammalian embryonic folding resulting in placement of the endoderm inside the body, the positions of CP cells are inverted. This results in FHF cells being positioned caudally with regards to the SHF cells. The FHF cells then fuse at the midline, forming the heart tube around day 28 - 30 post fertilization (CS10) in humans (O'Rahilly et al., 2010), E8.0 in mice (Evans et al., 2010) and 18 hpf in zebrafish (Staudt and Stainier, 2012). The cells coming from the first heart field have limited proliferative capabilities and, ultimately, only a small part of the heart can be traced back to this origin - the major part of the left ventricle cells in mammals and parts of the myocardial tissue in the atrium and ventricle in zebrafish (Felker et al., 2018).

Further heart development relies on the continuous addition of cells coming from the SHF, now located in the pharyngeal mesoderm. Concurrent with the addition of cells from the SHF, the heart tube undergoes looping, which marks the beginning of heart chamber formation. The morphogenic and molecular processes leading up to heart looping are virtually indistinguishable across vertebrate taxa (Warkman and Krieg, 2007, Xia et al., 2020). As a result of the looping, the vertebrate ventricle becomes positioned posteriorly to the atrium. CPs in the SHF can be further divided into anterior (a-) and posterior (p-) SHF, and each of these subpopulations goes on to make contributions to different structures of the heart. On the anterior (arterial) pole, SHF progenitors will constitute the myocardium of the forming ventricle as well as the outflow region - the smooth-muscular conus (in zebrafish - bulbus) arteriosus (Cai et al., 2003). On the posterior (venous) end of the tube, the SHF-derived cells are added to form the atrial myocardium and the inflow tract of the heart in mammals (Buckingham et al. 2005, van den Berg et al. 2009; Lescroart et al. 2012, Meilhac and Buckingham, 2018) as well as in zebrafish (Bertrand et al. 2011). The atrium and ventricle are separated by the atrioventricular canal (AVC), a constricted region housing the atrioventricular valves which regulate the blood flow, as well as the atrioventricular node of the cardiac conduction system (Staudt and Stainier, 2012).

In mammals, heart chamber formation starts with the ventricle developing from the ventral side of the primitive heart tube. As embryonic development progresses, the heart tube undergoes looping, and the definitive chamber formation begins. As the left and right ventricle become specified, the process of symmetric development of left and right atria progresses (Buijtendijk et al., 2020). In zebrafish, these later processes are limited to progressing septation and valve formation as the zebrafish has a two-chambered heart. However, the overall morphology of the maturing heart such as presence of the cardiac conduction system or trabeculation is conserved (Staudt and Stainier, 2012).

1.3. The heterogeneity and molecular markers of the second heart field

The existence of the second heart field was first established in 2001, when three research groups independently identified a population of progenitor cells that contributed to the heart development in chicken and mouse embryos after the initial heart tube was formed (Kelly et al., 2001, Mjaatvedt et al., 2001, Waldo et al., 2001). The zebrafish also has a SHF located in the anterior lateral plate mesoderm (ALPM), marked by *mesp1*, *isl1* and *nkx2.5* expression, similar to the SHF in amniotes (Laugwitz et al., 2008, Staudt and Stainier, 2012).

The SHF is regulated by several signaling pathways, including Wnt, Nodal, retinoic acid (RA) and BMP (Staudt and Stainier, 2012). In zebrafish, *ldb1* has been shown to regulate SHF-dependent cardiogenic processes by binding and stabilizing *isl1a*, enabling an *ldb1/isl1a* complex-dependent enhancement of transcription of key cardiac TFs, *mef2c* and *hand2* (Caputo et al., 2015). Conversely, RA-dependent LIM protein Ajuba represses the transcriptional activity of *isl1a*, limiting the activation of the cardiac program in the progenitor cells (Witzel et al., 2012). Likewise, *isl2b* has been shown to play a key role in SHF-dependent process during heart development as the loss of function of *isl2b* leads to a significant decrease in the number of SHF-derived cardiomyocytes (Witzel et al., 2017).

Nkx2-5 (NK2 homeobox 5) is a well-known early cardiac fate driver, playing a key role in establishing and maintaining cardiac cell identity (Lyons et al., 1995, George et al., 2014, Guner-Ataman et al., 2018). *nkx2.5* has also been shown to play a role in specifying the hemoangiogenic lineages, where a subpopulation of *nkx2.5*-expressing progenitors was found to give rise to the zebrafish pharyngeal arch endothelium (Paffett-Lugassy et al., 2015). Interestingly, other reports in zebrafish have shown that the heamoangiogenic cell program can be suppressed by *nkx2.5*, indicating a dual role that this transcription factor can play in this developmental process (Simões et al., 2011). In zebrafish, due to a teleost-specific genome duplication (Glasauer et al., 2014), there are two paralogues of the Nkx2-5 gene, *nkx2.5* and *nk2.7*, which have been shown to act redundantly during heart development. Upon loss of function of one, the other was capable of partially rescuing the phenotype (Tu et al., 2009, Targoff et al., 2013). The expression of both zebrafish *nkx2.5* and *nkx2.7* starts early after gastrulation within cells in the anterior lateral plate mesoderm (Swedlund and Lescroart, 2020) (Fig. 2).

Cells expressing *nkx2.5* are present in both the FHF and the SHF (Colombo et al., 2018, Duong et al., 2021). In zebrafish, both *nkx2.5* and *nkx2.7* act as key regulators of cell identity as they have been shown to limit the number of atrial cardiomyocytes and establish the proper number of CMs in the ventricle (Targoff et al., 2008).

In later stages of heart development, the activity of *nkx2.5* is necessary for establishing and maintaining atrial and ventricular identities of cardiomyocytes (Guner-Ataman et al., 2013, Targoff et al., 2013, Colombo et al., 2018). It was also found to be necessary for limiting the expansion of myocardium (Duong et al., 2021) and is crucial for maintaining pacemaker activity in the sinoatrial node (Ren et al., 2019).

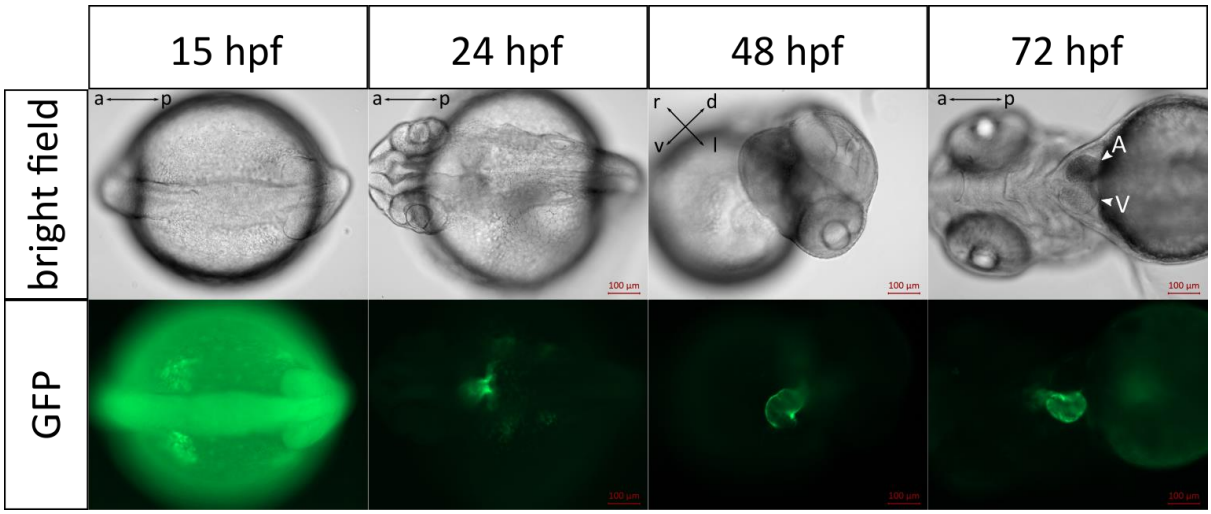


Figure 2. *nkx2.5* expression throughout zebrafish heart development. Zebrafish embryos from the Tg(nkx2.5:EGFP) line were imaged at key points during heart development. Body axes were marked in the upper left corner of each bright field micrograph: a – anterior, p – posterior, l – left, r – right, v – ventral, d – dorsal. A - atrium, V - ventricle.

The Isl1 (ISL LIM homeobox 1) transcription factor was initially identified as a second heart field marker in 2003 (Cai et al., 2003). However, it is also known to drive gene expression in the developing nervous and digestive systems in zebrafish (Thisse et al., 2005) as well as mammals (Ren et al., 2021), and was utilized as a reporter marker for these cell lineages in transgenic zebrafish lines (Higashijima et al., 2000). In the heart, *Isl1+* progenitors have been found to contribute extensively to various heart tissues, including cardiomyocytes, smooth muscle cells of aorta and coronaries as well as endothelial cells

in the heart vasculature (Moretti et al., 2006). In the absence of Isl1, SHF-derived structures are not formed (Cai et al., 2003). The zebrafish has one direct ortholog of the human *ISL1* gene, *isl1a*, with highly conserved coding sequence (97,99%) and function (Witzel et al., 2012). However, likely due to the aforementioned teleost genome duplication event, *isl1a* has 3 paralogs in the zebrafish genome: *isl1b*, *isl2a* and *isl2b*. Only one of these paralogs, *isl2b*, has been reported to share some of the function with *isl1a* in heart development. *isl2b* is co-expressed in the ALPM cells alongside *isl1a*, and its loss of function limits the abundance of SHF-derived cells in the developing heart, similarly to the *isl1a*-/- phenotype (Witzel et al., 2017).

The originally identified SHF cells expressing Isl1 were later recognized to constitute the anterior portion of the second heart field (Dodou et al., 2004). This characterization was then reinforced with several lines of study, deepening our understanding of the second heart field as a heterogeneous population of cardiac progenitors (van den Berg et al., 2009, Lescroart et al., 2012). Recently, a single cell study in mice has shown that some extent of molecular diversity can already be identified in the Mesp1+ progenitors (Lescroart et al., 2018). Studies utilizing later cardiac progenitors co-expressing Nkx2-5 and Isl1 have shown that these cells can generate colonies able to give rise to multiple cardiac cell types, indicating the plasticity of these cells (Moretti et al., 2006). The idea of heterogeneity of cells residing in the SHF was further reinforced by subsequent studies, identifying cells that will generate the facial muscles, the pulmonary artery and vein, and the lung mesenchyme in mammals (Lescroart et al. 2010, 2015; Peng et al. 2013). Interestingly, similar observations have been reported in invertebrate tunicates, where a group of common progenitors have been found for pharyngeal and cardiac tissues, suggesting extensive conservation of cell plasticity and molecular processes involved in organogenesis (Diogo et al., 2015). In addition, the progenitors residing in the second heart field have been recently reported to form a vasculogenic niche in mice, contributing to the lymphatic endothelial cells and ventral lymphatic vessels of the heart (Lioux et al., 2020).

1.4. Single cell RNA sequencing

In recent years, significant progress has been made in developing methods for in-depth transcriptomic studies. The increasing availability of single-cell sequencing allowed researchers to characterize the transcriptomes of different organs on a cellular level. The ability

to identify and categorize cells based on their individual molecular contents enabled discovery of cell types that were previously missed in marker-based bulk transcriptomic analyses (Skelly et al., 2018, Lescroart et al., 2018).

The possible applications of single cell RNA-seq extend beyond investigating cell heterogeneity and identification of novel cell types. Multiple research projects have leveraged time course analyses coupled with single cell sequencing to reconstruct molecular trajectories leading to cell type specification (Liu et al., 2017, Farrell et al., 2018, Street et al., 2018). Recently, lineage recording and tracing methods have been developed, combining single-cell RNA-seq with CRISPR/Cas9-based barcoding to trace cell type development (Raj et al., 2018, Kalhor et al., 2018). These approaches provide a powerful tool for developmental biologists. The ability to trace the history of individual cells making up an organ, linking them directly to progenitor cell populations they arise from can provide crucial insight into the mechanisms of cell diversification and organogenesis.

1.5. Project objectives

Although we now understand that the SHF progenitor population is essential for development of the cardiovascular system, there are still several knowledge gaps. Firstly, while we have the knowledge about the transcription factors expressed and involved in SHF specification, their precise mechanism is still largely uncharacterized. Secondly, the two best-known transcription factors implicated in this process, Isl1 and Nkx2-5, are rather pleiotropic in terms of expression and function. This raises a key question on how they exert their specific roles in the SHF. Lastly, despite the knowledge that the SHF consists of a diverse cell population and is capable of giving rise to diverse cell types, the molecular basis of this heterogeneity, as well as the timeline of its lineage evolution is still unknown.

The purpose of this project is to characterize the molecular mechanisms underlying the specification of the SHF, focusing on the role of *isl1a* and *nkx2.5*, using the zebrafish as a model organism. To address this, I pursued two avenues of investigation.

In the first part of the study, I ask what is the mechanism underlying the specific expression of *isl1a* in the heart. This question is especially important considering that *isl1a* is known to be expressed and implicated in the development of many other tissues, including neural

and pancreatic. To this end, I sought to identify the regulatory regions that could drive its expression in a cardiac-specific manner. Through *in silico* screening, I identified putative enhancers located in the proximity of the *isl1a* gene and validated them using an enhancer assay. If successful, a zebrafish transgenic line expressing a reporter gene co-regulated with *isl1a* solely in the heart could provide a powerful tool for further studies aiming to elucidate the mechanisms underlying early stages of heart development.

In the second line of research, I sought to establish an experimental toolbox which will enable me to determine the dual role of *nkx2.5* in driving FHF and SHF CP cells specification into various types of lineages. For this purpose, I optimized two experimental tools which would lay the foundation for cell lineage tracing of the CPs at high resolution. The first tool involves a time course single-cell transcriptome analysis aiming to establish the lineage evolution of *nkx2.5*-expressing cells starting from early progenitor stage throughout heart development. Utilizing single cell sequencing allows me to investigate *nkx2.5*+ cardiac progenitors earlier and with more granularity than previously possible with bulk analyses. This approach, coupled with single cell RNA-seq analysis of a *nkx2.5* loss of function model (Targoff et al., 2013), will provide functional insight into the role of *nkx2.5* activity in establishing the heterogeneity of cardiac progenitors as well as its dual role in adoption of different cell developmental trajectories. The second tool comprises a Cas9-based barcode editing system for cell lineage tracing (Raj et al., 2018). This system enables cell lineage recording and tracing through Cas9 editing of a synthetic barcode sequence incorporated in a transgene. In order to establish them in our lab, I generated two transgenic lines, each of which expresses the CRISPR sgRNAs and the genetic barcode respectively. If successful, it will allow me to link the cells making up a developing heart to their CP origins, providing further insight into the role of *nkx2.5* in driving cell fate in the heterogenous CPs population as well as molecular mechanisms governing heart organogenesis.

2. Search for early heart-specific developmental enhancers

2. 1. Materials and methods

2. 1. 1. Zebrafish husbandry

Adult zebrafish of wild-type and transgenic background were kept in standard conditions in the Zebrafish Core Facility of the International Institute of Molecular and Cell Biology in Warsaw (license no. PL14656251) in line with standard procedures and ethical guidelines. Briefly, dry feed was provided three times per day, supplemented with live *Artemia* sp. nauplii obtained from in-house food culture. The fish were kept at a density of up to 50 fish per 8 l of water at 28.5±0.5 °C. The photoperiod was set to 14/10 on/off with light at 300-330 lux intensity. Embryos were maintained in E3 medium (60 μg/ml of "Instant Ocean" sea salts in ddH_2O) and staged according to standard criteria (Kimmel et al., 1995).

2. 1. 2. *In silico* regulatory region prediction

In order to identify putative regions regulating gene expression in the SHF, publicly available data from ChIP-seq experiments focusing on epigenetic histone modifications, namely H3K27ac, deposited in the DANIO-CODE repository was referenced (GEO accession numbers: GSE32483, GSE35050, GSE70847, GSE48254). The data was visualized using the UCSC genome browser (https://genome-euro.ucsc.edu/index.html).

2. 1. 3. Enhancer cloning

Putative enhancer sequences, ranging from 296 to 1314 bp, were amplified from genomic DNA using specific pairs of primers (Table 1). Each amplified sequence was then re-amplified using primers containing an overhang of recognition sites for the BglII restriction enzyme. Then, the amplicons were BglII-digested and purified using E. Z. N. A. Plasmid MiniPrep kit (Omega Bio-Tek). Afterwards, each putative enhancer sequence was ligated into the e1b-GFP-Tol2 vector (Addgene, plasmid #37845), digested with the BglII enzyme, and transformed into Top10 competent cells. The bacteria were plated on Petri dishes containing LB-Agar and grown overnight at 37°C. Afterwards, single colonies were collected. DNA from each bacterial clone was used in colony PCR amplification with putative enhancer-specific primers. The colony PCR products were then loaded into a 1.5% agarose/Midori Green

gel to verify successful cloning. Bacterial clones containing the correct inserts were cultured in 5 ml liquid LB broth overnight in a shaking incubator at 37℃, 200 rpm. After overnight incubation, half of each culture was mixed with 50% glycerol and saved for long-term storage in -80℃, while the other half was used to extract plasmid DNA using the Plasmid Miniprep Kit (Promega) and suspended in nuclease-free water. Plasmid concentration and purity were estimated using NanoDrop and stored in -20℃.

Table 1. Basic information on genomic location of the putative enhancers tested, their sizes as well as the primer sequences used.

Name	Genomic location	Forward primer sequence [5'→3']	Reverse primer sequence [5'→3']	length [bp]
ChIP1U	chr5:40,734,722-40,735,017	GGGAAAGGCGCAAATCCGTA	GTTTAGCTCATATAGATCTATAGATC	296
ChIP2U	chr5:40,735,213-40,735,670	GCAACCGAGCATCGCTATAA	CAAATGGATACACAAACAAATCGCA	458
E3D	chr5:40,714,946-40,716,138	TCCTACGAGAGGAACGGAGAG	TAAGCTGACTGTCACATCCCG	1193
DC1	chr5:40,735,005-40,735,451	ATATGAGCTAAACTACGAGCTCAAACAGC	ATGCACATAAACAGCTGTAGCAGAAAC	447
DC4	chr5:40,773,247-40,774,328	AAACAAGATGGCACCACTTTGTACTTT	TGAAAACACATGTAGATGAGAACTGATG	1082
DC3	chr5:40,744,847-40,745,551	GGCATCTTTATTGTTCAACAATGATTTGG	GACAGATTATCCATCCCACAAATTTAAAGC	705
I3	chr5:40,730,772-40,732,085	TAAGATCTCACATATGTTAAGCGGTCAGCTG	ATAGATCTCTACAACTAACAAAGCATTGTTTG	1314

2. 1. 4. *Tol2* mRNA *in vitro* transcription

In vitro transcription of capped Tol2 transposase mRNA was performed using the mMESSAGE mMACHINE T3 kit (ThermoFisher). First, plasmid containing Tol2 transposase cDNA sequence (pT3TS_Tol2, Addgene) was linearized using the SpeI restriction enzyme. The resulting RNA was purified using the kit manufacturer's protocol and suspended in nuclease-free water. The RNA integrity was checked visually by heating the RNA sample at 95℃ for 5 minutes in a thermocycler and loading it on a denaturing agarose gel alongside RiboRuler (ThermoFisher), an RNA size marker. This was done to check whether the RNA formed a single, sharp band of the correct size, indicating that the *in vitro* transcription was successful. RNA concentration was then estimated using NanoDrop, aiming for an OD260/280 ratio of at least 2. The Tol2 transposase mRNA was then aliquoted and stored at -80℃.

2. 1. 5. Microinjection

The injection needles were prepared from borosilicate glass capillaries with an inner diameter of 0.75 mm and outer diameter of 1 mm (Sutter Instruments), pulled using the Micropipette Puller (Sutter Instruments) and opened using the Microforge (Narishige) so that the needle eye was 15 µm in diameter.

Injection plates were prepared by dissolving 1.5 g agarose in 100 ml of E3 medium in a microwave and pouring 20-25 ml of the forming gel per Petri dish that would serve as an injection plate. Then, a microinjection mold (World Precision Instruments) was placed on each dish so that no air bubbles were trapped between the mold and the surface of the liquid. The plates were kept at room temperature, uncovered, until the gel was set. Then the molds were removed and each dish was covered and stored at 4℃.

Injections were performed using the PicoPump microinjection pump (World Precision Instruments). First, a mix was prepared for each construct, containing 30 ng/µl of enhancer-containing plasmid and 25 ng/µl of Tol2 RNA. The solution was then loaded into a needle and mounted on a micromanipulator. Then, the injection volume was determined by dispensing it into mineral oil on a stage micrometer ruler and measuring the diameter of the formed droplet. Injection pressure and time were adjusted to calibrate droplet volume to 1 nl. After droplet calibration, the embryos were collected and loaded

onto an injection plate. A portion of embryos from each batch was set aside and left uninjected to serve as control. These embryos were otherwise treated the same as the injected ones. The embryos were then inspected to ensure that only embryos in the 1-cell developmental stage were injected. Any malformed embryos were discarded and not used in the experiment. Each embryo was then injected with 1nl of injection mix (30 pg of enhancer plasmid and 25 pg Tol2 transposase RNA). After injecting a batch of embryos, they were transferred into ventilated Petri dishes containing fresh E3 medium and put in an incubator at 28.5℃ to develop. After 3 hours of incubation, the embryos were screened to discard any unfertilized eggs. The remaining embryos were counted and placed back in the incubator.

2. 1. 6. Screening

At 24hpf, the embryos were screened for survival and GFP expression. To test for survival, the embryos were checked under a bright field microscope and the number of live embryos was counted. In order to rule out nonspecific effects of the microinjection procedure, phenotypic assessment was made to ensure no gross morphological defects were observed. In cases when the embryos had severe developmental defects resulting from the injection that a live-dead distinction could not be immediately made, the presence or lack of somites and/or central nervous system were used to make the distinction. The number of dead embryos was calculated by comparing the number of live embryos to the number of embryos noted down the day before, after the unfertilized eggs had been discarded. This was done to avoid overestimating the survival rate as dead embryos can disintegrate in the incubator fast enough that they can be easily missed at 24 hpf.

To check for GFP expression, the embryos were observed under a fluorescent microscope. If green fluorescence was observed in any part of the body of the embryo, it was considered GFP-positive, however, if the embryo exhibited any gross developmental perturbations, it would not be kept to adulthood. GFP-positive embryos from batches with an overall high survival rate (>60% alive at 24hpf) were collected and saved for rearing as F0.

2. 1. 7. F1 stable line generation

I established the stable transgenic line adapting the procedures described in our previously published work (Nieścierowicz et al., 2022). Briefly, after reaching maturity, the GFP-positive individuals were outcrossed to wild-type to check for germline transmission of the transgene.

Embryos from each crossed pair (F1) were screened for GFP expression and the consistency of its expression pattern. The individuals that produced the most tissue-specific, repetitive pattern of expression were then kept as positive founders. Their GFP-positive offspring (F2) were raised to adulthood and established as stable transgenic lines.

2. 1. 8. Data processing and statistical analysis

Data obtained throughout the course of this experiment was handled and processed using tidyverse (Wickham et al., 2019) and dplyr (Wickham et al., 2023) R packages. Performed statistical tests were indicated in the 2. 2. Results section where appropriate. Visualization of the results was done using ggplot2 (Wickham, 2016) and ggpubr (Kassambara, 2023).

2. 2. Results

2. 2. 1. *In silico* prediction of putative regulatory regions using publicly available epigenomic data

In order to identify regulatory regions that would drive *isl1a* expression in the developing heart, I employed an *in silico* approach. I utilized publicly available data from ChIP-seq experiments targeting H3K27ac, a known active enhancer marker (Creyghton et al., 2010). I used this marker to identify putative regulatory regions active at developmental time points when the earliest cardiac progenitors are specified, namely the bud (10 hpf) and 5-9 somite (~12 hpf) stages (Reifers et al., 2000, Deshwar et al., 2016, Budine et al., 2020). I performed a search within the genomic region ±10 kb of the body of the *isl1a* gene (GRCz11, chr5:40728613-40734027, Fig.).

Overall, I identified 7 regions that exhibit the characteristics of an enhancer. Six of the tested regions (ChIP1U, ChIP2U, DC1, DC3, DC4 and I3) were identified based on the presence of H3K27ac peaks at either 10 or 12 hpf (Fig. 3). One additional region, e3d (1193 bp, chr5:40,714,946-40,716,138), was found based on evolutionary conservation data from the Evolutionary Conservation of Genomes browser (ECR browser, access 03.10.2018). This region bears a 70% similarity to an *isl1a* regulatory region described in mice (Kappen et al., 2009). It is located downstream of the *isl1a* locus. The sequences of the peaks were investigated to make the final determination whether the sequence would be tested.

I excluded highly repetitive sequences from the experiment due to high error rates during amplification, cloning and sequencing of repetitive sequences, which made them unsuitable to my approach. All of the ChIP-seq-identified regions were located upstream of the *isl1a* locus with the exception of one, I3 (1314 bp, chr5:40,730,772-40,732,085), which was identified in the 3rd intron of the *isl1a* gene.

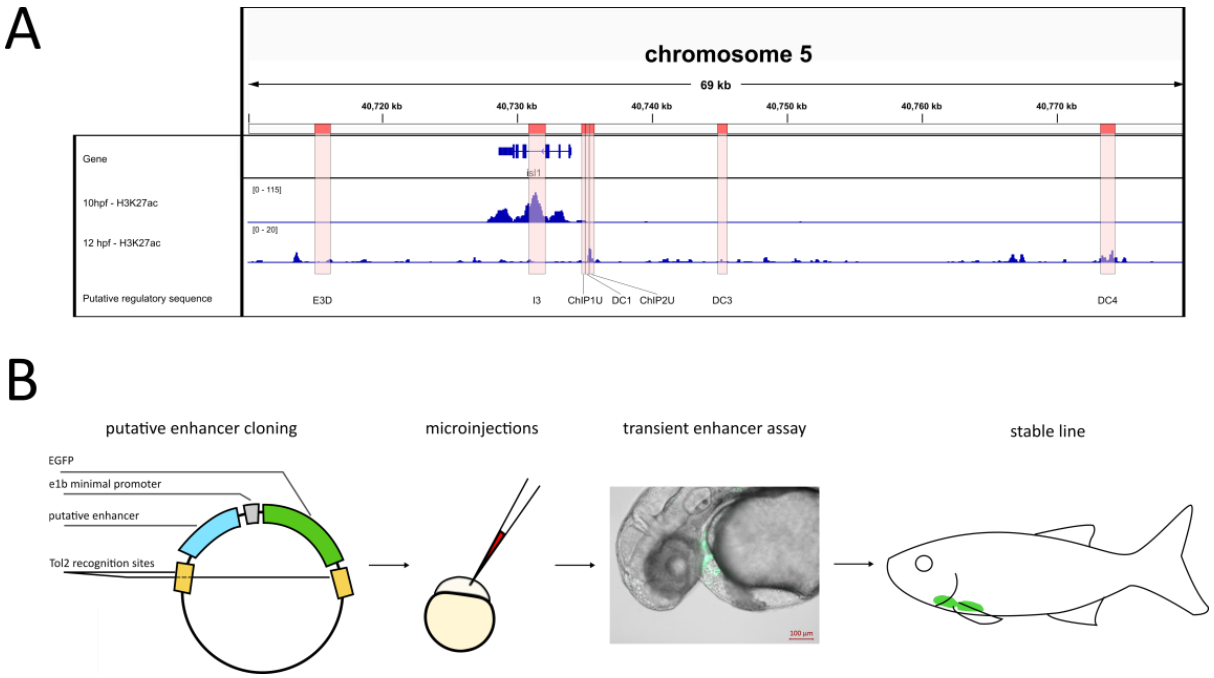


Figure 3. Graphical summary of the enhancer assay experiment. A. *in silico* search for putative regulatory regions surrounding the *isla* locus. "10 hpf - H3K27ac" and "12 hpf - H3K27ac" - tracks derived from publicly available ChIP-seq data. Red regions - sequences cloned into the enhancer assay vector. B. Summary of the experimental workflow. Putative enhancer sequences were cloned into a Tol2-compatible vector and injected into 1-cell stage zebrafish embryos. The embryos were screened for GFP expression. After identification of potential founder fish, stable transgenic lines expressing GFP under the control of putative enhancer sequences were generated.

2. 2. 2. Injected embryo survival is not enhancer construct-dependent

Following identification of the putative regulatory regions, I performed a Tol2-based enhancer assay. In this assay, a tested enhancer sequence is cloned upstream of a DNA sequence encoding GFP reporter under the control of a minimal e1b promoter (Birnbaum et al., 2012). The promoter-reporter construct is flanked by specific inverted terminal repeats, which allow Tol2 to integrate the flanked sequence into the zebrafish genome (Kawakami et al., 2007).

The vector contains the AmpR gene allowing for ampicillin selection of bacterial colonies. The construct is subsequently injected into the zebrafish embryo and the expression pattern of the reporter is then assessed to determine enhancer activity. I chose the e1b system (Li et al., 2010) as it was reported to exhibit low levels of background expression (Li et al., 2010, Begeman et al., 2022).

Each construct was co-injected with Tol2 transposase mRNA into at least 7 separate batches of embryos. For each batch prior to injections, I set aside a number of embryos as an uninjected control. The batches with survival rate under 60% in the uninjected control were discarded to rule out inherent developmental issues affecting this particular batch that might impact the result of the experiment.

I compared the mean survival rates of all of the viable injected construct batches with mean survival of uninjected embryos from each respective batch (Fig. 4). The survival of injected embryos was on average markedly lower than that of embryos that have not undergone the microinjection procedure. This is consistent with the expected outcome, since the injection itself as well as introduction of exogenous genetic material (plasmid and Tol2 mRNA) can significantly impact embryonic development and survival.

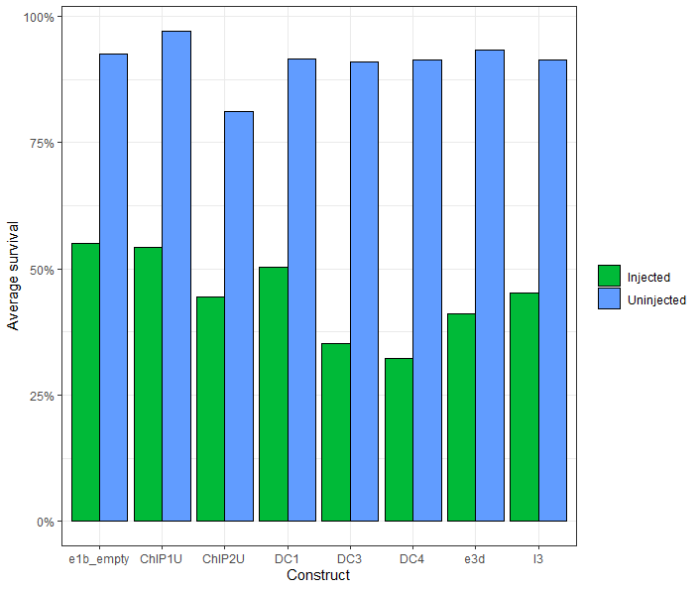


Figure 4. Comparison of mean survival rates for each enhancer-containing construct with matched uninjected embryos. The number of batches of embryos (n) analyzed varied per construct: n (e1b_empty) = 13; n (ChIP1U) = 7; n (ChIP2U) = 6; n (DC1) = 10; n (DC3) = 6; n (DC4) = 10; n (e3d) = 9; n (I3) = 10.

I then tested whether any specific regulatory region would be detrimental to embryo survival upon injection (Fig. 5). Out of the 7 enhancer constructs injected, 2 had on average higher survival rates than the embryos injected with the empty vector (e1b_empty), which served as injection control. The other 5 enhancer-containing constructs all had lower than average survival rates. However, I observed a large spread of survival rates within each condition, indicating that reliance on mean alone to gauge the construct-dependent survival was insufficient.

I compared each construct-injected group of embryos to those injected with empty e1b vector. The Shapiro-Wilk test for normality indicated that with the exception of the e3d construct all tested populations' survival rate follows the normal distribution. Therefore, for those constructs, I used pairwise comparison with Bonferroni correction applied to account for multiple comparisons. None of the comparisons produced significant ($p <= 0.05$) results, indicating that the mean survival of each construct does not differ significantly from the e1b_empty control. In case of the e3d construct I performed the Wilcoxon-Mann-Whitney test in which we do not assume the normality of data distribution. The result of the test ($p = 0.1508$) did not provide grounds for rejection of the null hypothesis, indicating that the difference in survival rates between the e3d and the e1b condition is not significant either.

Taken together, these results suggest that the sequence contained within each construct had no significant impact on the survival of the embryos. Rather, my observations indicate that all injected embryos suffered from a varying level of non-specific toxicity, likely resulting from injection of exogenous genetic material as well as the injection procedure itself. Nevertheless, this effect is generally low and a sufficient number of viable embryos could still be obtained for the experiment.

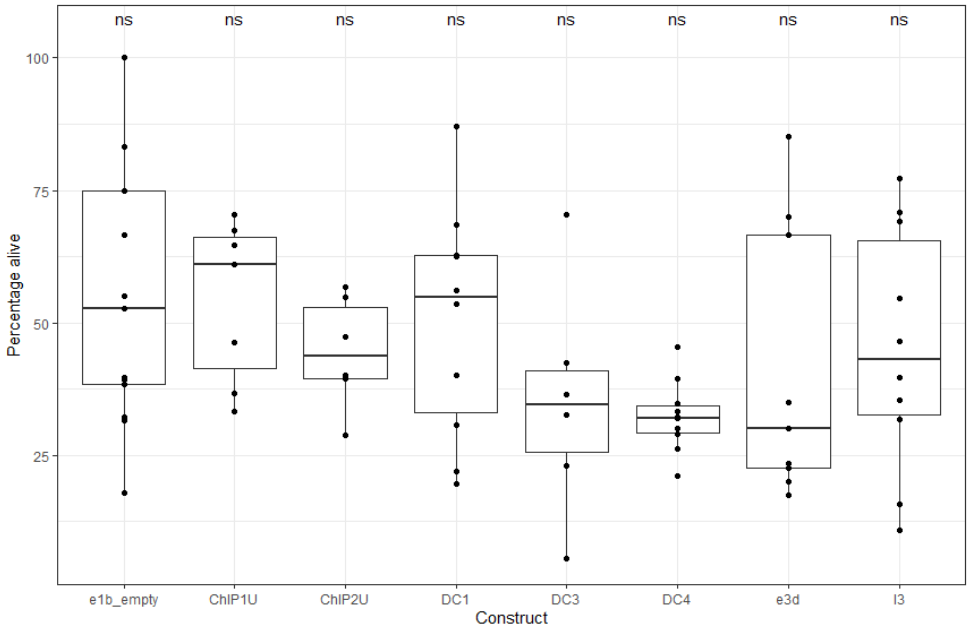


Figure 5. Percentage of surviving embryos depending on the regulatory sequence-containing vector injected. Dots represent batches of embryos injected with each construct. Line – median, whiskers – highest/lowest values within 1.5X interquartile range. The number of batches analyzed varied per construct: n (e1b_empty) = 13; n (ChIP1U) = 7; n (ChIP2U) = 6; n (DC1) = 10; n (DC3) = 6; n (DC4) = 10; n (e3d) = 9; n (I3) = 10.

2. 2. 3. Overall GFP expression is not predictive of specific enhancer activity

Next, I tested the effectiveness of inducing GFP expression across the putative regulatory sequences. To do this, embryos injected with each construct were compared to the GFP expression present in embryos injected with the empty e1b vector. The minimal promoter on its own should not be sufficient to drive GFP expression, barring events where the construct was integrated into the genome in a transcriptionally active *locus*. Therefore, I reasoned that a difference in frequency of GFP expression in the putative enhancer-injected zebrafish would be indicative of enhancer activity of the tested region.

I compared the expression pattern driven by the tested enhancers with that of the e1b_empty vector as background control. Of the 7 tested regions, 4 (DC1, DC3, DC4 and I3) resulted in more GFP-positive embryos on average than when injecting the vector not containing any putative enhancer sequence (e1b_empty) (Fig. 6). The Shapiro-Wilk test indicated that

except for ChIP1U (p = 0.017), data from all constructs was distributed normally. Bonferroni-corrected pairwise comparisons revealed that the average change in frequency of GFP-positive embryos from all tested constructs was not statistically significant compared to the e1b_empty vector. The ChIP1U construct was compared to the e1b_empty control using the Wilcoxon-Mann-Whitney test. The result of the test indicates that the difference between the two conditions was statistically significant (p = 0.007). Interestingly, 3 tested regions, ChIP1U, ChIP2U and e3d, produced on average lower than control levels of GFP expression. The difference was statistically significant (respective p-values from appropriate applied statistical tests: 0.007, 0.008, 0.001), suggesting that the regions tested might serve as silencers.

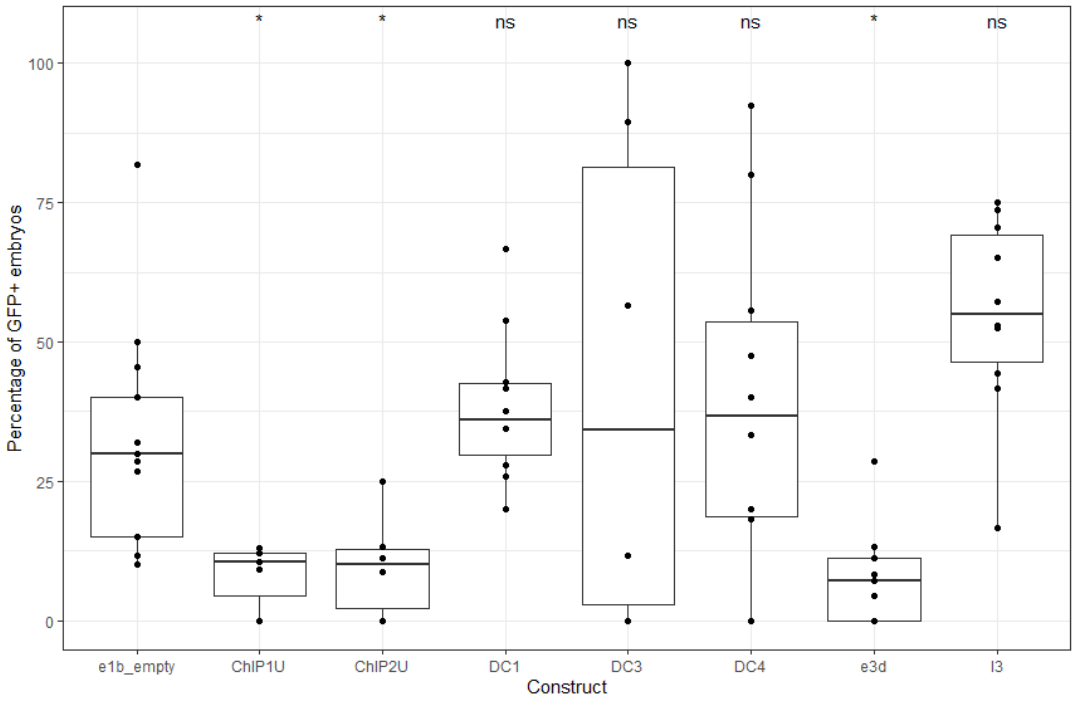


Figure 6. Percentage of GFP-expressing embryos depending on the putative enhancer sequence driving transgene expression. Each dot represents a batch of injected embryos. Line – median, whiskers – largest/lowest values within 1.5X interquartile range.

2. 2. 4. A regulatory region located in an intron of the *isl1a* gene induces heart-specific gene expression

Next, I tested whether the injected constructs drove GFP expression in the developing heart. I found that 4 out of 7 putative positive sequences showed some degree of GFP expression in the heart. However, none of them except one (I3), was able to drive heart GFP expression more frequently on average than the control construct containing only the minimal promoter

and GFP-encoding gene (Fig. 7). With the exception of the I3 construct, none of the tested enhancers produced normally distributed data. Pairwise comparison of e1b_empty and I3 indicates that the difference in GFP expression occurrence in the heart between these two constructs is statistically significant. The Wilcoxon-Mann-Whitney test indicates that none of the other constructs produced significantly different numbers of heart GFP-expressing embryos on average than the control condition.

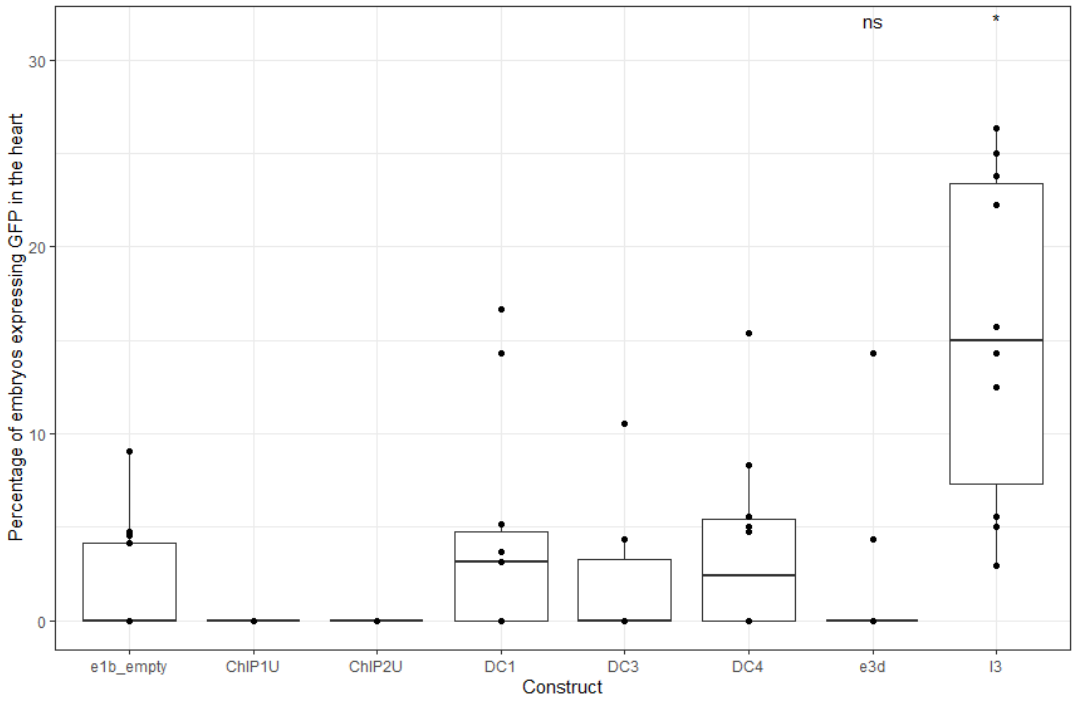
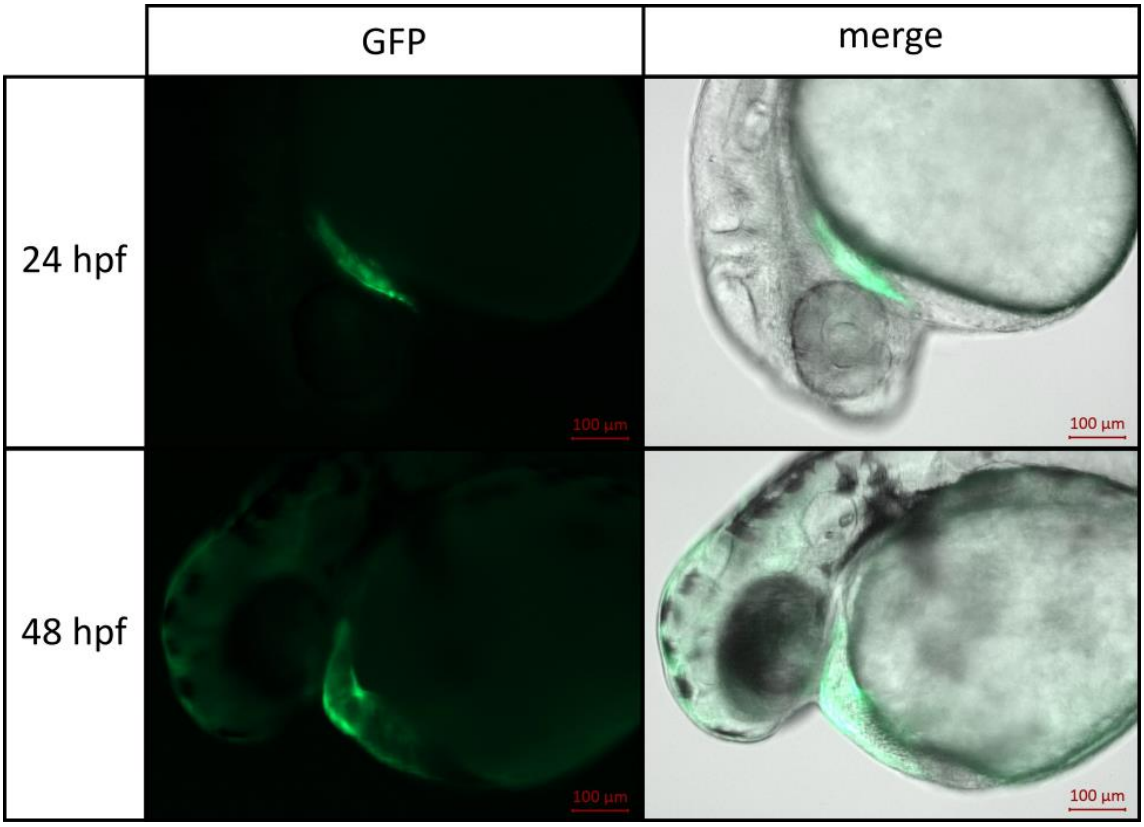


Figure 7. Percentage of embryos expressing GFP in the heart based on the putative enhancer sequence driving transgene expression. Dots represent batches of embryos injected with a given construct. . Line – median, whiskers – largest/lowest values within 1.5X interquartile range.

One of the putative enhancer constructs - I3, located within the 3rd intron of the *isl1a* gene (chr5:40,730,772-40,732,085, Fig. 8), stood out because it not only resulted in high overall GFP expression levels but also a high fraction of embryos with GFP expression limited to the developing heart (Fig. 8). This suggests that the I3 construct is capable of driving GFP expression in a tissue-specific manner, making it a highly favorable candidate for further investigation.



>5 dna:chromosome chromosome:GRCz11:5:40730772:40732085:1
CACATATGTTAAGCGGTCAGCTGTTAATTATTTTAACTCAGTTGTCTCCGAGGCCCATCTACGGTGGA
TTAACTTATGTGAAATAATAGGCAAATAATAATGTAGCCCATAGACTATAACAACAATAATAATACAT
TTAAAAACAATCACTCATGAAAAGGGAGTACTATTATTTTCAGCATATAATTTATGATCTGTTTAAGA
GACCGGTATTAAGAGGAAATCACTCCATACGCATAAAAATATATCCTTTAGCAATTATTTCTATATAG
AACTTTGGTCCTTCACTGATGTAGTATTTTTTCTGTATGTATTATGACTAAGAGAGCTTCCTTTTAAT
AAGTTTGAAACGCAGCCATGATCGCAATCCCTTTTCTTGCAATTATATGTTCCTTCTGCATGCAGTAA
ATTGTGAAGGGCGCGTATGACCCAGATAGGTTTACGACACTTAATTAAACCAATCTGTGCATACCTCT
GAATTTTAAATATTCAGGTGCGCGCGGTGTGCTGCCCATGACATTACTTTAAAGCTGCTGGTTATCAA
TAAACCTGCATTAACAACTGCACCATATAATTAAGATTATTATTTTATAAGAAATTAACCAAATAGGA
GGCATTAATGCAGTAAATGAAATAAAACACAGTAGTAGCCTACCATACTGCGGGTATTAATTAATTAA
TTTATTTTTAATGTTAAAACTCGATGCTTTTTTATTTAGACTGTTCTTAAATGCGGAATATTTCTCAT
GGTTGACACTGGTTTGTAAAAGCATAATAATTTAAAACGCATATATTAAAAGAGTAAAAAAAATCAAA
TGCTCGTTCATATGTAATTACACAAACACATAATATTGTGCTATAATATTTCTATAGCATATCAACAT
TTAGGATTGAATTACGTTTTCTTTGTATTTATTTGTGAATGGATCTTCTTTATTTCTCAAGTACATGC
ACAAGCTGACATTGTCGCGTGGTTGGCCTCTAATCAGCAGAGAATATCTATTTCATGCATGAATTGCT
GAATTCAATAACAGACAAACAAGAAAAATAAAGCGGGTATACTATCAAACCACCACACGCTTGTGGTT
TAGCTTGTTAATAAAATAATCATTCACGGATGGTTATGAGCCAGCATTCATTCATCATAAATCTCGGA
ATGCAGCAATATAATGTCATCTTAAACATTTGTGTTTATCTTAAAACATGCTAAAAATACTAGCCTAA
ATCTTCTGAGTAGATTTGGCTCTATAGAAATTCCTTCCCAAGACAAATAAACCCTGAGTCCTCGAGCA
AACAATGCTTTGTTAGTTGTAG

Figure 8. The I3 putative enhancer. Top: Micrographs of the heart and surrounding tissues in an I3-e1b:EGFP embryo at 24 and 48 hpf. GFP expression is defined and limited to the developing heart. Bottom: genomic sequence of the I3 putative enhancer.

2. 2. 5. I3-driven expression pattern marks the cardiomyocytes of the developing heart

In order to further characterize the expression driven by the I3 putative enhancer, I established stable transgenic lines. I selected 2 potential founders based on germline transmission of the transgene in the trial crosses. The 2 founders were each outcrossed with a wild-type fish and 50 embryos were selected based on presence of GFP expression to establish the stable F1 lines. Notably, the expression pattern observed even among the F0 fish was highly consistent, and after establishing the stable lines I did not observe any variation in GFP expression between individuals from each line. This might be caused by the fact that the Tol2 transposase facilitates rapid construct integration in the genome, resulting in little mosaicism among the transgenic embryos, as has been reported previously (Kawakami et al., 2007, Ni et al., 2016).

To determine the expression domain of the GFP reporter driven by the I3 enhancer, I crossed the heterozygous Tg(I3-e1b:EGFP) fish with a Tg(*myl7*:mRFP) reporter line, characterized by expression of mRFP protein in the cell membrane of early cardiac progenitors which later on develop into cardiomyocytes. I obtained a time series of confocal microscopy images from whole zebrafish embryos. Confocal imaging analysis revealed that the earliest I3 enhancer-driven GFP expression could be observed from 24 hpf, in the newly formed heart tube. By this time, the strongest GFP reporter expression could be observed in cardiomyocytes located throughout the developing heart tube, both in the atrium and the ventricle. Interestingly, the strongest expression of GFP was observed in the cardiomyocytes at the inner curvature of the atrioventricular canal. This pattern of expression was retained at 72 hpf when the distinctive heart chambers have developed. The GFP expression is retained at subsequent developmental stages, with the expression pattern remaining consistent for the remainder of embryonic development.

By 48 hpf, in addition to the strong expression in cardiomyocytes, GFP-positive cells were also observed in the inflow tract leading to the atrium (Fig. 9, arrowheads). These cells, while not as prevalent as those residing within the inner curvature of the atrioventricular (AV) canal, still contribute to the overall GFP expression in the heart.

Taken together, these observations suggest that the I3 putative enhancer located in intron 3 of the zebrafish *isl1a* gene drove gene expression in the early heart tube, particularly in cardiomyocytes which make up both cardiac chambers, with a higher expression in cells of the inner curvature of the AV canal. These results are consistent with the previous observations of SHF-derived cells, particularly near the poles of the heart and the ventricular myocardium (Witzel et al., 2017). However, the expression pattern I observed remains broader than the previously defined *isl1a* gene expression in the heart. This points to two likely reasons: either 1) the previously observed *isl1a* expression domain had been underestimated, or 2) the I3 enhancer-dependent expression is present in cells of FHF origin in addition to SHF-derived cells, suggesting that its regulatory function begins prior to the two lineages splitting.

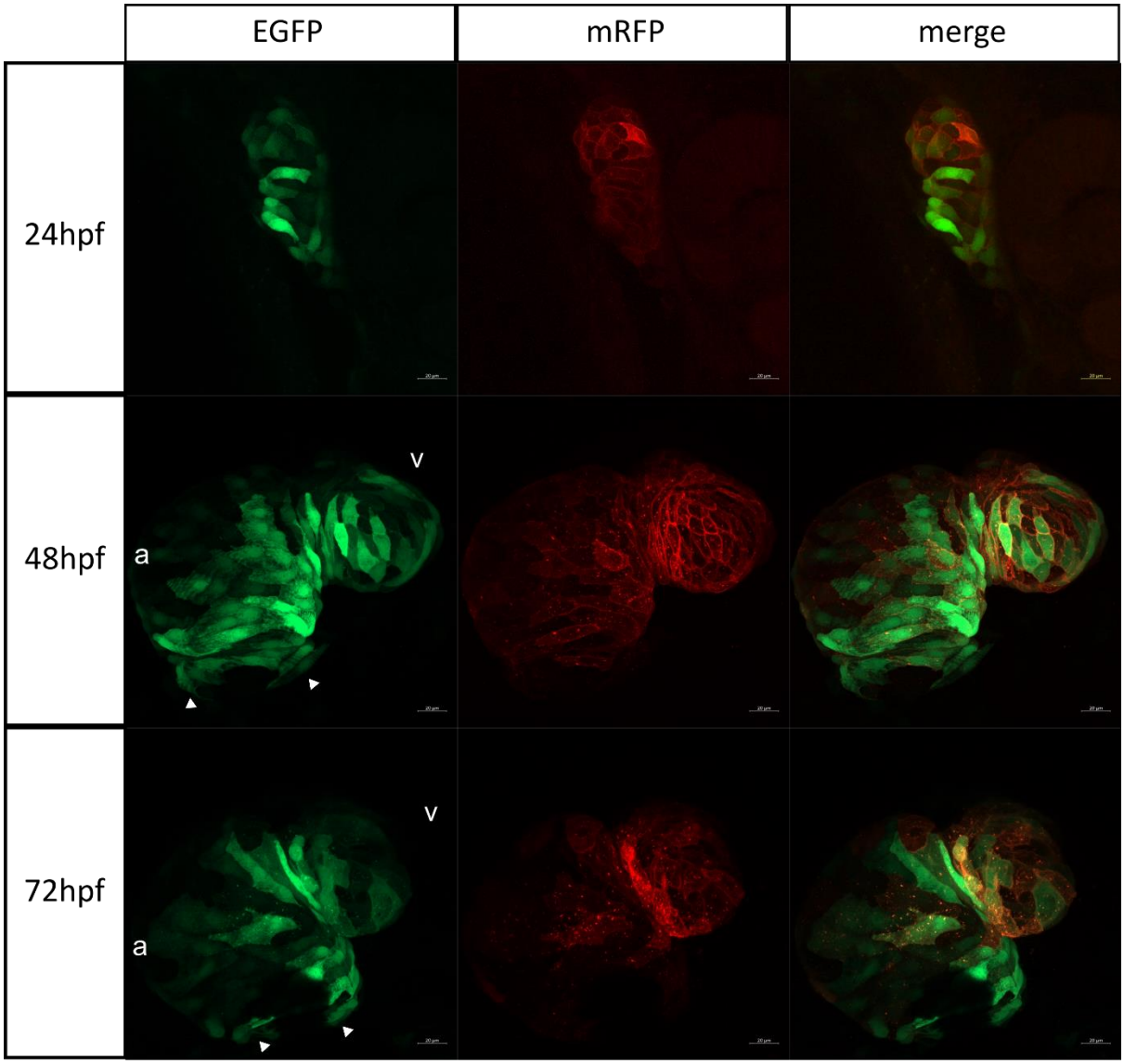


Figure 9. I3-driven GFP expression in the developing zebrafish heart. Live embryos were mounted in low-melting agarose and imaged using confocal microscopy at 24, 48 and 72 hpf. green - GFP expression driven by the I3 enhancer. red - mRFP expression driven in cardiomyocyte cell membrane by the *myl7* promoter in the Tg(*myl7*:mRFP) line. a - atrium; v - ventricle; arrowheads - inflow tract (IFT).

2. 2. 6. The I3 enhancer driven gene expression domain is enriched in SHF markers

To determine whether the expression domain of GFP driven by the I3 enhancer indeed encompasses the SHF, I examined the expression of known SHF marker genes in this cell population.

The *isl1a* gene is implicated in SHF specification and its expression is known to serve as a marker of this specialized cell population within the heart. Murine *Isl1* has been shown to mark cardiac progenitor populations contributing to the majority of heart structures, including the outflow tract, right ventricle and large parts of both atria (Cai et al., 2003). In zebrafish, *isl1a* has been identified as a key regulator of cardiomyocyte differentiation at the venous pole of the heart (de Pater et al., 2009, Witzel et al., 2017).

Nkx2.5 is another transcription factor implicated in cardiac progenitor specification. In vertebrates, it is considered one of the first markers of specified cardiac cells (Harvey, 2002, Witzel et al., 2012, Swedlund and Lescroart, 2020). *tinman*, the *Drosophila* ortholog of *nkx2.5*, is viewed as the master regulator of invertebrate cardiogenesis (Swedlund and Lescroart, 2020). However, in mice *Nkx2-5* mutants are able to form an early heart tube, indicating that the gene's role in FHF expansion and specification has been lost over the course of evolution, possibly due to redundancy introduced by the expansion of the Nkx family. On the other hand, it has been shown to regulate cardiac ion channel formation and function, influencing electrical activity of the myocardium (Furtado et al., 2016). *Nkx2-5* deficiency was also shown to cause outflow tract truncation due to lack of proliferation of SHF cells (Prall et al., 2007), indicating that its role in the SHF has been largely retained.

I isolated GFP-expressing cells from embryos of the transgenic line Tg(I3-e1b:EGFP) at 24 hpf using fluorescence-activated cell sorting (FACS). The protocol was adapted from the protocol described in our previously published work (Migdał and Tralle et al., 2021). For each run of the experiment, I used 500 GFP-positive embryos of the Tg(I3-e1b:EGFP) transgenic line. During the course of cell dissociation, I typically obtained 10-15 million cells of high viability, ranging from 87 to 98% of all cells. To help identify the GFP-positive cell population during the sorting, I simultaneously prepared a suspension of dissociated cells from wild-type embryos of the same developmental stage to serve as a negative fluorescence control.

The cells of interest typically constituted between 0.31% to 1.3% of the overall cell population (Fig. 10), which was consistent with results I obtained from a similar experiment using the Tg(*nkx2.5*:EGFP) line, which also expresses GFP in heart cells, described in Chapter 3.

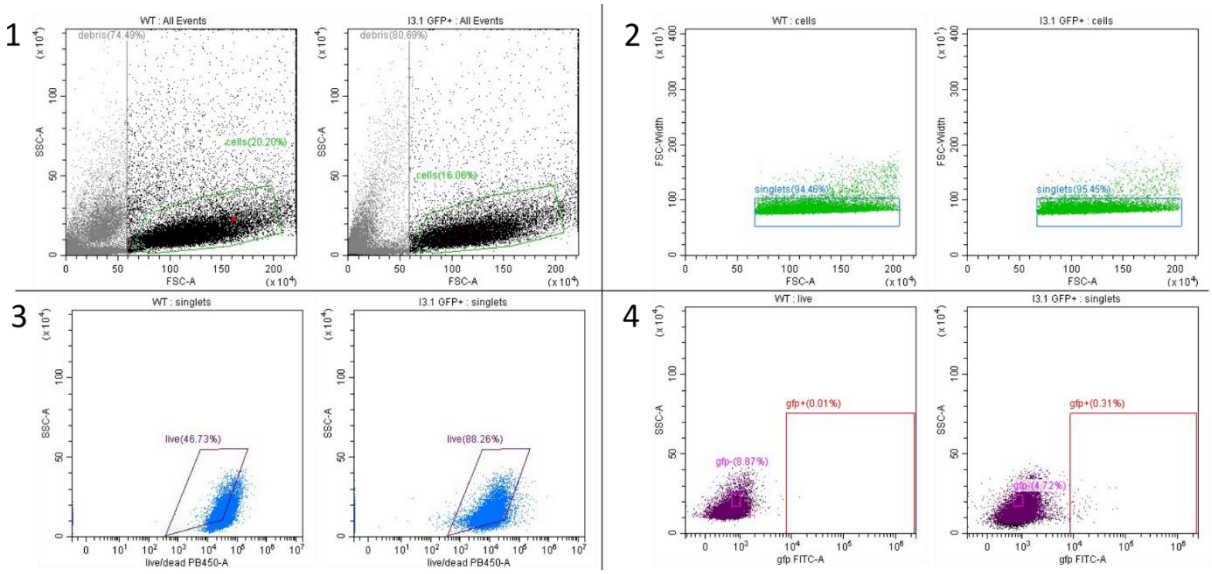


Figure 10. Sorting strategy for GFP+ cells from the I3 enhancer line. (1) Cells were identified according to their size and granularity. (2) Doublet discrimination was performed based on cell area to width ratio. (3) Dead cells were discarded based on PB450 fluorescence intensity, measured thanks to prior cell staining using LIVE/DEAD violet fixable stain. (4) GFP+ and GFP- cells were identified based on FITC intensity.

The sorted cells were then isolated and had their total RNA extracted. In order to determine their identity, I performed RT-qPCR and assessed the gene expression of selected markers in GFP+ cells as compared to the simultaneously sorted GFP- cells (Fig. 11). As described above, *isl1a* and *nkx2.5* are used to assess the presence of SHF cells in our sorted cell population. Additionally, to ascertain their identity as cardiac progenitors, I tested the expression of other key cardiac marker genes, namely *myl7* (Huang et al., 2003, Holtzmann et al., 2007, Capon et al., 2022), and *nkx2.7* (Targoff et al., 2013). As a negative control, I selected *nkx2.3*, a gene expressed by endodermal cells located in the pharyngeal region. As these cells are not directly contributing to heart muscle formation but rather are implicated in the development of craniofacial skeleton and thymus (Li et al., 2019), and are residing in close proximity to cardiac progenitors, the expression of *nkx2.3* could therefore identify potential issues in my cell dissociation and sorting approach.

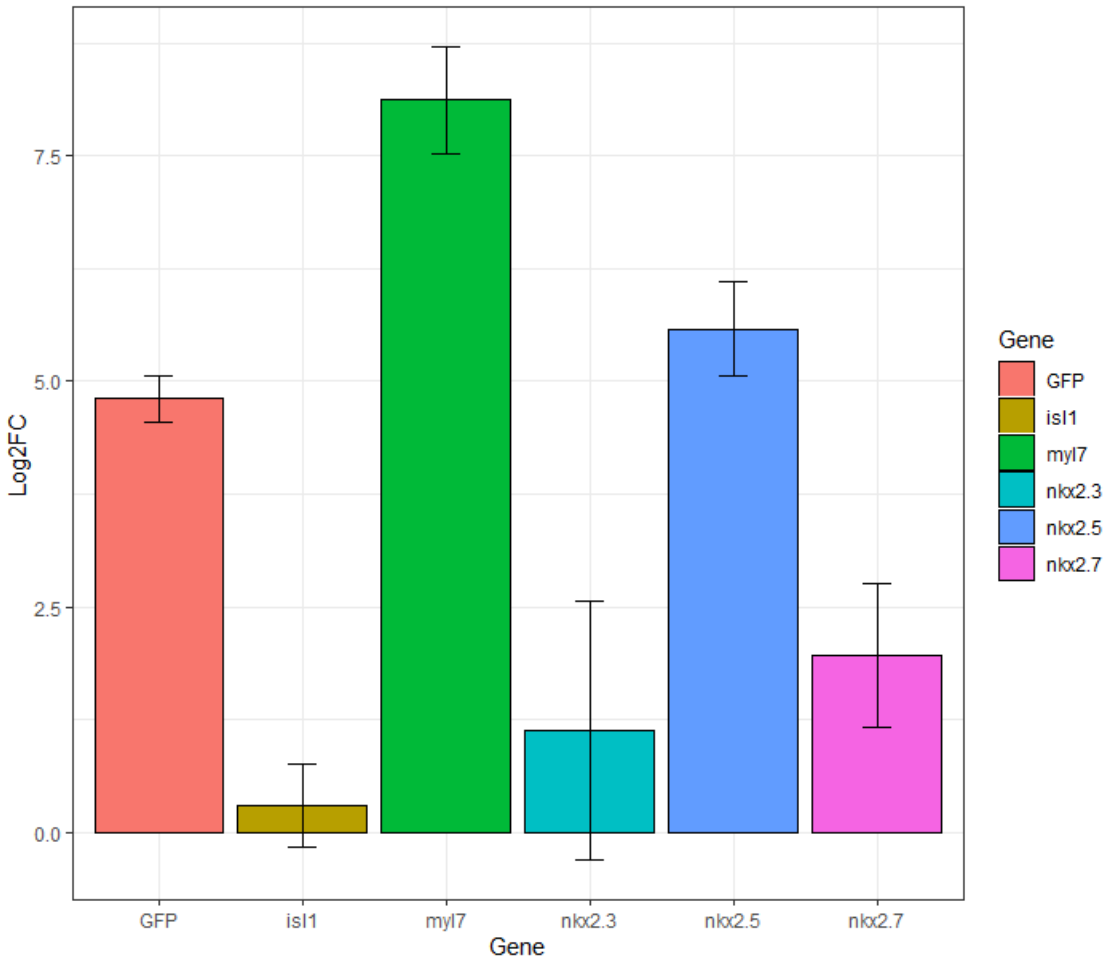


Figure 11. Key marker expression in GFP+ cells sorted from the I3 enhancer line. cDNA from each biological replicate ("Replicate group") was split into 3 technical replicates and compared to similarly split biological replicates of GFP- cells obtained from the same samples. Each subplot depicts LFC in expression of a key marker gene in GFP+ cells compared to the same gene's expression in GFP- cells (y = 0), calculated based on the ΔΔCt method. Error bars represent standard deviation.

Confirming our findings from confocal microscopy, the GFP+ cells from the I3-enhancer line express the canonical cardiomyocyte marker *myl7* more abundantly than the GFP- cells. Similarly, the key cardiac developmental factors *nkx2.5 and nkx2.7* were more highly expressed in the GFP+ cells compared to GFP-. The expression of *isl1a* which marks the SHF could be detected in the GFP+ cells. However, it is not highly enriched when compared to that of GFP-. This could be explained by the fact that *isl1a* is also expressed in other tissues including the nervous (Uemura et al., 2005) and endocrine (Wilfinger et al., 2013) systems, which also constituted the GFP- sorted cell population. On the other hand, *nkx2.3*, which is typically associated with the non-cardiac endodermal progenitor population

in the pharyngeal region, is present at lower levels than the cardiac progenitor and cardiomyocyte markers. The expression of these markers in the GFP+ cell population collectively suggest that the GFP+ cells representing the expression domain driven by the I3 enhancer are of cardiomyocyte origin and likely include the SHF due to the expression of *isl1a* and *nkx2.5* markers.

Taken together, the gene expression analysis of the GFP+ cell population suggests that they likely encompass the cell population representing the SHF. However, the absence of a specific SHF marker did not allow us to rule out the possibility that additional cells other than those of the SHF are also present within the GFP+ population.

2. 2. 7. I3 regulatory element contains binding sites for key cardiac transcription factors

Next, I investigated which transcription factors could potentially bind to my set of putative regulatory sequences and trigger related gene expression. To this end, I performed transcription factor binding motif search *in silico* using FIMO (Grant et al., 2011), part of the MEME suite of tools designed for motif-based sequence analysis. While none of the available databases contained information on zebrafish transcription factor binding motifs, I searched for general vertebrate transcription factor binding motifs using the JASPAR CORE motif database. This allowed me to scan the sequences for the presence of binding sites for 841 non-redundant TFs. This approach was informed by the fact that there is a high degree of conservation of developmentally relevant transcription factor binding sites (TFBSs) across vertebrate taxa (Taher et al., 2011, Hombach et al., 2016). I designed the search so that I could identify TFs that would be more likely to bind to my sequences than to a set of sequences resulting from scrambling and nucleotide-adjustment of my sequences.

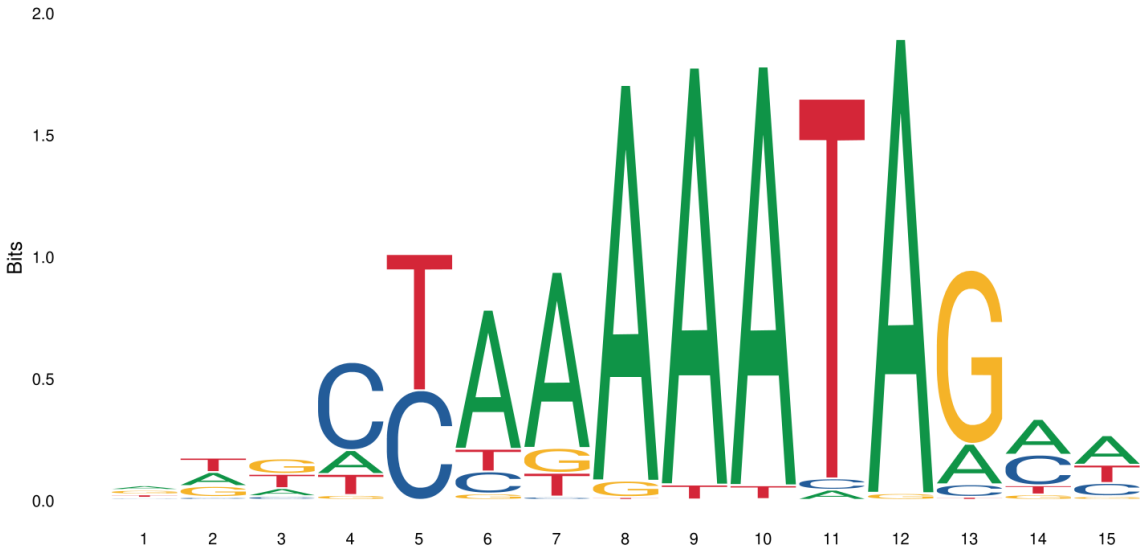
I identified 70 transcription factors that had a higher probability of binding to the I3 sequence than to any other of the tested elements around the *isl1a* locus, as well as the scrambled set. I compared the list of 70 TFs to the expression data in the ZFIN database. 8 transcription factors previously identified in the heart were present on the list: POU6F, HOXC9, SOX18, GFI1B, PAX9, TEAD3, HNF1B and MEF2C.

Additionally, since enhancer function is typically conserved across vertebrates, I decided to search for conserved transcription factor binding sites in other vertebrates. I obtained the Isl1 gene intron 3 sequences of human, mouse, western clawed frog (*Xenopus tropicalis*) and green anole (*Anolis carolinensis*). *X. tropicalis* and *A. carolinensis* are used as model organisms to study the biology of their respective evolutionary classes (amphibians and reptiles, respectively), and were chosen for this comparison as their genomes have been fully sequenced and assemblies are publicly available (Hellsten et al., 2010, Alföldi et al., 2011).

First, I performed local multiple sequence alignment using Kalign (Lassmann et al., 2005) with standard settings for nucleic acid alignment (gap open penalty = 217, gap extension penalty = 39,4, terminal gap penalties = 292,6). I decided to focus on local alignments as the intron 3 sequences from the compared species were varied in length, from 1417 to 8469 bp, making end-to-end alignments suboptimal. The overall sequence conservation was low, ranging from 37,05% between zebrafish and human, 37,54% between zebrafish and mouse, 39,69% between zebrafish and green anole to 57,46% between zebrafish and *Xenopus*.

Next, I scanned the intron 3 sequences for TF motifs that were present in the zebrafish I3 sequence as well as in at least one other vertebrate. The search resulted in a list of 54 TFs, including all of the cardiovascular TFs found in the I3 sequence except for HNF1B.

Of the identified binding sites, 4 - POU6F, SOX18, GFI1B and TEAD3 - were only present in tested sequences from zebrafish and *A. carolinensis*. PAX9 binding sites were present in zebrafish, *A. carolinensis* and *X. tropicalis*, whereas HOXC9 binding sites were identified in all tested vertebrates but mice. Interestingly, MEF2C binding sites were found in all introns 3 of the Isl1 gene in all vertebrates tested.
Mef2c (Fig. 12) is a pan-cardiac gene, expressed in the cells that contribute to the formation of the OFT and the right ventricle in mice (Dodou et al., 2004). In addition, its function in anterior SHF (a-SHF) was linked to *Isl1*, both as a direct transcriptional target of *Isl1* (Dodou et al., 2004), and through *Isl1*/*Ldb1* complex regulating *Mef2c* expression, among others (Caputo et al., 2015).



Motif ID	Alt ID	Sequence Name	Strand	Start	End	p-value	q-value	Matched Sequence
MA0497.1	MEF2C	I3	+	1209	1223	1.75e-05	0.145	ATGCTAAAAATACTA
MA0497.1	MEF2C	I3	-	173	187	9.2e-05	0.33	ATGCTGAAAATAATA

Figure 12. I3 enhancer contains binding sites for the MEF2C transcription factor. Top: Sequence logo of the human MEF2C transcription factor binding site. Bottom: output from the FIMO tool. The I3 sequence was compared to the other sequences tested as part of the enhancer assay for occurrence of common cardiac transcription factor binding sites.

2. 3. Discussion and future work

In this part of the project, I have shown that *in silico* approaches in combination with enhancer assays can be leveraged to gain valuable insights on gene expression in understudied cell populations. I sought to identify a regulatory region responsible for driving *isl1a* expression in the heart progenitors. I identified a regulatory region located in an intron of the *isl1a* gene that drives gene expression in a tissue-specific manner. This is consistent with the established view that tissue-specific enhancers can be often found in noncoding regions, including introns (Gillies et al., 1983, Borsari et al., 2021).

Notably, the expression triggered by the I3 enhancer was highly uniform even in the F0 specimens, indicating that the transgene insertion occurs relatively fast, resulting in a high number of transgenic cells in a given embryo, which in turn facilitates establishing of stable transgenic lines. This further confirmed the Tol2 transgenesis system as a viable and convenient

method of obtaining transgenic zebrafish as it was previously established (Parinov et al., 2004, Kawakami, 2007).

Several lines of evidence suggest that the I3 enhancer drives gene expression in the SHF-derived cardiac progenitors. Confocal microscopy observations revealed that the heterozygous Tg(I3-e1b:EGFP) x Tg(*myl7*:mRFP) embryos exhibit a heart-specific GFP expression pattern. The GFP expression is prevalent in cardiomyocytes, most notably in the cells located in the inner curve of the atrioventricular canal, with additional GFP+ cell population residing in the inflow tract. This expression pattern largely overlaps with the previously described domain of *isl1a/isl2b* expression in the heart, as reported by Witzel et al. (Witzel et al., 2012, Witzel et al, 2017). The expression of *isl1a* and *isl2b* indicate cells originating from the SHF which are incorporated into the elongating heart tube. The aforementioned works utilize antibody staining to visualize *isl1a* expression, which, as the authors note, recognize both *isl1a* and *isl2b*. Upon *isl1a* knockout, the *isl1a/isl2b*-positive cell population at the venous pole of the heart is lost, without impacting the atrioventricular cell population. This indicates that the group of cells marked by GFP expression in my transgenic enhancer line likely represent SHF originating progenitors. On the other hand, the I3 enhancer drives gene expression across the majority of the myocardium, suggesting that while it encompasses the SHF population, it is also capable of driving gene expression in the FHF.

A second line of evidence was provided by the RT-qPCR gene expression analysis of GFP+ cell population. Here, I found that the GFP-expressing cells are expressing the canonical SHF markers *isl1a* and *nkx2.5*, as well as cardiomyocyte progenitor marker *myl7*. This further indicates that the cells marked by GFP expression encompass, but are likely not limited to, the SHF.

Next, I performed TF motif analysis to identify potential TFs that could bind to the putative enhancer sequence located in intron 3 of the *isl1a* gene. I compared this sequence to all the other putative enhancer sequences tested in my assay that did not drive heart-specific gene expression, reasoning that the TFs that bind to the I3 sequence and not the others could influence the observed enhancer-dependent gene expression pattern. I found 70 transcription factors that can bind the I3 sequence but not the others. From this list, 8 TFs were previously identified in the zebrafish heart.

Then, I conducted a comparative analysis of the I3 sequence and Isl1 intron 3 sequences from other vertebrates, namely human, mouse, western clawed frog and green anole. Overall, the sequences are not well-conserved, with all percentage identity scores lower than 70%. However, scanning them for transcription factor binding sites identified 54 transcription factors that can bind to the I3 sequence as well as the Isl1 intron 3 sequence of at least one other of the tested vertebrates. Interestingly, from the 8 cardiac TFs identified in the previous analysis only one, MEF2C, had its binding sites conserved across all tested taxa.

Mef2c is a well-characterized cardiac transcription factor with an established role in murine OFT and ventricle development (Dodou et al., 2004). It has been described as a pan-cardiac factor necessary for embryonic development, as its early onset loss of function in the myocardium results in severe cardiovascular defects and embryonic lethality (Materna et al., 2019). Its role in heart development has also been described in zebrafish, where due to the genome duplication two partially redundant copies of the gene, *mef2ca* and *mef2cb* are present (Hinits et al., 2012). Additionally, its interaction with *Isl1* has been studied in mice, indicating its *Isl1*-dependent role in early heart development (Dodou et al., 2004) as well as in zebrafish (Witzel et al., 2017).

Taken together, these results have presented evidence supporting the function of the I3 enhancer in driving gene expression in the heart, and more specifically, in the SHF. However, it is important to note several limitations of my approach.

Firstly, while I show that the expression of GFP driven by the enhancer in the I3 line is limited to the developing heart, direct evidence on co-localization of GFP and *isl1a* within the same cell population is yet to be observed. A possible approach is to perform double-color whole mount in situ hybridization to visualize GFP and *isl1a* expression simultaneously. However, this technique proved to be challenging to optimize and is not yet established in our lab. To circumvent this issue, I isolated the GFP+ cells and performed RT-qPCR to assess the expression of various markers including *isl1a.* While I did not observe a significant enrichment of *isl1a* in the GFP+ cells due to its expression in other cell types, my analyses indicated a high level of *isl1a* expression which provided strong support for the co-localization of the two markers.

While I observed a clear cardiac expression pattern of GFP, both on the protein as well as RNA level starting at 24 hpf, I was unable to detect GFP expression from earlier timepoints. The expression of GFP at earlier stages, particularly prior to the formation of the heart tube, would have provided crucial information on the SHF origin of GFP+ cells in my transgenic line and further link the observed expression pattern with the identity of GFP+ cells. One possible way to obtain the information on GFP expression at the earlier time point is to perform the assessment at the level of its mRNA by whole mount in situ hybridization. It is known that there is a delay between transcription and GFP maturation and accumulation, where the expression of GFP is initiated hours earlier than the mature protein accumulation to the point where it becomes observable under a confocal microscope (Balleza et al., 2018). Therefore, an *in situ* RNA hybridization experiment would enable the detection of the GFP transcripts earlier than the observable fluorescence from the mature protein.

A second limitation to my analysis was posed by the limited knowledge on protein interactions. Although there is existing experimental data suggesting a relationship between *Isl1* and *Mef2c*, *Isl1* was rather known as a co-regulator, together with *Ldb1*, of *Mef2c* (Caputo et al., 2015). This existing knowledge does not provide a clear explanation for the presence of a MEF2C binding site in a regulatory region of *isl1a*. The zebrafish genome contains two orthologues of *MEF2C* - *mef2ca* and *mef2cb*. Possible future investigations could involve *mef2ca/b* loss of function experiment focusing on their impact on *isl1a* expression and the phenotype in my enhancer transgenic line. This would allow us to gain a more comprehensive understanding of the relationship between *mef2ca/b* and *isl1a* as well as its role in heart development.

Another possible approach that would further our understanding of the regulatory mechanism of the I3 enhancer is to systematically delete fragments of the enhancer, specifically targeting the TFBSs identified using FIMO scanner and JASPAR CORE motifs. This part of my analysis could be extended to include other vertebrates, especially those of which whole genome assemblies are more complete since the conservation of some TFBSs between zebrafish and green anole but not western clawed frog could indicate issues with the genome assemblies. Deletion of binding sites for other conserved TFs that had previously been reported as active in heart development could provide especially interesting insight into the function of the I3 enhancer and how it has changed throughout evolution. Further insight into conservation

of function of this enhancer could be gained by testing the ability of TFs from other vertebrates to activate the I3 enhancer in zebrafish embryos.

Overall, these additional experimental approaches would provide crucial evidence to support and validate my findings, helping to enhance our understanding of the mechanisms governing gene expression in the developing heart, specifically the SHF. Additionally, a more in-depth functional analysis of the I3 enhancer would provide further insight into the general mechanisms of gene expression regulation during embryonic development.

3. Time-course transcriptome analysis of the developing zebrafish cardiac progenitors at single-cell resolution

3. 1. Materials and methods

3. 1. 1. Zebrafish embryo collection and staging

Zebrafish embryos utilized in this experiment were collected immediately after fertilization and split into ventilated Petri dishes filled with E3 medium, 50 embryos per dish. The dishes were then put in an incubator at 28.5℃ with light/dark cycles of 14/10 h. After 3 hours of incubation, the embryos were checked for proper development and any unfertilized eggs were discarded. Then, one hour before the set dissociation time, the embryos were screened under the microscope and the abnormally developing embryos were discarded. Embryo staging was done according to standard criteria (Kimmel et al., 1995).

3. 1. 2. Morpholino-mediated *nkx2.5* and *nkx2.7* knockdown

The *nkx2.5* and *nkx2.7* knockdown was achieved by microinjecting morpholinos (MOs) against *nkx2.5* and *nkx2.7* (GeneTools, USA). The morpholino stock concentration was measured with NanoDrop, following the morpholino manufacturer's guidelines. The morpholino injection experiment and phenotype analysis was conducted following the methods described in our previously published work (Nieścierowicz et al., 2022).

Prior to injection, a mix was prepared containing 1.5 ng/µl of each morpholino, as well as nuclease-free water and 0.1 X Phenol Red, which serves as both a pH indicator in the mix and a visual cue that a droplet has been successfully deposited inside the embryo. Before being loaded into the needle, the mix was incubated at 65°C and cooled down to room temperature, according to the manufacturer's instructions.

Each embryo was injected at 1-cell stage with 1 nl of the morpholino mix. A part of each batch was set aside prior to injection to serve as control. Aside from not being injected, these embryos were treated in the same manner as the injected embryos. After injections, the embryos were transferred into ventilated Petri dishes containing fresh E3 and incubated at 28.5°C.

After 3 hours of incubation, the dishes were checked, any unfertilized eggs were discarded and the number of embryos was noted.

At 24 hpf, survival of the embryos was estimated as described in 2. 1. 5. The surviving embryos were transferred to fresh E3 and maintained at 28,5 °C. At 48 hpf, the injected and control embryos were screened for phenotypes to assess knockdown penetrance.

3. 1. 3. Embryonic cell dissociation

Five hundred embryos of the Tg(*nkx2.5*:EGFP) transgenic line at the desired developmental stage were pooled in a glass beaker and dechorionated by treatment with 0.2 mg/ml Pronase (Sigma-Aldrich) in E3 medium for 5 minutes. Afterwards, the embryos were washed 5-10 times with E3 medium, discarding the broken chorions. Dechorionated embryos were transferred into 5 ml Eppendorf Lo-Bind tubes and the E3 medium was drained. Then, 2 ml of de-yolking buffer (55 mM NaCl, 3.6 mM KCl, 1.25 mM $NaHCO_3$) was added to the tube and manually aspirated with a 1000 µl low retention pipette tip (Axygen) 10-20 times, until the solution turned cloudy, but the embryonic bodies were still intact. The tube was then centrifuged at 500 rcf in 4°C for 30 seconds.

The supernatant was discarded, and the embryos were resuspended in a dissociation mix, comprising 1920 µl of 1X Trypsin (Gibco) in PBS and 80 µl of 100 mg/ml collagenase (Sigma-Aldrich). The embryos were then dissociated in a Thermomixer (Eppendorf) at 30°C, 300 rpm, for 5-10 minutes. Every 1 minute the contents of the tube were manually disrupted by aspirating with a 1000 µl low retention tip mounted on an automatic pipette set to 700 µl to avoid flooding. Once the clumps of tissue were no longer visible, an equal volume of stopping solution (L-15 medium, 10% FBS) was added to the tube and mixed thoroughly. Then, the tube was transferred to the centrifuge set to 4°C and centrifuged for 10 minutes at 700 rcf. After centrifugation, the supernatant was discarded, and the cell pellet was resuspended in 3 ml of 1X PBS and passed through a 40 µm cell strainer (VWR) onto a glass Petri dish set on ice.

The cell suspension was then placed in a 5 ml Lo-Bind Eppendorf tube and LIVE/DEAD violet fixable stain (ThermoFisher) was added. The staining was conducted according to the manufacturer's recommendations. Afterwards, the cell solution was centrifuged again

at 4°C, 700 rcf for 10 minutes, followed by discarding the supernatant and resuspension of the cell pellet in 0.4% BSA in 1X PBS. The solution was then strained through a 35 µm nylon mesh strainer (Falcon) into a 5 ml round bottom polystyrene tube (Falcon) and placed on ice.

3. 1. 4. Cell sorting and viability assessment

The dissociated cells were sorted using the CytoFlex SRT (Beckman Coulter, USA) equipment. First, the cells were distinguished from debris using the FSC-A/SSC-A parameters. Then, the identified cells were size-selected to exclude multiplets using the FSC-A/FSC-W parameters. The live cell population was identified according to the intensity of PB450 signal, reacting to the LIVE/DEAD violet fixable stain (ThermoFisher) used in the dissociation step. Finally, the GFP+ and GFP- populations were identified based on the FITC signal. For single-cell library preparation, live GFP+ cells were sorted using the purity mode with a sort guard band at 25%, which sorts a droplet only if the 25% of the volume of the two droplets immediately preceding and following the investigated droplet does not contain any particles deemed as potential contaminants. The cells were sorted into 600 µl of 0.4% BSA in PBS in a 1.5 ml Lo-Bind Eppendorf tube.

After sorting, the cells were concentrated by centrifugation in 4°C at 700 rcf for 10 minutes. Then, the supernatant was discarded, leaving 40 - 50 µl of the solution at the bottom of the tube. This volume was used to resuspend the cells. To quantify the cell number, 10 µl of concentrated cells were transferred with the same pipette tip into a 0.2 ml Lo-Bind Eppendorf tube and mixed with an equal volume of Trypan Blue (NanoEnTek). This mixture was then split and loaded into two chambers of a cell counting slide (NanoEnTek, South Korea). The slides were then loaded into the Countess automated cell counter (Thermo Fisher, USA). The number of live cells was estimated by averaging the readouts from the two chambers used.

3. 1. 5. qPCR-based quality control of cell sorting

In order to validate my cell sorting strategy, an equal number of GFP+ and GFP- cells was collected into separate tubes, each containing 600µl of Trizol LS (Thermo Fisher, USA). The RNA from each sample was extracted using the Direct-zol RNA MicroPrep kit (Zymo Research, USA). RNA yield and integrity was estimated with the TapeStation system (Agilent, USA).

Next, cDNA was reverse-transcribed using Maxima H- RT (Thermo Fisher, USA) from equal amounts of RNA across GFP+ and GFP- samples. Finally, the amount of cDNA was estimated based on TapeStation (Agilent, USA) results and the cDNA concentration across samples was adjusted once again.

3. 1. 6. qPCR efficiency assessment

In order to establish the reaction efficiency for each primer pair used in the qPCR reactions, a separate experiment was set up using the same lots of reagents and disposables, as well as the same instrument as used in the rest of the experiment. In addition, a similar biological sample was used for primer efficiency estimation as that used to generate the single-cell libraries, being cDNA from GFP+ 24 hpf zebrafish embryonic hearts.

The standard cDNA was serially diluted 5 times to range from 4 ng/µl to 0.0003 ng/µl per reaction in order to span across the potential range of concentrations in the final samples. Each reaction was then set up using the FastStart Essential DNA Green Master (Roche, Switzerland), a set amount of cDNA, and a pair of primers to be tested. After amplification, a standard curve was calculated for each primer pair using the LightCycler software (Roche, Switzerland), which then served as a basis for reaction efficiency estimation. Utilizing the efficiency estimate established, sorted samples from each time point were tested prior to following the 10X protocols in order to verify that the cell sorting resulted in successful enrichment of the GFP+ population.

3. 1. 7. GEM generation, library preparation and sequencing

The single cell libraries were prepared using the Chromium Next GEM Single Cell 3′ LT Kit v3.1 (low throughput), or the Chromium Next GEM Single Cell 3’ GEM, Library & Gel Bead Kit v3.1 (10X Genomics, USA) according to the manufacturer’s protocol. Briefly, live cells were loaded together with oligo-coated gel beads as well as lysis and RT reagents into the 10X Controller (10X Genomics, USA), where encapsulation into Gel Beads in-emulsion (GEMs - a term used by 10X Genomics to indicate a droplet containing a bead, a cell and reagents) took place. This results in generation of cDNA molecules from single cells that share a common barcode. Additionally, each individual cDNA molecule carries a unique molecular identifier (UMI). cDNA is the PCR-amplified to generate sufficient amounts of genetic material for library construction. Library preparation begins by fragmentation,

end-repair, ligation of sample indexes and Illumina sequencing adapters, followed by final PCR amplification. Throughout library preparation stages, cDNA concentration and library size are estimated using the TapeStation 2100 (Agilent).

The libraries were sequenced at the Genomics Core Facility (CeNT, Warsaw) using the Illumina NovaSeq 6000 paired-end 2x100 bp sequencing to a depth of 50000 reads per cell. Read 1 was used to sequence the 16-bp 10X barcode and 12-bp UMI of each molecule, whereas Read 2 was used to sequence the cDNA fragment.

3. 1. 8. scGESTALT lineage tracing system

The scGESTALT system was established in our lab following the protocol detailed by Raj and others (Raj et al., 2018). Briefly, I used two plasmids, pTol2-hsp70l:Cas9-t2A-GFP, 5xU6:sgRNA (Addgene, plasmid #108871) and pTol2-hspDRv7_scGstlt (Addgene, plasmid #108870), together with *in vitro*-transcribed Tol2 mRNA. These constructs were each injected to 1-cell stage embryo as described previously in chapter 2. 1. 5. to establish two transgenic lines, each carrying one of the constructs. After the F0 fish reached maturity, each line was outcrossed with wild-type fish and the resulting embryos were screened for transgene expression. The transgene-positive embryos were kept as F1 fish.

The F1 fish were genotyped to determine the copy number of transgenic insertion. For this purpose, 3-month-old individuals were fin clipped to obtain genomic DNA following standard procedures (Xing et al., 2014). Fish were anesthetized in MS-222 solution (Sigma). Then, each fish was placed on a Petri dish and a small 1-2 mm piece of its tail fin was cut using a sterile scalpel. The piece was then transferred to a 0.2 ml sterile tube containing 100 µl of 50 mM NaOH. The fish was placed in a tank containing fresh water and kept in separation until full recovery.

The genotyping of fish was conducted following the methods used in our previously published work (Nieścierowicz et al., 2022). Briefly, the DNA from fin clippings was extracted using the modified HotShot method (Meeker et al., 2007). Briefly, after the fin piece was put in 50 mM NaOH, it was heated to 95°C for 20 minutes. The sample was then cooled to 4°C and 10 µl of 1M Tris-HCl was added to neutralize the solution. The sample was then centrifuged to pellet the debris and the supernatant was used for PCR reactions (Table 2).

To establish the functionality of the inducible expression system, a trial heat shock experiment was performed on 24 hpf embryos for 1 hour at 37℃, after which the embryos were transferred to 28.5℃ egg water and maintained for 2 hours in standard conditions before beginning observation. Injected larvae were observed under fluorescent stereomicroscope (Olympus, Japan) at 26-28 hpf to establish GFP expression. Images were taken with Zeiss Axio Imager 2 (Zeiss, Germany).

Table 2. Primers used during establishing of the scGESTALT system.

Name	Sequence [5'→3']	Purpose
qPCRctrl F	TCAGTCAACCATTCAGTGGCCCAT	Amplification of a control region for copy number determination
qPCRctrl R	CAGGAAAGGGAATGCAGGGTTTGT	Amplification of a control region for copy number determination
qPCRdsRed F	GAGCGCGTGATGAACTTCGAGG	Amplification of the insert for copy number determination
qPCRdsRed R	CAGCCCATAGTCTTCTTCTGCATTACG	Amplification of the insert for copy number determination
scGESTALT F	TCGAGCTCAAGCTTCGG	Barcode amplification for dropout assessment
scGESTALT R	CTGCCATTTGTCTCGAGGTC	Barcode amplification for dropout assessment

3. 2 Results

3. 2. 1. *nkx2.5/nkx2.7* morpholino-mediated knockdown recapitulates mutant phenotype

As an initial step towards single cell transcriptomic analysis to investigate the role of *nkx2.5/nkx2.7* in driving cell lineage diversity in CPs, I optimized a loss of function system for these two transcription factors. I modeled my approach on a previously published study that indicated their key role in sustaining the ventricular identity of cardiomyocytes (Targoff et al., 2013), and opted for morpholino-mediated knockdown of both transcription factors.

To establish the effects of *nkx2.5/nkx2.7* knockdown in zebrafish heart development, I first assessed the overall survival rates of morpholino-injected embryos compared to uninjected and mock-injected ones (Fig. 13). Mock injections were performed with a mix of water and Phenol Red to identify potential issues with embryo survival resulting from the injection procedure itself. Uninjected embryos served as batch quality control, with high mortality rates among uninjected embryos indicating underlying problems of the embryos used for experiment, in which case the whole batch would be excluded from further analysis.

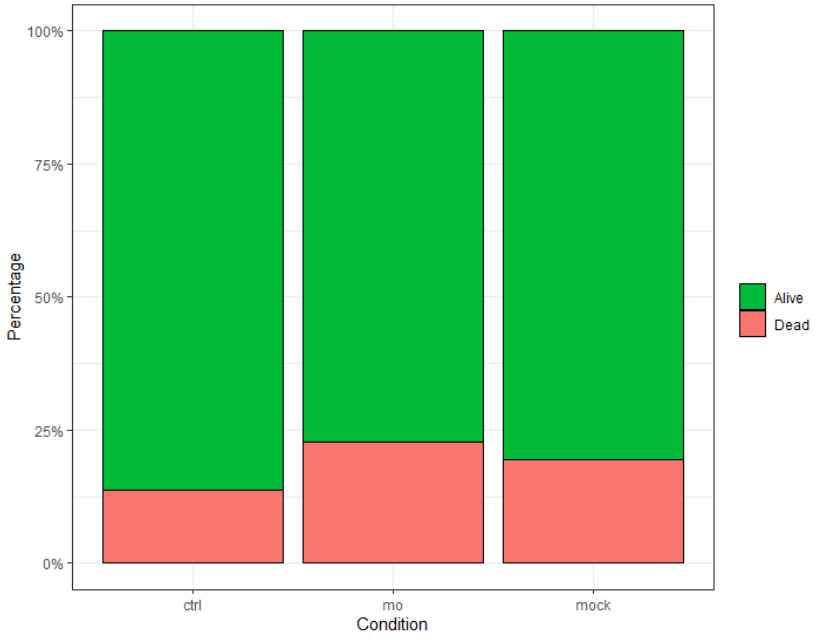


Figure 13. Summary of the survival rates of morpholino-injected embryos across conditions. Each embryo injected at 1-cell stage, injection into the yolk, injection solution: 1.5ug/µl of each MO, injection volume: 1nl; mock - water and Phenol Red, ctrl - uninjected control.

Average survival in morpholino-injected embryo batches was lower (77,2%) than in control (86,3%) and mock-injected (80,5%) embryos. Nevertheless, the number of surviving embryos was sufficient for further experiments.

Based on a previous study of *nkx2.5*/*nkx2.7* mutants, the expected phenotype upon loss of function of both *nkx2.5* and *nkx2.7* is heart looping failure, followed by severe loss of ventricular cardiomyocyte identity leading to atrium overgrowth and presence of diminished ventricle observed as early as 48 hpf (Targoff et al., 2013). To assess the percentage of morphants exhibiting the specific phenotype I conducted a phenotype penetrance experiment (Fig. 14).

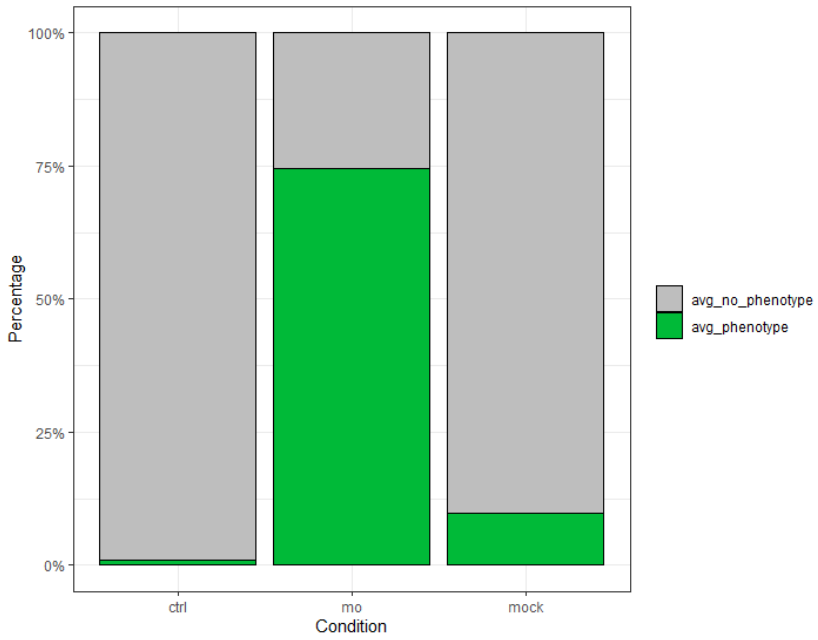


Figure 14. Average percentage of embryos exhibiting the heart looping failure and atrium overgrowth phenotypes across conditions. ctrl - uninjected embryos; mo - embryos injected with an *nkx2.5*/*nkx2.7* morpholino; mock - embryos injected with diluted Phenol Red.

My results indicate that, while the penetrance of the phenotype is not complete, I was able to replicate the observed mutant phenotype in a majority of embryos (74,6%). On the other hand, only a small percentage (0,8%) of embryos exhibited heart-looping defects in the control condition, whereas a slightly higher proportion (9,7%) of embryos with similar defects in the mock-injected batches points to the "background" developmental issues that result from handling and experimental manipulation.

3. 2. 2. Optimized dissociation and sorting protocol enables capture of cells enriched in GFP and *nkx2.5* markers

I developed a new protocol for single cell dissociation tailored to process samples from early zebrafish heart developmental stages which takes into account specific issues posed by features of the early zebrafish embryo (see Chapter 3.3. Discussion and future work).

To validate the efficacy of the dissociation and sorting method in isolating the desired cells, I initially employed my protocol to extract RNA samples from both the GFP+ and GFP- cell populations collected from 5-somite, 15-somite, 18-somite, and 24 hpf time points. Each time point sample consisted of 500 dissociated Tg(*nkx2.5*:EGFP) embryos.

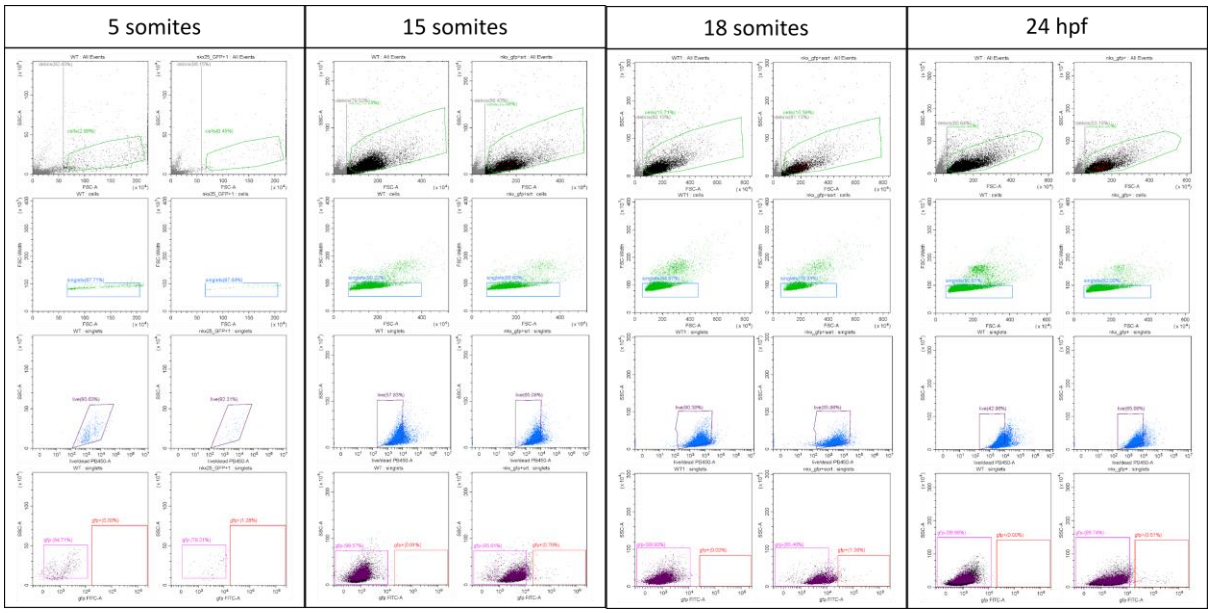


Figure 15. Summary of the gating strategy for sorting Tg(*nkx2.5*:EGFP) cells across different stages of embryonic development. First, intact cells are identified. Then, single cells are discriminated from doublets. Next, live cells are identified among the single cells. Finally, single, live cells are sorted based on the GFP fluorescence.

The GFP+ cell population does not form a separate cell cluster on the FITC-A plots due to low overall GFP intensity at early stages of development (Fig. 15). This is consistent with previous studies utilizing this transgenic line (Pawlak et al., 2019). Therefore, I employed a gating strategy where cells from dissociated wild-type embryos at matching developmental stages were used to set the GFP- and GFP+ gates prior to sorting. This allowed us to collect GFP+ cells based on their absence in the wild-type sample rather than fluorescence intensity alone.

After reverse-transcribing the cDNA from the samples, I performed an RT-qPCR experiment focusing on amplification of GFP and *nkx2.5* transcripts (Fig. 16).

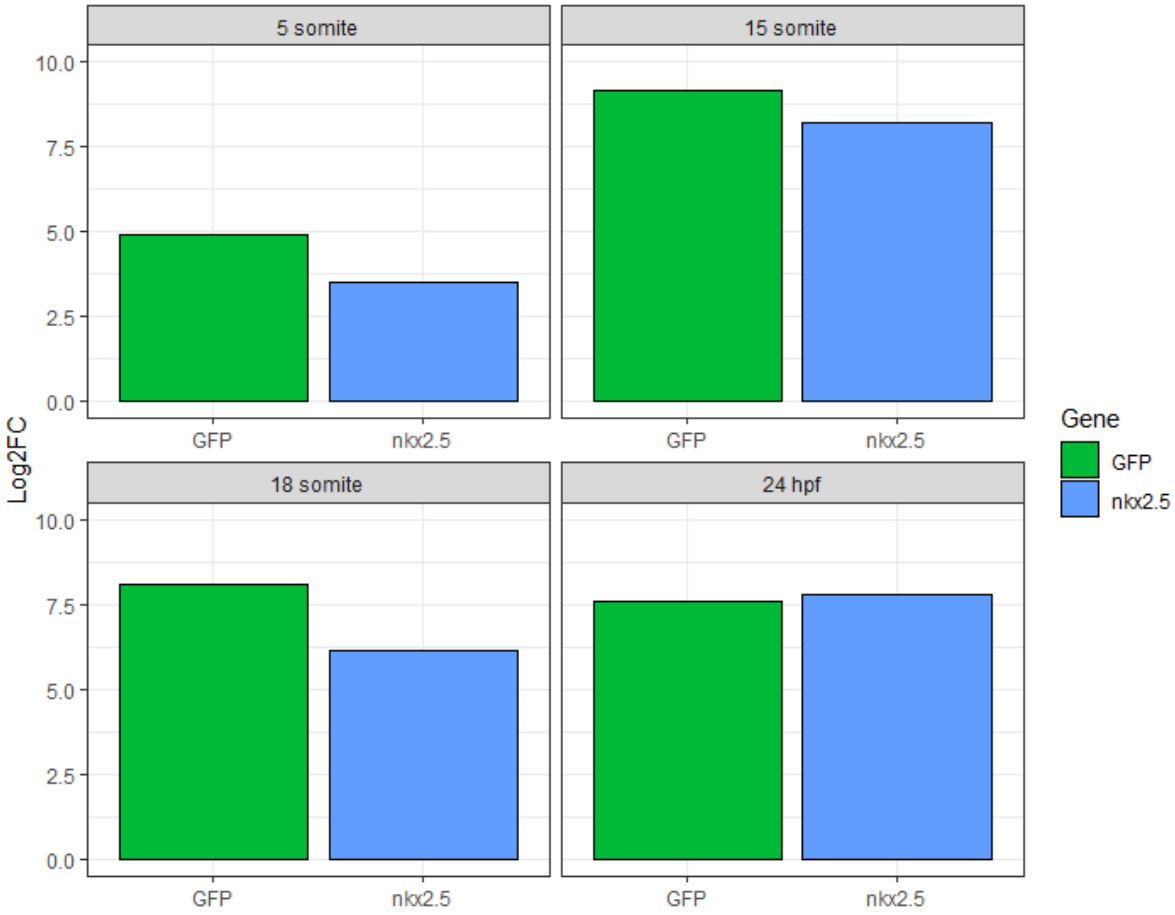


Figure 16. Log2 fold difference of *GFP* and *nkx2.5* levels between GFP+ and GFP- cells across the stages of embryonic development. Each bar represents results of a triplicate qPCR reaction.

The results show that the samples sorted from each of the timepoints were enriched in both the GFP and *nkx2.5* transcripts, indicating that my dissociation and sorting strategy were successful. This allows me to obtain sufficient numbers of *nkx2.5*-expressing cells for the transcriptomic analysis of the developing zebrafish heart.

3. 2. 3. Optimization and generation of single cell libraries on the 10X Genomics platform

To investigate the role of *nkx2.5*-expressing cells in heart development I performed a morpholino-mediated knockdown of *nkx2.5/nx2.7*, described in a previous subchapter. Then, I utilized my cell dissociation and sorting protocol to obtain GFP+ cell samples from both unmodified and morphant Tg(*nkx2.5*:EGFP) embryos.

These cells were then used to prepare single-cell RNA-seq libraries using the Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1, which allows for encapsulation and retrieval of transcriptomes from up to 10000 cells. I sorted 50000 and 47000 events from the unmodified and morphant embryos, respectively. This was done to account for significant loss of cell numbers during cell concentration and survival assessment, which in my hands ranged from 10 to 37%. Both samples were characterized by high cell viability, with 86% of cells from the unmodified embryos and 91% of cells from morphant embryos surviving the experiment up to the encapsulation step.

After sequencing, the initial data quality control was performed and summarized with FastQC (Andrews, 2010) and the 10X Genomics Cell Ranger 6.1.2. set of pipelines. The Cell Ranger analysis was performed with the help of dr S. Mondal. Overall, the data was of high quality (Fig 17.). For the 24 hpf sample approximately 70 (13,6%) million unique reads were identified before deduplication. The *nkx2.5/nkx2.7* knockdown sample consisted of approximately 110 (21,4%) million unique reads. It is important to note that this estimate was based on FastQC deduplication estimate, which is less effective than utilizing UMI information for the deduplication process. Based on the sequencing saturation metric produced by Cell Ranger, the fractions of unique reads were 52,4% and 41,4% in 24 hpf and *nkx2.5/nkx2.7* KD samples, respectively. This indicates a nearly 4- and 2-times higher fraction of unique reads based on the barcode and UMI information than that of FastQC estimate.

Figure 17. Mean Phred quality scores across all bases (blue line); 10^{th}:90^{th} percentile range (black whiskers), A - 24 hpf; B - *nkx2.5/nkx2.7* knockdown.

GC content was estimated at 49% and 47% for the 24 hpf and *nkx2.5/nkx2.7* KD samples, respectively. The distribution of GC content across all sequences deviated from the expected distribution (Fig. 18), however this could be the result of a number of factors, including cell enrichment strategy or rRNA presence. The source and impact of this deviation on the data quality should be estimated during downstream data processing and analysis.

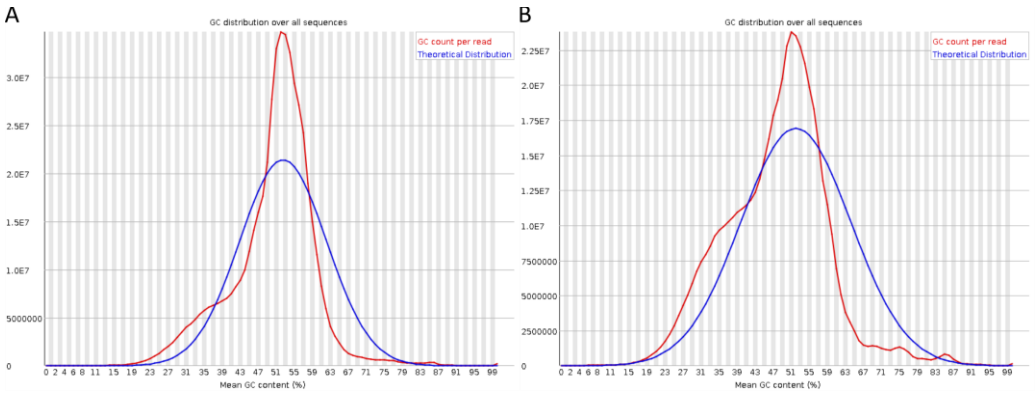


Figure 18. GC distribution across sequences. Red line - GC count per read; Blue line - estimated theoretical distribution; A - 24 hpf sample; B - *nkx2.5/nkx2.7* KD sample.

9434 cells were identified in the 24 hpf data, with 55000 mean reads per cell and 1942 median genes per cell. In the *nkx2.5/nkx2.7* KD data, 5778 cells were identified with 89751 reads per cell and 1778 genes per cell. In both datasets, the overall number of genes detected was similar, with 24182 total genes in the 24 hpf data and 24212 in *nkx2.5/nkx2.7* KD.

Overall, the sequencing data generated from this experiment was of high quality, allowing for downstream functional analysis of *nkx2.5/nkx2.7* role in heart development at a single-cell level.

3. 2. 4. Construction of single cell low throughput libraries for time course analysis of cardiac progenitors

To further assess the contribution of *nkx2.5*-expressing cells in the developing heart as well as investigate the time of emergence of heterogeneity and its extent in the SHF, we designed a time course experiment leveraging the previously described optimized cell dissociation and sorting protocol combined with scRNA-seq protocols, using the Chromium Next GEM Single Cell 3' LT (low throughput) v3.1 kit, which allows for encapsulation and capture of up to 1000 cells. The reason for adopting the low throughput approach is due to the high costs of running the standard throughput analysis and the very low number of cardiac progenitors available.

In order to cover the developmental stages from the earliest stage of cardiac progenitors' specification up to the formation of the heart tube, I sorted GFP+/*nkx2.5*+ cells from the 5-somite (5000 events), 15-somite (13452) and 18-somite (23737) stages. Cell survival was then confirmed with Trypan Blue staining. Overall, cell survival was high following the sort, with 80%, 87% and 91% live cells from the 5-, 15- and 18-somite stages respectively. After cell concentration adjustment, approximately 3100 - 3900 live cells were used for encapsulation in the 10X Controller, following the manufacturer's protocol with the aim of recovering 1000 cells after droplet generation. The downstream sample processing and sequencing for the LT kit is as described in subchapter 3. 1. 7. The final library size and concentration were within acceptable ranges provided by the manufacturer.

Initial quality control of the sequencing data was conducted as described in subchapter 3. 2. 4. Overall, the sequencing was of high quality, with Phred scores per read across all samples

averaging at 36. There is a pronounced drop in quality of 3 bases in 5- and 18-somite stage sequencing (these samples were sequenced in the same instrument run, Fig. 19), likely caused by a passing obstruction in the flow cell, which needs to be taken into account during downstream analysis. Based on FastQC estimation, the majority of reads in each sample were duplicates, with ~ 13,5 million (27%) of unique reads sequenced in the 5-somite sample, ~ 13 million (26,5%) in the 15-somite sample and ~ 8 million (15%) in the 18-somite stage.

Figure 19. A, B, C: mean Phred quality scores across all bases (blue line), 10^{th}: 90^{th} percentile range in quality (black whiskers); A', B', C': per tile sequence quality, heatmap adjusted for sequencing cycle; A, A; - 5-somite sample; B, B' - 15-somite sample; C, C' - 18-somite sample.

GC content across all samples was stable, estimated at 46% in the 5-somite sample, 45% in the 15-somite stage and 44% in the 18-somite sample (Fig. 20). Similarly to the standard throughput samples described in subchapter 3. 2. 4., GC content distribution per sequence deviates from the theoretical distribution and can be caused by the same issues such as cell enrichment strategy or rRNA presence in the samples. The exact source and impact of this issue should be determined during downstream analysis.

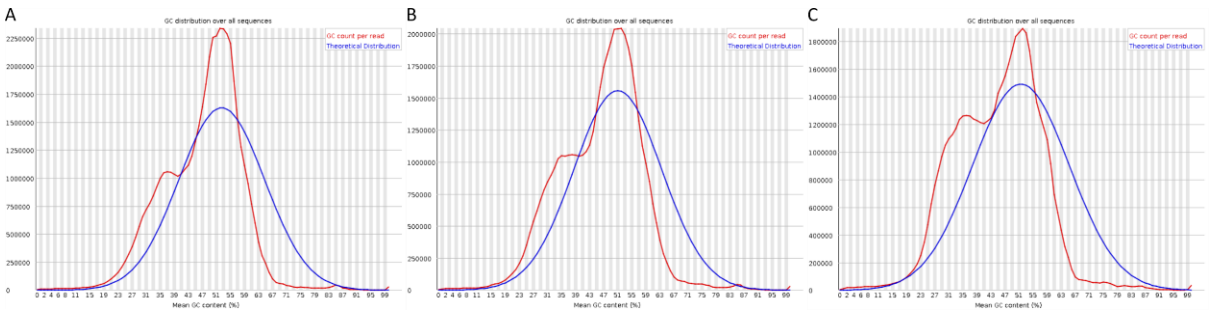


Figure 20. GC content across sequences; Red line - GC count per read; Blue line - estimated theoretical distribution; A - 5-somite sample; B - 15-somite sample; C - 18-somite sample.

Initial analysis conducted with the Cell Ranger pipeline identified 9207 cells in the 5-somite sample, 2768 cells in the 15-somite sample and 2312 cells in the 18-somite sample. In each sample, this number is remarkably higher than the 1000 expected cells recovered using the low-throughput chemistry. The low median number of identified genes per cell (350 in 5-somite, 364 in 15-somite and 251 in 18-somite sample) further indicates issues with the sequenced samples. Total number of genes detected was lower in the LT samples than in the ST samples described in subchapter 3. 2. 4., ranging from 16,346 to 19,692.

Barcode rank plots provide a good explanation for the unexpectedly high number of cells identified and low median number of genes per cell. The barcode rank plot is generated by the Cell Ranger as a way to visualize its cell calling algorithm. Briefly, the barcodes (each GEM has a unique barcode) are arranged in a descending order according to the number of UMIs (unique molecular identifiers, in principle only shared between molecules originating from the same RNA molecule) associated with each barcode. In my LT samples, only a low fraction of barcodes has high numbers of UMIs associated with them, meaning that only a small fraction of the sample should be considered as “true” single cells. However, the algorithm assigns the cell status to a larger fraction of barcodes, effectively including ambient RNA in the cell count, leading to an inflated reported cell number (Fig. 21).

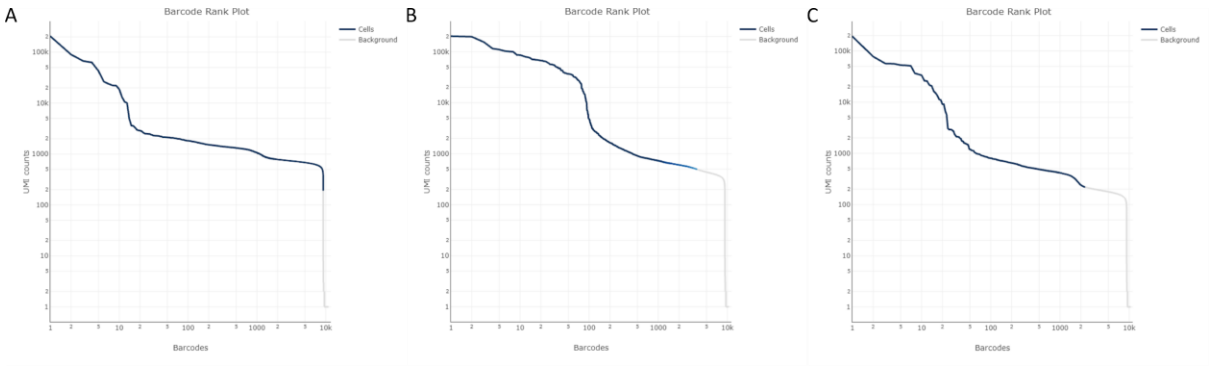


Figure 21. Estimated number and molecular complexity of sequenced cells. On barcode rank plots, barcodes are arranged along the *x*-axis in the descending order of the number of UMIs associated with them (*y*-axis). A: 5 somites; B: 15 somites; C: 18 somites.

The actual number of cells sequenced was lower - if we identify as single cells only the barcodes with associated UMI numbers similar to the ST samples (median UMI counts per cell: 8,772 for 24 hpf and 5,988 for *nkx2.5/nkx2.7* KD), we are able to confidently assign the cell status only to 10-100 barcodes in each sample, which is not enough to capture the complexity of cardiac progenitor population at these stages. This issue could stem from a number of factors during sample collection and GEM generation.

Overall, the results obtained so far indicate that in order to conduct the time course experiment, a repeat of the sample collection and sequencing is necessary to obtain meaningful insight into the heterogeneity of *nkx2.5+* cells at these stages of zebrafish heart development.

3. 2. 5. Establishing the scGESTALT system for lineage recording and tracing

Simultaneously to the time course experiment detailed in section 3. 2. 4., I established scGESTALT, first published in 2018 (Raj et al., 2018), in our lab. The basic principle of this system is enabling cell lineage recording and tracing through Cas9 editing of a synthetic barcode sequence incorporated in a transgene. Due to its potential use to study the mechanism underlying the diversification of cardiovascular progenitors in the future, I decided to establish this valuable system in our lab and perform initial validation of its functionality.

The scGESTALT system utilizes two transgenic zebrafish lines in order to record and trace cell lineage across development. The first line, carrying the pTol2-hsp70l:Cas9-t2A-GFP, 5xU6:sgRNA construct (hereafter referred to as hsp70l:Cas9) encodes a Cas9-t2A-GFP fusion protein driven by the heat-shock promoter hsp70l together with 5 unique sgRNA sequences, each driven by a U6 promoter sequence. The second construct, pTol2-hspDRv7_scGstlt (hereafter referred to as Gestalt), carries a heat-shock inducible mRFP followed by the scGESTALT barcode, as well as GFP sequence driven by the *myl7* promoter. The barcode sequence is unique in the context of the zebrafish genome and is recognized by the sgRNAs carried by the hsp70l:Cas9 line. The *myl7*:EGFP sequence is present for ease of screening for genomic integration of the construct. Upon crossing these two lines, embryos can be heat-shocked at a desired time point, inducing Cas9 expression which, guided by the sgRNAs, edits the barcode DNA sequence in each cell. After the desired developmental stage is reached, the embryos can be heat shocked again, inducing the barcode expression so that its sequence can be obtained together with the cellular transcriptome. During transcriptomic data analysis the lineage history of each cell can be reconstructed based on the editing patterns in the barcode present in the cell.

To generate and establish stable lines of the two scGESTALT transgenics, I performed microinjection of embryos as described in section 2.1.4. To validate the functionality of the hsp70l:Cas9, I first tested whether the integrated Cas9 and barcode expression can be induced using heat-shock in embryos injected with this construct. Following the original protocol, I heat-shocked 24 hpf F0 embryos at 37°C for 1 hour. Afterwards, the embryos were transferred to standard conditions and the induction of expression of GFP and dsRed was monitored. After 2-4 hours the expression of transgenic reporter proteins in each line was observed under a fluorescent microscope (Fig. 22).

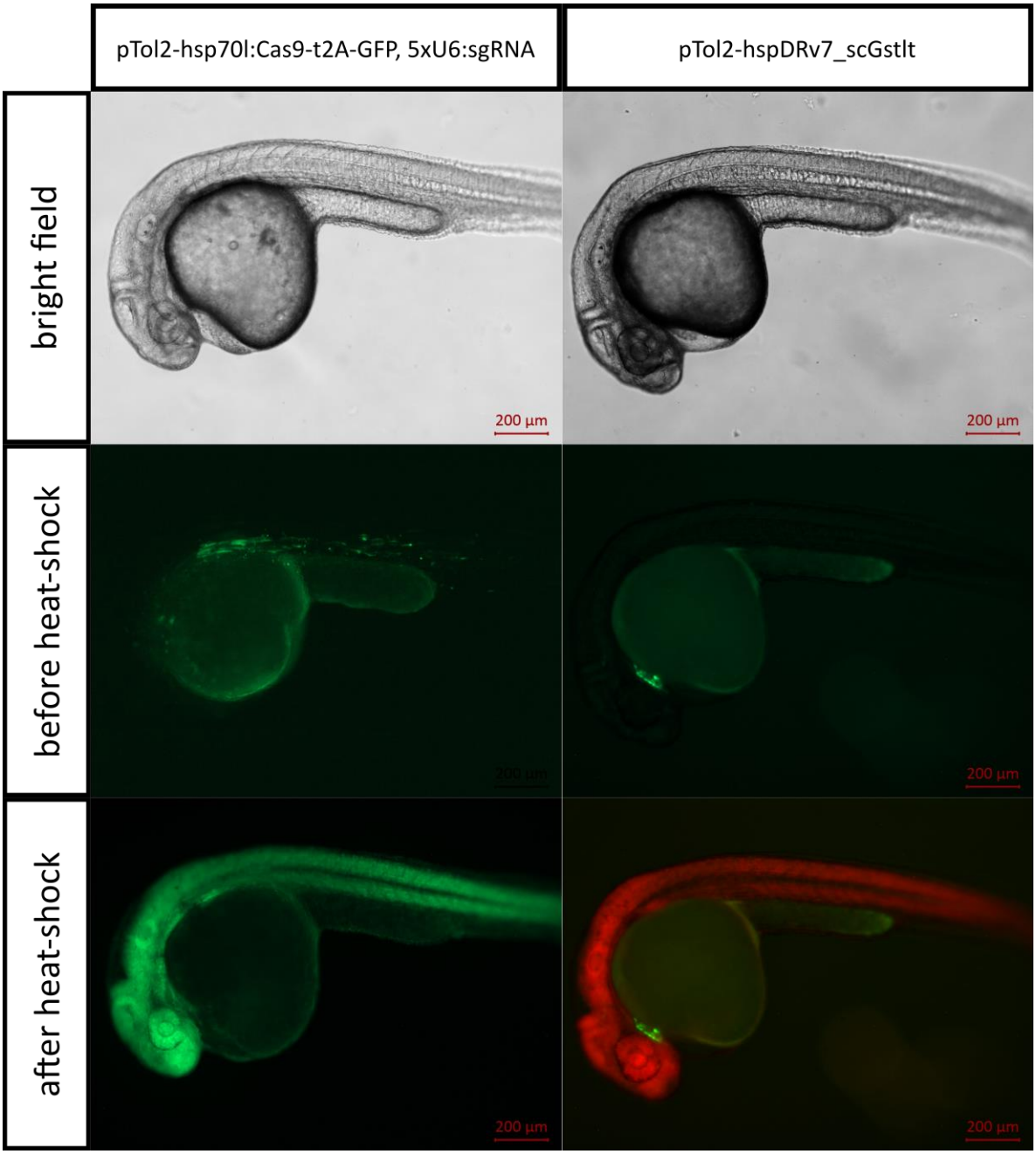


Figure 22. Reporter transgene expression in scGESTALT lines. Micrographs taken before and after heat shocking the 24 hpf embryos at 37°C for 1 hour. pTol2-hsp70l:Cas9-t2A-GFP, 5xU6:sgRNA: ubiquitous induction of the Cas9-t2A-GFP construct was visible within 2 hours after heat shock; pTol2-hspDRv7_scGstlt: the stable heart-specific GFP expression is observed thanks to the inclusion of a *myl7*:EGFP reporter into the construct. Ubiquitous dsRed expression is visible 2-3 hours after heat shock.

Next, I checked the transgene copy number in the second transgenic line carrying the barcodes - the Gestalt line, as it is crucial that the barcode sequence is only present in one copy in each cell. To do this, I performed a qPCR assay following the established protocol and previously published primers (Raj et al., 2018). I PCR-amplified genomic DNA obtained from F1 Gestalt fish, and compared the number of dsRed sequences to a heterozygous Tg(*fabp10a*:dsRed) stable fish which carries only a single copy of the dsRed transgene. I used a wild-type fish as a negative control, as it does not carry the dsRed transgene. I fin-clipped and tested 20 Gestalt fish to identify potential single-copy transgene carriers to serve as founders for the next generation. To determine the transgene copy number I followed the established protocol (Raj et al., 2018). Briefly, I set up triplicate amplification reactions for each fish tested. I then compared the average dsRed Ct to the average control region Ct within each set of triplicates. Next, I compared the differences between each Gestalt fish and the control WT (negative) and dsRed (standard) fish. In my experiment, the resulting comparison values ranged from 0 (indicating lack of transgene integration) to 2 (indicating 2 copies of the transgene), with a cutoff for a single copy set in the range of 0,75 to 1,25 to account for variability in amplification (Fig. 23). Out of the 20 fish tested, 11 carried a single copy of the transgene and were selected as founders of the next generation.

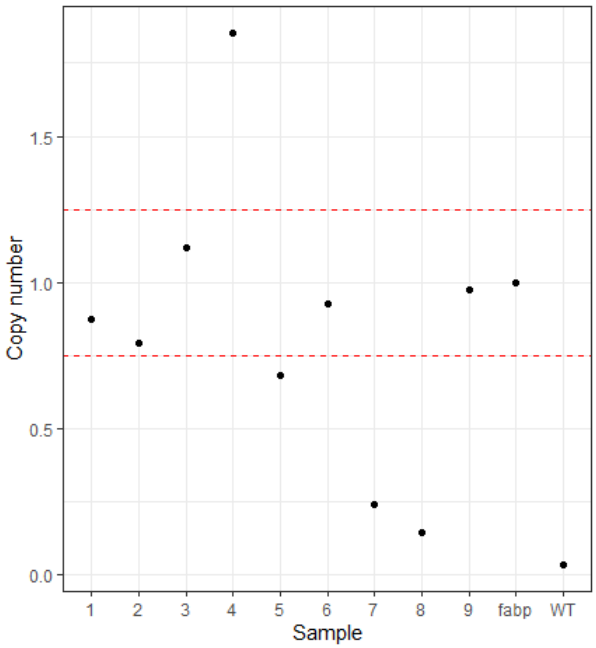


Figure 23. Scatterplot representing the transgene copy number in 9 randomly selected Gestalt F1 fish. Red dashed line: 0.75-1.25 cutoff for single copy integration; 1-9: individual Gestalt fish; fabp - positive control Tg(*fabp10a*:dsRed) fish; WT - negative control fish.

I then assessed the possibility of barcode dropout. Since the sgRNAs are expressed continuously throughout the embryonic development and the heat shock activation leads to a long period of Cas9 expression, it is possible that Cas9 cleaves two or more *loci* within the barcode simultaneously, causing the loss of the sequence between the cleaved *loci* and deletion of the lineage record of the affected cell (Salvador-Martínez et al., 2019, Zafar et al., 2020). To test this, I crossed the hsp70l:Cas9 fish with the Gestalt fish and heat shocked the resulting embryos for 1 hour at 37°C. I then isolated the genomic DNA from each embryo, PCR-amplified the barcode sequences and resolved them on a 1,5% agarose gel (Fig. 24). The product electrophoresis indicates that it is possible for the barcode length to decrease due to Cas9 editing. In some cases, the band representing the barcode of correct length is still visible, indicating that not all cell barcodes were edited in a way that causes their length to noticeably decrease, but most amplification products lack a discernible barcode band. This demonstrates that barcode dropout can be a limiting factor for utilization of scGESTALT in lineage tracing experiments.

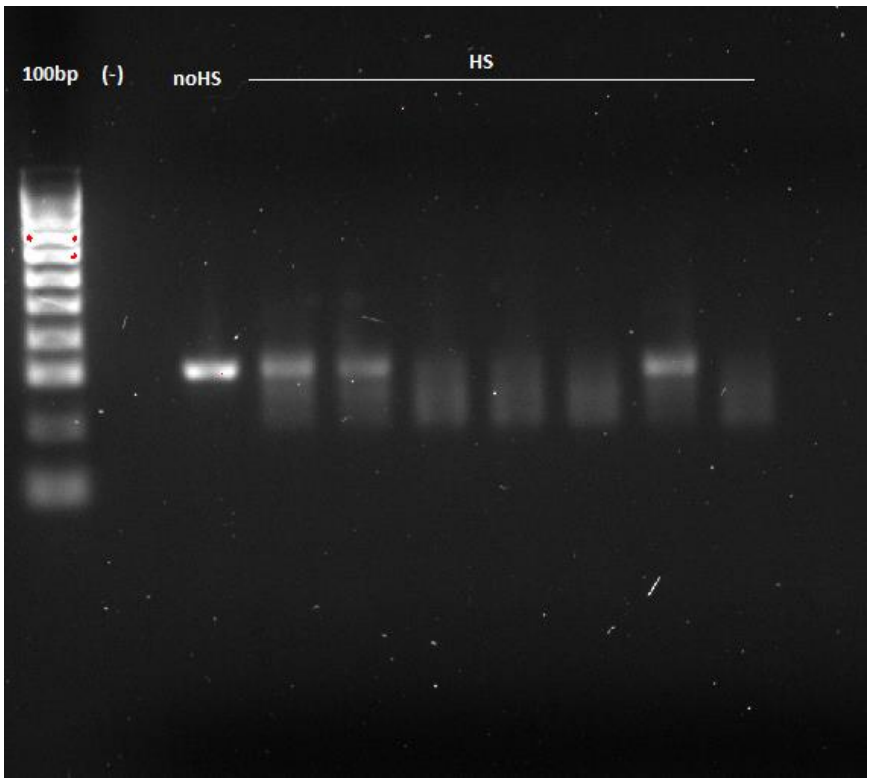


Figure 24. The Gestalt barcode sequence lengths before (noHS) and after heat shock (HS). Each well contains the DNA amplification product from a single 24 hpf embryo. 100 bp: size marker. (-): negative control (PCR product from a WT embryo).

I have successfully established the scGESTALT transgenic lines that can be utilized in a cell lineage recording and tracing experiment. The system would be a valuable tool which can supplement my time course experimental approach by providing information on the likely fate of each subpopulation of cardiac progenitors. Additionally, preliminary experiments suggest the possible limitations of the system due to barcode dropout which needs to be considered when interpreting the results.

3. 3. Discussion and future work

Established protocols for zebrafish embryo dissociation typically focus on obtaining cells from developmental stages later than that required of my study (Bresciani et al., 2018). My samples therefore presented several key issues that would make these established protocols unsuitable for this experiment. Firstly, my sample would need to be further processed after cell dissociation - I needed to enrich the final sample with GFP-positive cells, which necessitated FACS sorting. Moreover, the majority of cells (>80%) would need to remain intact after the sort. A high number of non-viable cells impacts the single cell RNA-seq library preparation steps and can cause significant bias following data analysis. Secondly, early-stage embryos contain large amounts of yolk and chorion which could negatively impact the downstream steps by causing false positive signal due to autofluorescence during sorting. Thirdly, the GFP-positive cell population is relatively scarce at these stages, constituting only around 1% of all cells of an embryo. This, combined with the presence of the yolk, could lead to failure of GFP-positive cell enrichment. The necessary high cell viability also meant that increasing the sorting speed or time would not be possible to circumvent these issues, as it would invariably result in lower cell survival. I therefore optimized my own protocol adapted for single cell isolation from early-stage zebrafish embryos.

Firstly, to tackle the chorion remnants present in the samples, I performed dechorionation using Pronase, which enables easy separation of the chorion from the embryos. Then, to address the abundance of yolk matter in the samples, I incorporated a de-yolking step by adding a de-yolking buffer (Link et al., 2006), manually disrupting the embryos and subsequently washing them with E3. Finally, I modified the dissociation step present in all established protocols by reducing the severity of manual tissue disruption. To counteract the possible presence of clumps of tissue resulting from this gentler tissue dissociation, I introduced an additional filtration step. In order to further minimize the damage to the cells during FACS,

I used a larger nozzle size - 100 µm - which, while extending sorting time, subjects the cells to the least amount of pressure, increasing overall cell survival (Higdon et al., 2019). This modified protocol enables me to reliably capture GFP+ cell population from early heart developmental stages, starting at 5 somites.

I successfully established standard-throughput single cell encapsulation and library preparation using the commercially available solutions provided by 10X Genomics. The libraries prepared using the 10X Genomics standard throughput (Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1) chemistry were of high quality and, after further analysis, is expected to provide valuable insights into the biology of *nkx2.5*-expressing cells at 24 hpf. I am currently optimizing protocols for another single-cell experiment, aiming to uncover chromatin accessibility in *nkx2.5*+ cells at 24 hpf using scATAC-seq. This will supplement my transcriptome data, providing a more complete molecular landscape of cells contributing to heart formation at this developmental stage.

In order to gain insight into the extent and underlying mechanism of the development of cell heterogeneity among cardiac progenitors residing in the SHF, I performed a time course experiment, isolating *nkx2.5*-expressing cells from 5-, 15- and 18-somite stages. I opted for the 10x Genomics low-throughput (LT) chemistry for this experiment, as at these early stages each embryo only contains a small number of *nkx2.5*+ cells, allowing me to capture a diversity of cells from large numbers of embryos within a smaller number of cells. Additionally, the low cell number requirement for the LT chemistry enabled me to maximize cell survival during sorting, as attempting to sort larger numbers of cells would extend the sorting time, leading to a drop in cell survival. Unfortunately, libraries obtained from the early time points failed to produce data which meets the required quality metrics, as the library preparation and sequencing failed to capture the number of cells required for further analysis.

The low number of actual cells in my LT samples can stem from multiple sources. One possible explanation was an overestimation of the number of cells initially used for encapsulation. However, this estimate was based on a robust cell counting protocol, using an average of two separate automated measurements from each cell sample. Another potential problem could occur at the single cell encapsulation step, as it is possible that the cells did not survive the master mix preparation before loading into the 10X Controller.

Low number of unique reads could be attributed to overamplification of the cDNA before and during library preparation, and could be inherent to the LT protocols, which were discontinued at the time of writing this thesis. In order to address these issues, I will conduct the time course experiment again, utilizing standard throughput chemistry.

Once the relevant data is obtained, we will conduct time course data analysis to identify and characterize different subpopulations of cells contributing to SHF-dependent processes in heart development. This analysis will be connected to the *nkx2.5* 24 hpf functional analysis, as the 24 hpf transcriptome will serve as the fourth and final time point. We will then validate the results of our analysis by performing targeted knock-out or knock-down experiments utilizing the insight gained during data analysis.

Simultaneously, I established two transgenic lines serving as the basis of the scGESTALT lineage tracing system first published in 2018 (Raj et al., 2018). My initial experiments indicate that barcode dropout can be an important bottleneck when using this system, as information encoded by the scGESTALT barcode can be lost during development due to Cas9 activity. This can potentially limit the number of cells carrying relevant information on their lineage at the end of the experiment (VanHorn et al., 2021). It is important to note that the results I obtained are not quantitative - we do not know how many cells are subject to barcode dropout and what the dropout rate is. A computational approach could potentially be implemented to address this issue, as in recent years a number of computational tools have been developed for this purpose (Salvador-Martínez et al., 2019, Zafar et al., 2021). If successful, the lineage recording and tracing experiment enabled by the scGESTALT system will provide further insight into the fate of early cardiac progenitor subpopulations identified using my time course experiments.

4. Conclusions

Our understanding of vertebrate heart development has increased significantly over the last decades of continuous research. Starting with the first descriptions of the second heart field in the early XXI century, many research projects have successfully broadened our understanding of the molecular processes involved. Molecular pathways and transcription factors crucial to heart development have been extensively described. At the same time, the molecular mechanism of two best-known transcription factors implicated in SHF development, Isl1 and Nkx2.5, remains largely unknown, which raises the question of how these pleiotropic transcription factors exert their specific function on the SHF. Furthermore, recent reports have indicated a previously uninvestigated cell heterogeneity within the SHF population, enabling them to give rise to various cell types. The molecular basis of this diversity, as well as the timeline of its development remain unknown.

Over the course of this project, I sought to investigate the underlying mechanism leading to the specification of the SHF, focusing on the roles of *nkx2.5* and *isl1a*. I used the zebrafish as the molecular mechanisms of heart development are largely conserved across vertebrates, making it a valuable model organism.

During embryonic development, the expression of Isl1, as well as *isl1a* - its zebrafish ortholog, is not limited to the heart. Isl1/*isl1a* is also known to play a role in the development of nervous and digestive systems. Therefore, I aimed to investigate the mechanism driving heart-specific expression of *isl1a*. Through *in silico* screening I identified 7 putative enhancers in proximity of the *isl1a locus*. I used an enhancer assay to investigate the function of these putative enhancers and found one, located in an intron of the *isl1a* gene, that was capable of driving reporter gene expression specifically in the heart. I characterized the cells expressing the enhancer-dependent reporter and found them to be enriched in canonical markers for cardiac progenitors and cardiomyocytes, namely *nkx2.5*/*nkx2.7* and *myl7*. I also investigated the presence and cross-species conservation of transcription factor binding sites located in the enhancer. I found potential binding sites of several cardiac transcription factors in this enhancer. One of them, MEF2C, was likely to bind the corresponding genomic region in other vertebrates including human, mouse, Western clawed frog and green anole, despite low overall similarity of the investigated sequence. This part of the project lays the groundwork for the future investigation of the relationship between *mef2ca/b* and *isl1a*. Additionally,

the transgenic line generated in this study serves as a useful tool for future research focusing on the development of the SHF.

nkx2.5 plays a multifaceted role in driving FHF and SHF progenitor specification. It can seemingly both promote and restrict hemoangiogenic cell fates, as well as ensure heart function by establishing the proper numbers of atrial and ventricular cardiomyocytes. I aimed to establish and optimize a set of experimental tools to enable investigation of these roles of *nkx2.5* at a single-cell resolution throughout embryonic development. I successfully established a protocol for dissociation and FACS of zebrafish embryonic cells from early developmental stages. I obtained transcriptomic data on *nkx2.5*-expressing cells from a series of samples, encompassing key stages throughout heart development. I complemented this approach with single cell transcriptome data from an *nkx2.5/nkx2.7* loss of function experiment which will provide further insight into the role of *nkx2.5* in fate determination of heart field progenitor cells. The analysis of data obtained through these experiments is ongoing.

Furthermore, I established a Cas9-based lineage tracing system which enables reconstruction of the developmental trajectory of heart cells. In order to establish this system in our lab, I generated two stable transgenic lines, each of which carry either Cas9 and sgRNAs or a synthetic barcode which is randomly edited by Cas9, creating a lineage record of the cell. I performed a set of preliminary validation experiments to ensure the viability of this system for heart-targeted discovery. I ensured that each of the barcode fish carries only a single copy of the transgene, limiting the number of confounding factors in the experiment, as well as assessed the possibility of barcode dropout. This tool will allow us to obtain single cell transcriptome data from cells constituting a developed heart coupled with their lineage records. This will enable us to link these cells to their ancestral CPs, providing insight into the timeline of establishment and changes in heterogeneity of early cardiac progenitor cells.

Taken together, the results presented in this work provide key insight into the regulatory processes governing the specification of the heart progenitors population. I established a set of tools and obtained valuable data that serve as solid groundwork for future investigations into the mechanism of early heart progenitors' specification and its subsequent development at a single cell resolution.

5. References

1. David M Parichy (2015) The Natural History of Model Organisms: Advancing biology through a deeper understanding of zebrafish ecology and evolution eLife 4:e05635.
2. Nelson, J.S. Fishes of the World. 4th Edition, John Wiley & Sons (2006).
3. Veldman MB, Lin S. Zebrafish as a developmental model organism for pediatric research. Pediatr Res. 2008 Nov;64(5):470-6.
4. Chia K, Klingseisen A, Sieger D, Priller J. Zebrafish as a model organism for neurodegenerative disease. Front Mol Neurosci. 2022 Oct 13;15:940484.
5. MacRae CA, Peterson RT. Zebrafish as tools for drug discovery. Nat Rev Drug Discov. 2015 Oct;14(10):721-31.
6. Bakkers J. Zebrafish as a model to study cardiac development and human cardiac disease. Cardiovasc Res. 2011 Jul 15;91(2):279-88.
7. Zang L, Maddison LA and Chen W (2018) Zebrafish as a Model for Obesity and Diabetes. Front. Cell Dev. Biol. 6:91.
8. Astell KR, Sieger D. Zebrafish In Vivo Models of Cancer and Metastasis. Cold Spring Harb Perspect Med. 2020 Aug 3;10(8):a037077.
9. Katoch S, Patial V. Zebrafish: an emerging model system to study liver diseases and related drug discovery. J Appl Toxicol. 2020; 41: 33-51.
10. Westerfield, M. and Eisen, J.S. (1988) Neuromuscular specificity: pathfinding by identified motor growth cones in a vertebrate embryo. Trends in neurosciences. 11:18-22.
11. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. Stages of embryonic development of the zebrafish. Dev Dyn. 1995 Jul;203(3):253-310.
12. Westerfield, M. (2000) The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio), 4th Edition. University of Oregon Press, Eugene.
13. Beier DR. Zebrafish: genomics on the fast track. Genome Res. 1998 Jan;8(1):9-17.
14. Mullins MC, Navajas Acedo J, Priya R, Solnica-Krezel L, Wilson SW. The zebrafish issue: 25 years on. Development. 2021 Dec 15;148(24):dev200343.
15. Nasevicius A, Ekker SC. Effective targeted gene 'knockdown' in zebrafish. Nat Genet. 2000 Oct;26(2):216-20.
16. Parinov S, Kondrichin I, Korzh V, Emelyanov A. Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes in vivo. Dev Dyn. 2004 Oct;231(2):449-59.

17. Kawakami, K. Tol2: a versatile gene transfer vector in vertebrates. Genome Biol 8 (Suppl 1), S7 (2007).
18. Howe, K., Clark, M., Torroja, C. et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496, 498–503 (2013).
19. Garcia, G.R., Goodale, B.C., Wiley, M.W., La Du, J.K., Hendrix, D.A., Tanguay, R.L. (2017) In Vivo Characterization of an AHR-Dependent Long Noncoding RNA Required for Proper Sox9b Expression.. Molecular pharmacology. 91(6):609-619.
20. Paone, C., Rudeck, S., Etard, C., Strähle, U., Rottbauer, W., Just, S. (2018) Loss of zebrafish Smyd1a interferes with myofibrillar integrity without triggering the misfolded myosin response. Biochemical and Biophysical Research Communications. 496(2):339-345.
21. Gore AV, Pillay LM, Venero Galanternik M, Weinstein BM. The zebrafish: A fintastic model for hematopoietic development and disease. Wiley Interdiscip Rev Dev Biol. 2018 May;7(3):e312.
22. Liu J., Stainier D. Y. R. (2012). Zebrafish in the study of early cardiac development. Circ. Res. 110, 870–874.
23. Tu S, Chi NC. Zebrafish models in cardiac development and congenital heart birth defects. Differentiation. 2012;84(1):4–16.
24. Wessels A, Sedmera D. Developmental anatomy of the heart: a tale of mice and man. Physiol Genomics. 2003 Nov 11;15(3):165-76.
25. Evans SM, Yelon D, Conlon FL, Kirby ML. Myocardial lineage development. Circ Res. 2010; 107:1428–1444
26. Staudt D, Stainier D. Uncovering the molecular and cellular mechanisms of heart development using the zebrafish. Annu Rev Genet. 2012;46:397-418.
27. Saga Y, Miyagawa-Tomita S, Takagi A, Kitajima S, Miyazaki J, Inoue T. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. Developmen (1999) 126(15): 3437-3447.
28. Bondue A, Tännler S, Chiapparo G, Chabab S, Ramialison M, Paulissen C, Beck B, Harvey R, Blanpain C. Defining the earliest step of cardiovascular progenitor specification during embryonic stem cell differentiation. J Cell Biol. 2011 Mar 7;192(5):751-65.
29. Chan SS, Shi X, Toyama A, Arpke RW, Dandapat A, Iacovino M, Kang J, Le G, Hagen HR, Garry DJ, Kyba M. Mesp1 patterns mesoderm into cardiac, hematopoietic, or

skeletal myogenic progenitors in a context-dependent manner. Cell Stem Cell. 2013 May 2;12(5):587-601.

30. W Patrick DevineJoshua D WytheMatthew GeorgeKazuko Koshiba-TakeuchiBenoit G Bruneau (2014) Early patterning and specification of cardiac progenitors in gastrulating mesoderm eLife 3:e03848.
31. Lescroart F, Chabab S, Lin X, Rulands S, Paulissen C, Rodolosse A, Auer H, Achouri Y, Dubois C, Bondue A, Simons BD, Blanpain C. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. Nat Cell Biol. 2014 Sep;16(9):829-40.
32. Kirby ML. Cardiac Development. Oxford University Press (2007).
33. Kelly RG, Brown NA, Buckingham ME. The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. Dev Cell. 2001 Sep;1(3):435-40.
34. Meilhac SM, Esner M, Kelly RG, Nicolas JF, Buckingham ME. The clonal origin of myocardial cells in different regions of the embryonic mouse heart. Dev Cell. 2004 May;6(5):685-98.
35. Marques SR, Yelon D. Differential requirement for BMP signaling in atrial and ventricular lineages establishes cardiac chamber proportionality. Dev Biol. 2009 Apr 15;328(2):472-82.
36. Gilbert SF, Barresi MJF. Developmental biology, 11th Edition, Sinauer Associates (2016).
37. O'Rahilly R, Müller F. Developmental stages in human embryos: revised and new measurements. Cells Tissues Organs. 2010;192(2):73-84.
38. Felker, A., Prummel, K.D., Merks, A.M. et al. Continuous addition of progenitors forms the cardiac ventricle in zebrafish. Nat Commun 9, 2001 (2018).
39. Warkman AS, Krieg PA. Xenopus as a model system for vertebrate heart development. Semin Cell Dev Biol. 2007 Feb;18(1):46-53.
40. Xia J, Meng Z, Ruan H, Yin W, Xu Y and Zhang T (2020) Heart Development and Regeneration in Non-mammalian Model Organisms. Front. Cell Dev. Biol. 8:595488.
41. Cai CL, Liang X, Shi Y, Chu PH, Pfaff SL, Chen J, Evans S. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. Dev Cell. 2003 Dec;5(6):877-89.
42. Buckingham M, Meilhac S, Zaffran S. Building the mammalian heart from two sources of myocardial cells. Nat Rev Genet. 2005 Nov;6(11):826-35.

43. van den Berg G, Abu-Issa R, de Boer BA, Hutson MR, de Boer PA, Soufan AT, Ruijter JM, Kirby ML, van den Hoff MJ, Moorman AF. A caudal proliferating growth center contributes to both poles of the forming heart tube. Circ Res. 2009 Jan 30;104(2):179-88.
44. Lescroart F, Wang X, Lin X, Swedlund B, Gargouri S, Sànchez-Dànes A, Moignard V, Dubois C, Paulissen C, Kinston S, Göttgens B, Blanpain C. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. Science. 2018 Mar 9;359(6380):1177-1181.
45. Meilhac SM, Buckingham ME. The deployment of cell lineages that form the mammalian heart. Nat Rev Cardiol. 2018 Nov;15(11):705-724.
46. Bertrand N, Roux M, Ryckebüsch L, Niederreither K, Dollé P, Moon A, Capecchi M, Zaffran S. Hox genes define distinct progenitor sub-domains within the second heart field. Dev Biol. 2011 May 15;353(2):266-74.
47. Buijtendijk, M.F.; Barnett, P.; van den Hoff, M.J. Development of the human heart. Am. J. Med. Genet. Part C Semin. Med. Genet. 2020, 184, 7–22.
48. Mjaatvedt CH, Nakaoka T, Moreno-Rodriguez R, Norris RA, Kern MJ, Eisenberg CA, Turner D, Markwald RR. The outflow tract of the heart is recruited from a novel heart-forming field. Dev Biol. 2001 Oct 1;238(1):97-109.
49. Waldo KL, Kumiski DH, Wallis KT, Stadt HA, Hutson MR, Platt DH, Kirby ML. Conotruncal myocardium arises from a secondary heart field. Development. 2001 Aug;128(16):3179-88.
50. Laugwitz KL, Moretti A, Caron L, Nakano A, Chien KR. Islet1 cardiovascular progenitors: a single source for heart lineages? Development. 2008 Jan;135(2):193-205.
51. Caputo L, Witzel HR, Kolovos P, Cheedipudi S, Looso M, Mylona A, van IJcken WF, Laugwitz KL, Evans SM, Braun T, Soler E, Grosveld F, Dobreva G. The Isl1/Ldb1 Complex Orchestrates Genome-wide Chromatin Organization to Instruct Differentiation of Multipotent Cardiac Progenitors. Cell Stem Cell. 2015 Sep 3;17(3):287-99.
52. Witzel HR, Jungblut B, Choe CP, Crump JG, Braun T, Dobreva G. The LIM protein Ajuba restricts the second heart field progenitor pool by regulating Isl1 activity. Dev Cell. 2012 Jul 17;23(1):58-70.
53. Witzel, H., Cheedipudi, S., Gao, R. et al. Isl2b regulates anterior second heart field development in zebrafish. Sci Rep 7, 41043 (2017).

54. Lyons I, Parsons LM, Hartley L, Li R, Andrews JE, Robb L, Harvey RP. Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev. 1995 Jul 1;9(13):1654-66.
55. George V, Colombo S, Targoff KL. An early requirement for nkx2.5 ensures the first and second heart field ventricular identity and cardiac function into adulthood. Dev Biol. 2015 Apr 1;400(1):10-22.
56. Guner-Ataman, B., González-Rosa, J.M., Shah, H.N., Butty, V.L., Jeffrey, S., Abrial, M., Boyer, L.A., Burns, C.G., Burns, C.E. (2018) Failed Progenitor Specification Underlies the Cardiopharyngeal Phenotypes in a Zebrafish Model of 22q11.2 Deletion Syndrome. Cell Reports. 24:1342-1354.e5.
57. Paffett-Lugassy N, Singh R, Nevis KR, Guner-Ataman B, O'Loughlin E, Jahangiri L, Harvey RP, Burns CG, Burns CE. Heart field origin of great vessel precursors relies on nkx2.5-mediated vasculogenesis. Nat Cell Biol. 2013 Nov;15(11):1362-9.
58. Simões, F.C., Peterkin, T., and Patient, R. (2011) Fgf differentially controls cross-antagonism between cardiac and haemangioblast regulators. Development (Cambridge, England). 138(15):3235-3245.
59. Glasauer SM, Neuhauss SC. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol Genet Genomics. 2014 Dec;289(6):1045-60.
60. Tu CT, Yang TC, Tsai HJ. Nkx2.7 and Nkx2.5 function redundantly and are required for cardiac morphogenesis of zebrafish embryos. PLoS One. 2009;4(1):e4249.
61. Targoff KL, Colombo S, George V, Schell T, Kim SH, Solnica-Krezel L, Yelon D. Nkx genes are essential for maintenance of ventricular identity. Development. 2013 Oct;140(20):4203-13.
62. Swedlund B, Lescroart F. Cardiopharyngeal Progenitor Specification: Multiple Roads to the Heart and Head Muscles. Cold Spring Harb Perspect Biol. 2020 Aug 3;12(8):a036731.
63. Colombo S, de Sena-Tomás C, George V, Werdich AA, Kapur S, MacRae CA, Targoff KL. Nkx genes establish second heart field cardiomyocyte progenitors at the arterial pole and pattern the venous pole through Isl1 repression. Development. 2018 Feb 5;145(3):dev161497.
64. Duong TB, Holowiecki A, Waxman JS. Retinoic acid signaling restricts the size of the first heart field within the anterior lateral plate mesoderm. Dev Biol. 2021 May;473:119-129.

65. Targoff, K.L., Schell, T., and Yelon, D. (2008) Nkx genes regulate heart tube extension and exert differential effects on ventricular and atrial cell number. Developmental Biology. 322(2):314-321.
66. 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 (Cambridge, England). 140(6):1353-1363.
67. Ren J, Han P, Ma X, Farah EN, Bloomekatz J, Zeng XI, Zhang R, Swim MM, Witty AD, Knight HG, Deshpande R, Xu W, Yelon D, Chen S, Chi NC. Canonical Wnt5b Signaling Directs Outlying Nkx2.5+ Mesoderm into Pacemaker Cardiomyocytes. Dev Cell. 2019 Sep 23;50(6):729-743.e5.
68. Thisse, C., and Thisse, B. (2005) High Throughput Expression Analysis of ZF-Models Consortium Clones. ZFIN Direct Data Submission. (http://zfin.org).
69. Ren J, Miao D, Li Y, Gao R. Spotlight on Isl1: A Key Player in Cardiovascular Development and Diseases. Front Cell Dev Biol. 2021 Nov 25;9:793605.
70. Higashijima S, Hotta Y, Okamoto H. Visualization of cranial motor neurons in live transgenic zebrafish expressing green fluorescent protein under the control of the islet-1 promoter/enhancer. J Neurosci. 2000 Jan 1;20(1):206-18.
71. Moretti A, Caron L, Nakano A, Lam JT, Bernshausen A, Chen Y, Qyang Y, Bu L, Sasaki M, Martin-Puig S, Sun Y, Evans SM, Laugwitz KL, Chien KR. Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. Cell. 2006 Dec 15;127(6):1151-65.
72. Dodou E, Verzi MP, Anderson JP, Xu SM, Black BL. Mef2c is a direct transcriptional target of ISL1 and GATA factors in the anterior heart field during mouse embryonic development. Development. 2004 Aug;131(16):3931-42.
73. Lescroart F, Mohun T, Meilhac SM, Bennett M, Buckingham M. Lineage tree for the venous pole of the heart: clonal analysis clarifies controversial genealogy based on genetic tracing. Circ Res. 2012 Oct 26;111(10):1313-22.
74. Lescroart F, Kelly RG, Le Garrec JF, Nicolas JF, Meilhac SM, Buckingham M. Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. Development. 2010 Oct;137(19):3269-79.

75. Lescroart F, Hamou W, Francou A, Théveniau-Ruissy M, Kelly RG, Buckingham M. Clonal analysis reveals a common origin between nonsomite-derived neck muscles and heart myocardium. Proc Natl Acad Sci U S A. 2015 Feb 3;112(5):1446-51.
76. Peng T, Tian Y, Boogerd CJ, Lu MM, Kadzik RS, Stewart KM, Evans SM, Morrisey EE. Coordination of heart and lung co-development by a multipotent cardiopulmonary progenitor. Nature. 2013 Aug 29;500(7464):589-92.
77. Diogo R, Kelly RG, Christiaen L, Levine M, Ziermann JM, Molnar JL, Noden DM, Tzahor E. A new heart for a new head in vertebrate cardiopharyngeal evolution. Nature. 2015 Apr 23;520(7548):466-73.
78. Lioux G, Liu X, Temiño S, Oxendine M, Ayala E, Ortega S, Kelly RG, Oliver G, Torres M. A Second Heart Field-Derived Vasculogenic Niche Contributes to Cardiac Lymphatics. Dev Cell. 2020 Feb 10;52(3):350-363.e6.
79. Skelly DA, Squiers GT, McLellan MA, Bolisetty MT, Robson P, Rosenthal NA, Pinto AR. Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. Cell Rep. 2018 Jan 16;22(3):600-610.
80. Liu, Z., Lou, H., Xie, K. et al. Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. Nat Commun 8, 22 (2017).
81. Jeffrey A. Farrell et al. ,Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis.Science360,eaar3131(2018).
82. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics. 2018 Jun 19;19(1):477.
83. Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, Gagnon JA, Schier AF. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat Biotechnol. 2018 Jun;36(5):442-450.
84. Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, Mali P, Church GM. Developmental barcoding of whole mouse via homing CRISPR. Science. 2018 Aug 31;361(6405):eaat9804.
85. Nieścierowicz K, Pryszcz L, Navarrete C, Tralle E, Sulej A, Abu Nahia K, Kasprzyk ME, Misztal K, Pateria A, Pakuła A, Bochtler M, Winata C. Adar-mediated A-to-I editing is required for embryonic patterning and innate immune response regulation in zebrafish. Nat Commun. 2022 Sep 20;13(1):5520.
86. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K,

Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” Journal of Open Source Software, 4(43), 1686.

87. Wickham H, François R, Henry L, Müller K, Vaughan D (2023). dplyr: A Grammar of Data Manipulation.
88. Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4
89. Kassambara A (2023). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.0
90. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010 Dec 14;107(50):21931-6.
91. Reifers, F., Walsh, E.C., Léger, S., Stainier, D.Y.R., and Brand, M. (2000) Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (fgf8/acerebellar). Development (Cambridge, England). 127(2):225-235.
92. Deshwar AR, Onderisin JC, Aleksandrova A, Yuan X, Burrows JTA, Scott IC. Mespaa can potently induce cardiac fates in zebrafish. Dev Biol. 2016 Oct 1;418(1):17-27.
93. Budine, T.E., de Sena-Tomás, C., Williams, M.L.K., Sepich, D.S., Targoff, K.L., Solnica-Kreze, L. (2020) Gon4l/Udu Regulates Cardiomyocyte Proliferation and Maintenance of Ventricular Chamber Identity During Zebrafish Development. Developmental Biology. 462(2):223-234.
94. Kappen C, Salbaum JM. Identification of regulatory elements in the Isl1 gene locus. Int J Dev Biol. 2009;53(7):935-46.
95. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, Pappalardo Z, Clarke SL, Wenger AM, Nguyen L, Gurrieri F, Everman DB, Schwartz CE, Birk OS, Bejerano G, Lomvardas S, Ahituv N. Coding exons function as tissue-specific enhancers of nearby genes. Genome Res. 2012 Jun;22(6):1059-68.
96. Li, Q., Ritter, D., Yang, N., Dong, Z., Li, H., Chuang, J.H., and Guo, S. (2010) A systematic approach to identify functional motifs within vertebrate developmental enhancers. Developmental Biology. 337(2):484-495.
97. Begeman IJ, Emery B, Kurth A, Kang J. Regeneration and developmental enhancers are differentially compatible with minimal promoters. Dev Biol. 2022 Dec;492:47-58.

98. Ni J, Wangensteen KJ, Nelsen D, Balciunas D, Skuster KJ, Urban MD, Ekker SC. Active recombinant Tol2 transposase for gene transfer and gene discovery applications. Mob DNA. 2016 Mar 31;7:6.
99. de Pater E, Clijsters L, Marques SR, Lin YF, Garavito-Aguilar ZV, Yelon D, Bakkers J. Distinct phases of cardiomyocyte differentiation regulate growth of the zebrafish heart. Development. 2009 May;136(10):1633-41.
100. Harvey RP, Lai D, Elliott D, Biben C, Solloway M, Prall O, Stennard F, Schindeler A, Groves N, Lavulo L, Hyun C, Yeoh T, Costa M, Furtado M, Kirk E. Homeodomain factor Nkx2-5 in heart development and disease. Cold Spring Harb Symp Quant Biol. 2002;67:107-14.
101. Furtado MB, Wilmanns JC, Chandran A, Tonta M, Biben C, Eichenlaub M, Coleman HA, Berger S, Bouveret R, Singh R, Harvey RP, Ramialison M, Pearson JT, Parkington HC, Rosenthal NA, Costa MW. A novel conditional mouse model for Nkx2-5 reveals transcriptional regulation of cardiac ion channels. Differentiation. 2016 Jan-Mar;91(1-3):29-41.
102. Prall OW, Menon MK, Solloway MJ, Watanabe Y, Zaffran S, Bajolle F, Biben C, McBride JJ, Robertson BR, Chaulet H, Stennard FA, Wise N, Schaft D, Wolstein O, Furtado MB, Shiratori H, Chien KR, Hamada H, Black BL, Saga Y, Robertson EJ, Buckingham ME, Harvey RP. An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. Cell. 2007 Mar 9;128(5):947-59.
103. Migdał M, Tralle E, Abu Nahia K, Bugajski Ł, Kędzierska KZ, Garbicz F, Piwocka K, Winata CL, Pawlak M. Multi-omics analyses of early liver injury reveals cell-type-specific transcriptional and epigenomic shift. BMC Genomics. 2021 Dec 18;22(1):904.
104. Huang, C.-J., Tu, C.-T., Hsiao, C.-D., Hsieh, F.-J., and Tsai, H.-J. (2003) Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. Developmental Dynamics : an official publication of the American Association of Anatomists. 228(1):30-40.
105. Holtzman, N.G., Schoenebeck, J.J., Tsai, H.J., and Yelon, D. (2007) Endocardium is necessary for cardiomyocyte movement during heart tube assembly. Development (Cambridge, England). 134(13):2379-2386.

106. Capon SJ, Uribe V, Dominado N, Ehrlich O, Smith KA. Endocardial identity is established during early somitogenesis by Bmp signalling acting upstream of npas4l and etv2. Development. 2022 May 1;149(9):dev190421.

107. Li L, Ning G, Yang S, Yan Y, Cao Y, Wang Q. BMP signaling is required for nkx2.3-positive pharyngeal pouch progenitor specification in zebrafish. PLoS Genet. 2019 Feb 14;15(2):e1007996.

108. Uemura O, Okada Y, Ando H, Guedj M, Higashijima S, Shimazaki T, Chino N, Okano H, Okamoto H. Comparative functional genomics revealed conservation and diversification of three enhancers of the isl1 gene for motor and sensory neuron-specific expression. Dev Biol. 2005 Feb 15;278(2):587-606.

109. Wilfinger, A., Arkhipova, V., and Meyer, D. (2013) Cell type and tissue specific function of islet genes in zebrafish pancreas development. Developmental Biology. 378(1):25-37.

110. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011 Apr 1;27(7):1017-8.

111. Taher L, McGaughey DM, Maragh S, Aneas I, Bessling SL, Miller W, Nobrega MA, McCallion AS, Ovcharenko I. Genome-wide identification of conserved regulatory function in diverged sequences. Genome Res. 2011 Jul;21(7):1139-49.

112. Hombach D, Schwarz JM, Robinson PN, Schuelke M, Seelow D. A systematic, large-scale comparison of transcription factor binding site models. BMC Genomics. 2016 May 21;17:388.

113. Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L, Blitz IL, Blumberg B, Dichmann DS, Dubchak I, Amaya E, Detter JC, Fletcher R, Gerhard DS, Goodstein D, Graves T, Grigoriev IV, Grimwood J, Kawashima T, Lindquist E, Lucas SM, Mead PE, Mitros T, Ogino H, Ohta Y, Poliakov AV, Pollet N, Robert J, Salamov A, Sater AK, Schmutz J, Terry A, Vize PD, Warren WC, Wells D, Wills A, Wilson RK, Zimmerman LB, Zorn AM, Grainger R, Grammer T, Khokha MK, Richardson PM, Rokhsar DS. The genome of the Western clawed frog Xenopus tropicalis. Science. 2010 Apr 30;328(5978):633-6.

114. Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, Russell P, Lowe CB, Glor RE, Jaffe JD, Ray DA, Boissinot S, Shedlock AM, Botka C, Castoe TA, Colbourne JK, Fujita MK, Moreno RG, ten Hallers BF, Haussler D, Heger A, Heiman D, Janes DE, Johnson J, de Jong PJ, Koriabine MY, Lara M, Novick PA, Organ

CL, Peach SE, Poe S, Pollock DD, de Queiroz K, Sanger T, Searle S, Smith JD, Smith Z, Swofford R, Turner-Maier J, Wade J, Young S, Zadissa A, Edwards SV, Glenn TC, Schneider CJ, Losos JB, Lander ES, Breen M, Ponting CP, Lindblad-Toh K. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature. 2011 Aug 31;477(7366):587-91.

115. Lassmann T, Sonnhammer EL. Kalign--an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics. 2005 Dec 12;6:298.

116. Gillies SD, Morrison SL, Oi VT, Tonegawa S. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell. 1983 Jul;33(3):717-28.

117. Borsari B, Villegas-Mirón P, Pérez-Lluch S, Turpin I, Laayouni H, Segarra-Casas A, Bertranpetit J, Guigó R, Acosta S. Enhancers with tissue-specific activity are enriched in intronic regions. Genome Res. 2021 Aug;31(8):1325-1336.

118. Materna SC, Sinha T, Barnes RM, Lammerts van Bueren K, Black BL. Cardiovascular development and survival require Mef2c function in the myocardial but not the endothelial lineage. Dev Biol. 2019 Jan 15;445(2):170-177.

119. Hinits Y, Pan L, Walker C, Dowd J, Moens CB, Hughes SM. Zebrafish Mef2ca and Mef2cb are essential for both first and second heart field cardiomyocyte differentiation. Dev Biol. 2012 Sep 15;369(2):199-210.

120. Balleza E, Kim JM, Cluzel P. Systematic characterization of maturation time of fluorescent proteins in living cells. Nat Methods. 2018 Jan;15(1):47-51.

121. Raj B, Gagnon JA, Schier AF. Large-scale reconstruction of cell lineages using single-cell readout of transcriptomes and CRISPR-Cas9 barcodes by scGESTALT. Nat Protoc. 2018 Nov;13(11):2685-2713.

122. Xing L, Quist TS, Stevenson TJ, Dahlem TJ, Bonkowsky JL. Rapid and efficient zebrafish genotyping using PCR with high-resolution melt analysis. J Vis Exp. 2014 Feb 5;(84):e51138.

123. Meeker ND, Hutchinson SA, Ho L, Trede NS. Method for isolation of PCR-ready genomic DNA from zebrafish tissues. Biotechniques. 2007 Nov;43(5):610, 612, 614.

124. Pawlak M, Kedzierska KZ, Migdal M, Nahia KA, Ramilowski JA, Bugajski L, Hashimoto K, Marconi A, Piwocka K, Carninci P, Winata CL. Dynamics of cardiomyocyte transcriptome and chromatin landscape demarcates key events of heart development. Genome Res. 2019 Mar;29(3):506-519.

125. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online].
126. Salvador-Martínez I, Grillo M, Averof M, Telford MJ. Is it possible to reconstruct an accurate cell lineage using CRISPR recorders? Elife. 2019 Jan 28;8:e40292.
127. Zafar H, Lin C, Bar-Joseph Z. Single-cell lineage tracing by integrating CRISPR-Cas9 mutations with transcriptomic data. Nat Commun. 2020 Jun 16;11(1):3055.
128. Bresciani E, Broadbridge E, Liu PP. An efficient dissociation protocol for generation of single cell suspension from zebrafish embryos and larvae. MethodsX. 2018 Oct 10;5:1287-1290.
129. Link V, Shevchenko A, Heisenberg CP. Proteomics of early zebrafish embryos. BMC Dev Biol. 2006 Jan 13;6:1.
130. Higdon LE, Cain CJ, Colden MA, Maltzman JS. Optimization of single-cell plate sorting for high throughput sequencing applications. J Immunol Methods. 2019 Mar;466:17-23.
131. VanHorn S, Morris SA. Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development. Dev Cell. 2021 Jan 11;56(1):7-21.

6. List of publications

1. Migdał M*, Tralle E*, Abu Nahia K, Bugajski Ł, Kędzierska KZ, Garbicz F, Piwocka K, Winata CL, Pawlak M. Multi-omics analyses of early liver injury reveals cell-type-specific transcriptional and epigenomic shift. BMC Genomics. 2021 Dec 18;22(1):904.
2. Nieścierowicz K*, Pryszcz L*, Navarrete C*, Tralle E*, Sulej A*, Abu Nahia K, Kasprzyk ME, Misztal K, Pateria A, Pakuła A, Bochtler M, Winata C. Adar-mediated A-to-I editing is required for embryonic patterning and innate immune response regulation in zebrafish. Nat Commun. 2022 Sep 20;13(1):5520.

* equal first authorship