Technische Universität München
Fakultät für Medizin

Defining early steps of human cardiac progenitor lineage segregation, fate decision and gene regulatory elements

Dorota Marta Zawada

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität

München zur Erlangung einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Peter Ewert

Prüfer*innen der Dissertation:

1.   Prof. Dr. Alessandra Moretti

2.   Prof. Dr. Julien Gagneur

3.   Prof. Dr. Christian Wahl-Schott

Die Dissertation wurde am 26.01.2023 bei der Technischen Universität München eingereicht

und durch die Fakultät für Medizin am 13.06.2023 angenommen.

Abstract

Cardiogenesis is a highly dynamic process that relies on the precise spatiotemporal coordination of lineage commitment, proliferation and migration of progenitor cell populations. These events are orchestrated by an interplay of transcription factors, their targets, chromatin modifications and non-coding RNAs that work together in complex gene regulatory networks. Lineage tracing analysis in mice identified two main sources of early mesodermal cardiac progenitors that differentially contribute to specific heart compartments - first and second heart fields (FHF and SHF). Recently, a novel population of cardiac progenitors has been described, named juxta-cardiac field (JCF), that contributes not only to the myocardium but also to the epicardium of the heart.

Defects at many stages of cardiac development can lead to congenital heart diseases (CHDs), which are one of the most-common birth malformations. The majority of CHDs have a multifactorial etiology, with both genetic and environmental factors playing a recognized role. Some CHDs are characterized by defects in heart structures that originate specifically from FHF or SHF, suggesting that perturbations within distinct pools of cardiac progenitors could lead to these diseases. Nevertheless, given the technical and ethical issues related to the accessibility of human foetal tissues, the precise molecular and cellular understanding of CHDs remains elusive.

Human induced pluripotent stem cells (hiPSCs) or human embryonic stem cells (hESCs) are characterized by the ability to give rise to any cell building a human body, by nearly unlimited renewal potential and proliferation capacity. Generation of hiPSC lines from patients affected by CHDs allows, at least to some degree, to recapitulate the the pathophysiological feature of the disease in the contexst of the patient-specific genetic background. Although several experimental protocols currently allow for differentiation of human pluripotent stem cells (hPSCs) into cardiomyocytes, the controlled generation of different progenitors resembling mouse FHF and SHF and their systematic characterization is still missing. This would have tremendous implications for advancing our capability of modelling human cardiogenesis and CHDs as well as for informing novel preventive and therapeutic approaches for these diseases.

Here, we used a combination of genetic labelling, sequential cell sorting, and time-course bulk and single-cell transcriptional profiling to thoroughly characterize various

mesodermal progenitor pools (corresponding to human heart field-like progenitors) and their definitive progenies emerging during direct cardiac differentiation of hPSCs. These analyses allowed us to monitor the occurrence of heart field-like specification, deconstruct heterogeneity of human heart field-like progenitors including cells resembling the progenitors of the JCF, and define conditions for their specific induction *in vitro*. We identified ITGA8 as a putative surface marker for isolation of JCF-like cells and showed that key markers of JCF could be potentially directly regulated by retinoic acid (RA) signalling through RA response elements. To explore the utility of these *in vitro*-derived defined progenitors in studying developmental defects, we used hiPSC lines derived from a patient with hypoplastic left heart syndrome, one of the most severe CHDs. Within this study, we also analysed lncRNAs signatures of the different heart field-like progenitors. We showed that deletion of a promoter region consisting of the transcriptional start site of the lncRNA *HANDSDOWN,* which is essential for heart development in mice, does not affect human cardiomyocytes formation *in vitro*, suggesting the existence of divergent regulatory mechanisms during mouse and human cardiogenesis.

Together, this work provides a comprehensive molecular roadmap defining early steps of human cardiac progenitor lineage segregation and fate decision and describes ways to generate well-characterized specific cardiovascular progenitors from hPSCs to study human heart development and disease.

Zusammenfassung

Die Kardiogenese ist ein hochdynamischer Prozess, der auf der präzisen räumlich-zeitlichen Koordinierung von Spezifizierung, Proliferation und Migration von Vorläuferzellpopulationen beruht. Diese Vorgänge werden durch ein Zusammenspiel von Transkriptionsfaktoren, ihren Zielgenen, Chromatinmodifikationen und nicht-kodierenden RNAs orchestriert, die in komplexen Genregulationsnetzwerken zusammenwirken. Bei der Analyse von Abstammungslinien in Mäusen wurden zwei Hauptquellen früher mesodermaler Herzvorläuferzellen identifiziert, die in unterschiedlicher Weise zu bestimmten Herzkompartimenten beitragen - dem ersten und zweiten Herzfeld (FHF und SHF). Kürzlich wurde eine neue Population von Herzvorläuferzellen beschrieben, die als „juxta-cardiac field" (JCF) bezeichnet wird und nicht nur zum Myokard, sondern auch zum Epikard des Herzens beiträgt.

Defekte in vielen Stadien der Herzentwicklung können zu angeborenen Herzfehlern (AHF) führen, die zu den häufigsten Fehlbildungen bei der Geburt gehören. Die meisten AHF haben eine multifaktorielle Ätiologie, bei der sowohl genetische als auch Umweltfaktoren eine anerkannte Rolle spielen. Einige AHF sind durch Defekte in Herzstrukturen gekennzeichnet, die spezifisch von FHF oder SHF stammen, was darauf hindeutet, dass Störungen innerhalb verschiedener Pools von Herzvorläuferzellen zu diesen Krankheiten führen könnten. Angesichts der technischen und ethischen Probleme im Zusammenhang mit der Zugänglichkeit von menschlichem fötalem Gewebe ist das genaue molekulare und zelluläre Verständnis von AHF jedoch nach wie vor schwer fassbar.

Humane induzierte pluripotente Stammzellen (hiPSCs) oder humane embryonale Stammzellen (hESCs) zeichnen sich durch die Fähigkeit aus, jede Zelle zu erzeugen, die einen menschlichen Körper bildet, sowie durch ein nahezu unbegrenztes Erneuerungspotenzial und Proliferationsvermögen. Die Erzeugung von hiPSC-Linien von AHF-Patienten ermöglicht es zumindest bis zu einem gewissen Grad, die pathophysiologischen Merkmale der Krankheit im Kontext des patientenspezifischen genetischen Hintergrunds zu rekapitulieren. Obwohl mehrere experimentelle Protokolle derzeit die Differenzierung humaner pluripotenter Stammzellen (hPSCs) in Kardiomyozyten ermöglichen, fehlt es noch an der kontrollierten Erzeugung verschiedener Vorläuferzellen, die der FHF und SHF der Maus ähneln, und an deren systematischer Charakterisierung. Dies hätte enorme Auswirkungen auf unsere Fähigkeit,

die menschliche Kardiogenese und AHF zu modellieren und neue präventive und therapeutische Ansätze für diese Krankheiten zu finden.

In dieser Arbeit wurde eine Kombination aus genetischer Markierung, sequentieller Zellsortierung, und Massen- und Einzelzell-Transkriptomik im Zeitverlauf verwendet, um verschiedene mesodermale Vorläuferpools (die den menschlichen herzfeldähnlichen Vorläufern entsprechen) und ihre endgültigen Nachkommen, die während der direkten kardialen Differenzierung von hPSCs entstehen, gründlich zu charakterisieren. Diese Analysen ermöglichten es uns, das Auftreten der herzfeldähnlichen Spezifizierung zu verfolgen, die Heterogenität der menschlichen herzfeldähnlichen Vorläuferzellen zu dekonstruieren, einschließlich der Zellen, die den Vorläuferzellen des JCF ähneln, und die Bedingungen für ihre spezifische Induktion *in vitro* zu definieren. Wir haben ITGA8 als mutmaßlichen Oberflächenmarker für die Isolierung von JCF-ähnlichen Zellen identifiziert und gezeigt, dass Schlüsselmarker des JCF möglicherweise über RA-Reaktionselemente direkt durch Retinsäure (RA) reguliert werden. Um den Nutzen dieser *in vitro* gewonnenen definierten Vorläuferzellen für die Erforschung von Entwicklungsdefekten zu untersuchen, haben wir hiPSC-Linien verwendet, die von einem Patienten mit hypoplastischem Linksherzsyndrom, einer der schwersten AHF, stammen. Im Rahmen dieser Studie wurden auch die lncRNA-Signaturen der verschiedenen herzfeldartigen Vorläuferzellen analysiert. Wir konnten zeigen, dass die Deletion einer Promotorregion, die aus der Transkriptionsstartstelle der lncRNA *HANDSDOWN* besteht, die für die Herzentwicklung bei Mäusen essenziell ist, die Bildung menschlicher Kardiomyozyten *in vitro* nicht beeinträchtigt, was auf die Existenz unterschiedlicher epigenetischer Mechanismen während der Kardiogenese von Maus und Mensch hindeutet. Insgesamt liefert diese Arbeit einen umfassenden molekularen Fahrplan, der die frühen Schritte der Abstammung und Schicksalsentscheidung menschlicher kardialer Vorläuferzellen definiert, und beschreibt Wege zur Erzeugung gut charakterisierter spezifischer kardiovaskulärer Vorläuferzellen aus hPSCs, um die Entwicklung und Erkrankung des menschlichen Herzens zu untersuchen.

Table of Contents

Table of contents

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
A-CM	Atrial cardiomyocyte
ANOVA	Analysis of variance
aSHF	Anterior second heart field
AVC	Atrio-ventricular canal
BMP	Bone morphogenetic protein
bp	Base pair
BSA	Bovine serum albumin
CDM	Chemically defined medium
cDNA	Complementary DNA
CHD	Congenital heart disease
CHIR	CHIR-99021
CM	Cardiomyocyte
CPC	Cardiac progenitor cell
CRISPR	Clustered regularly interspaced short palindromic repeats
CTRL	Control
CVM	Conoventricular cardiomyocyte
CVP	Conoventricular progenitor
DEGs	Differentially expressed genes
DNA	Deoxyribonucleic acid
DR	Direct repeat
E8	Essential 8
EC	Endothelial cell
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein

Abbreviations

EGW	Estimated gestation week
EMT	Epithelial-to-mesenchymal transition
EndoMT	Endothelial-to-mesenchymal transition
EndoPC	Endothelial progenitor cell
EPDC	Epicardium-derived cell
FACS	Fluorescence activated cell sorting
FB	Fibroblast
FBS	Fetal bovine serum
FGF	Fibroblast growth factor
FHF	First heart field
GEM	Gel bead-in-emulsion
GO	Gene ontology
GRN	Gene regulatory network
gRNA	Guide RNA
hESC	Human embryonic stem cell
hiPSC	Human induced pluripotent stem cell
hiPSC-CM	hiPSC-derived cardiomyocyte
hPSC	Human pluripotent stem cell
hPSCreg	Human Pluripotent Stem Cell Registry
IFT	Inflow-tract
JCF	Juxta-cardiac field
kB	Kilobase
KCl	Potassium chloride
LA	Left ventricle
lncRNA	Long non-coding RNA
LV	Left ventricle
LY	LY-294002
Mb	Megabase
MC	Mesenchymal cell
MgCl2	Magnesium chloride
mRNA	Messenger RNA
NaOH	Sodium hydroxide
ncRNA	Non-coding RNA

OFT	Outflow-tract
OFT-CM	Outflow-tract cardiomyocyte
PCA	Principle component analysis
PCR	Polymerase chain reaction
PE	Proepicardium
PFA	Paraformaldehyde
Poly-HEMA	Polyhydroexyethylmethacrylate
PS	Primitive streak
pSHF	Posterior second heart field
qPCR	Quantitative real-time PCR
R&D	Research and development
RA	Retinoic acid
RALDH	Retinaldehyde dehydrogenase
RAOH	Retinol
RAR	Retinoic acid receptor
RARE	Retinoic acid response element
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
RNP	Ribonucleoprotein
RT	Room temperature
RT-PCR	Reverse transcription PCR
RV	Right ventricle
SCENIC	Single-cell regulatory network inference and clustering
scRNA-Seq	Single-cell RNA sequencing
SHF	Second heart field
SHH	Sonic hedgehog protein
SMC	Smooth muscle cell
SSC-A	Side scatter - area
TGFβ	Transforming growth factor β
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier
V-CM	Ventricular cardiomyocyte
VC	Valvular cell

Abbreviations

WNT	Wingless-related integration site proteins
WT	Wild type
YS	Yolk sac
yylncRNAs	Yin yang long non-coding RNAs

Introduction

1.1 Early heart development

Cardiogenesis is orchestrated by complex gene regulatory networks (GRNs) of transcription factors, their targets, chromatin modifications and non-coding RNAs. However, fundamental understanding of the early stages of human heart development that occur around two weeks after gestation (Estimated Gestation Week (EGW): 4) is limited due to the technical and ethical inaccessibility of human foetal tissues. Thus, most of our knowledge comes from lineage tracing studies in mice and other animal models. Regardless of the high conservation of developmental events that occur during mouse and human heart morphogenesis, there are still differences between the species in terms of timing of developmental processes, patterns of gene expression, and organ size (Wessels and Sedmera, 2003; Cui et al., 2019).

1.1.1 Origin and lineage specification of cardiovascular progenitors

In mice, the origin of cells that will form the future heart can be mapped in the epiblast around the onset of gastrulation (around embryonic day 6.5 (E6.5); in humans - in EGW 4; or at the estimated embryonic age: 16 days post fertilization; Lescroart et al., 2014). Over the next days, these cells ingress sequentially, in a timely controlled manner, through the primitive streak (Fig. 1). Recent studies suggest that cardiogenic progenitor populations emerging at different times and domains of the primitive streak along the distal-proximal axis already have a lineage potential restricted to distinct cardiac regions (Lescroart et al., 2014; Ivanovitch et al., 2021).

Upon leaving the primitive streak, cardiogenic precursors undergo an epithelial-to mesenchymal-transition (EMT), and migrate towards the most anterolateral region of the embryo. Cells that migrate first from the early/mid-primitive streak become the first heart field (FHF) cell population (Fig. 1). At about E7.5 (in human EGW 4; 20 days post fertilization; Ivanovitch et al., 2021) these FHF cells form a horseshoe-shaped structure located between the headfolds and extra-embryonic regions, known as the cardiac crescent. Progenitors of the

prospective FHF are unipotent and once deployed in the cardiac crescent region rapidly differentiate into cardiomyocytes and converge at the midline to form a primitive heart tube (Lescroart et al; 2014). Ultimately, FHF cells mainly contribute to the left ventricle (LV) and to portions of the conductive system, atria and atrioventricular canal (AVC; Meilhac et al., 2004; Spaeter et al., 2013).

Cardiac precursors of the prospective second heart field (SHF) delaminate from the primitive streak later than the FHF, starting at the late primitive streak stage till no bud – early bud stages (Fig. 1) (Ivanovitch et al., 2021). Ultimately, SHF cells position medially and posteriorly to the crescent in the splanchnic mesoderm and are progressively added to the arterial and venous poles of the primitive heart tube. SHF cells at the arterial pole (named anterior SHF - aSHF) give rise to the future right ventricle (RV) and outflow-tract (OFT), while those allocated at the venous pole (named posterior SHF - pSHF) contribute to the atria, the inflow-tract (IFT) and parts of the OFT (Kelly et al., 2001; Meilhac et al., 2004; Zaffran et al. 2004; Galli et al., 2008; Rana et al., 2014). SHF populations maintain an elevated proliferation capacity and undifferentiated characteristics longer than the FHF, and ultimately specify into multiple cell types of the heart, not only myocardial but also endocardial, vascular endothelial and smooth muscle cells (Cai et al., 2003; Dyer and Kirby, 2009; Chen et al., 2009; Lescroart et al., 2014).

Recently, studies by Tyser et al. (2021) and Zhang et al. (2021) introduced a novel population of progenitors identified as a distinct subset of FHF cells in the mouse. These cells, named juxta-cardiac field (JCF) by Tyser et al., originate from precursors mapped at the embryonic and extraembryonic mesoderm boundary at the primitive streak stage (Fig. 1). During the cardiac crescent stage, they occupy the ventral domain overlaying the cardiac crescent, corresponding to the primitive septum transversum. The JCF constitutes a multipotent population of cells that contributes not only to a portion of FHF-derived myocardial cells, such as LV and AVC cardiomyocytes, but also to cells within the septum transversum, the proepicardium (PE), that forms a transient structure of mesothelial cells emerging from the septum transversum at the base of the heart tube, the epicardium, and the extraembryonic mesoderm. To date, it is not known if this population also exists in the human developing heart.

~E6.75   ~E7.25   ~E7.5   ~E7.75

Mid-Streak   No bud   Late bud   Late headfold

Hand1$^+$ (Zhang et al., 2021)

Mab21l2$^+$ JCF (Tyser et al., 2021)

FHF

SHF
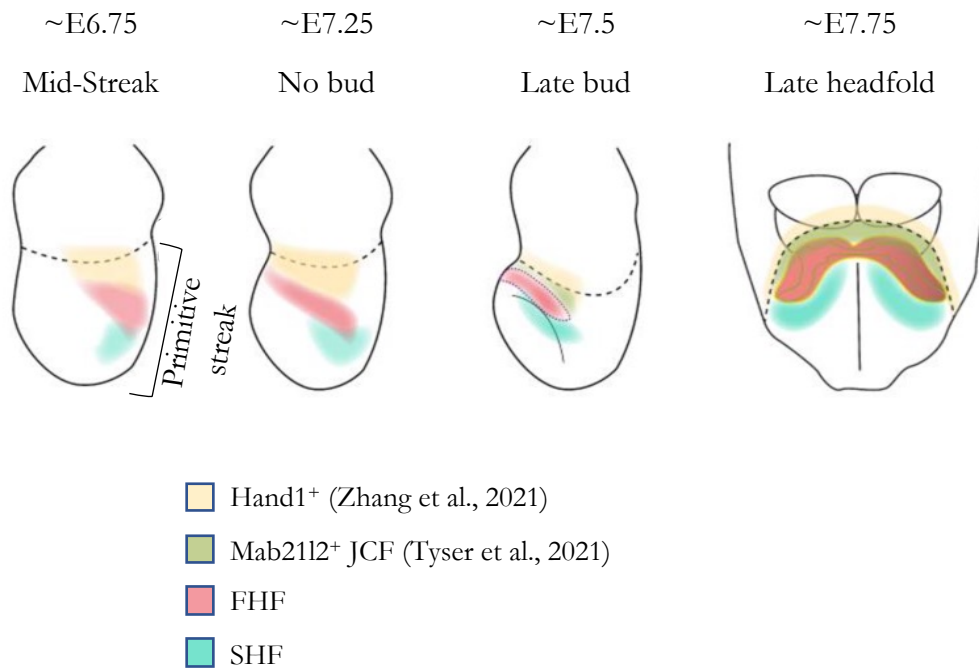
Figure 1: Deployment of cardiac progenitors during mouse cardiogenesis.

Positioning of early cardiac progenitors within heart field domains in the developing mouse heart. Embryos at ~E6.75, ~E.725 and ~E7.5 in the sagittal view; embryo at ~E7.75 in the frontal view. The upper dotted line depicts the border between embryonic and extraembryonic domains. FHF: first heart field; SHF: second heart field; JCF: juxta-cardiac field. Adapted from Sendra et al., 2022.

1.1.2 Formation of the functional heart

The early heart tube consists of two layers: the myocardium – the outer layer of cardiomyocytes, and the endocardium – the inner layer of endothelial cells. The endocardium will remain the innermost layer of the heart and become the source of the coronary endothelial cells and mesenchymal cells of the cardiac valves – separating the atria and ventricles – and the semilunar valves of the OFT, which are required for maintaining the unidirectional blood flow (Zhang et al., 2018). The third layer of the heart, the epicardium, forms the outermost epithelial layer of the heart and gives rise to the majority of non-myocardial cell populations (Meilhac and Buckingham, 2018). It originates from cells arising at E8.5 in the PE. At around E9.5, migratory cells from the PE cover the outer surface of the heart (Fig. 2). As mentioned above, studies in the mouse have recently shown that a subset of JCF-derived cells also contribute to

the PE and subsequently to the epicardium (see 1.1.1) (Tyser et al., 2021; Zhang et al., 2021). The subset of epicardial cells undergo an EMT giving rise to epicardial-derived cells (EPDCs), which can be detected within subepicardial space. Subsequently, those cells migrate into the myocardium to give rise to several cardiac cell types including vascular smooth muscle cells and cardiac fibroblasts, and presumably a subpopulation of cardiomyocytes, and endothelial cells (Gittenberger-de Groot et al., 2010).

Between E8.25 and E10.5 (in humans EGW 6.5-7.5 weeks; Krishnan et al., 2014), the heart tube folds upon itself in the process known as cardiac looping. Even though the primitive heart starts to beat at around E8.5 (in humans EGW 5; 23 days post-fertilization), cardiogenesis is still ongoing, with significant morphogenetic changes and extensive growth. The four chambers with partial septation and immature valves start to be identifiable at around E10.5 (in humans EGW 7.5 weeks) and all definitive structures become distinguishable at E14.5 – 9 EGW in humans (Fig. 2; Krishnan et al., 2014).
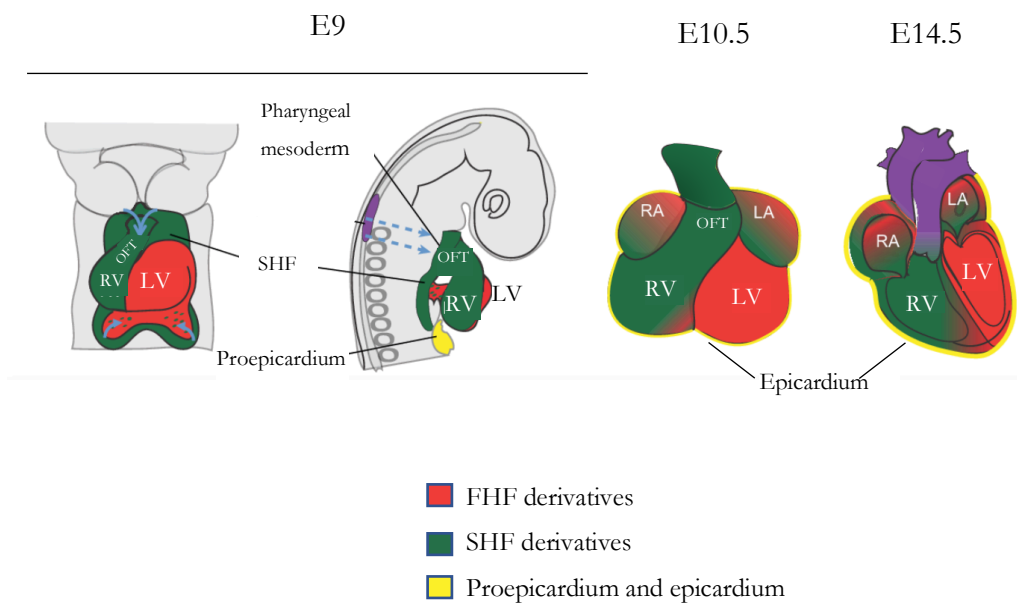
Figure 2: Progenitors of the FHF, SHF and PEO contribute to distinct structures of the developing heart.

Lineage contributions of the first heart field (FHF) (in red), second heart field (SHF) (in green), and proepicardial cells (in yellow) during the formation of the looping heart tube (E9) and the chambered heart (E10.5, E14.5). LA: left atrium; LV: left ventricle; RA: right atrium; RV: right ventricle; OFT: outflow-tract. Adapted from Santini et al., 2016.

1.2 Gene regulatory networks orchestrating early heart development

During development, pluripotent cells are specified, meaning that their ability to form alternative fates is progressively restricted. Along lineage commitment trajectories, cells transition through certain transcriptional states that define different cell types. Spatial and temporal control of gene expression levels is crucial for the proper specification of cardiogenic precursors towards more restricted cell types. A central role in integrating cues from the microenvironment, signalling cascades and genetic programs is played by transcriptional factors and their target genes, working together in complex GRNs.

1.2.1 Signalling pathways and key transcription factors involved in the lineage specification of cardiovascular progenitors

Cell patterning and lineage decisions during all stages of cardiogenesis, from the primitive streak to the looping heart tube, are coordinated by the interplay of diverse extracellular signalling pathways and regulatory genes. The timing, duration, and local concentration of opposing gradients of different signals orchestrate distal-proximal, anterior-posterior and dorsal-ventral patterning of the primitive streak and mesodermal lineages. These complex signalling networks include signal transducers such as Wingless-related integration site proteins (WNT); the TGF-β superfamily, such as Activin/Nodal; and Bone morphogenetic proteins (BMPs), Fibroblast growth factors (FGFs), Retinoic acid (RA) and Sonic hedgehog proteins (SHH) (reviewed by Witman et al., 2020; Parikh et al., 2015).

A crosstalk between WNT/β-catenin, Activin/Nodal and BMP signals is involved in the formation of the primitive streak (Funa et al., 2015; Ben-Haim et al., 2006; reviewed by Arnold and Robertson, 2009). During primitive streak elongation, Activin/Nodal activity is limited to the distal part of the streak around the node while BMP activity remains high in the proximal domain (Fig. 3). This leads to differing gradients of BMP and Activin/Nodal along the distal-proximal axis of the primitive streak, which in turn specify the fate of cells depending on their allocation while ingressing through the primitive streak (Parameswaran and Tam, 1995; Kinder et al., 1999; reviewed by Riviera-Perez and Hadjantonakis, 2014). The interplay of WNT/β-catenin, Activin/Nodal and BMP signals results in the emergence of cells expressing the T-box transcription factor *T* (T/TBXT/Brachyury) and Eomesodermin (*Eomes*) at the

primitive streak stage; these are the cells that will form the future heart. Almost all cardiovascular progenitors arise from population of cells expressing *Mesp1* (Kitajima et al., 2005), which is directly upregulated by T and EOMES (Costello et al., 2011; Bondue et al., 2008). Cells migrate from the primitive streak expressing *Mesp1* and downregulate it upon reaching the anterior-most domain of the embryo, where the activity of WNT signalling is inhibited by signals coming from the anterior endoderm (Costello et al., 2011; Bondue et al., 2008; Arai et al., 1997; WNT signalling during heart development reviewed by Gessert and Kuhl, 2010).
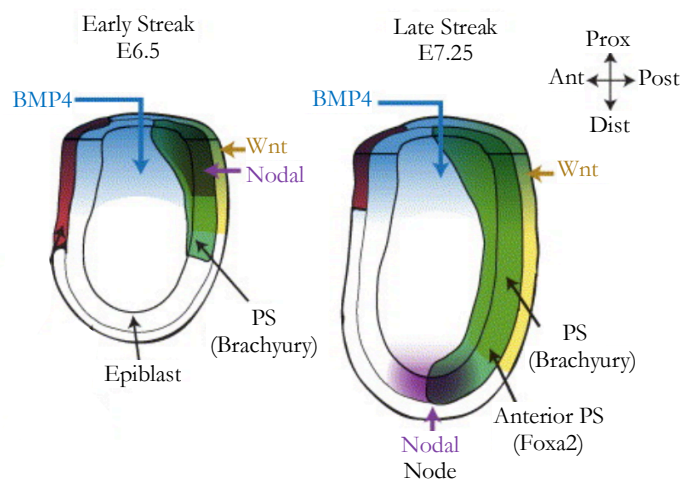
Figure 3: Signalling pathways involved in the patterning of mouse cardiac mesoderm.

Cardiogenesis is driven by gradients of various signalling pathways controlling the specification of cells in the primitive streak. Mouse embryos at E6.5 and E7.25 in sagittal view. PS: primitive streak; Ant: anterior; Prox: proximal; Dist: distal; Post: posterior. Adapted from Gadue et al., 2005.

Cardiogenic progenitors of different fields arising sequentially from *Mesp1*-positive cells express several overlapping markers, but in a defined temporal manner. The first wave of *Mesp1*-expressing FHF cells upregulate the expression of *Nkx2.5*, *Gata4* and *Tbx5* upon exposure to BMP signalling and inhibition of WNT signals while forming the cardiac crescent. This leads to the initiation of a cardiomyocyte differentiation gene program, including genes coding for contractile apparatus proteins such as myosin light and heavy chains (Devine et al., 2014; Klaus et al., 2007; Lien et al., 2002; Bondue et al., 2008; Luna-Zurite et al., 2016; Klaus et al., 2007; BMP signalling in cardiogenesis reviewed by Schultheiss et al., 1997). The second wave of *Mesp1*-expressing prospective SHF cells start to express *Isl1* upon exposure to WNT signalling when reaching the pharyngeal mesoderm. ISL1 is a key factor required for the

proliferation and maintenance of SHF progenitors, i.e., the prevention of their differentiation (Cai et al., 2003; Kwon et al., 2009; Cohen et al., 2012). SHH signals produced by the pharyngeal endoderm then activate the expression of *Tbx1* in the pharyngeal mesoderm (Yamagishi et al., 2002). In the aSHF, TBX1 activates the expression of *Fgf10* and *Fgf8* (Hu et al., 2004; Aggarwal et al., 2006). On the other hand, BMP signalling represses *Tbx1,* leading to myocardial differentiation of the OFT (Wang et al., 2011; Klaus et al., 2007).

While migrating from the primitive streak towards the anterior-most region of the developing embryo, cells are also exposed to various anterior-posterior patterning cues, including FGF and RA (Ciruna and Rossant, 2001; Sirbu et al., 2008; Martínez-Morales et al., 2011). The positioning of anterior and posterior subpopulations of the SHF develops with changing gradients of FGF and RA signalling. RA signalling notably induces expression of *Tbx5* in the posterior domain, which results in the downregulation of aSHF markers, e.g., *Tbx1* and *Fgf10* (Rana et al., 2014; De Bono et al., 2018).

1.2.2 Role of retinoic acid signalling in early cardiogenesis

Signalling mediated by RA differs from other signal transducers such as FGF, WNT, and TGF-β. Instead of initiating a signal by binding to receptors on the cell surface, RA diffuses inside the cell nucleus and forms heterodimers with RA receptors (RARα, RARβ, RARγ) and retinoid X receptors (RXRα, RXRβ, RXRγ). Unless liganded and activated by RA, these receptors bind to regulatory sequences of DNA known as RA-response elements (RARE) and repress the transcription of target genes. The binding of RA to heterodimers results in the recruitment of coactivators that facilitate the initiation of transcription. More rarely, RA binding can also lead to the recruitment of histone deacetylases, which results in the transcriptional repression of associated genes, as in the case of the regulation of *Fgf8* expression (reviewed by Cunningham and Duester, 2015).

The regulation of RA timing and dosage during development is of major importance, as the misdistribution of enzymes involved in both RA synthesis and degradation has been proven to have deleterious effects on embryos (Niederreither et al., 1999; Abu-Abed et al., 2001). During cardiogenesis, RA signalling influences heart tube patterning and looping, the septation of chambers and the OFT, and the trabeculation of ventricles (reviewed by Stefanovic and Zaffran, 2016).

Interestingly, RA seems to have a different effect on heart development depending on the species. The disruption of RA signalling in zebrafish influences both heart fields and the atrial and ventricular cardiomyocytes that derive from them, while in mice it primarily affects SHF derivatives (D'Aniello et al., 2013; Waxman et al., 2008; Keegan et al., 2005). In mice, loss of RA signals results in the expansion of the anterior domain of the heart tube at the expense of the posterior domain (Fig. 4a'-a''; Hochgreb et al., 2003). Limiting the *Fgf8-Isl1* expression domain through RA repression of *Fgf8* is required for proper heart tube patterning (Sirbu et al., 2008). However, RA also plays an instructive role by inducing *Hox* genes expression. *Hox* genes are conserved homeodomain transcription factors which are expressed sequentially within the primitive streak and play a role in determining the positional identity of cells along the anterior-posterior axis. (Langston et al., 1997; Stefanovic et al., 2020).
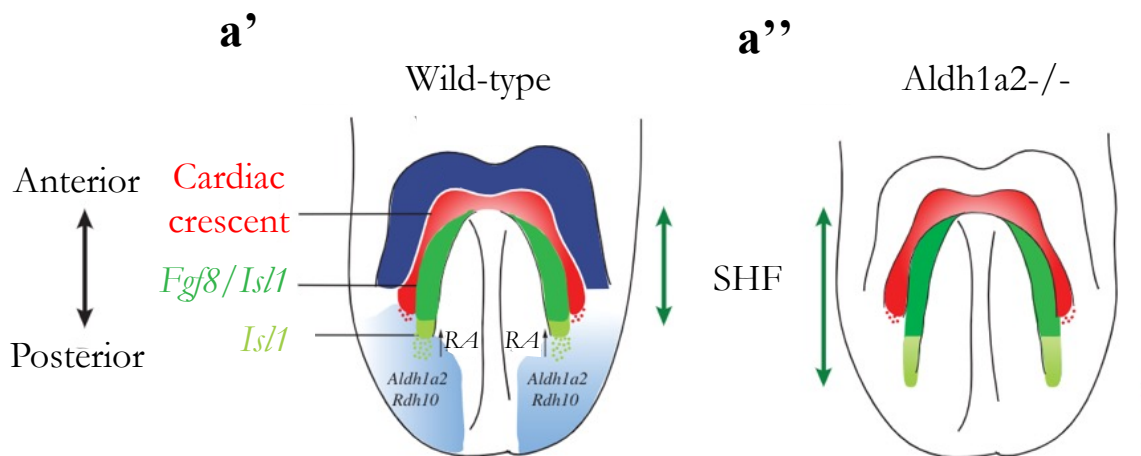
Figure 4: The anterior-posterior patterning of the mouse heart depends on the antagonism of RA and FGF signalling.

(a') Correct patterning of differentiated FHF progenitors within the cardiac crescent (in red) and SHF progenitors (in green) located between domains of RA degradation (dark blue) and RA synthesis (light blue) in wild-type mouse embryos at E8. **(a'')** Posterior expansion of the SHF domain expressing *Fgf10* and *Isl1* in absence of RA synthesis by ALDH1A2. RA: retinoic acid; SHF: second heart field. Adapted from Bernheim and Meilhac, 2020.

1.2.3 lncRNAs involved in cardiogenesis

In recent years, long non-coding RNAs (lncRNAs) have emerged as important regulators of gene expression. lncRNAs are more than 200 base pairs (bp) long RNA molecules with limited protein-coding potential which hold versatile, but still mainly unknown, functions in gene regulation (e.g., chromatin decompaction, recruitment/sequestering of chromatin modifiers to/from promoters of associated genes to induce or inhibit their transcription, interfering with the transcription machinery, splicing and translation of mRNA, and many others) and genome homeostasis (e.g., DNA damage and repair response). Genes that code for lncRNAs are less conserved than those coding for proteins and tend to be more specific to cellular types and states (The ENCODE Project Consortium, 2012; The ENCODE Project Consortium et al., 2020; reviewed by Statello et al., 2021).

Several lncRNAs with important roles in heart development have been identified, including Braveheart, Fendrr, Upperhand (Uph) and Handsdown (Hdn), among others (Fig. 5; Klattenhoff et al., 2013; Grote et al., 2013; Ritter et al., 2019). Braveheart (Bvht) is expressed in early cardiogenic progenitors and neonatal cardiomyocytes in mice but not in humans. Knockdown of *Bvht* in mouse embryonic stem cells leads to the downregulation of *Mesp1* and upregulation of *Brachyury* and *Eomes* and, subsequently, to a failure of cardiac gene program initiation (Klattenhoff et al., 2013). Another lncRNA important for early cardiogenesis, Fendrr, is transiently expressed in the caudal end of the lateral plate mesoderm and plays a role in the lineage commitment of this structure towards heart and trunk muscles (Grote et al., 2013). Many lncRNAs also modulate later cardiovascular gene programs. One example is Hdn, which regulates the expression of *Hand2*, an important factor for heart chamber morphogenesis. When *Hdn* transcription is activated, looping of chromatin enables the interaction between the *Hdn* promoter and *Hand2* enhancers, which subsequently leads to the unavailability of the *Hand2* enhancer for activation of Hand2 (Ritter et al., 2019). This exquisite regulatory mechanism allows for precise control of gene expression and proper development of the heart.
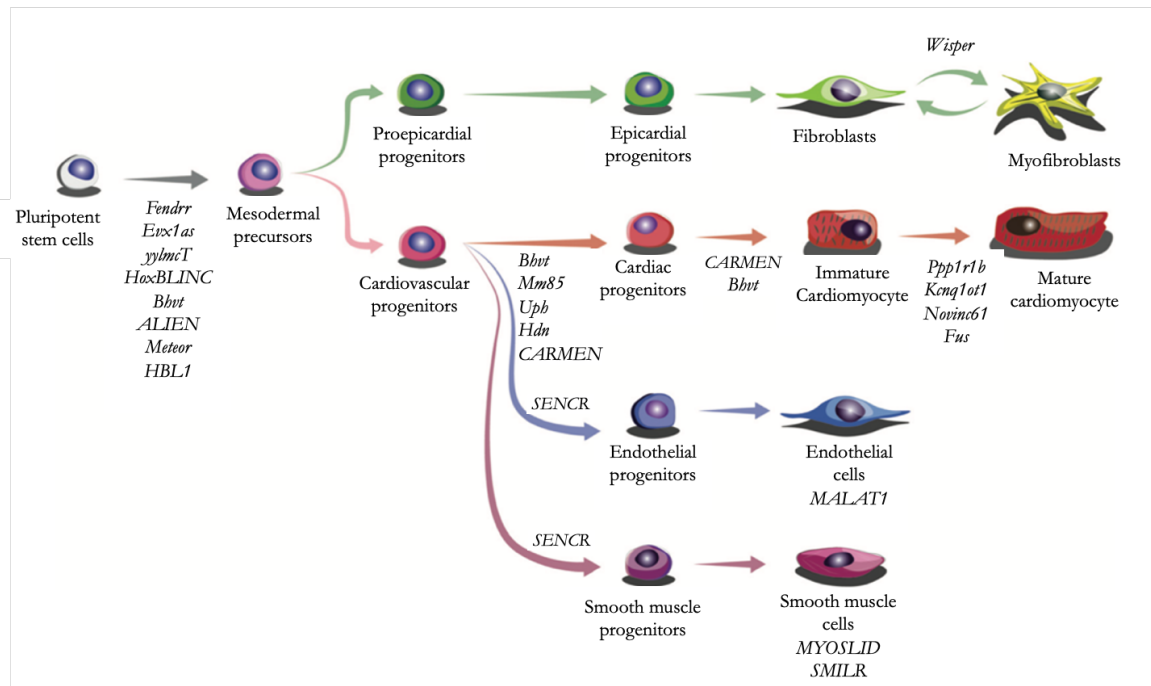
Figure 5: LncRNAs are involved in cardiac cell lineage commitment at various stages of heart development.

Schematic representation of cardiac lncRNAs that have been functionally characterized so far and the associated cardiovascular cell types. Adapted from Alexanian and Ounzain, 2020.

1.3 Congenital heart diseases

Defects in many processes of cardiac development can lead to congenital heart diseases (CHDs). CHDs represent a wide spectrum of structural heart defects, ranging from minor septal defects to complex life-threatening manifestations consisting of multiple anomalies (reviewed by Lin et al., 2021). The prevalence of CHDs is estimated to be ~1% of live births (Hoffman & Kaplan, 2002). They are the most common type of birth malformation and represent a major cause of infant mortality and morbidity in developed countries. In the last decades, survival rates of newborns have increased due to advances in pre- and postnatal diagnostics, early interventions, and others (Knowles et al., 2012). This has provided further motivation to elucidate the mechanisms of CHDs and discover novel therapeutic strategies for affected adults.

The majority of CHDs are caused by multifactorial (genetic and environmental) or unidentified causes. Some are attributed to monogenic or chromosomal anomalies and copy

number variants. Manifestations of CHD phenotypes might be affected by variance in genetic penetrance, a multitude of misexpressed genes or mechanobiological stress affecting structures of the developing heart (Pierpont et al., 2008). Advances in next-generation sequencing technologies and analyses of large cohorts of patients have accelerated the identification of genes associated with CHDs (e.g., whole-exome sequencing of patients affected by CHD undertaken by the Pediatric Cardiac Genomics Consortium (Pediatric Cardiac Genomics Consortium et al., 2013; Jin et al., 2017; Sifrim et al., 2016). Many of these genes encode transcription factors (e.g., *GATA4, NKX2.5, TBX5*) and transcriptional and epigenetic modulators (e.g., *CHD7, KMT2D*) related to heart development (Ang et al., 2016; Garg et al., 2003; Basson et al., 1997; Benson et al., 1999; Corsten-Janssen et al., 2013; Digilio et al., 2017). Transcriptomic analyses of single cells from healthy and patient tissues have additionally indicated that the expression of genes linked with CHDs can be dysregulated in specific cardiac cell types rather than ubiquitously (Miao et al., 2020b). Nevertheless, the precise understanding of many of these defects remains elusive due to the technical and ethical inaccessibility of human foetal tissues, which has resulted in a need to explore other models.

1.3.1 *In vitro* modelling of CHDs using human induced pluripotent stem cells

Human induced pluripotent stem cells (hiPSCs) are a type of pluripotent stem cell that are generated through reprogramming of human somatic cells. Their capacity to self-renew indefinitely and ability to differentiate into various cell types in nearly unlimited amounts make them a powerful tool for modelling human diseases. hiPSCs can be used to study the underlying mechanism of disease, its pathogenesis and progression, as well as to test possible treatments (Ang et al., 2016; Miao et al., 2020b). Importantly, the generation of hiPSC lines from patients affected by CHDs opens the possibility to study the disease phenotype in the patient-specific genetic background, which was not feasible with animal models. Combining hiPSC-derived *in vitro* models with genome editing tools (e.g., the CRISPR/Cas9 system) that can introduce or correct specific gene mutations has become a gold standard for precisely dissecting the underlying causes of disease. So far, hiPSCs have been used to model supravalvular aortic stenosis (Ge et al., 2012), calcific aortic valve disease (Theodoris et al., 2015), cardiac septal defects (Ang et al., 2016), Barth syndrome (Wang et al., 2014), and Hypoplastic left heart syndrome (HLHS) (Hrstka et al., 2017; Yang et al., 2017; Miao et al., 2020b), among others.

1.3.2 *In vitro* modelling of hypoplastic left heart syndrome

HLHS, which is primarily characterised by an underdeveloped left ventricle and malformations of the left ventricular outflow tract, is one of the most complex and severe birth defects. The underlying mechanisms are still poorly understood, partly because of the high genetic heterogeneity of the disease (Yagi et al., 2018). Due to some differences between human and mouse cardiovascular development as well as the heterogenous phenotype of the disease, not all types of HLHS can be replicated using mouse models (Krishnan et al., 2014). Initial approaches to model HLHS using cardiomyocytes generated from patient hiPSC lines revealed impaired cell lineage commitment and differentiation as well as higher immaturity in gene expression profiles compared to healthy hiPSC-derived cells (e.g., reduced expression of key mesodermal markers (*MESP1*) at early stages and of genes important for cardiomyocyte function later on (*TNNT2*, *GJA1*; Jiang et al., 2014; Kobayashi et al., 2014). Overall, these defects led to reduced yield and functionality of the obtained cardiomyocytes, including disrupted sarcomere organisation and reduced contractile force and beating frequency. Recently, Krane et al., (2021) used scRNA-Seq and 3D modelling of HLHS-hiPSC derived cells to uncover dysregulations in cell cycle, unfolded protein response and autophagy as well as a premature cell cycle exit of ventricular cardiomyocytes as probable causative factors resulting in defective differentiation, maturation and exacerbated apoptosis of cardiomyocytes in HLHS (Krane et al., 2021), while Miao et al. reported defects in endocardial function as another pathogenetic factor (Miao et al. 2020b).

1.4 Aims

Given our limited understanding of early human cardiogenesis, the *in vitro* generation of various human cell types and states faithfully corresponding to their *in vivo* counterparts would be of tremendous advantage for modelling processes of cardiac development and disease, ultimately facilitating the discovery of novel therapeutic approaches.

To address this need, this thesis had the following aims:

1. Establish and optimize protocols to direct human pluripotent stem cells (hPSCs) towards the cardiac lineage and obtain different cardiac mesodermal populations, including early multipotent cardiac progenitors and their late derivatives.

2. Monitor the emergence of the various cell populations using a reporter cell line and isolate distinct subpopulations of human cardiac progenitors, with a specific focus on progenitors of the first and second heart fields (FHF and SHF).

3. Thoroughly characterize the different cardiac progenitor populations and their derivatives using single-cell transcriptomics, among others, with the goal to capture and define dynamic cellular states of human cardiogenesis.

4. Explore the utility of these *in vitro*-derived defined progenitor pools in addressing questions related to developmental malformations by using HLHS patient-specific hiPSC lines.

5. Identify gene regulatory elements important for early cardiogenesis with a focus on lncRNAs and RA-response elements and perform a functional assessment of selected elements.

2 Materials and methods

2.1 Cell culture

All cell culture experiments were done as described in Zawada et al., 2022 (preprint).

2.1.1 Cell lines

All human pluripotent stem cell lines used in experimental procedures are listed in Tables 1 and 2. Identification numbers assigned through Human Pluripotent Stem Cell Registry (hPSCreg) can be referred to for detailed descriptions. Experimental procedures using human embryonic stem cell line ES03 (hPSCreg ID ESIBIe003) were authorised by Robert Koch Institute (AZ 3.04.02/0131). This study was approved by the Ethics Commission of the TUM Faculty of Medicine (# 447/17 S).

Table 1: Human pluripotent stem cell lines used in this study.

Cell type	Human Pluripotent Stem Cell Registry ID	Identifier in this work	Sex	Comments
hiPSC	MRIi003-A	hiPSC	Male	
hiPSC	MRIi003-A-8	hiPSC eGFP+	Male	Parental line: MRIi003-A
hESC	ESIBIe003	H3 WT	Female	
hESC	N/A	H3 TN	Female	Parental line: ESIBIe003
hiPSC	N/A	HDN $^{\lrcorner TSS}$	Male	Parental line: MRIi003-A

Table 2: Patient-derived pluripotent stem cell lines used in this study.

Cell type	KaBi-DHM ID / Helmholtz ID	Identifier in this work	Sex	Health status
hiPSC	375 / 69270	HLHS	Male	HLHS

The H3 WT line was a kind gift of Dr. David Elliott (MCRI, Australia). H3 TN line was generated by Dr. Daniel Ortmann and Dr. Maraiestela Ortiz (University of Cambridge, UK) for this project using homologous recombination resulting in one allele of TBX5 and NKX2.5 being replaced with the sequence encoding for the fluorescent protein. hiPSC GFP+ line was generated in house by Fangfang Zhang (Klinikum Rechts der Isar, Germany).

2.1.2 hESC/iPSC cell culture

All pluripotent stem cell lines were maintained in E8 medium (A1517001, Gibco/Invitrogen) containing 0.5% Penicillin/Streptomycin (Gibco; 15140-122) on Geltrex (A1413302, Gibco/Invitrogen) coated dishes under standard culture conditions (37°C, 5% CO_2); medium was changed daily. Cells were non-enzymatically passaged every 4-5 days using 0.5 mM EDTA (AM9912, Invitrogen) in PBS solution without Ca^{2+} or Mg^{2+} (PBS(-/-); Gibco; 10010023). To promote survival, Thiazovin (SML1045, Sigma) was added at a concentration of 2 µM for 24 h after passaging. For long-term storage, cells were frozen in a cryopreservation medium consisting of 10% DMSO (Sigma-Aldrich; D2650) diluted in E8 at - 80°C before transferring to liquid nitrogen.

Karyotype analysis was performed at the Institute of Human Genetics of the Technical University of Munich using standard methodology.

2.1.3 Cardiac differentiation

Cardiac differentiation protocol was adapted from Mendjan et al., 2014, Hofbauer et al., 2021 and described in Zawada et al., 2022 (preprint). Briefly, for cardiac differentiation, cells were passaged normally and plated on Geltrex-coated 24-well plates at a density of 200 000 cells/well in E8 medium containing 2 µM Thiazovin (SML1045, Sigma; day -1). 18 - 24 hours later, media was changed for E8 medium without TV. And 6 - 10 hours later media was changed for 1:1 DMEM/F-12 with Glutamax (31331028, Gibco/Invitrogen) and IMDM (21980032, Gibco/Invitrogen) media containing 0.1 g/ml BSA (A9647, Sigma), 30 mg/ml transferrin (10652202001, Roche), 1% chemically defined lipid concentrate (11905031, Gibco/Invitrogen), ~0.46 mM (0.004%) of thioglycerol (M6145, Sigma (referred here as CDM-BSA media), with addition of 10 ng/ml recombinant human BMP4 (314-BP, R&D) and 1.5 µM CHIR99021 (4423, R&D) together with 50 ng/ml Activin A (SRP3003, Sigma), 30 ng/ml bFGF (233-FB, R&D), 5 µM LY 294002 hydrochloride (1130, R&D). After 40 h, media was exchanged for CDM-BSA with addition of 10 ng/mL recombinant human BMP4, 8 ng/mL bFGF, 10 µg/ml insulin (11376497001, Roche/Sigma), 5 µM IWP2 (72122, Stem Cells; referred here as CDM-Meso media). Additionally, media was supplemented (or not) with RA (R2625, Sigma) in a concentration of 0.5 or 1 µM depending on the treatment. CDM-Meso media was changed for the next 4 days at exactly 24h intervals. Depending on the treatment RA was added to the media for 2 or 4 days. On days 5 and 6, media was changed for CDM-BSA media containing 10 ng/ml recombinant human BMP4, 8 ng/ml bFGF, and 10 µg/ml insulin. From day 7 on, media was changed every second day for CDM-BSA media with the addition of 10 µg/ml insulin (referred to as CDM-Maintenance media).

2.1.4 Aggregation of flow cytometry-based sorted cardiac progenitors

For aggregation, at d4.5 of differentiation cells were washed two times with PBS(-/-), dissociated with Accutase (A11105013, Gibco/Invitrogen 5min, 37°C), centrifuged at 1200 rpm for 5min, resuspended in 2% FCS in PBS, filtered through a 40 µm filter and subjected to a sorting procedure on FACS Aria III Cell Sorter (BD). DAPI (D3571, ThermoFisher Scientific) staining was used to discriminate between dead and live cells (final concentration 0.01 ng/µl). Cells were sorted based on mCherry and eGFP expression and collected into 50-

100% FCS in PBS. Then, cells were centrifuged, resuspended in CDM-BSA media, counted, centrifuged again, and resuspended in CDM-Meso including 0.5% Penicillin/Streptomycin (15140122, Gibco/Invitrogen), 10 µM of Rock Inhibitor Y- 27632 (688000, Calbiochem) and supplemented with or without RA in concentration depending on the treatment in the density of 100 000 cells per 200 µl per well U-shaped 96-well plate covered with 5% Poly(2-hydroxyethyl methacrylate (P3932, Sigma). Plates were centrifuged for 2 min at 1200 rpm and transferred to an incubator. From the next day on, the differentiation protocol as described above was applied. At d7 aggregates were transferred into wells of a 48-well plate covered with 5% Poly(2-hydroxyethyl methacrylate) and put on a shaker. Also, from d7 on, CDM-Maintenance media was additionally supplemented with 50 ng/µl VEGF (293-VE, R&D).

2.1.5 Epicardial differentiation

For epicardial differentiation, d4.5 FACS sorted ITGA8+ cells were re-plated on 0.1% gelatin (G1393, Sigma Aldrich) coated wells of 12-well chamber slides (81201, Ibidi) in a density of 20 000 cells per 1 cm^2 and subjected to a modified protocol of Bao et al. (2017). Briefly, cells were re-plated in Advanced DMEM F12 (12634010, Gibco/Invitrogen) containing Glutamax (35050061, Gibco/Invitrogen), Ascorbic Acid (0.1 mg/ml; A5960, Sigma) and 0.5% Penicillin/Streptomycin (LaSR media) with addition of 1% FCS and 10 µM of Rock Inhibitor Y- 27632. On d6 and d7 media was changed for LaSR media with the addition of 3 µM CHIR99021. From d8 onwards media was changed daily for LaSR media till d12.

2.1.6 Epicardial – myocardial differentiation

For simultaneous epicardial and myocardial differentiation, hiPSC eGFP- derived d4.5 cardiac progenitors obtained through cardiac differentiation protocol described above were replated on Geltrex-coated wells of 12-well chamber slides (81201, Ibidi) and 96-well plates (89646, Ibidi) in density of 40 000 cells per 1 cm^2. Immediately, FACS sorted d4.5 eGFP+/ITGA8+ and eGFP+/ITGA8- cells were added in the number of 5-10 cells per well resulting in 0-5 cells per well on the next day. Cells then were subjected to modified cardiac differentiation protocol described above. Modulations of BMP, RA and WNT signaling cues

were described previously to favor either myocardial or epicardial lineages (Wiesinger et al., 2022) and thus we decided to use low levels of WNT signaling inhibition with concomitant addition of BMP4 and RA. Briefly, cells were re-plated in CDM-BSA media supplemented with 20 ng/ml BMP4, 8 ng/ml bFGF, 10 μg/ml insulin, 2.5 μM IWP2, 1 μM RA with addition of 0.5% Penicillin/Streptomycin, 1% FCS, 10 μM of Rock Inhibitor Y-27632. On d6 and d7 the medium was replaced with CDM-BSA supplemented with 20 ng/ml BMP4, 8 ng/ml bFGF, 10 μg/ml insulin, and 1 μM RA with addition of 0.5% Penicillin/Streptomycin. From d8 onwards CDM-Maintenance with addition of 0.5% Penicillin/Streptomycin media was replaced daily till d15.

2.1.7 Dissociation of cells and aggregates

For fluorescence-activated flow cytometry analysis and RNA collection cells were dissociated using Accutase up to day 8 and using papain from d8 on. When dissociating aggregates, aggregates were transferred into tubes and media was removed from above; when dissociating cells in dish plates - media was removed. Then aggregates/cells were washed 2 times with 1 ml of 2 mM EDTA. Dissociation was carried out using papain solution prepared in advance (Fischer et al., 2018). Briefly, papain solution was prepared by diluting 1:1 with a PBS(-/-) solution containing diluted in PBS(-/-) papain (LS003124, Worthington Biochemical Corporation) in a concentration of 40 U/ml and 10 μl/ml L-cysteine (stock 0.2 M; C6852, Sigma) that was incubated for 10 min in 37°C. Cells/aggregates were incubated with such prepared papain solution for 20 - 40 min at 37°C (the optimal dissociation time was previously determined for each cell line/timepoint to obtain single-cell suspension without compromising cell quality and survival). Aggregates were incubated on a shaker at 37°C. After that time, the solution containing 1 mg/ml trypsin inhibitor (T9253, Sigma) was added. Cells/aggregates were dissociated by pipetting, transferred to a tube containing PBS(-/-), centrifuged at 1200 rpm for 5 min and resuspended in buffers/solutions depending on the downstream analysis.

2.1.8 Cardiomyocytes replating for immunofluorescence analysis

At d30 of cardiac differentiation, media was changed to EB2 media consisting of DMEM/F12 (21331-020, Gibco), 2% FBS, 1% L-glutamine (25030-081, Gibco), 1% non-essential aminoacids (11140-050, Gibco), 0.5% Penicillin/Streptomycin and 0.1mM beta-mercaptoethanol and changed every second day. At d40 of cardiac differentiation, cells were subjected to papain-based dissociation as described above. Cells were replated on fibronectin (F1141, Sigma) coated wells of 12-well chamber slides in a density of 140 000 – 200 000 cells per 1 cm^2 in EB20 media consisting of consisting of DMEM/F12, 20% FBS, 1% L-glutamine, 1% non-essential aminoacids, 0.5% Penicillin/Streptomycin and 0.1mM beta-mercaptoethanol. The next day media was changed for EB2 and changed daily till day 45.

2.2 Transcriptomic analysis

2.2.1 RNA isolation, reverse transcription PCR (RT-PCR), and quantitative real-time PCR (qPCR)

RNA isolation, RT-PCR and qPCR were done as described in Zawada et al., 2022 (preprint). For RNA isolation cells/aggregates were dissociated as described previously. In the case of d0 - d8 RNA collection was performed 2 hours after media change. In the case of aggregates, 3 - 5 aggregates were pulled from each differentiation for RNA collection. In case of sorted cells, cells were subjected to the sorting procedure as described before. After centrifugation, cell pellets were lysed and RNA was isolated using the Absolutely RNA Microprep Kit (400805, Agilent Technologies). In the case of aggregates and lower amounts of sorted cells Absolutely RNA Nanoprep Kit (400753, Agilent Technologies) was used. RNA was quantified using Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific) and analysed with ND1000 software. 0.3 - 0.5 μg of RNA was used to synthesize cDNA with the HighCapacity cDNA Reverse Transcription kit (4368813, Applied Biosystems; 0.3 μg for cardiomyocytes; 0.5 μg for d0-d8). Samples were diluted 1:2 and the gene expression was quantified by qPCR using 1 μl cDNA and the Power SYBR Green PCR Master Mix (4367659, Applied Biosystems) and 7500 Real-Time PCR System (Applied Biosystems), using the following program:

Materials and methods

Stage	Temperature	Duration	Cycles
Holding stage	50°C	20 s	1
	95°C	10 min	
Cycling stage	95°C	15 s	40
	60°C	1 min	
Melting curve	95°C	15 s	1
	60°C	1 min	
	95°C	30 s	
	60°C	15 s	

Sequences of primers used for qPCR analysis are listed in Table 3. Gene expression levels were quantified relative to *GAPDH* expression using the ΔCt method unless otherwise indicated.

Table 3: qPCR primer sequences.

Gene	Forward primer sequence	Reverse primer sequence
BMP4	CACAGCACTGGTCTTGAGT	TGGTCCCTGGGATGTTCT
RSPO3	CCAGAAGGGTTGGAAGCCAACA	CCTTCTTCGTGCATGGACTCCA
SOX4	TCTGCACCCCCAGCAAGA	CACCCCGGAGCCTTCTGT
ISL1	AAAGTTACCAGCCACCTTGGA	ATTAGAGCCCGGTCCTCCTT
FGF10	CATCAGTAGAAATCGGAGTTGTTG	CCTCTCCTTCAGCTTACAGTCATT
TBX1	GTGTGAGCGTCAGCTAGAG	TCACTTGGAAGGTGGGAAAC
WNT5A	CTGCAGAGAGGCTGTGCTCC	TTCTCCTTCGCCCAGGTTGT
ITGA8	GCTGCTGGGGAGTTTACTGG	GATGCCATCTGTTCTCCCGTG
CXCR4	CCCTCCTGCTGACTATTCCC	TAAGGCCAACCATGATGTGC

HOXB1	CTCCTCTCCGAGGACAAGGAA	CTGTCTTGGGTGGGTTTCTCTTAA
HAND1	AACTCAAGAAGGCGGATGG	CGGTGCGTCCTTTAATCCT
mCherry	GAACGGCCACGAGTTCGAGA	CTTGGAGCCGTACATGAACTGAGG
HAND2	AGAAGACCGACGTGAAAGAGG	TTCTTGTCGTTGCTGCTGCYCAC
eGFP	GGAGCGCACCATCTTCTTCA	AGGGTGTCGCCCTCGAA
FN1	CTGGCCAGTCCTACAACCAG	CGGGAATCTTCTCTGTCAGCC
LTBP3	CCAATGGCTCCTACAGATGTC	GGCTGCACTCATCTATGTCTTG
KCNJ3	TCATCAAGATGTCCCAGCCCAAGA	CACCCGGAACATAAGCGTGAGTTT
KCNA5	AGTGTAACGTCAAGGCCAAGAGCA	TCACAAATCTGTTTCCCGGCTGGT
THBS4	ACCGACAGTAGAGATGGCTTCC	CGTCACATCTGAAGCCAGGAGA
MESP1	GTGCTGGCTCTGTTGGAGA	CAGAGACGGCGTCAGTTGT
IRX4	GGCTATGGCAACTACGTGACCT	AGAGCTGGCTCGTAAGGGTAGT
LGR5	CTTCCAACCTCAGCGTCTTC	TTTCCCGCAAGACGTAACTC
WNT2	TAGTCGGGAATCTGCCTTTG	TTCCTTTCCTTTGCATCCAC
NKX2.5	CAAGTGTGCGTCTGCCTTTC	CGCACAGCTCTTTCTTTTCGG
T	TGTTTATCCATGCTGCAATCC	CCGTTGCTCACAGACCACAG
TNNT2	AGCATCTATAACTTGGAGGCAGAG	TGGAGACTTTCTGGTTATCGTTG
GAPDH	TCCTCTGACTTCAACAGCGA	GGGTCTTACTCCTTGGAGGC
MYL2	TACGTTCGGGAAATGCTGAC	TTCTCCGTGGGTGATGATG
MYL3	AAGGAGGTCGAGTTTGATGCT	TCCTTGAACTCTTCAATCTGCTC
TBX5	GGGCAGTGATGACATGGAG	GCTGCTGAAAGGACTGTGGT
FHL2	GTGGTGTGCTTTGAGACCCTGT	GAGCAGTGCAAACAGGCTTCATG

21

MYH7	TCGTGCCTGATGACAAACAGGAGT	ATACTCGGTCTCGGCAGTGACTTT
BNC2	ACATGGCAAAACGCTGATACTAA	TTCCCTGGCTGAAAACATTC
MAB21L2	GAGTGCTACTCGCTGACCG	CACTGAGAGGCACTTGTTTC
HOXB6	CACTCCGGTCTACCCGTGGATGCA	CATATCTTGATCTGCCTCTCCGTCAG
POSTN	CCGAGCCTTGTATGTATGTTATG	TTT GTA CAC CAA GCA CCT ATT T
MYL7	CCGTCTTCCTCACGCTCTT	TGAACTCATCCTTGTTCACCAC
TNNI1	GCTCCACGAGGACTGAACAA	CTTCAGCAAGAGTTTGCGGG
TNNI3	CCTCCAACTACCGCGCTTAT	CTGCAATTTTCTCGAGGCGG

2.2.2 RNA isolation, reverse transcription PCR (RT-PCR), and quantitative real-time PCR (qPCR) of sorted cells at d4.5

For RNA isolation of ITGA8+ sorted cells, after sorting and centrifugation in PBS, cell pellets of 20 000 cells were lysed and RNA was isolated using the Rneasy Microkit (74004, Qiagen). 40 ng was used to synthesize cDNA with the SuperScript IV Vilo Master Mix (11756050, Invitrogen). Gene expression was quantified by qPCR using 1µl cDNA and the Power SYBR Green PCR Master Mix. Gene expression levels were quantified relative to *GAPDH* expression using the ΔCt method unless otherwise indicated.

2.2.3 RNA-Seq

For RNASeq at d5.5 and d6.5 of differentiation, cells were dissociated with Accutase as described above, resuspended in 2% FCS in PBS, filtered through a 40 µm filter and subjected to sorting procedure on FACS Aria III Cell Sorter (BD). DAPI staining was used to discriminate dead and live cells (final concentration 0.01 ng/µl). Cells were sorted based on mCherry and eGFP expression and 500 000 cells per sample were collected and processed using

RNeasy Mini Kit. For each experimental conditions two biological replicates were prepared from 1 µg of total RNA. RNA was depleted of rRNA using Ribo-Minus technology. Libraries were prepared from purified RNA using ScriptSeq v2 (SSV21106, Invitrogen) and were sequenced on Illumina HiSeq platform. We obtained 120×10^6 paired-end reads of 150 bp length. Library preparation and sequencing was performed by Novogene (Cambridge, UK).

2.2.4 Single-cell RNA sequencing (scRNAseq)

scRNA-Seq was done as described in Zawada et al., 2022 (preprint). Briefly, for scRNASeq at d1.5, cells were washed two times with PBS, dissociated with Accutase (3-5 min, 37°C), centrifuged at 1200 rpm for 5 min, resuspended in 0.04% BSA in PBS, filtered through a 40 µm filter and counted. For scRNASeq at d4.5 of differentiation, cells were dissosciated with Accutase as described above, resuspended in 2% FCS in PBS, filtered through a 40 µm filter and subjected to sorting procedure on FACS Aria III Cell Sorter (BD). DAPI staining was used to discriminate dead and live cells (final concentration 0.01 ng/µl). Cells were sorted based on mCherry and eGFP expression and collected into 10% FCS in PBS. Then cells were centrifuged, washed in 0.04% BSA in PBS, centrifuged again, filtered through a 40 µm filter and resuspended in 0.04% BSA in PBS for counting. For scRNASeq at d30 of aggregates differentiation, aggregates were dissociated with papain, filtered through 40 µm filter, centrifuged at 1000 rpm for 3 min and resuspended in 0.04% BSA in PBS for counting. After counting, 10 000 cells for each sample were processed using Chromium Single Cell 3' Library & Gel Bead Kit v3.1 (1000075, 10x Genomics) Chromium Single Cell B Chip Kit (1000073, 10x Genomics) and Chromium i7 Multiplex Kit (220103, 10x Genomics) to generate Gel Bead-In-EMulsions (GEMs) and single cell sequencing libraries. Libraries were pooled and sequenced using the NextSeq 500/500 (Ilumina; High Output v2 kit 75 cycles v2.5 flow cell) with 28 cycles in read1 for the 10x barcodes and UMIs in and 8 cycles i7 index read and 58 cycles for cDNA in read2 with a read depth of at least 20 000 pair reads per cell.

Quality control of cDNA samples using Bioanalyzer (Agilent, Germany) and sequencing was performed by Dr. Rupert Öllinger (TUM, Germany).

2.3 Flow cytometry analysis

All flow cytometry analyses were done as described in Zawada et al., 2022 (preprint).

2.3.1 Flow cytometry of live cells

For live flow cytometry analysis, cells were washed two times with PBS, dissociated with Accutase (3 - 5 min, 37°C) or papain as described before, centrifuged at 1200 rpm for 5 min, resuspended in 2% FCS in PBS, filtered through a 40 μm filter and subjected to live flow cytometry analysis procedure on Cytoflex S (Beckman-Coulter). Analysis was performed using Kaluza software (Beckman-Coulter).

2.3.2 Flow cytometry of intracellularly/intranuclearly stained cells

For intracellular/intranuclear staining for flow cytometry cells were dissociated, counted and distributed in equal numbers per sample in 15 ml tubes (mainly 2 x 10^6 cells per sample per tube). Cells were fixed with 4% PFA for 7 min at room temperature (500 μl / 10^6 cells), then centrifuged 3 min at room temperature at 300 - 400 g. Next cells were washed 3 times with PBS (shaked for 5 min and centrifuged 5 min at 300 – 400 g between each wash). Then cells were stored in 2% FCS in PBS+/+ at 4°C or right away incubated with blocking/permeabilization buffer containing 10% FCS, 0.1% Triton-X-100 (X100, Sigma), 0.1% saponin (47036, Sigma) in PBS(+/+) (for intranuclear staining) or 10% FCS, 0.1% saponin in PBS+/+ (for intracellular/membrane staining) for 1h at room temperature on a shaker (1 ml / 10^6 cells). Then cells were centrifuged and incubated with the primary antibody (listed in Table 4) diluted in 1% FCS, 0.1% saponin with or without 0.1% Triton-X-100 in PBS+/+ (500 μl / 10^6 cells), overnight at 4°C on a shaker. Then cells were washed three times with 0.1% saponin in PBS+/+ with or without 0.1% Triton-X-100 for a total of 45 min on a shaker (cells were centrifuged between washes). Then the secondary antibody (listed in Table 5) was added diluted in 1% FCS, 0.1% saponin in PBS+/+ with or without 0.1% Triton-X-100, (500 μl / 10^6 cells) and cells were incubated for 1 hour at room temperature, protected from light, on a shaker. After that time, cells were washed three times with 0.1% saponin in PBS+/+ with or without 0.1% Triton-X-100 for a total of 45 min on a shaker (cells were

centrifuged between washes). Next, cells were resuspended in 2% FCS in PBS+/+ (100 μl / 10^6 cells), passed through a 40 μm strainer and subjected to the analysis on CytoFlex S or Gallios (Beckman-Coulter). Analysis was performed using Kaluza software (Beckman Coulter). No-primary antibody, no-secondary antibody, and IgG antibody controls were performed.

2.3.3 Flow cytometry of surface protein-stained cells

For surface staining for FACS, cells were dissociated, counted and distributed in equal numbers per sample in 15 ml tubes (mainly 5 x 10^6 cells per sample per tube). Samples were incubated with ITGA8 or IgG antibodies conjugated with APC diluted in the FACS buffer containing 2% FCS in PBS (10 μl antibody / 100 μl buffer / 10^6 cells) for 30 min on ice. Then cells were washed three times with FACS buffer and resuspended in FACS buffer for sorting. Before sorting DAPI was added to samples to final concentration 0.01 ng/μl do discriminate dead and live cells. Cells were sorted into tubes containing 1 ml of 50% FCS in PBS, washed with PBS, centrifuged and resuspended in lysis buffer for RNA extraction or cell culture media for further differentiation.

2.4 Immunofluorescence analysis

All immunofluorescence analyses were done as described in Zawada et al., 2022 (preprint).

2.4.1 Immunofluorescence staining of cells cultured in monolayer

At d10 cells cultured in wells of 12-well chamber slides (81201, Ibidi) were fixed with 4% PFA for 10 min at room temperature, washed three times with PBS and permeabilized with 0.1% Triton-X-100 in PBS for 15 min. After blocking with 10% FCS in 0.1% Triton-X-100 in PBS for 60 min, samples were subjected to immunostaining using primary antibodies (listed in Table 4) diluted in 0.1% Triton-X-100 PBS containing 1% FCS overnight at 4°C. The samples were then washed 5 times for 5min with 0.1% Triton-X-100 PBS and incubated with secondary antibodies (listed in Table 5) diluted 1:500 in 0.1% Triton-X-100 PBS containing 1% FCS for 1 hour at room temperature. Nuclei were detected with 1 μg/ml Hoechst 33258.

2.4.2 Immunofluorescence staining of epicardial cells

Cells at d12 were fixed with 4% PFA for 15 min at room temperature, washed 3 times with PBS, blocked for 1 hour in 0.1% Triton-X-100 PBS containing 3% BSA and subjected to immunostaining using primary antibodies (listed in Table 4) diluted in 0.1% Triton-X-100 PBS containing 0.5% BSA overnight at 4°C. The samples were then washed three times for 5 min with 0.1% Triton-X-100 PBS and incubated with secondary antibodies (listed in Table 5) diluted 1:500 in 0.1% Triton-X-100 PBS containing 0.5% BSA for 1 hour at room temperature. Specimens were washed three times with 0.1% Triton-X-100 PBS. Nuclei were detected with 5 µg/ml Hoechst 33258 (5 min). Slides were washed once again with PBS, covered with mounting medium and cover slip and stored at 4°C.

2.4.3 Aggregates embedding, cryosectioning and immunofluorescence staining

In preparation for cryosectioning and immunofluorescence staining aggregates were subjected to modified protocol of Lancaster & Knoblich (2014). Briefly, aggregates were transferred to wells of a 24-well plate containing 1 ml PBS+/+ using a cut 1000 µl pipette tip. PBS+/+ was replaced with 4% PFA, and aggregates were incubated for 30 min at room temperature. After that time, aggregates were washed three times for 5 min with PBS+/+. Then, 30% sucrose in PBS+/+ was added for overnight in 4°C. For aggregates embedding, the 30% in PBS+/+ sucrose on aggregates was replaced with warmed up to 37°C 1 ml 10% sucrose and 7.5% gelatin in PBS+/+ solution and allowed to equilibrate for 15 min at 37°C on aggregates. Meanwhile, the bottom of a cryomold was covered with 400 µl sucrose/gelatin solution and placed at 4°C to solidify. Next aggregates were transferred to molds using a cut 1000 µl tip, and placed for 3-5 min at 4°C, then the cryomold was filled with 500 µl sucrose/gelatin solution and placed at 4°C for 20 min to solidify. For aggregates freezing, such prepared molds were placed in cold 2-methylbutane (M32631, Sigma) for 1 - 2 min and stored at -80°C. Molds were cryosected using Cryostat (Microm HM 560, Thermo Scientific) into 12-16 µm slices on poly-L-lysine coated slides (J2800AMNZ, Thermo Scientific), dried for 30 min at room temperature and stored at -80°C. For immunofluorescence analysis, slides were dried for 30 min at room temperature after taking out of -80°C freezer, fixed with 4°C PFA for 10 min at room temperature, washed three times for 5 min with PBS+/+. Then permeabilized

with 0.25% Triton X-100 for 15 min at room temperature, washed again three times for 5 min with PBS, blocked with 3% BSA in 0.05% Tween-20 (P2287, Sigma) in PBS+/+ (PBST) for 1h. Then primary antibodies (listed in Table 4) diluted in 0.5% BSA in PBST were added for overnight at 4°C. Next day, slides were washed five times for 10 min with PBST. Then secondary antibodies (listed in table 5) diluted in 0.5% BSA in PBST were added and slides were incubated for 2h at room temperature. After that time, slides were washed 5 times for 10 min with PBST. Next, Hoechst 33258 (94403, Sigma) at 5 µg/mL diluted in PBS was added for 15 min at room temperature. Slides were washed once with PBS, covered with mounting medium and cover slip and stored at 4°C.

Images were acquired using a DMI6000B and TCS SP8 confocal laser-scanning microscopes Leica Microsystems, Wetzlar, Germany). Images were assigned with pseudo-colours and processed with Leica Application Suite software from Leica Microsystems.

Table 4: Primary antibodies used for immunofluorescence staining.

Target	Host species	Manufacturer	Catalog number	Concentration
CK-18	Mouse	Abcam	ab668	1:100
cTnT	Rabbit	Abcam	ab92546	1:500
E-cadherin	Mouse	Abcam	ab1416	1:100
Fibronectin	Rabbit	Abcam	ab2413	1:250
ISL1	Mouse	Developmental Studies Hybridoma bank	39.4D5	1:100
NKX2.5	Rabbit	Novus Biologicals	NBP1-31558	1:200
TCF21	Rabbit	Sigma-Aldrich	HPA013189	1:100
GFP	Chicken	Abcam	ab13970	1:500
Anti-Rabbit IgG isotype control	Rabbit	Abcam	ab27478	1 µg / 100 µl / 10^6 cells
mCherry	Goat	Sicgen	ab0040-200	1:200
TBX5	Rabbit	Abcam	ab18531-100	1:100

CD31	Sheep	R&D systems	AF806	1:200
Anti-Mouse IgG2b isotype control	Mouse	Sigma Aldrich	M5534	1 μg / 100 μl / 10^6 cells
ITGA8-Alexa-Fluor-647-conjugated	Mouse	R&D systems	fab9194r	10 μl / 100 μl / 10^6 cells
ZO-1	Mouse	Invitrogen	33-9100	1:100
Myosin Light Chain 2/MLC-2V	Rabbit	Proteitech	10906-1-AP	1:100
Myosin light chain 2/MLC-2A	Mouse	Synaptic Systems	311 011	1:200
WT1	Rabbit	Abcam	ab89901	1:100

Table 5: Secondary antibodies used for immunofluorescence staining.

Target species	Host species	Conjugate	Manufacturer	Catalog number	Concentration
Rabbit	Donkey	Alexa Fluor 647	Invitrogen	A31573	1:500
Mouse	Donkey	Alexa Fluor 647	Invitrogen	A31571	1:500
Chicken	Donkey	Alexa Fluor 488	Jackson Immuno Research	703-545-155	1:500
Sheep	Donkey	Alexa Fluor 647	Invitrogen	A21448	1:500
Goat	Donkey	Alexa Fluor 594	Invitrogen	A11058	1:500
Rabbit	Goat	Alexa Fluor 594	Invitrogen	A11012	1:500
Mouse	Donkey	Alexa Fluor 594	Invitrogen	A21203	1:500

2.5 Handsdown Transcriptional Start Site Knockout cell line generation using CRISPR/Cas9

2.5.1 Plasmid generation

Cloning of the HDN knock-out guides into Puro-Cas9 Plasmid was done according to Ran et. al, (2014). Briefly, sgRNAs oligos were phosphorylated and annealed by mixing 1 µl of 100 µM forward and reverse sgRNAs oligos with 1 µl of 10x T4 ligation buffer (B0202S, NEB), 1 µl of T4 PNK (M0201S, NEB) and 6 µl of ddH$_2$O, using the following parameters: 37°C for 30 min; 95°C for 5 min; ramp down to 25°C. Such prepared oligos were diluted 1:200 in ddH$_2$O and 2 µl of oligos were mixed with 2 µl of 10x Tango buffer (BY5, ThermoFisher Scientific), 1 µl of 10 mM of DTT (R0862, ThermoFisher Scientific), 1 µl of 10 mM of ATP (P0756S, NEB), 1 µl of FastDigest *BbsI* (FD1014, ThermoFisher Scientific), 0.5 µl of T7 ligase (L602L, Enzymatics), and pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene #62988), filled up with ddH$_2$O up to 20 µl and ligated using following set-up: 1h reaction including 1-6 cycles of incubation at 37°C for 5 min followed by incubation at 21°C for 5 min. Ligation mixture was then treated with PlasmidSafe exonuclease (E3101K, Epicentre) to digest any residual linearized DNA at 37°C for 30 min, followed by 70°C for 30 min. Such PlasmidSafe-treated plasmids were transferred into competent E. coli strains. Plasmid DNA from cultures were isolated using a QIAprep Spin Miniprep Kit (27104, Qiagen) according to the manufacturer's instructions.

Following gRNA primers were used:

HDN_HGLibA_2_guide: TAACTAGTACTTGGCGAAGA	
gRNA_HDN_HGLibA_2_Fw	CACCGAACTAGTACTTGGCGAAGA
gRNA_HDN_HGLibA_2_Rv	AAACTCTTCGCCAAGTACTAGTTC

HDN_HGLibA_3_guide: AGTCTACGATCCTAGAACTT	
gRNA_HDN_HGLibA_3_Fw	CACCGAGTTCTAGGATCGTAGACT
gRNA_HDN_HGLibA_3_Rv	AAACAGTCTACGATCCTAGAACTC

2.5.2 Plasmid lipofection

Day before lipofection, not fully confluent hiPSC line cultures were washed twice with PBS, dissociated with Accutase for 5 min at 37°C, centrifuged in E8 media supplemented with 10 µM Rock inhibitor Y- 27632, resuspended in 1 ml of E8 media, counted and plated in number of 10-20 000 cells per well per 200 µl E8 of 96-well plate covered with Geltrex. Next day Rock inhibitor Y- 27632 was removed. For lipofection, cells were transfected with 20 µl of plasmid lipofection solution containing 50 ng of HDN guide 2 plasmid in 0.4 µl and 50 ng of HDN guide 3 plasmid in 0.4 µl, 0.2 µl Lipofectamine Stem Transfection Reagent (STEM00015, Invitrogen) and 19 µl Opti-Mem (31985062, Gibco/Invitrogen). Solution was incubated in room temperature for 20 min and added to wells. After 24h, media was changed for E8 containing 0.2 µg/ml Puromycin (p9620, Sigma). After 46h media was changed. When cells regrew, they were passaged into 10 cm^2 plate dish at the density of 1000-2000 cells per plate. After around 10 days colonies were picked and screened. After around 5 days clones with confirmed knockout were transferred into one well of 12 well plate and passaged as usual. Derived clones were subsequently screened for off-targets and subjected to karyotype analysis.

2.5.3 Genomic PCR

For colony screening half of every analysed colony was transferred into lysis buffer (300 µl of 1 M Tris (T2194, Sigma) of pH 8.0, 60 µl of 0.5 M EDTA, 1.5 ml of 1 M KCl (60135, Fluka Analytical) 60 µl of 1 M MgCl$_2$ (M1028, Sigma), 28 ml H2O) supplemented with proteinase K (670 µg/ml; P2308, Sigma) and RNAse A (200 µg/ml; 158922, Qiagen), incubated for 3h in 60°C and 2 min in 90°C and subjected to the PCR reaction using Polymerase Taq DNA High Fidelity Polymerase (11304011, Invitrogen). PCR products were evaluated by

agarose gel electrophoresis separation at 230 V in 1% agarose gels containing the DNA intercalator ethidium bromide followed by imaging with a Biovision 3000 WL transilluminator (Peqlab, Germany; primers used listen in Table 6). For sequencing of DNA fragments containing modified loci, total DNA was extracted from cells with the Gentra Puregene kit (Qiagen; 158722) according to the manufacturer's instructions. PCR was performed with the Q5 High-Fidelity DNA Polymerase (NEB; M0491S).

Table 6: Primers used to identify heterozygous and homozygous clones and used for sequencing.

	Primer forward sequence	**Primer reverse sequence**	**PCR product size**
Unsuccessful cutting with guide 2	CCAGCTGAGATCCTTTACAA	TATTAGCAGCCAGGTAGCCC	~1kb
Unsuccessful cutting with guide 3	TTTCTGTTTTGCTCAATGGG	GGGTCTGTTTATTGCCGCAA	~1kb
Successful cutting with both guides	CCAGCTGAGATCCTTTACAA	TTTCTGTTTTGCTCAATGGG	300-500bp

2.6 Statistical analysis

Statistical analysis was done as described in Zawada at el., 2002 (preprint). Briefly, statistical analysis was performed with GraphPad Prism version 5 and 8 (La Jolla California, USA). Bar graphs indicate the mean ± SEM with all data points displayed separately, unless otherwise indicated. Data from two experimental groups were compared either by unpaired Student's *t*-test or by Mann-Whitney-Wilcoxon test depending on the assumed distribution. For more than two experimental groups one-way analysis of variance (ANOVA) was used first, followed by Student's *t*-test or by Mann-Whitney-Wilcoxon for groupwise comparisons in case

of a statistically significant result from the overall analysis. A p-value < 0.05 was considered statistically significant, unless otherwise indicated.

2.7 Data analysis

2.7.1 RNA-Seq analysis

Bioinformatical analysis of deep RNA-Seq data was performed by Dr. Gianluca Santamaria (TUM, Germany).

The reads were filtered for low-quality, contaminating 5'adapters, homopolymers and trimmed for 3'adapters. Quality control analysis was performed using *FastQC*, trimming of small 3'RNA adapter sequences, using *Trimmomatic v0.36* with default parameters. Quality checked reads were then aligned to the human genome (GRCh37 assembly) using *TopHat2* (default parameters). Gene annotation was obtained for all known genes in the human genome, as provided by Ensemble (GRCh37; support.illumina.com/sequencing/sequencing_software/igenome.ilmn). Using the reads mapped to the genome, the number of reads mapping to each transcript was calculated with *HTSeq-count* (default parameters). Raw read counts were then used as input to *DESeq2* v1.20.0. (default parameters) for calculation of size factor and scaling factor of normalized signal to bring the count values across all the samples to a common scale for each transcript. For the analysis, genes with the adjusted *p*-value ≤ 0.05 were deemed DEGs. Gene Set Enrichment Analysis (GSEA) was performed using an algorithm implemented in the R Software Environment for Statistical Computing. Statistical significance was assessed through data set randomization by permuting gene sets 1000 times and considering only gene sets with adjusted *p*-value ≤ 0.05. Heat-maps and gene ontology (GO) terms from differentially expressed genes (DEGs) was performed using *clusterProfile*.

2.7.2 LncRNA detection

LncRNA detection was performed by Dr. Ali Tamer (Goethe University Frankfurt, Germany).

Read mapping was done with STAR aligner using the option --outSAMtype BAM SortedByCoordinate (Dobin et al., 2013) with default settings. For known transcript models we used GRCh38.101 ensembl annotations downloaded from ensembl repository (Zerbino et al., 2018). Coding and non-coding transcripts are filtered using home-made python script. Counting reads over gene model was carried out using GenomicFeatures Bioconductor package (Lawrence et al., 2013).The aligned reads were analysed with custom R scripts in order to obtain gene expression measures. For normalization of read counts and identification of differentially expressed genes we used DESeq2 with padj < 0.01 and LFC=0.59 cutoff (Love et al., 2014).

2.7.3 scRNAseq analysis

Bioinformatical analysis of scRNA-Seq data was performed by Dr. Gianluca Santamaria and Dr. Alexander Goedel as described in Zawada et al. 2022 (preprint)..

The Cell Ranger pipeline (v6.1.1) was used to perform sample demultiplexing, barcode processing and generate the single-cell gene counting matrix. Briefly, samples were demultiplexed to produce a pair of FASTQ files for each sample. Reads containing sequence information were aligned using the reference provided with Cell Ranger (v6.1.1) based on the GRCh37 reference genome and ENSEMBL gene annotation. PCR duplicates were removed by matching the same UMI, and 10x barcode and gene were collapsed to a single UMI count in the gene-barcode UMI count matrix. All the samples were aggregated using Cell Ranger with no normalization and treated as a single dataset. The R statistical programming language (v3.5.1) was used for further analysis. The count data matrix was read into R and used to construct the Seurat object (v4.0.1).

The Seurat package was used to produce diagnostic quality control plots and select thresholds for further filtering. The filtering method was used to detect outliers and high numbers of mitochondrial transcripts. These pre-processed data were then analysed to identify variable genes, which were used to perform principal component analysis (PCA). Statistically significant PCs were selected by PC elbow plots and used for UMAP analysis. Clustering parameter resolution was set to 1 for the function FindClusters() in Seurat. For sub-clustering analysis, we used the clustree package (v0.4.3). The clustering analysis of the d30 dataset resulted in three

clusters that were characterized solely by very low read counts (5, 15, 17) and were removed from further analyses.

All DEGs were obtained using the Wilcoxon rank-sum test using as threshold p-value ≤ 0.05. We used an adjusted p-value based on Bonferroni correction using all features in the dataset. For the cell type-specific analysis, single cells of each cell type were identified using the FindConservedMarkers function as described within the Seurat pipeline.

TF regulatory network analysis was performed using SCENIC workflow (version 1.1.2.2) with default parameters. Gene regulatory networks were inferred with GENIE3 (version1.6.0) based on the samples: FACS sorted cells for eGFP at d4.5 within noRA condition and FACS sorted cells for eGFP and mCherry at d4.5 within RA condition. Enriched TF-binding motifs, predicted candidate target genes (regulons), and regulon activity was inferred by RcisTarget. The transcription regulatory network was visualized by a binary regulon activity heatmap.

For all the gene signatures analysed we used a function implemented in the *yaGST* R package (https://rdrr.io/github/miccec/yaGST/).

Comparison of generated for this projects scRNA-Seq datasets with published scRNA-sequencing datasets were carried as follows: 1) for datasets from mouse (Tyser et al., 2021) IDs were converted to Homo sapiens gene symbols according to the ortholog list from Ensemble (only one-to-one orthologues were considered; (Cunningham et al., 2022); 2) the datasets were combined using the CCA workflow implemented in the Seurat package (Hao et al., 2021) based on 10 dimensions and 2000 anchor features; 3) after integration the Euclidean distances between the different clusters were calculated based on the averaged, corrected, normalized counts followed by hierarchical clustering using Ward's hierarchical agglomerative clustering method (Murtagh & Legendre, 2014). The results were plotted as a dendrogram.

2.7.4 Identification of retinoic acid response elements in the promoter region of the gene of interest

Analysis and consensus promoter algorithm developed by Dr. Gianluca Santamaria

Gene flanking promoter region (5kb) for all genes of interest were downloaded from Ensembl87 using the Biomart web interface. We developed a consensus promoter algorithm,

implemented in the R Software Environment, able to scan for RA response elements (RAREs) spaced by n nucleotides, expanding the spacing to 9 nucleotides, and considering both inversion and reverse complement associated with a direct repeat. Subsequently, genes of interest were classified based on the number of spacing nucleotides in direct repeats (DR), everted repeats (ER), and inverted repeats (IR).

3 Results

3.1 Characterization of *in vitro* generated human cardiac progenitors and their derivatives

One of the major aims of cardiac developmental biology is to better understand the mechanisms that govern the differentiation paths of embryonic cells towards the many adult cells building a functional heart. Unfortunately, early events of human cardiogenesis are difficult to study due to limited access of tissue *in vivo*. Here, we took advantage of direct differentiation protocols that recapitulate key stages of cardiac development to improve our understanding of the emergence and characteristics of cardiovascular progenitors (Mendjan et al., 2014; Hofbauer et al., 2021).

3.1.1 Retinoic acid signalling affects cardiovascular progenitor specification

RA has many different functions during cardiac development: it has notably been shown to accelerate and boost cardiomyocyte generation and maturation (Wobus et al., 1997, Miao et al., 2020) and facilitate atrial cardiomyocytes differentiation (Zhang et al., 2011; Cyganek et al., 2018). Recently, RA has also been used to obtain hPSCs-derived cardiac organoids of FHF origin (Hofbauer et al., 2021). To get better insight into the effect of RA on cardiomyocyte formation, the cardiac differentiation protocol previously described by Mendjan et al., 2014 and Hofbauer et al., 2021 was adopted with slight modifications, resulting in the introduction of two versions: with supplementation of 0.5 μM of RA daily for 4 days between days 1.5 and 4.5 of differentiation (RA condition) or without RA addition (noRA) (Fig. 6a). More specifically, to obtain mid-anterior primitive streak-like cells serving as a source of prospective cardiac mesoderm, hESCs were exposed to Activin A, BMP4, and bFGF with inhibitors of phosphatidylinositol 3-kinase (LY294002), and GSK3-b (CHIR99021). This led to the induction of well-known primitive streak/early mesoderm markers (*TBXT*, *MESP1*) between days 0 and 2 of differentiation (Fig. 6b). *TBXT* was downregulated on day 3, followed

by *MESP1*. Further inhibition of WNT signalling together with the addition of BMP4, bFGF, and insulin with or without RA led to the upregulation of early cardiac progenitor markers (*NKX2-5*, *ISL1)* within 24 hours. Of note, *MESP1* expression persisted slightly longer in the noRA condition.
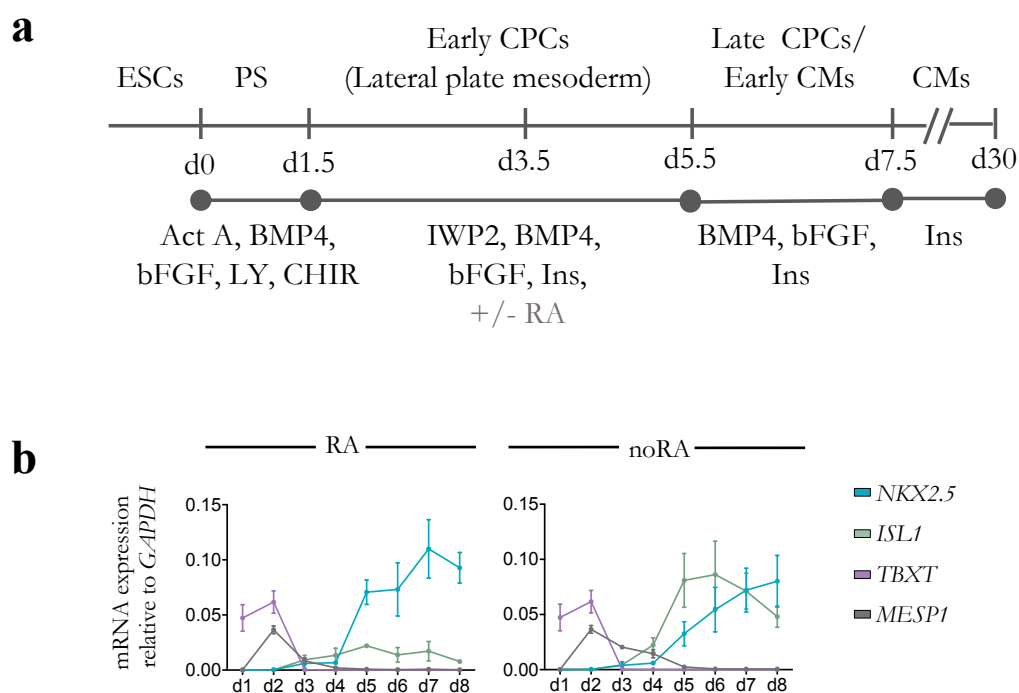
Figure 6: Recapitulation of the key stages of early cardiogenesis during *in vitro* cardiac induction.

(a) Schematic representation of the cardiovascular differentiation protocol used to induce human embryonic stem cells (ESCs) into cardiomyocytes (CMs) through defined steps of primitive streak/early mesoderm (PS), cardiac mesoderm and cardiac progenitor cells (CPCs), with the addition of retinoic acid (RA) or without (noRA). Act A: Activin A; CHIR: CHIR-99021; LY: LY-29004; Ins: Insulin. **(b)** Time-course of mRNA expression of *TBXT, MESP1* (primitive streak/early mesoderm), *NKX2.5*, and *ISL1* (cardiac progenitors) between day 1 and day 8 of noRA and RA differentiation. mRNA expression relative to *GAPDH*. Data are mean ± SEM; n = 3 differentiations per time point.

Studies in mouse embryos have indicated that RA provides a permissive environment for FGF signalling repression and TBX5 activation (Sirbu et al., 2008). TBX5 upregulation also leads to a direct and indirect inhibition of several aSHF-associated markers (Sirbu et al., 2008; De Bono et al., 2018; Rankin et al., 2021). Consistent with this, a time course transcriptional

analysis from day 1 till day 8 of differentiation revealed that application of RA during cardiac mesoderm patterning induced expression of *TBX5* and *WNT2* – markers indicating the posteriorization of cardiac precursors as well as FHF identity (Miyamoto et al., 2021, preprint; de Soysa et al., 2019) – as early as days 3 and 4, respectively, at the expense of aSHF commitment (*ISL1*, *FGF10*, *BMP4;* de Soysa et al., 2019; High et al., 2009) (Fig. 7a-b).

By contrast, we hypothesized that the lack of RA could lead to the prolongation of *ISL1* and *FGF10* expression and the emergence of aSHF progenitors. Indeed, in the absence of RA, a high expression of *ISL1*, *FGF10* and *BMP4* was maintained (Fig. 7a-b). In addition, the expression of a novel human OFT marker, *LGR5* (Sahara et al., 2019), was maintained at the cardiac progenitor stage, between days 5 and 7, while markers of the FHF and posterior cardiac progenitor lineage (*TBX5*, *WNT2*) were absent (Fig. 7a-b). Thus, the lack of RA created a permissive environment for the emergence of aSHF-like progenitors.
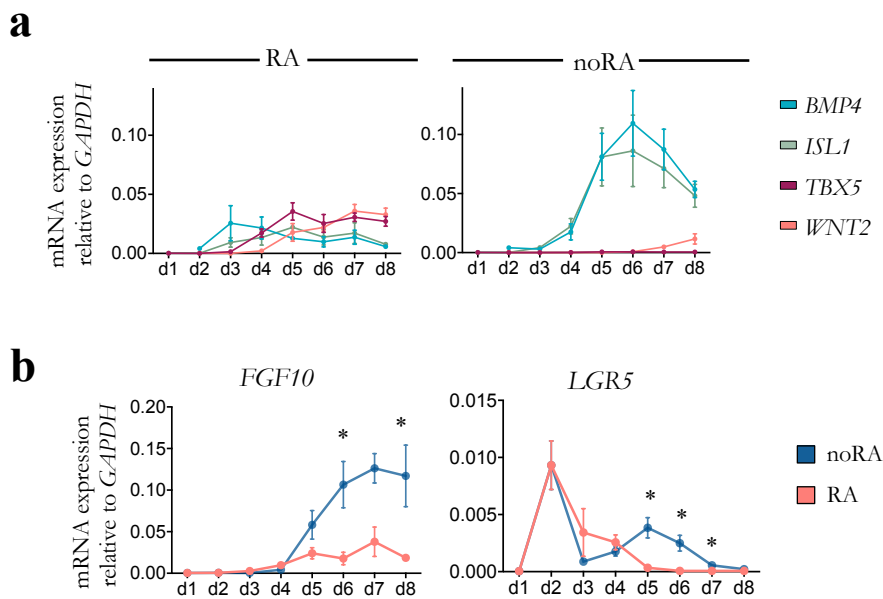
Figure 7: Manipulation of RA signalling allows the selective enrichment of FHF/pSHF or aSHF-like cardiac progenitors.

(a) Time-course of mRNA expression of *BMP4, ISL1* (aSHF), *TBX5, WNT2* (FHF and posterior cardiac progenitors) between day 1 and day 8 of noRA and RA differentiation. mRNA expression relative to *GAPDH*. Data are mean ± SEM; n = 3 differentiations per time point. **(b)** Time-course of mRNA expression of *FGF10* and *LGR5* between d1 and day 8 of noRA and RA differentiation. mRNA expression relative to *GAPDH*. Data are mean ± SEM; n = 3 differentiations per time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed *t*-test between RA and noRA at a given day).

In mice, FHF-like cells enter cardiomyocyte commitment first, followed by SHF cells (Cai et al., 2003; van den Berg et al., 2009). Consistent with this, the cardiomyocyte markers *TNNT2* and *MYL3* were already present on day 5 of differentiation using RA, while in noRA condition these markers appeared one and two days later, respectively (Fig. 8a). This correlated with the appearance of beating foci: upon RA treatment a partially beating monolayer was already visible in the culture dishes at day 6.5, 1.5-2 days before the appearance in untreated cultures (Fig. 8b).
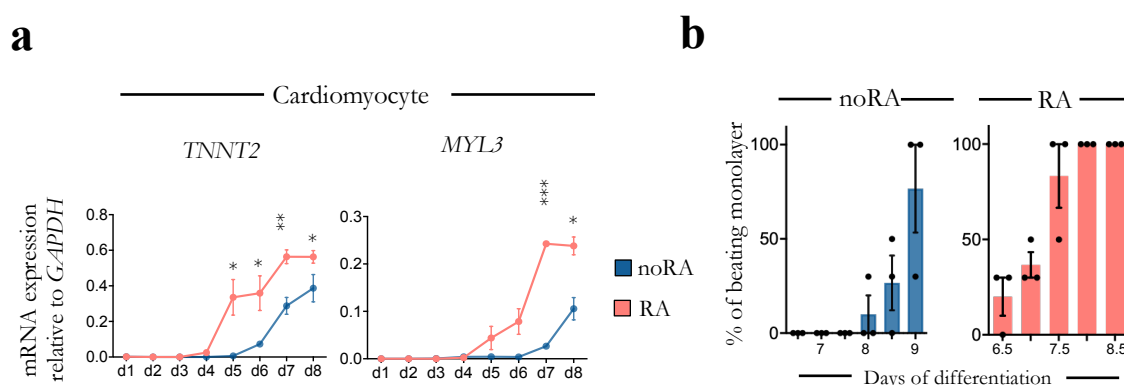
Figure 8: Effect of RA on cardiomyocyte commitment.

(a) Analysis of mRNA expression of *TNNT2, MYL3* between day 1 and day 8 of noRA and RA differentiation. mRNA expression relative to *GAPDH*. Data are mean ± SEM; n = 3 differentiations/time point; *$p<0.05$, **$p<0.005$, ***$p<0.001$ (unpaired two-tailed *t*-test between RA and noRA at a given day). **(b)** Analysis of onset of spontaneous beating during noRA and RA differentiation. Percentage of beating monolayer: 20-50% partial, beating foci in parts of monolayer; 100% full beating monolayer. Data are mean ± SEM; n = 3 differentiations.

Taken together, these data show that RA signalling affects cardiac progenitor specification: the lack of RA allows for upregulation of aSHF markers, while the presence of RA instructs induction of posterior and FHF markers. Moreover, RA-treated cells enter cardiomyocyte commitment faster than in the absence of RA, which points to a FHF-like identity of these progenitor cells.

3.1.2 Retinoic acid acts in a dose- and time-dependent manner on the enrichment of aSHF and FHF progenitors and their derivatives

Previous studies have shown that a higher concentration of RA (1 μM) applied between days 3 to 6 or the same concentrations of RA as used in this study (0.5 μM) but maintained for longer (12 days between days 3 and 14; Cyganek et al., 2018; Lee et al. 2017) can lead to the generation of atrial-like cardiomyocytes, which arise from the pSHF in mice (Galli et al., 2008). By contrast, low to middle concentrations of RA have been used for the generation of FHF lineage organoids (Hofbauer et al., 2021). To test the effect of timing and concentration of RA on the enrichment of different progenitor pools in our system, various treatments were applied: 1) noRA, as previously described; 2) addition of 0.5 μM RA for 2 days between days 1.5 and 3.5 (referred to as S-RA); 3) 0.5 μM RA for 4 days, as previously described; and 4) application of 1 μM RA for 4 days between days 1.5 and 5.5 (referred to as H-RA) (Fig. 9).
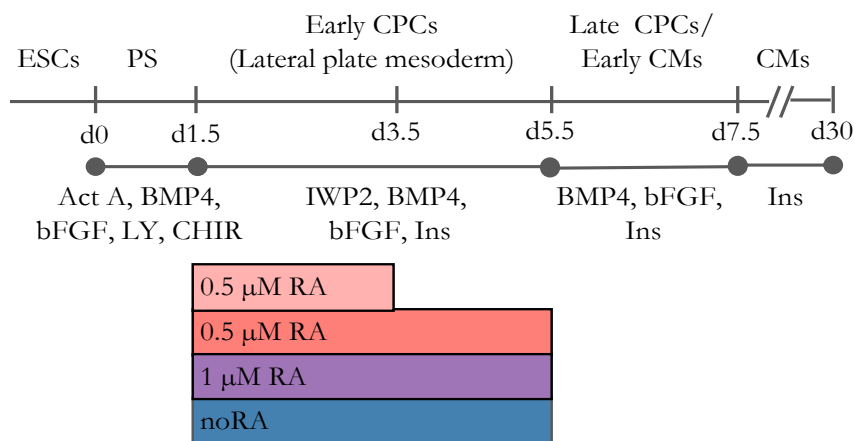
Figure 9: Graphic representation of different treatments applied to modulate the RA signaling pathway.

Schematic representation of the cardiovascular differentiation protocol used to induce human embryonic stem cells (ESCs) into cardiomyocytes (CMs) through defined steps of primitive streak/early mesoderm (PS), cardiac mesoderm and cardiac progenitor cells (CPCs), with different conditions of addition of retinoic acid (RA) or lack of RA (noRA). Act A: Activin A; CHIR: CHIR-99021; LY: LY-29004; Ins: Insulin. Adapted from Zawada et al., 2022 (preprint).

We assessed the expression of FHF markers in these conditions through RT-qPCR, including *THBS4,* which was found to be highly enriched in FHF cells in scRNA-Seq datasets of mouse embryo (Tyser et al., 2021; online search tool). We observed a dose-dependent induction of FHF and posteriorization markers: the expression of *TBX5, THBS4, WNT2* and *HOXB1* was not observed in the absence of RA, while a slight increase in transcripts of *TBX5* and *WNT2,* but not of *THBS4,* was visible in S-RA (Fig. 10). Upon exposure for 4 days, regardless of RA concentration, the expression *TBX5* increased around two-fold compared to S-RA and the expression of *THBS4* appeared. aSHF markers exhibited opposite trends. *BMP4, FGF10, WNT5A* were elevated in the absence of RA. Cells in S-RA still exhibited a slight activation of *BMP4,* which was almost absent in samples treated with RA for full 4 days. Only in the noRA condition could we detect the expression of the well-established aSHF marker *TBX1* (Fig. 10).
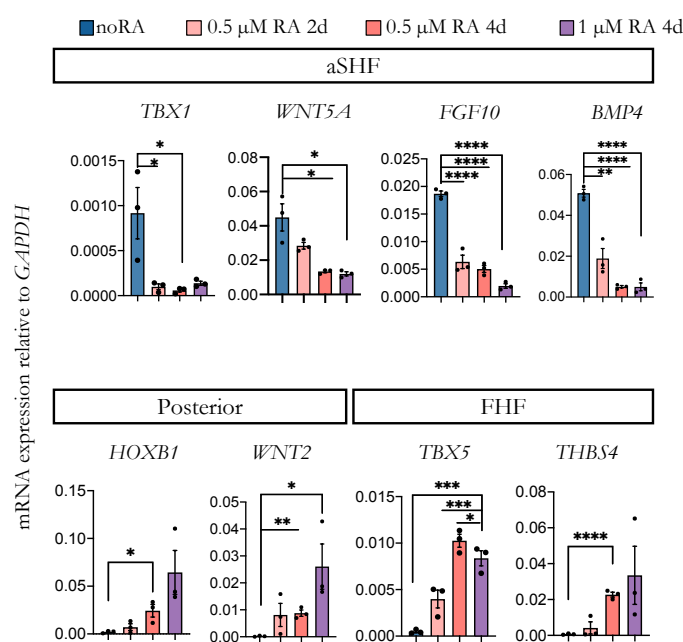
Figure 10: Modulation of RA signaling results in a RA dose- and time-dependent expression of aSHF and FHF markers.

Analysis of mRNA expression of aSHF and FHF markers at day 5 of each variant of differentiation. Colours correspond to different variants of differentiation as indicated in Fig. 9. mRNA expression relative to *GAPDH.* Data are mean ± SEM; n = 3 differentiations/time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed *t*-test). Adapted from Zawada et al., 2022 (preprint).

Results

As the initial analysis suggested that RA modulation guides the specification of FHF- and aSHF-like cardiac progenitors, it was hypothesized that differences in the enrichment of different types of cardiomyocytes may be observed at later stages of differentiation, according to the lineage potential of the respective progenitors (ventricular and OFT cardiomyocytes are derived from aSHF; atrial cardiomyocytes arise mainly from pSHF, and mainly ventricular cardiomyocytes from FHF; Meilhac et al., 2004). On day 30, cells from all treatments exhibited comparable levels of *TNNT2* transcripts, indicating equivalent efficiencies in generating cardiomyocytes (Fig. 11). A RA dose-dependent decrease in the expression of the OFT/conoventricular cardiomyocytes markers *LTBP3* and *RSPO3* was noted (Fig. 11) (Sahara et al., 2019). Atrial markers (*KCNJ3, KCNA5;* Asp et al., 2019; Cui et al., 2019) were almost absent, though a slight increase in their expression was noticeable when applying RA for four days (Fig. 11). Ventricular cardiomyocyte markers were shared by both treatment groups (noRA and various RA treatments), nevertheless, expression of *FHL2*, which was reported to be enriched in adult left ventricular cardiomyocytes, was increased upon longer RA treatment (Fig. 11) (Litvinukova et al., 2020). Taken together, different subtypes of cardiomyocytes indeed emerged from pools of progenitors characterized by different molecular signatures early on.
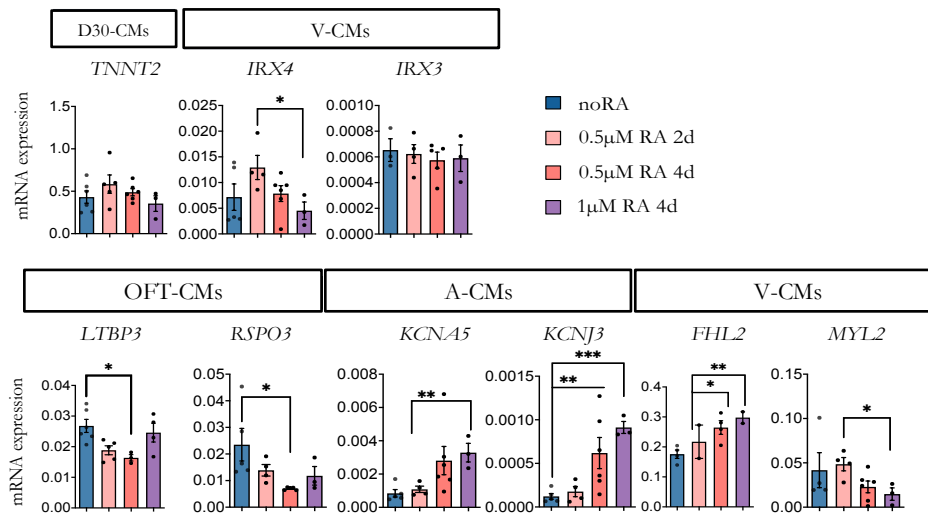
Figure 11: Modulation of RA signaling results in a RA dose- and time dependent aupregulation of ventricular, atrial and OFT cardiomyocyte markers.

Analysis of mRNA expression of ventricular, OFT and atrial cardiomyocytes markers at day 30 of each variant of differentiation. mRNA expression relative to *TNNT2* and *GAPDH* for all genes with exception of *TNNT2*. For *TNNT2* mRNA expression relative to *GAPDH*. Data are mean ± SEM; n

= 2 - 6 differentiations/time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed *t*-test). Adapted from Zawada et al., 2022 (preprint).

To confirm these observations on the protein level, immunofluorescence analysis was performed for two isoforms of myosin light chain 2 proteins – one specifically marking ventricular cardiomyocytes (MLC2v), and the other one specific for atrial cardiomyocytes (MLC2a). This showed that all differentiation conditions primarily resulted in cells positive for MLC2v but not MLC2a (Fig. 12).
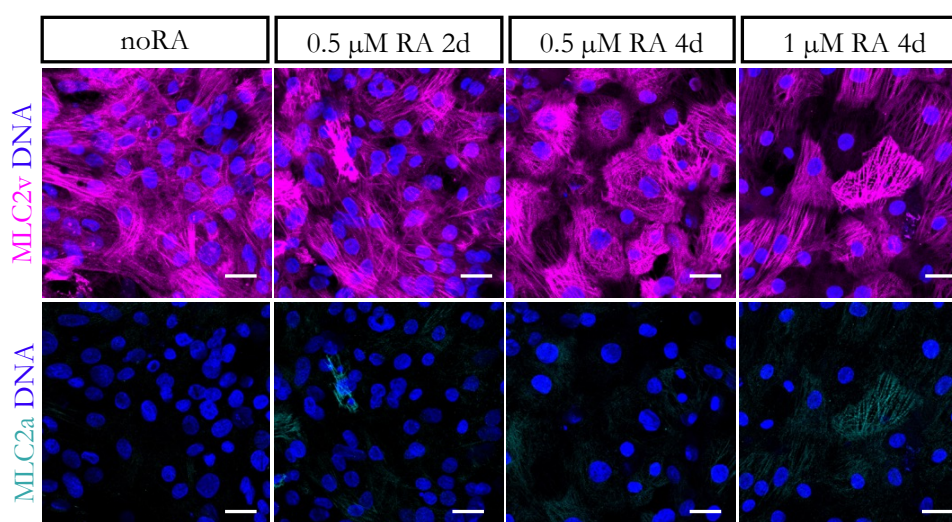
Figure 12: Short exposure to RA from day 1.5 mainly generates ventricular cardiomyocytes.

Representative images of cells stained for MLC2v (magenta) and MLC2a (turquoise) at day 45 in the indicated differentiations conditions. Size bar = 50 μm. Adapted from Zawada et al., 2022 (preprint).

Taken together, early addition of RA in the middle/low concentration range can induce progenitors with FHF-like characteristics that ultimately mainly give rise to ventricular cardiomyocytes and a small portion of cardiomyocytes expressing atrial cardiomyocytes related transcripts. On the other hand, in the absence of RA, progenitors with aSHF-like gene profile arise and later become ventricular cardiomyocytes as well as cells characterized by expression of OFT/conoventricular cardiomyocyte associated genes.

The initially used dose and timing of RA (0.5 μm, 4 days) was sufficient to suppress aSHF commitment and induce genes associated with FHF fate, and we therefore decided to use this variation of the protocol in further experiments.

3.1.3 Dynamics of emergence and characteristics of early cardiovascular progenitors

In mice, *Nkx2.5* marks all cardiac progenitors, whereas *Tbx5* is one of the key markers of early FHF commitment (Bruneau et al., 1999; Lescroart et al., 2018). As the timing of *TBX5* expression seemed to be one of the main differences between progenitors obtained through the two protocols (Fig. 13a), a hESC double reporter expressing eGFP and mCherry under the control of the endogenous NKX2.5 and TBX5 loci, respectively, was used to monitor the kinetics of these markers during cardiac differentiation (hereafter referred to as the H3 TN cell line; Fig. 13b).
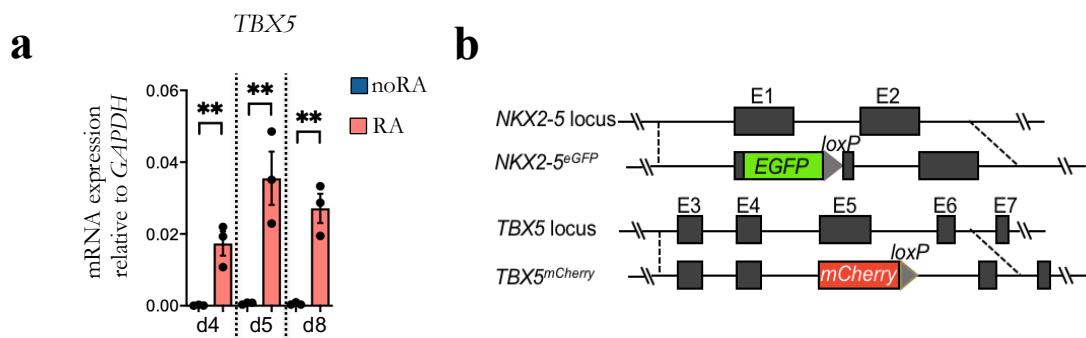
Figure 13: Generation of the TBX5^mCherry and NKX2-5^eGFP hESC double reporter line.

(a) Analysis of mRNA expression of *TBX5* on days 4, 5 and 8 of noRA and RA differentiation. mRNA expression relative to *GAPDH*. Data are mean ± SEM; n = 3 differentiations per time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed *t*-test between RA and noRA). **(b)** Schematic representation of the modifications in TBX5^mCherry and NKX2-5^eGFP hESC double reporter (H3 TN). Adapted from Zawada et al., 2022 (preprint).

To confirm proper reporting of the H3 TN cell line, the expression of mCherry and eGFP was correlated with the expression of TBX5 and NKX2.5, respectively, using RT-qPCR and immunofluorescence staining (Fig 14a, 14c). A normal karyotype was also validated (Fig. 14c).
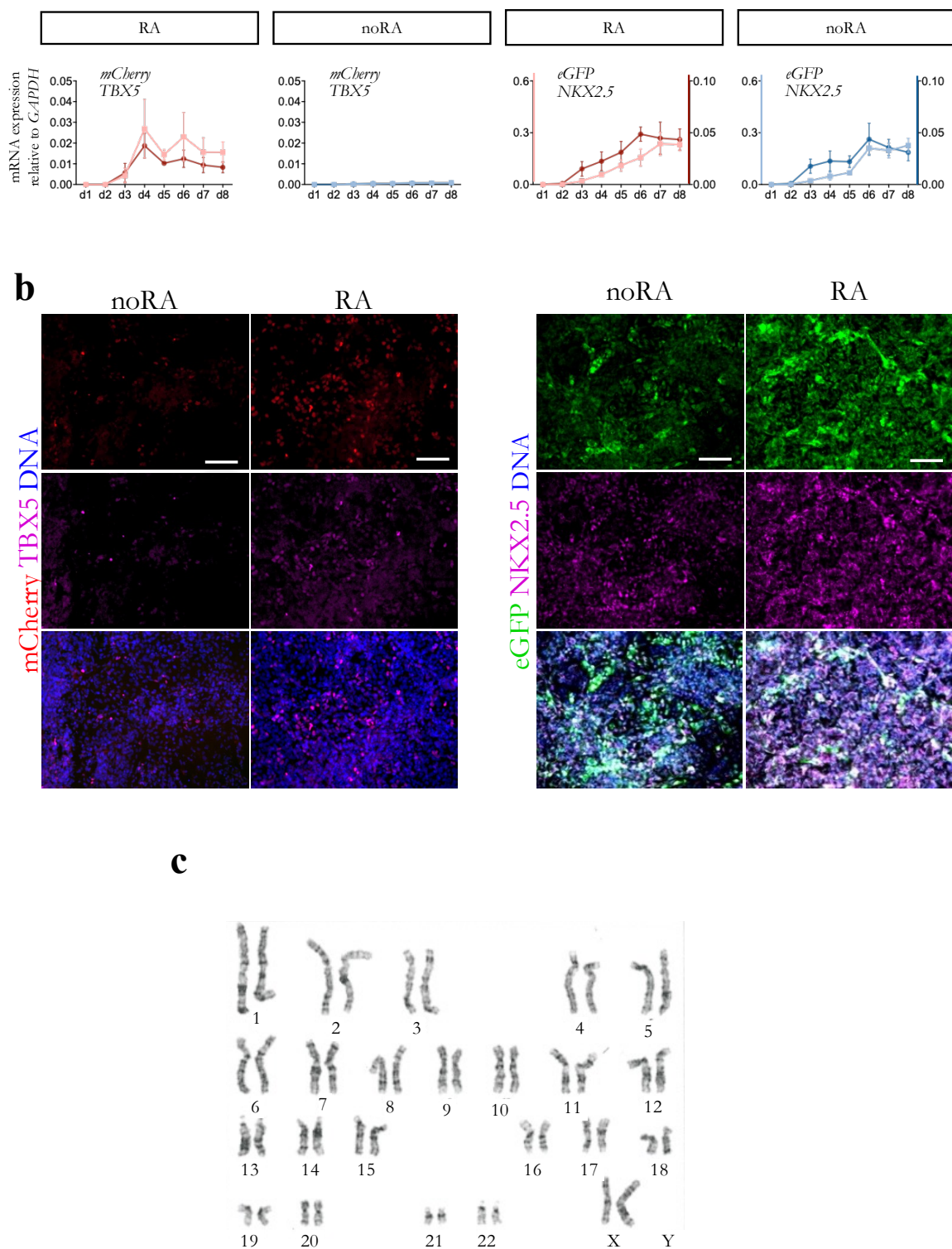
Figure 14: Validation of TBX5^mCherry and NKX2-5^eGFP hESC double reporter line.

(a) Analysis of mRNA expression of *mCherry, TBX5* and *eGFP*, *NKX2.5* between day 1 and day 8 of noRA and RA differentiation of the H3 TN line. mRNA expression relative to *GAPDH*. Data are mean

± SEM; n = 3 differentiations per time point. **(b)** Representative immunofluorescence images of day 10 cells of noRA and RA differentiation stained using antibodies against mCherry (red), TBX5 (magenta); eGFP (green), NKX2.5 (magenta). Nuclei were counterstained with Hoechst (blue). Scale bar: 100 μM. **(c)** Karyotype analysis of the H3 TN line. Adapted from Zawada et al., 2022 (preprint).

Transcriptional analysis of key genes marking all stages of cardiovascular differentiation using the RA/noRA protocols confirmed that transcript levels and gene expression patterns were comparable between the reporter line (H3 TN), its parental line (H3 WT) and an unrelated hiPSC line, regardless of haploinsufficiency of the reporter cell line (Fig.15-16).
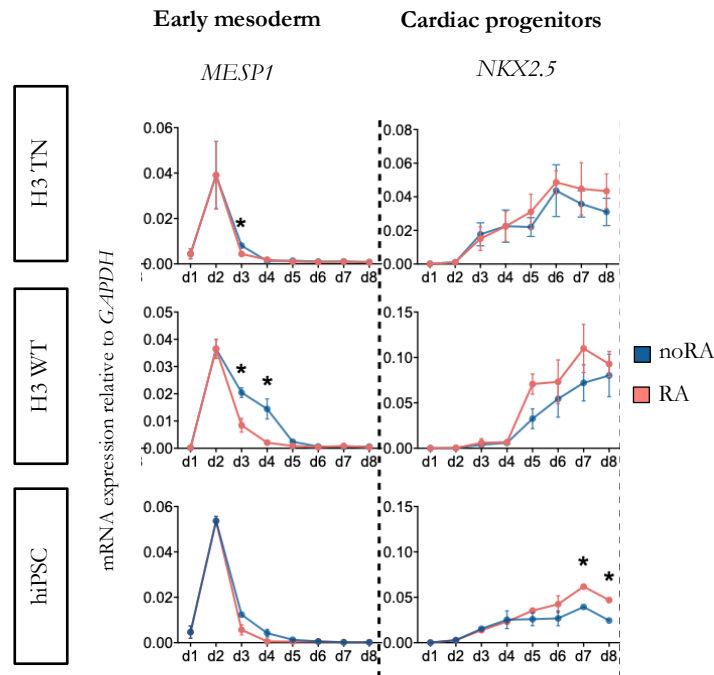
Figure 15: Various human pluripotent cell lines show similar expression patterns of key markers of early cardiac commitment during RA and noRA differentiation.

Analysis of mRNA expression of *MESP1* and *NKX2.5* between day 1 and day 8 of noRA and RA differentiation. mRNA expression relative to *GAPDH*. Data are mean ± SEM; n = 3 differentiations/time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed *t*-test between RA and noRA at a given day). Adapted from Zawada et al., 2022 (preprint).
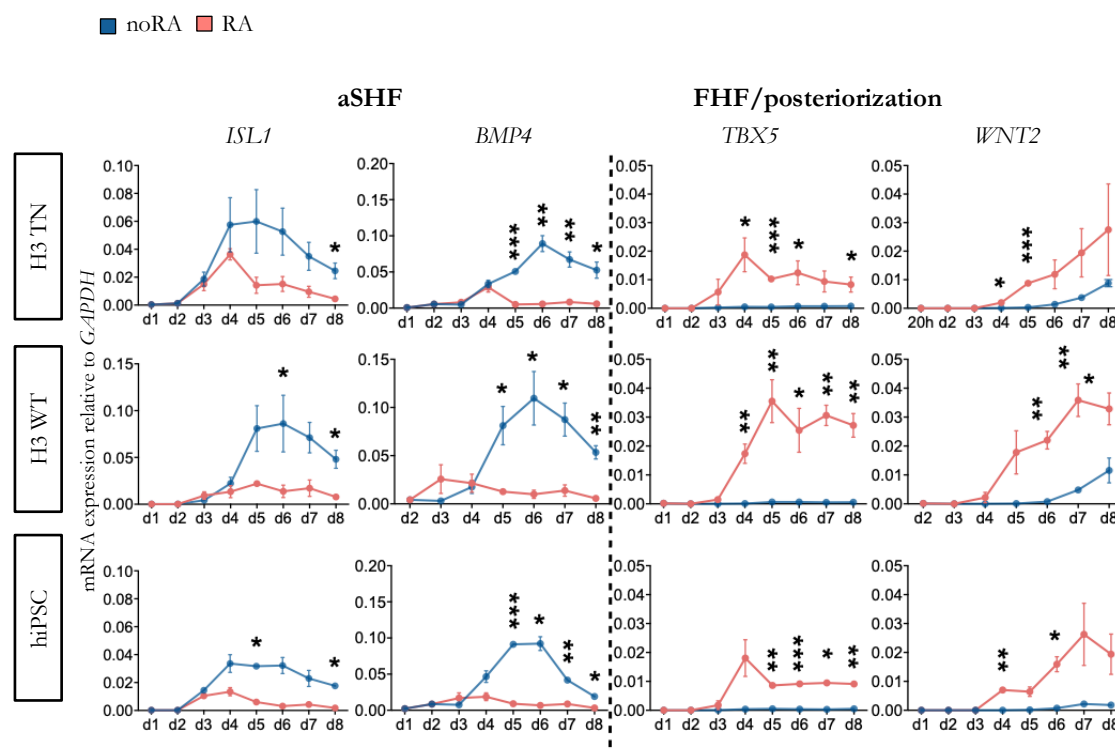
Figure 16: Various human pluripotent cell lines show similar expression patterns of key aSHF, FHF and posterior markers during RA and noRA differentiation.

Analysis of mRNA expression of *ISL1, BMP4* (aSHF) and *TBX5, WNT2* (FHF/posteriorization) between day 1 and day 8 of noRA and RA differentiation. mRNA expression relative to *GAPDH*. Data are mean ± SEM; n = 3 differentiations per time point; $*p<0.05$, $**p<0.005$, $***p<0.001$ (unpaired two-tailed *t*-test between RA and noRA at a given day). aSHF: anterior second heart fiels; FHF: first heart field. Adapted from Zawada et al., 2022 (preprint).

The dynamics of the emergence of progenitors expressing TBX5 and NKX2.5 were investigated using flow cytometry. This confirmed the appearance of NKX2.5 (eGFP) expressing cells on day 3, and the lack of TBX5 (mCherry) expressing progenitors at early differentiating timepoints in the noRA condition. Only after day 8 of noRA condition, some eGFP+ cells activated mCherry expression (Fig. 17a-b). The addition of RA led to the appearance of eGFP+ and eGFP+/mCherry+ cell populations on day 3 and day 5, respectively. On day 4, a subset of cells that were exclusively positive for mCherry was also detected (Fig. 17a-b).
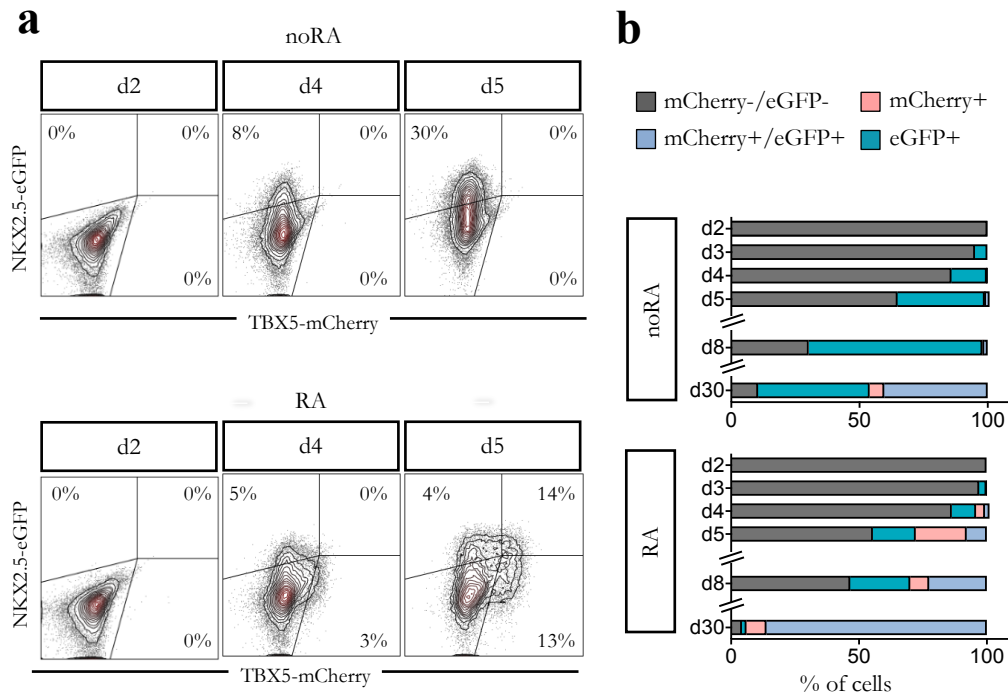
Figure 17: Tracking the emergence of distinct pools of progenitors with the TBX5^mCherry and NKX2.5^eGFP hESC double reporter line.

(a) Representative plots of live flow cytometry time-course analysis of cells expressing mCherry (TBX5) and (eGFP) NKX2.5 of noRA and RA differentiation. **(b)** Summary of live flow cytometry time-course analysis of cells expressing mCherry (TBX5) and eGFP (NKX2.5) of noRA and RA differentiation. Data are mean; n > = 3 differentiations. Adapted from Zawada et al., 2022 (preprint).

In mice, SHF progenitors are characterized by a prolonged expression of *Isl1* and slower upregulation of myocytic markers compared to FHF progenitors (Cai et al., 2003; van den Berg et al., 2009). Thus, trying to distinguish between aSHF and FHF identity, an analysis of cTnT and ISL1 was performed by flow cytometry. Starting between days 4 and 5, we observed a rapid downregulation of ISL1 and upregulation of cTnT in all analysed populations obtained in the RA condition, compared to NKX2.5+ cells generated in absence of RA, suggesting again that cardiac progenitors emerging upon RA treatment quickly enter myocytic differentiation, a hallmark of the FHF (Fig. 18). Interestingly, while eGFP+ (NKX2.5+) and mCherry+/eGFP+/ (TBX5+/NKX2.5+) populations in RA treatment demonstrated an increase in the number of cTnT expressing cells, which was maintained over the analysed days,

the mCherry+ (TBX5+) population slightly downregulated cTnT from day 6 (Fig. 18). As findings of Zhang et al. (2019) suggested the existence of epicardial-like cells expressing *TBX5*, *WT1* and *TBX18* at day 10 of WNT-modulation-based cardiac differentiation, these results could indicate that early mCherry+ progenitors in RA treatment could be a dynamic population that contributes to the myocardial fate at early stages but becomes committed to other lineages with time.
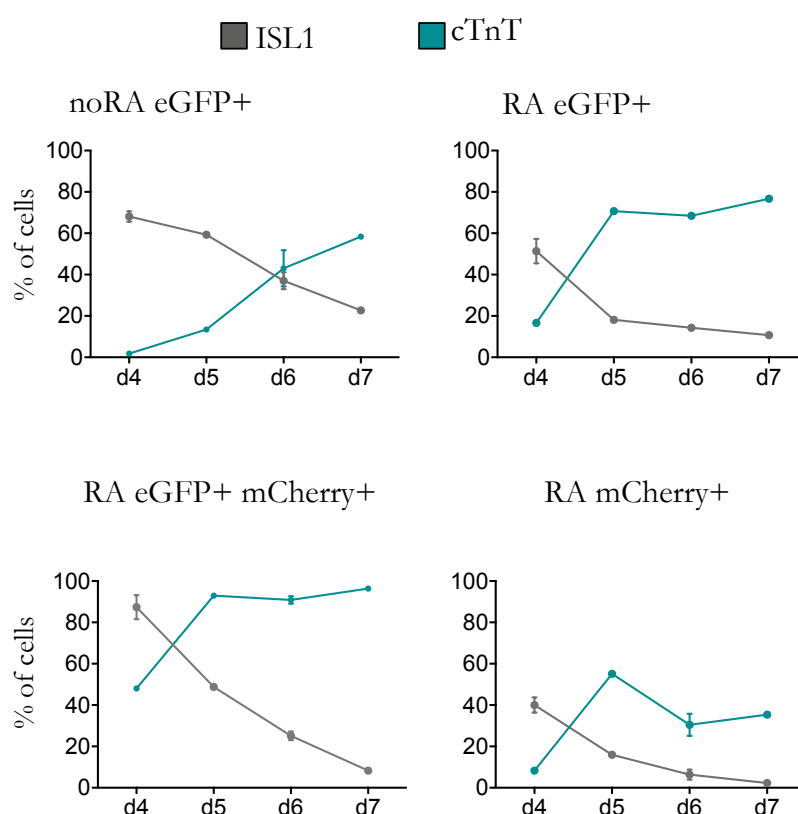
Figure 18: Human heart field-like progenitor pools marked by TBX5 (mCherry) and NKX2.5 (eGFP) are characterized by distinct dynamics of ISL1 and cTnT expression.

Summary of flow cytometry time-course analysis of cells expressing ISL1 and cTnT within NKX2.5+ (eGFP+), TBX5+/NKX2.5+ (mCherry+/eGFP+) and TBX5+ (mCherry+) populations of RA differentiation and within NKX2.5+, population of noRA differentiation. Data are mean ± SEM; n = 2 differentiations. Adapted from Zawada et al., 2022 (preprint).

Overall, using the double reporter cell line allowed us to identify, on the one hand, a population of cells arising in RA treatment characterized by fast upregulation of TBX5, rapid downregulation of ISL1 and commitment to a myocytic fate that could correspond to human FHF-like progenitors, and, on the other hand, a population of cells arising in the absence of RA characterized by lack of TBX5 expression, prolonged maintenance of ISL1 expression and slower commitment to the myocytic fate, suggesting human aSHF-like progenitors.

3.2 Single-cell transcriptomics of early human cardiac progenitor populations

Single-cell RNA sequencing (scRNA-Seq) has become a widely used technique to evaluate transcriptional and cellular heterogeneity, identify rare populations of cells that might not be detectable when analyzing pools of cells, and study lineage relationships between different cellular states. Immunofluorescence analysis at day 45 of differentiation showed that the studied differentiation conditions mainly give rise to ventricular cardiomyocytes (see 3.1.2). However, transcriptomic analyses revealed low expression levels of markers specific to other cardiomyocyte subtypes (see 3.1.2), which suggests that small populations of progenitors of varying potential could also be present. In addition, flow cytometry analysis of ISL1 and cTnT expression within NKX2.5+ (eGFP+), TBX5+/NKX2.5+ (mCherry+/eGFP+) and TBX5+ (mCherry+) populations of RA differentiation and within NKX2.5+ population of noRA differentiation suggested that distinct populations of progenitors characterized by different maturation states or varying fates might co-exist during the studied differentiations (see 3.1.3).

We therefore performed scRNA-seq to more precisely define the identity of generated progenitors, determine the degree of heterogeneity of putative human FHF- and aSHF-like progenitors emerging in the presence and absence of RA and confirm their lineage potential. Discrete populations were compared: NKX2-5+ cells present in the absence of RA (noRA-N+) and NKX2-5+, TBX5+, TBX5+/NKX2-5+ cells emerging with RA treatment (RA-N+, RA-T+, RA-T+/N+, RA-T-/N-). Cells were collected at three time points: on day 1.5, just before the onset of RA exposure; on day 4.5 when four populations noRA-N+, RA-N+, RA-T+, and RA-T+/N+ were sorted; and on day 30, from aggregates formed from the four sorted populations which were allowed to differentiate for further 25 days (Fig. 19; Table 7). Day 4.5

was chosen as a day when all three populations emerging within RA treatment (RA-N+, RA-T+, RA-T+/N+) are present together for the first time.
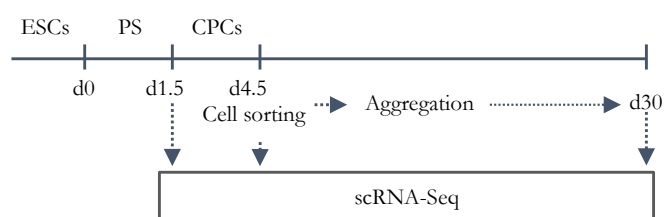

Figure 19: Experimental design applied for scRNA-Seq analysis of cardiac progenitor populations and their derivatives.

Graphic representation of the timing used for collecting samples for scRNA-Seq. ESCs: embryonic stem cells; CPCs: cardiac progenitors.

Unbiased clustering of cells was performed using the standard Seurat workflow and dimensionality reduction was performed through UMAP (Uniform Manifold Approximation and Projection; Stuart et al., 2019). Clusters were annotated based on the expression of well-known marker genes.

Table 7: Recovered samples parameters after scRNA-Seq.

Sample	Estimated number of cells	Median reads per cell	Median genes per cell
Day 1.5	13253	10118	1862
Day 4.5 noRA-N+	8589	13715	2088
Day 4.5 RA-N+	9677	11366	1954
Day 4.5 RA-T+	9644	11548	1975
Day 4.5 RA-T+/N+	9821	13489	1701
Day 30 Aggregates of noRA-N+	9729	26440	2561
Day 30 Aggregates of RA-N+	7842	24911	2701

Day 30 Aggregates of RA-T+	10510	17284	2135
Day 30 Aggregates of RA-T+/N+	4826	49 114	3144

3.2.1 RA signalling is induced in early cardiac mesodermal cells emerging from mid-anterior primitive streak

First, cells of day 1.5 were merged with the sorted cell populations of day 4.5 (noRA-N+, RA-N+, RA-T+/N+, RA-T+). Cells segregated into nineteen clusters that formed five major groups: early mesodermal cells (day 1.5), endothelial/endocardial progenitor cells (Endo-PCs), endodermal cells, and cardiovascular progenitor cells derived in the absence (CPC-noRA) or presence of RA (CPC-RA; Fig. 20a-b).
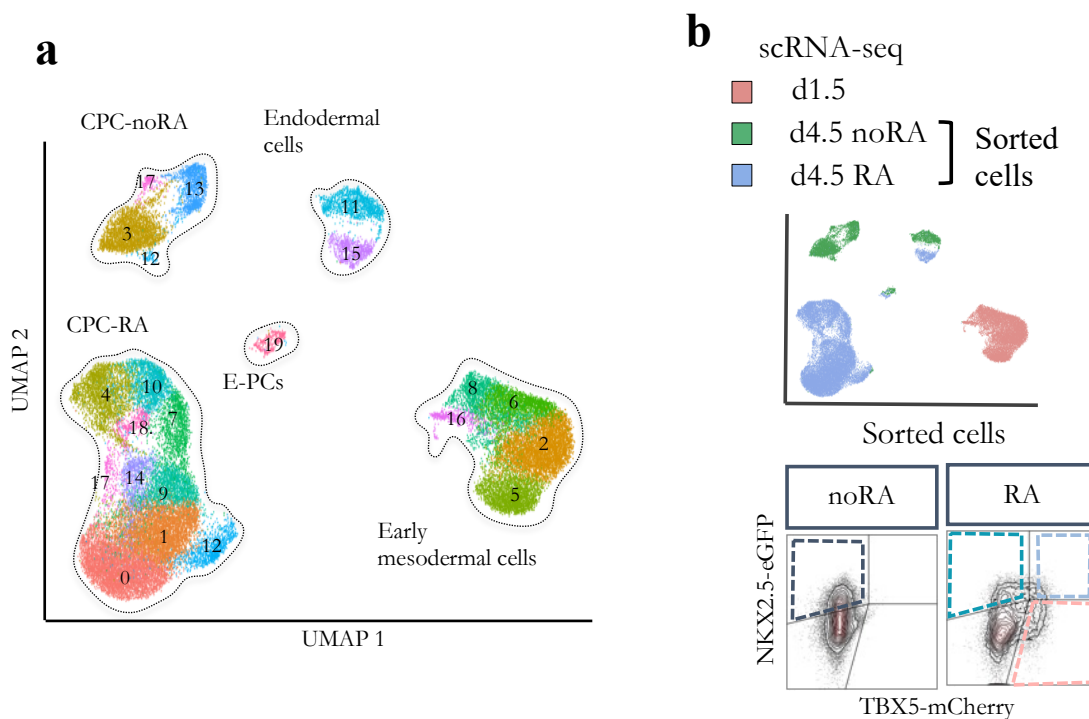
Figure 20: ScRNA-seq analysis reveals heterogeneity within human heart-field like progenitors.

(a) UMAP clustering of single cells captured at day 1.5, and FACS sorted cells expressing mCherry (TBX5), eGFP (NKX2.5) or both at day 4.5 of 0.5 µM 4d RA; and eGFP (NKX2.5) at day 4.5 of noRA differentiation. **(b)** Upper panel: UMAP plot showing contribution from all 3 sorted samples of 0.5 µM 4d RA and one sample of noRA differentiations and day 1.5. Bottom panel: Representative flow

cytometry analysis plots showing gating strategy for sorting cells at day 4.5. Adapted from Zawada et al., 2022 (preprint).

Cells from day 1.5 seemed to still be fairly homogenous - they formed one big cluster that could be divided into five sub-clusters (Fig. 20a). At this stage a vast majority of cells were *MESP1+/EOMES+/GSC+/MIXL+/T+* pre-cardiac mesodermal cells corresponding to mid-anterior primitive streak-like cells, as described previously by Mendjan et al. (2014; Fig. 21a-b).
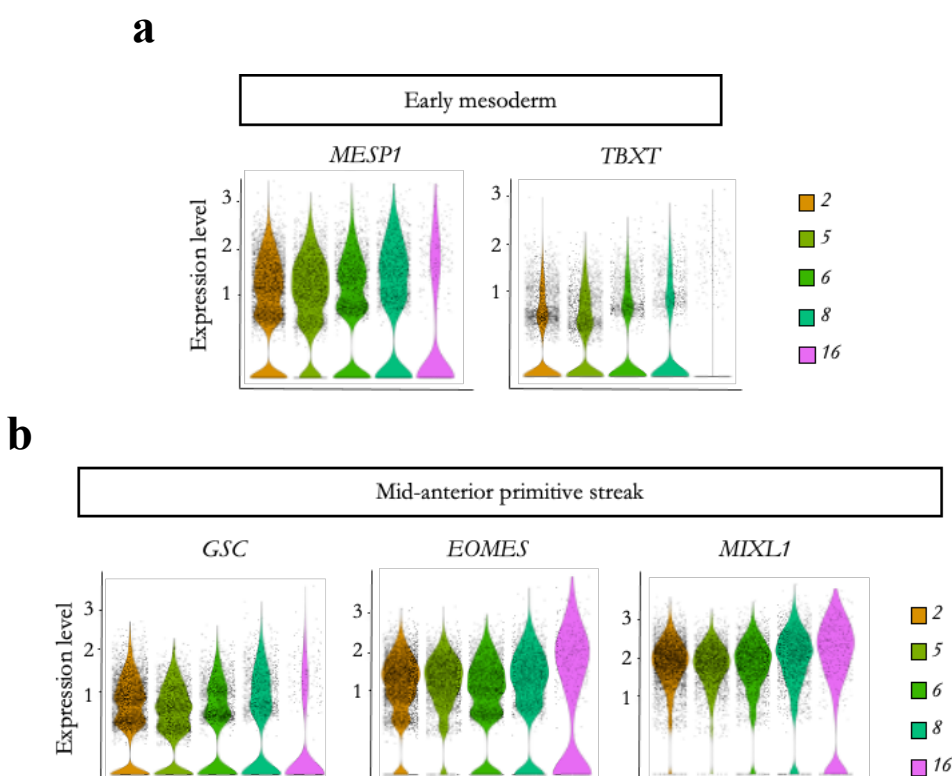
Figure 21: Early mesodermal cells of day 1.5 correspond to population of cells emerging mid-anterior primitive streak.

(a) Violin plots showing the expression level of early mesoderm genes. **(b)** Violin plots showing the expression level of genes defining the mid-anterior primitive streak stage, as defined by Mendjan et al., (2014) in d1.5 clusters depicted in UMAP plot in Fig. 20a. Adapted from Zawada et al., 2022 (preprint).

To more precisely stage these cells, the dataset of cells collected on day 1.5 was integrated with the dataset of human cells undergoing gastrulation *ex vivo* (Tyser et al., 2021b; Fig. 22). This analysis positioned the majority of day 1.5 cells between nascent and emergent mesodermal cell populations in the process of emerging from the primitive streak.
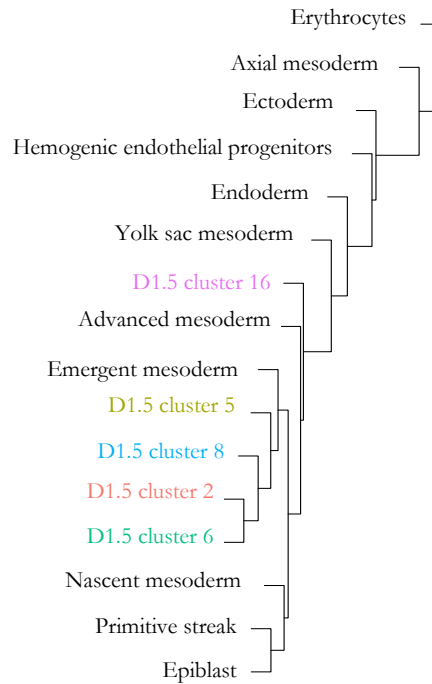
Figure 22: Comparison of d1.5 cells with cell populations from a human gastrulating embryo.

Dendrogram showing the integration of day 1.5 cell clusters with clusters of *ex vivo* human gastrulation scRNA-Seq dataset of Tyser et al. (2021b).

Previously, Lee et al. (2017) reported that at the early stages of cardiogenic mesoderm formation it is possible to distinguish between different mesoderm subtypes: CD235A/CYP26A1 expressing cells that differentiate towards ventricular cardiomyocytes and RALDH2-expressing cells that differentiate towards atrial cardiomyocytes. Interestingly, we observed a high expression of *CYP26A1* in day 1.5 early mesodermal cells, suggesting that we may have generated the mesoderm subtype that mainly gives rise to ventricular cardiomyocytes. (Fig. 23). s
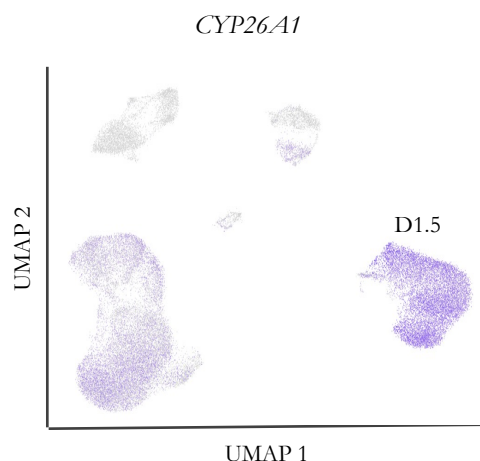
Figure 23: d1.5 mesodermal cells are marked by high expression of *CYP26A1*.

Feature plot showing expression of *CYP26A1* in cells collected at day 1.5 and FACS-sorted progenitors at day 4.5 .

3.2.2 Heterogeneity within human heart field-like progenitors

Next, we assessed the heterogeneity of cardiac progenitors sorted on day 4.5. Most of noRA-N+ progenitors separated from precursors obtained with RA treatment and allocated in clusters 3 and 13 formed by *ISL1+/FGF10+/BMP4+* aSHF-like cells (Fig. 24a-c). Inferring genes associated with different cell cycle phases revealed that cells in clusters 13 were in G2/M phase, forming a proliferating part of aSHF-like cells (Fig. 25).

Progenitors from RA treatment showed a shared contribution to clusters 0, 1, 7, 9, 10, 14, and 18, which had enriched expression of *WNT2* and *THBS4* and lower *ISL1* levels, a feature indicative of FHF identity. Clusters 0 and 1 represented the most differentiated subsets based on high expressions of sarcomere protein genes, such as *MYL3*, *MYL7,* and *PLN,* which denoted fast myocardial specification characteristic of FHF progenitors. Interestingly, markers characterising the newly identified heart field JCF (*MAB21L2*, *HOXB6*) were found expressed in clusters 18 and 14. The relatively small cluster 9 showed elevated expression of *TBX5, WNT2, HOXB1* and lower expression of genes related to contractile activity and *NKX2.5* compared to other clusters shared by RA samples. Interestingly, we noted in this cluster a slight expression of *MAB21L2,* indicating that this cluster might consist of early JCF cells (Fig. 24c). In line with mouse data, the majority of cells in this cluster were coming from the RA-T+ sample (~75%; Fig. 24b).
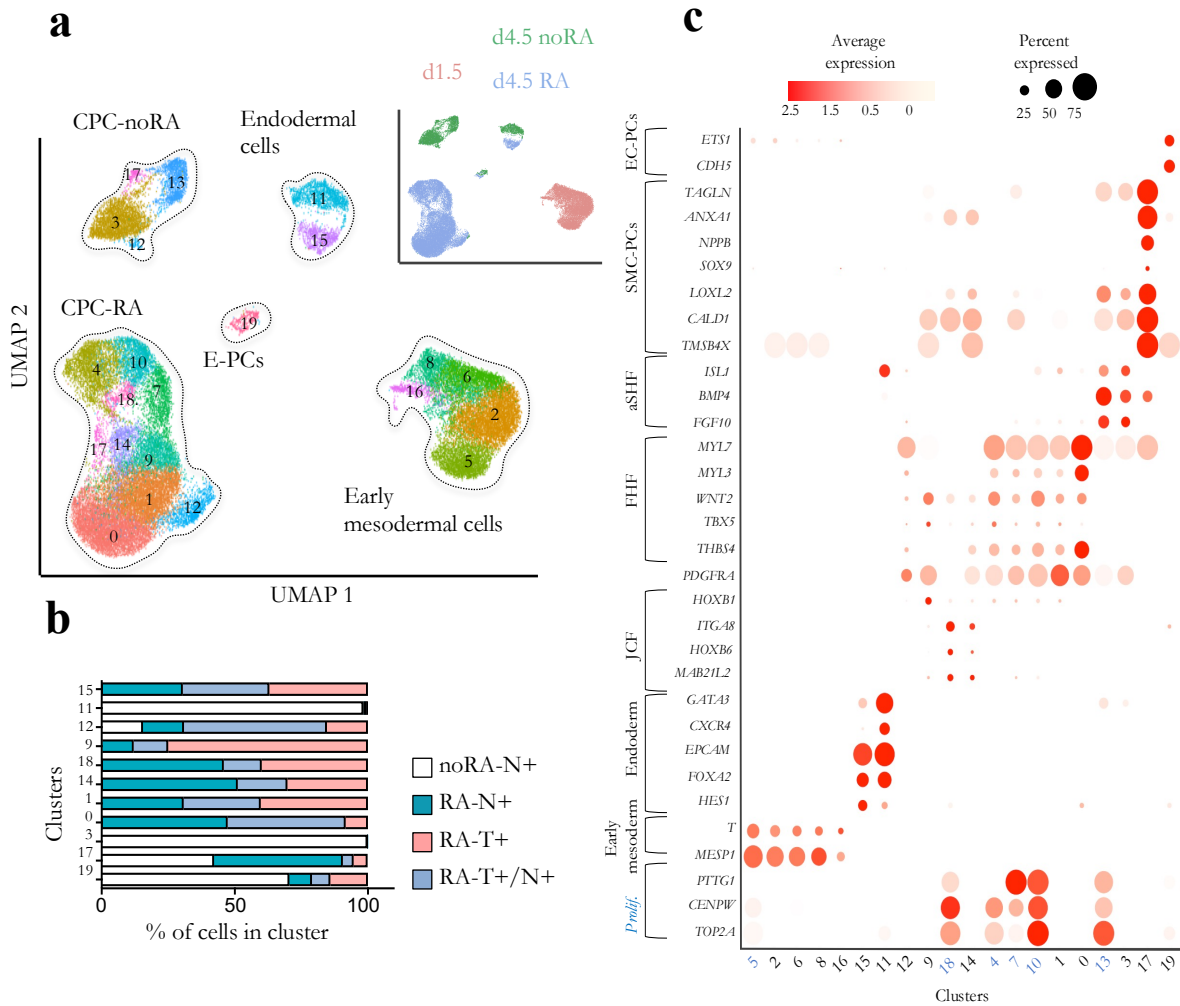
Figure 24: Heterogeneity of cell types and states within human heart field-like progenitors.

(a) UMAP clustering of single cells captured at day 1.5, and FACS sorted cells expressing mCherry (TBX5), eGFP (NKX2.5) or both at day 4.5 of RA and noRA differentiation. Inset shows contribution from RA and noRA differentiations and d1.5. CPCs: cardiac progenitor cells; aSHF: anterior second heart field; FHF: first heart field; EndoPCs: endocardial progenitors. **(b)** Percentage of cells contributing to given clusters relative to all cells in the sample: FACS sorted cells expressing mCherry (TBX5), eGFP (NKX2.5) at day 4.5 of RA differentiation and eGFP (NKX2.5) at d4.5 of noRA differentiation. T = TBX5; N = NKX2.5. **(c)** Dot plot showing selected differentially expressed genes for clusters identified as 15, 11 - endoderm; 12 – early CPCs; 1, 0 – FHF; 19 – EndoPCs; 18 - proliferating juxta-cardiac field (JCF); 14, 9 – JCF. Adapted from Zawada et al., 2022 (preprint).

Clusters 12 included cells from all sorted samples and was characterised by high expression of *PDGFRA*, which is an early myocytic marker, while cluster 17 was shared only between noRA-N+ and RA-N+ and could be distinguished by elevated expression of *TAGLN*, *NPPB*, *SOX9*, *LOXL2*, *TMSB4X*, *CALD1*, indicative of SMC-like characteristics (Tsuji-Tamura et al., 2021; Varona et al., 2008; Mayanagi and Sobue, 2011; Saunders et al., 2018). Endocardial-like progenitors expressing *KDR*, *CDH5*, *ETS1, SOX7* were identified in cluster 19, which emerged mainly, but not exclusively, from noRA-N+ (~66%; Fig. 24b).

The small clusters 11 and 15 were classified as endodermal cells based on the presence of *FOXA2* and *EPCAM* transcripts (Nowotschin et al., 2019; Sarrach et al., 2018). Furthermore, cluster 11, to which contributed exclusively cells from noRA-N+, was characterised by expression of *CXCR4* as well as *ISL1*, *NKX2.5* and *GATA3,* which in mouse are expressed in the pharyngeal endoderm (Fig. 24c) (online search tool: https://marionilab.cruk.cam.ac.uk/MouseGastrulation2018; Pijuan-Sala et al., 2018). The pharyngeal endoderm is closely positioned to mesodermal cells of aSHF throughout the early stages of mouse development and is the source of many signal transducers. Signals produced by the endoderm and the mesoderm are known to reciprocally influence each other for proper patterning and development (Gouon-Evans et al., 2006). It can therefore not be excluded that the developmental cues used in this protocol could instruct a small portion of cells to become endodermal cells.
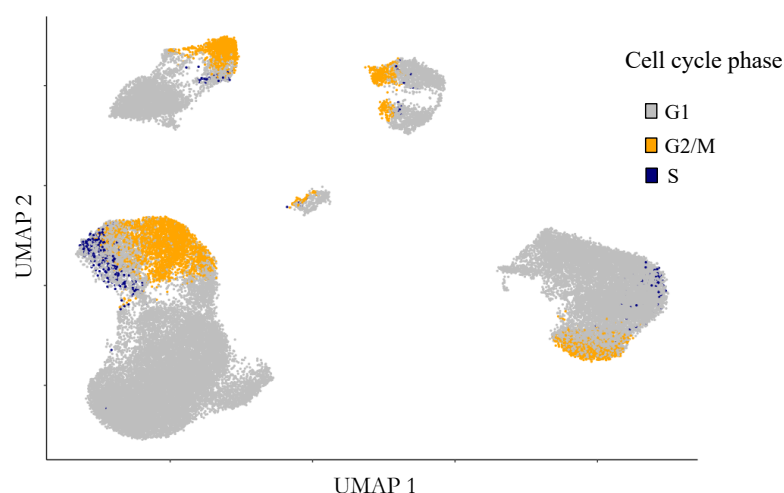
Figure 25: Identification of cells at different cell cycle phases.

Feature plot showing specific gene signatures associated with different cell cycle phases – of G2/M phase (yellow), of S phase (blue), of G1 (grey).

It is worth noting that while *TBX5* expression was almost restricted to RA-T+ and RA-N+/T+ populations at day 4.5, *NKX2-5* expression was detectable in all sorted populations. Presence of *NKX2.5* transcripts in RA-T+ cells could indicate that these cells were already entering cardiomyocyte differentiation (Fig. 26).
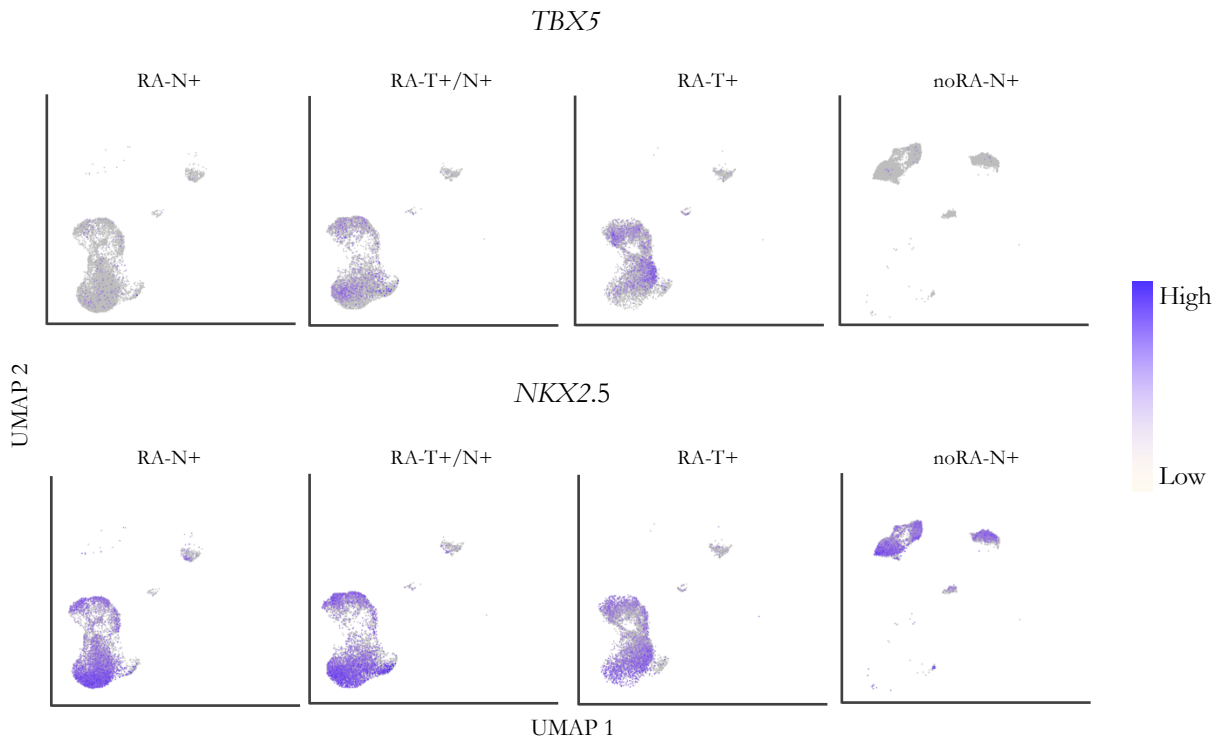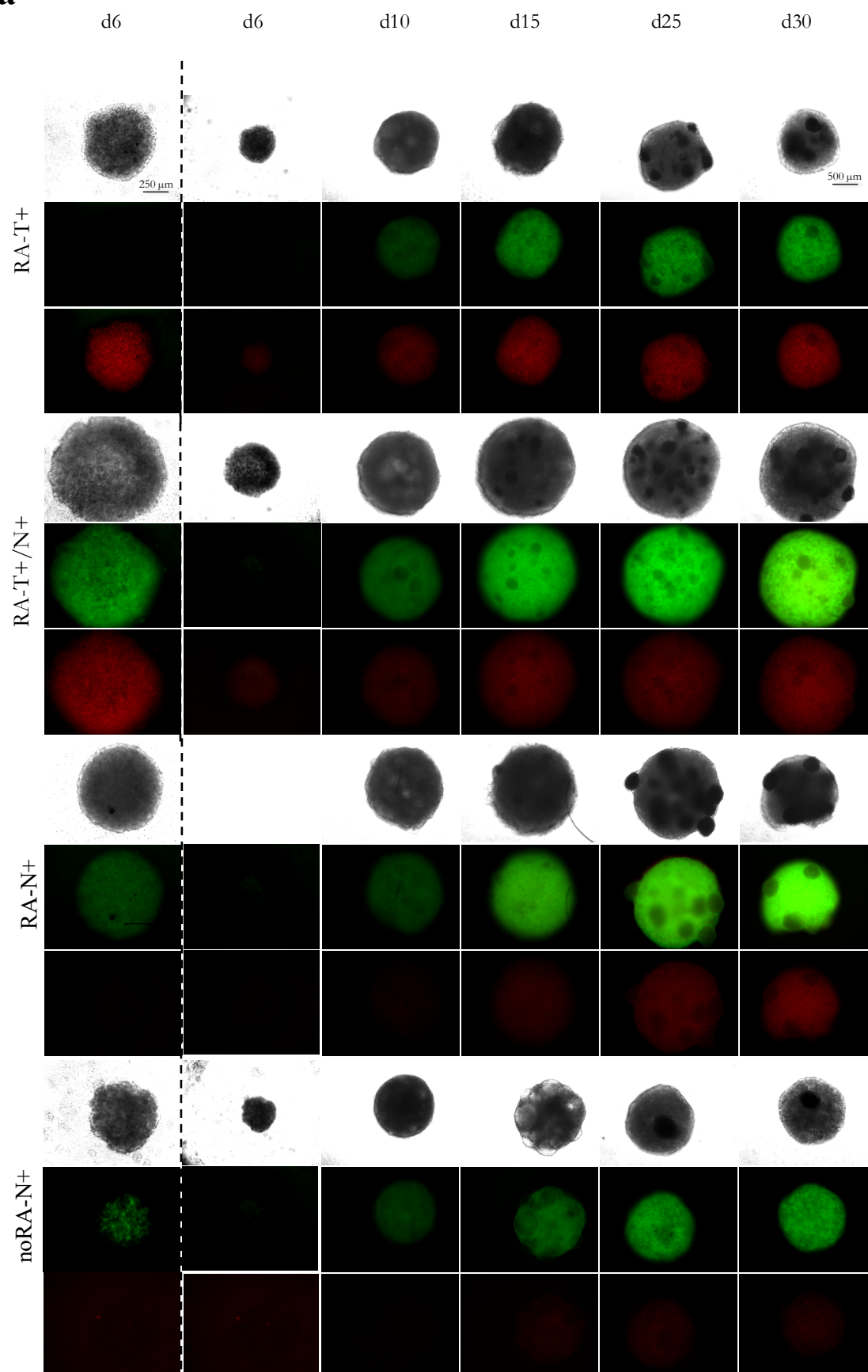

Figure 26: Dynamics of NKX2.5 and TBX5 expression within human heart field-like progenitors.

Feature plots showing *TBX5* and *NKX2.5* expression within FACS sorted cells expressing mCherry (TBX5), eGFP (NKX2.5) at d4.5 of RA differentiation and eGFP (NKX2.5) at day 4.5 of noRA differentiation. N = NKX2.5, T = TBX5.

To further confirm findings on the transcriptional level showing expression of *NKX2.5* within *TBX5* population, the potential of RA-T+ and RA-N+ cells to mature into RA-T+/N+ was investigated. To do this, live imaging analysis of aggregated RA-T+ and RA-N+ progenitors was performed. Indeed, during differentiation, the sorted RA-T+ fraction of mCherry+ (TBX5+) cells switched on eGFP (NKX2.5), and the majority of sorted RA-N+ cells switched on mCherry (TBX5; Fig. 27a). This was corroborated by flow cytometry analysis of aggregates on day 30, which showed that aggregates generated by RA-T+/N+ and RA-N+ progenitors were consisting of ~90% eGFP+/mCherry+ cells, while aggregates derived from

RA-T+ were constituted by ~60% eGFP+/mCherry+ and ~8% mCherry+/eGFP-low/- (Fig. 27b).

a
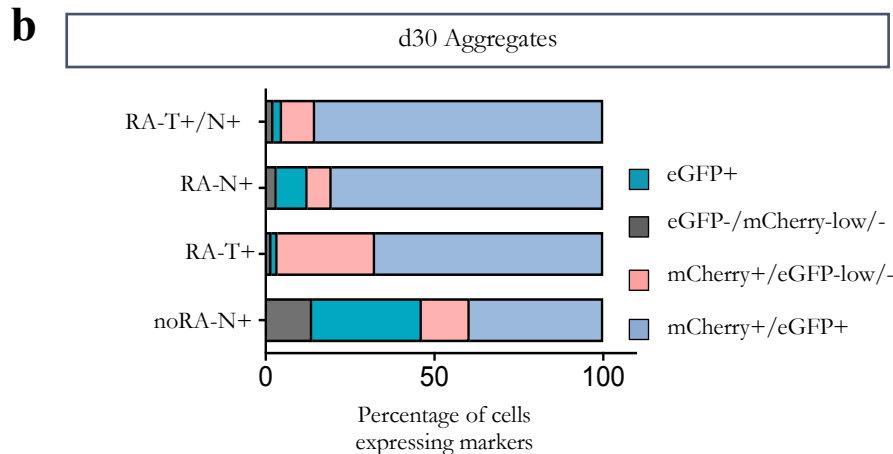
Figure 27: aSHF and FHF-like progenitors are characterized by different dynamics of NKX2.5 and TBX5 expression over time.

(a) Representative live images of days 6 to 30 aggregates derived from FACS sorted cells expressing eGFP (NKX2.5), mCherry (TBX5) or both at day 5 of noRA or RA differentiation. eGFP (NKX2.5) in green, mCherry (TBX5) in red. Scale bars: 250 μm (first column from the left) and 500 μm (columns 2 – 6). T = TBX5, N = NKX2.5. **(b)** Summary of live flow cytometry analysis of day 30 aggregates derived from FACS sorted cells expressing eGFP (NKX2.5), mCherry (TBX5) or both at day 5 of noRA or RA differentiation for a number of cells expressing mCherry (TBX5), (eGFP) NKX2.5, both or none within them. Data are mean. n = 3 differentiations.

3.2.3 Gene regulatory networks within human heart field-like progenitors

To get a better understanding of the activity of gene regulatory networks in each cell of different cell populations within human aSHF-like noRA-N+ and FHF-like RA-T+/N+ progenitors single-cell regulatory network inference and clustering was performed using SCENIC (Aibar et al., 2017). This analysis allows for matching gene expression with *cis*-regulatory analysis to identify transcription factors characteristic for distinct cell states. SCENIC analysis follows four main steps: 1) detection of co-expression of genes with transcription factors; 2) *cis*-regulatory motif analysis within each co-expression module to obtain regulons; 3) evaluation of the activity of each regulon in each cell; 4) generation of a binary activity matrix indicating a shared activity of a regulatory subnetwork (Aibar et al., 2017).

SCENIC analysis of the RA-T+/N+ sample provided 44 regulons that allowed for classification of three distinct cell states: cardiomyocyte committed progenitors; endodermal

progenitors and within each of them proliferative progenitors (Fig. 28). Progenitors that already committed towards the cardiomyocyte fate had a high activity of regulons controlled by master regulators such as *GATA4* (which was predicted to control expression of 21 co-expressed genes within cells of that cluster), while the endoderm cluster was enriched in activity of regulons controlled by *FOXA3* (11 genes), *FOXA2* (252 genes), and *RXRA* (15 genes), among others. *GATA4* and *FOXA2* have already been described as important drivers of myocardial and endodermal commitment, respectively (Ang et al., 2016; Warren et al., 2020, preprint).
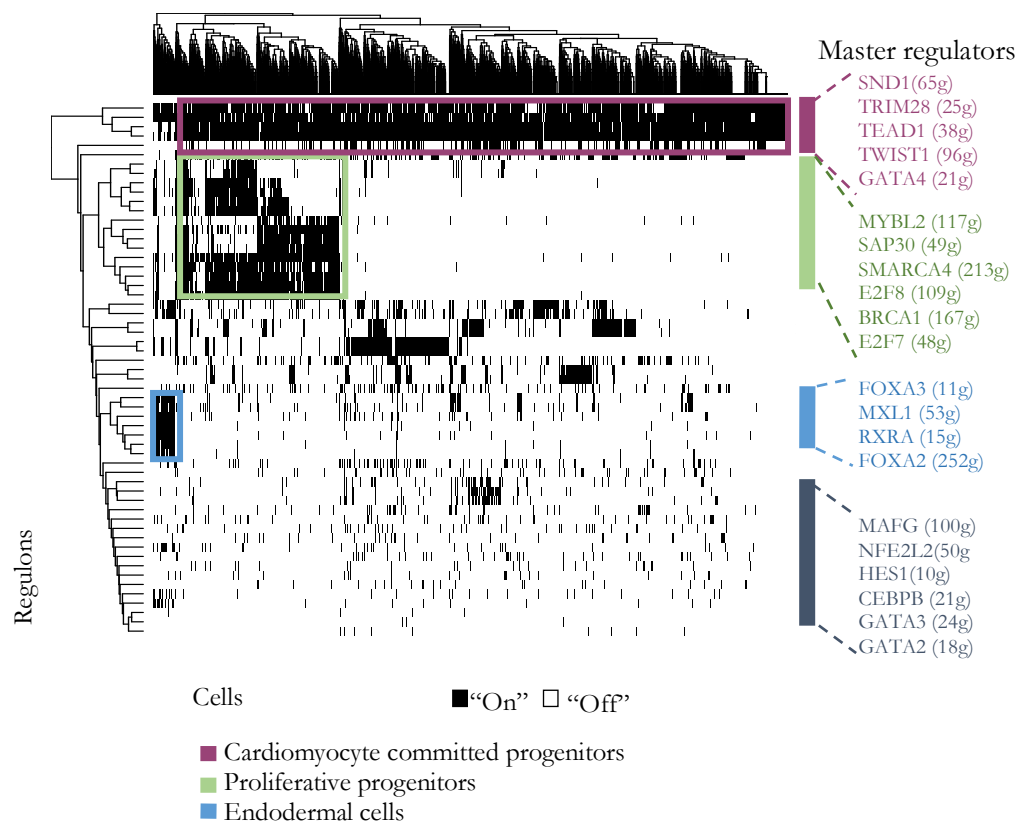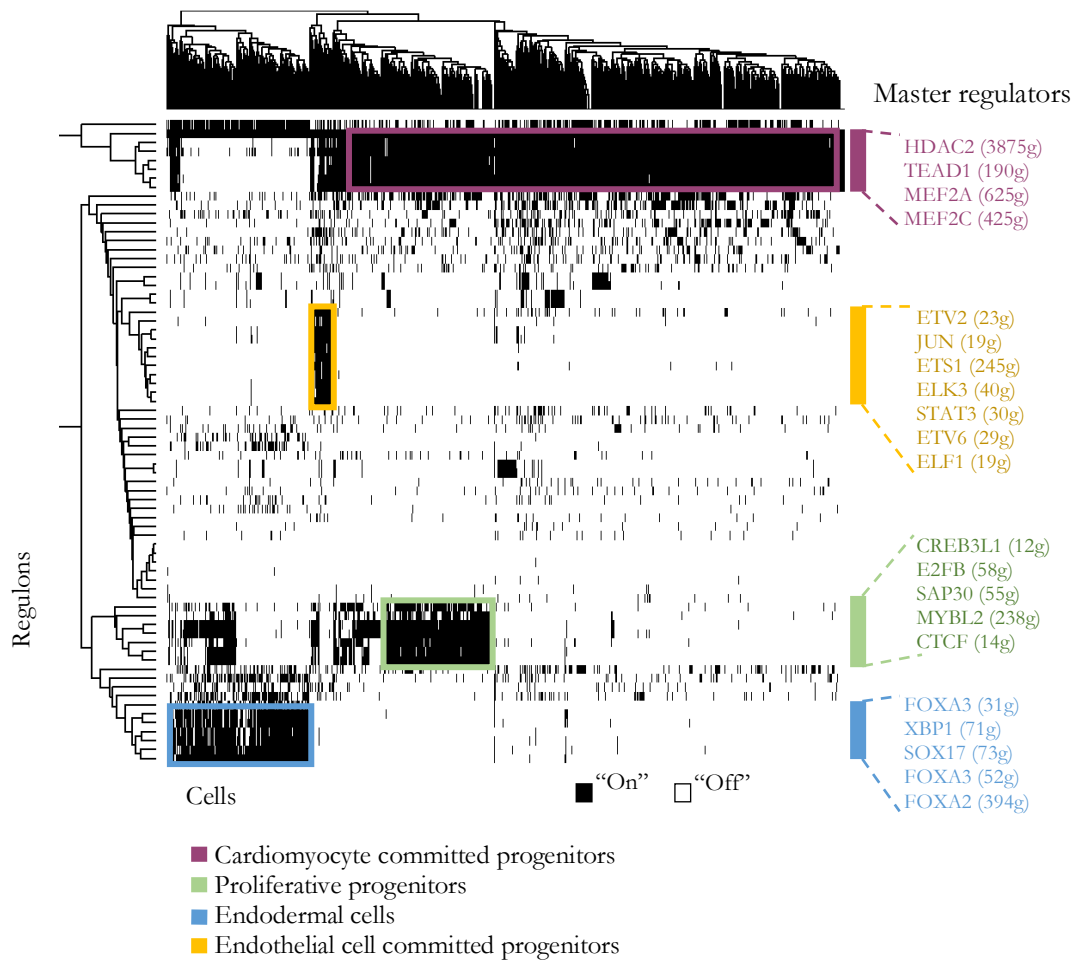

Figure 28: Distinct regulon activity driven by predicted master regulators characterize different cell types and states recovered within RA-T+/N+ progenitors.

Heatmap showing the activity of regulons (on or off) of given master regulators in cells recovered from sample FACS-sorted for TBX5 and NKX2.5 at d4.5 of RA differentiation. g: genes.

A parallel analysis of the noRA-N+ sample allowed the distinction of four distinct cell states based on 53 regulons: an early endodermal population; a cardiomyocyte committed population, and proliferative cells within these groups as well as an endothelial cell population (Fig. 29). Endodermal cells had a high activity of regulons controlled by master regulators such as *FOXA3* (which was predicted to control expression of 13 co-expressed genes within cells of

that cluster), *SOX17* (52 genes), and *FOXA2* (394 genes). Endothelial cells showed, among others, high activity of the *ETV2* regulon consisting of 23 target genes, including ones important for endocardial/endothelial cells development, such as *SOX7*, *NR4A1*, *KDR* and *KLF7* (Hong et al., 2021, Cheng et al., 2021). Other transcription factors that were predicted to govern endocardial/endothelial cells development included *ETS1* (245 genes), *ELK3* (40 genes), *ETV6* (29 genes), while myocardial progenitors were characterised by presence of *MEF2C* (425 genes), *MEF2A* (625 genes), and others. In line with the more heterogenous nature of SHF progenitors, this population of cells seems to have a higher number of regulons governing the specification of cells towards different fates.

Figure 29: Distinct regulon activity driven by predicted master regulators characterize different cell types and states recovered within noRA-N+ progenitors.

Heatmap showing the activity of regulons (on or off) of given master regulators in cells recovered from sample FACS-sorted for NKX2.5 at d4.5 of noRA differentiation. g: genes.

Many master regulators predicted through SCENIC were already reported to be essential for controlling developmental gene regulatory networks, such as *ETV2* for endothelial specification (Koyano-Nakagawa and Garry, 2013) or *FOXA1/2* for endodermal differentiation (Warren et al., 2020, preprint). Further studies would be required to confirm whether other, so far unvalidated master regulators are indeed essential for specification into particular lineages (for example, *ELK3* and *TRIM28* have not yet been studied in the context of regulating endothelial and cardiomyocyte gene programs, respectively).

3.2.4 *In vitro* generated human heart field-like progenitors give rise to distinct cardiac cell types according to their predicted lineage potential

To investigate the lineage potential and fate decisions of the progenitors described above, on day 4.5 cells differentiated in presence or absence of RA were sorted based on TBX5/mCherry and NKX2.5/eGFP expression, aggregated and allowed to differentiate for another 25 days. 3D approach was adopted to allow for more physiological and less invasive cell culture condition. Datasets coming from day 30 were then merged and analysed (Fig. 30).
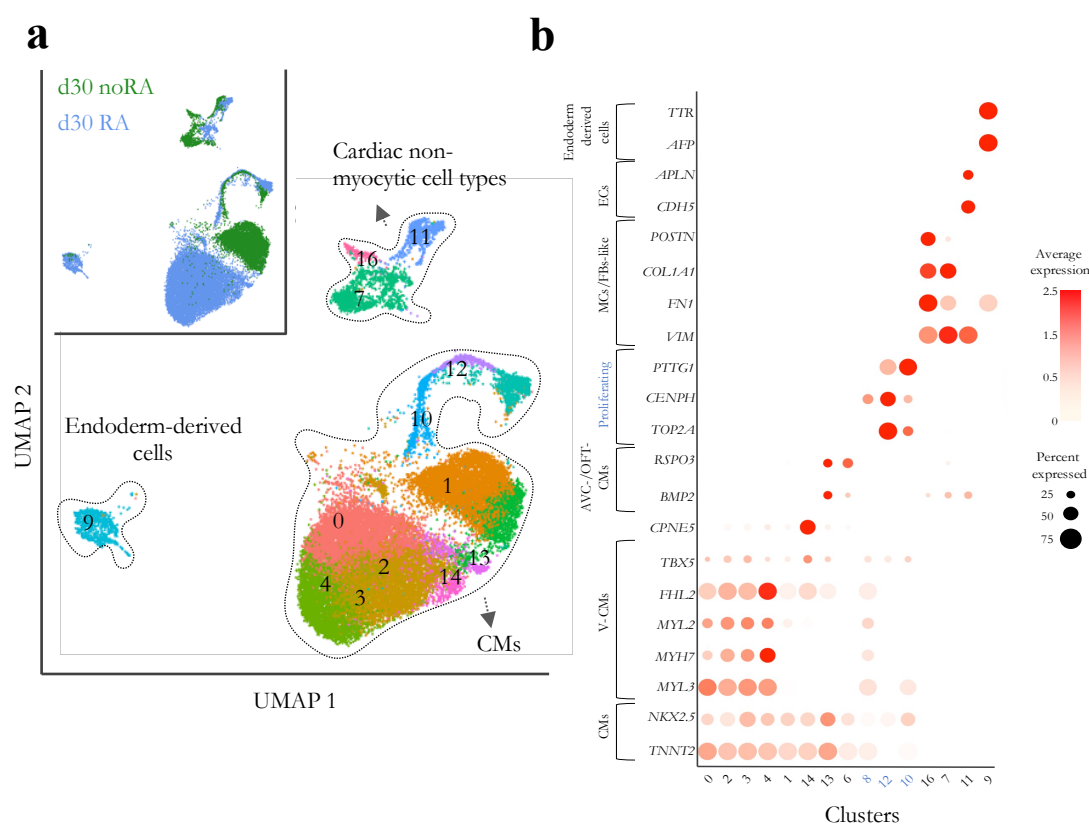
Figure 30: Human heart field-like progenitors give rise to distinct cardiac cell types.

(a) UMAP clustering of single cells captured at d30. Merged image of all analysed populations at d30. CMs: cardiomyocytes. Inset shows contribution from RA and noRA differentiations. **(b)** Dot-plot showing selected differentially expressed genes for clusters shown in panel a. CMs: cardiomyocytes; V-CMs: ventricular cardiomyocytes; AVC/OFT-CMs: atrioventricular canal/outflow-tract cardiomyocytes; MCs/FBs: mesenchymal cells/fibroblasts; ECs: endothelial cells. Clusters enriched in transcripts upregulated during proliferation are highlighted in blue. Adapted from Zawada et al., 2022 (preprint).

On day 30, the vast majority of cells were *TNNT2+* cardiomyocytes (clusters 0, 1, 2, 3, 4, 8, 10, 12, 13, 14), but transcripts typical of endothelial/endocardial cells and fibroblasts-/mesenchymal cells- like cells (clusters 7, 11 and 16; *CDH5+*; *VIM+*), and *AFP+* endoderm-derived cells (cluster 9) were likewise recovered from all samples, although with different contribution (Fig. 31a-b).
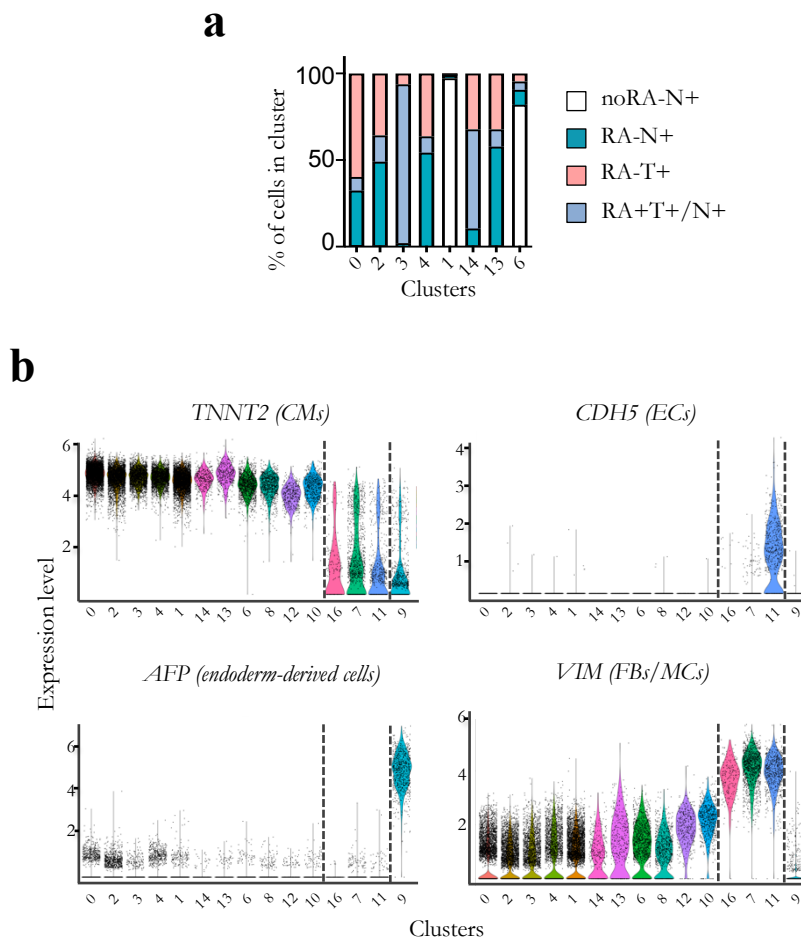
Figure 31: Human heart field-like progenitors give rise to three major groups of cells.

(a) Percentage of cells contributing to given clusters relative to all cells in the sample – aggregates from FACS sorted cells expressing mCherry (TBX5) and/or eGFP (NKX2.5) at day 4.5 of RA differentiation and eGFP (NKX2.5) at day 4.5 of noRA differentiation. Data are normalized to the total number of cells in each population **(b)** Violin plots showing gene expression level in chosen clusters. 0, 2, 3, 4, 1 – Ventricular-CMs; 6 – Outflow-tract CMs; 13 – Atrioventricular canal-CMs; 14 – Nodal transitional-like CMs; 8, 10, 12 – proliferating CMs. 16, 7, 11 – non-myocytic cardiac cell types. 9 – endoderm derived cells. MCs: mesenchymal cells; FBs: fibroblasts; ECs: endothelial cells; CMs: cardiomyocytes.

Remarkably, most cardiomyocytes emerging from noRA-N+ progenitors clustered separately from those obtained from RA-N+, RA-T+, and RA-T+/N+ progenitors, suggesting intrinsic differences in their transcription profiles (Fig. 30). Clusters 0, 2, 3, 4 were exclusively composed of cells generated in the presence of RA (Fig. 31a). They were all characterised by expression of *MYL2* and *MYH7,* and as such could be assigned as ventricular cardiomyocytes (Fig. 32). Cells from aggregates of noRA-N+ origin contributed exclusively to cluster 1 consisting of *MYL2*+ ventricular cardiomyocytes (Fig. 32). Cells in cluster 6 displayed elevated levels of *RSPO3* and *BMP2* transcripts - genes described to be enriched in conoventricular cardiomyocytes (late OFT cardiomyocytes; Sahara et al., 2019). This subset of cells was also characterised by higher expression of SMC markers (e.g., *ACTA2*, *TAGLN*, *COL1A2*) and relatively lower expression of *MYL2* (Fig. 32). This corresponds well with the transcriptional profile of myocardial subpopulations building the OFT, as reported in mice (Liu et al., 2019). *TBX5* expression was not detected in this subset of cells, in line with the pattern of *TBX5* expression in human foetal tissues (Hatcher et al., 2000). Overall, this classification proves that human aSHF-like progenitor derivatives indeed adopt myocytic lineage fates corresponding to the ones described in animal models *in vivo* - myocardial cells of the OFT and the ventricle.

Cluster 13, which was mainly composed of cells from RA-T+ and RA-T+/N+ samples, could be identified as AVC cardiomyocytes (Fig. 32) Indeed, it shared some similarities in gene expression profile to cells of cluster 6 (e.g., expression of *BMP2, ACTA2, COL1A2*), but differed by its relatively high *TBX5* expression (Fig. 32). This is line with the transcriptional profiles of the AVC and the OFT *in vivo* (Vrlijcak et al., 2012, Hatcher et al., 2000).

To cells of the small cluster 14, to which contributed cells derived from progenitors generated in the presence of RA, we assigned the identity of cells in the process of transitioning

into nodal CMs (Fig. 31a). These cells were enriched in *CPNE5* and *IGFBP5* transcripts similarly to population of cells identified by Wiesinger et al. (2022) during directed differentiation of hPSCs towards pacemaker-like cells (Fig. 32).
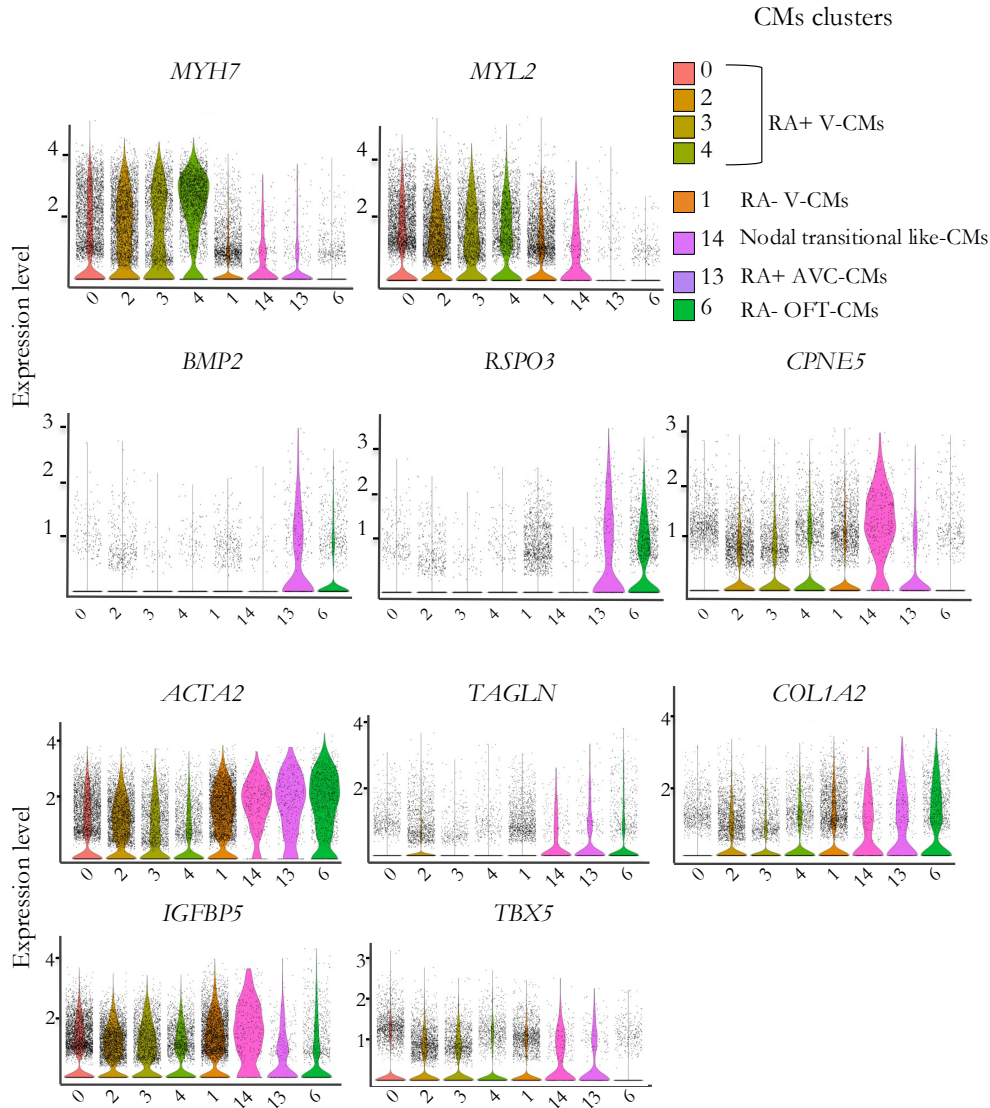

Figure 32: Human heart field-like progenitors give rise to distinct cardiomyocyte subtypes according to their predicted lineage potential.

(a) Violin plots showing selected differentially expressed genes for clusters identified as 13 – AVC CMs; 6 – OFT CMs; 0, 1, 2, 3, 4 – V-CMs **(b)** Feature plots of selected differential expressed genes for given clusters. UMAP plot shown in Fig. 30. V: Ventricular; OFT: Outflow-tract; AVC: Atrioventricular canal-CMs. Adapted from Zawada et al., 2022 (preprint).

To refine our understanding of the cell types generated *in vitro*, a comparison was performed between scRNA-Seq datasets obtained from human foetal tissues and our datasets. Co-expression of top differentially expressed genes in ventricular cardiomyocytes provided by Cui et al. (2019) and Asp et al. (2019) (*MYH7, VCAN, MYL2, MASP1, S100A4, LDHA, FHL2*) specifically matched cluster 2, 3, 4, as well as some cells of cluster 0 and part of cluster 1 (Fig. 33). As scRNA-Seq analyses clearly defining human OFT cells have not been exhaustively covered so far, the mouse dataset provided by Li et al. (2016) was used to obtain the OFT markers (*COL1A2, RSPO3, CXCL12, ACTA2, TAGLN*). In line with our classification, this OFT signature mainly matched cells in cluster 6 (Fig. 33).
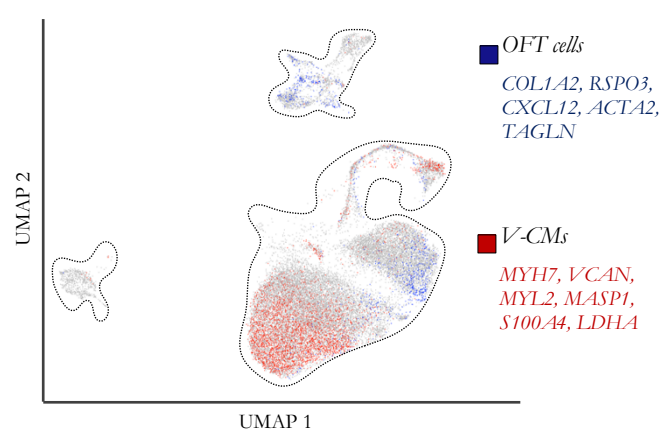
Figure 33: Human heart field-like progenitors' derivatives are characterized by gene signatures of different populations identified by Cui et al. (2019) and Asp et al. (2019), and Li et al. (2016).

Feature plot showing V-CMs: ventricular cardiomyocyte signature – co-expression of top differentially expressed genes defining ventricular cardiomyocytes as described by Cui et al., (2019) and Asp et al. (2019), and confirmed with specific expression in ventricular clusters in Asp et al. (2019); and OFT-CMs – co-expression of key genes defining OFT cardiomyocytes as described by Li et al. (2016). OFT: outflow-tract. V: ventricular. Adapted from Zawada et al., 2022 (preprint).

To get a better resolution of cardiac *VIM+* or *CDH5+* non-myocytic clusters, a second round of clustering was performed. Subclustering of clusters 7, 11 and 16 recovered five groups of cell types (Fig. 34a-b): endothelial cells expressing *APLN* and *CDH5* in cluster 2; myofibroblasts expressing *MYL2* and *MYOZ2* in cluster 3; smooth muscle cells- like cells expressing *TAGLN, ACTA2* and *CALD1* in cluster 1; cells enriched in the fibroblasts markers *THY1, COL1A1* as well as markers of endothelial-to-mesenchymal transition (Endo-MT), such

as *ZEB2* and *LUM* in cluster 0; and cells expressing genes associated with more advanced Endo-MT processes and valvular markers such as *SOX9, POSTN* in cluster 4 (Fig. 34d). Also, co-expression of key markers of valvular-like cells described by Mikyrukov et al. (2021) and Neri et al. (2019) – *SOX9, POSTN, LUM, CDH11, COL3A1, VIM* – specifically highlighted cells in cluster 4 (Fig. 34b). The vast majority of clusters 4 and 0 was composed of cells from the noRA-N+ sample, which is in line with SHF-derived endocardial progenitors' contribution to valve formation in mice (Fig. 34d; Verzi et al., 2005; Crucean et al., 2017).
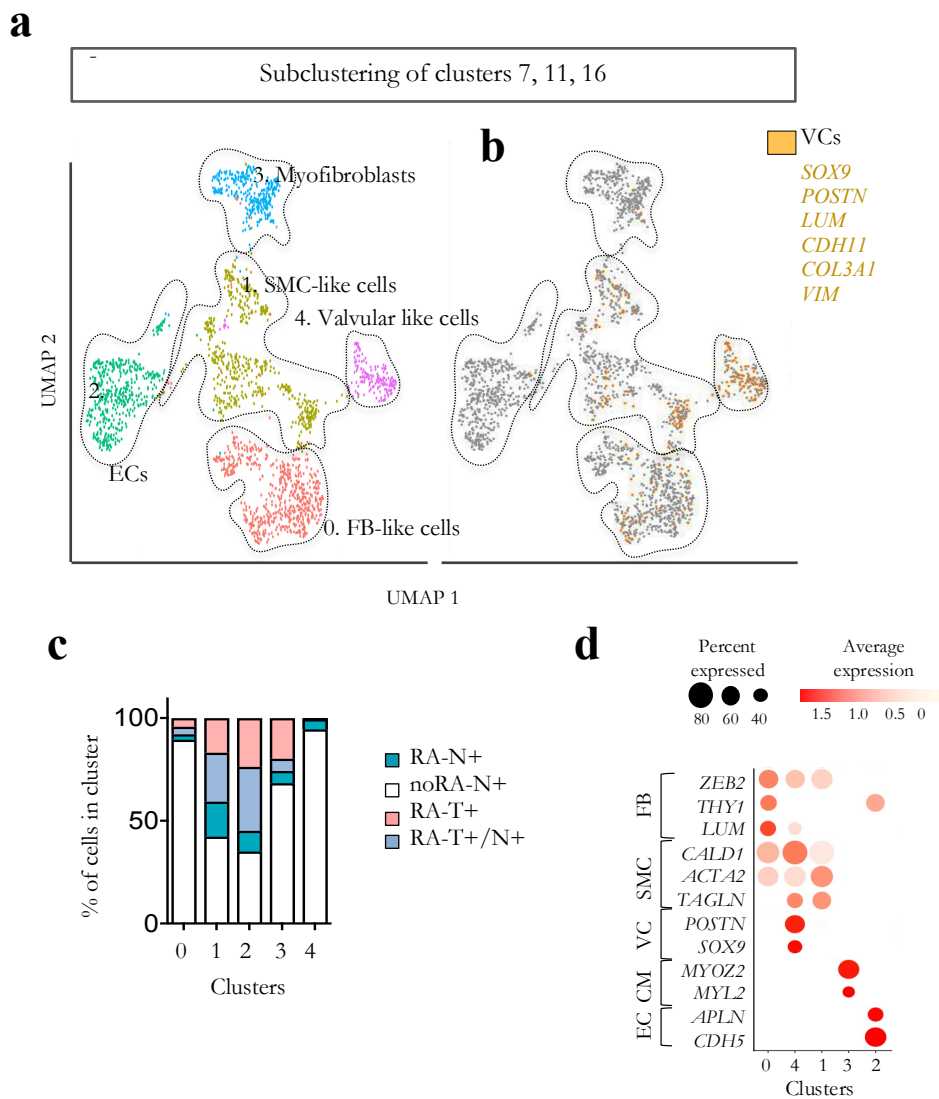
Figure 34: Human heart field-like progenitors give rise to distinct non-myocardial cell types and states according to their predicted lineage potential.

(a) UMAP subclustering of cardiac non-myocytic clusters (7, 11, 16) of single cells coming from FACS-sorted cells reaggregated on day 4.5 and captured 25 days after reaggregation at day 30. **(b)** Feature plot showing cells matching a valvular cell (VC) signature, i.e., the co-expression of key VC genes described by Neri, et al. (2019) and Mikryukov et al. (2020) and Cui et al. (2019), namely, *SOX9*, *POSTN*, *LUM*, *CDH11*, *COL3A1*, *VIM*. **(c)** Percentage of cells contributing to given clusters relative to all cells in the cluster: FACS-sorted cells expressing mCherry (TBX5), eGFP (NKX2.5) or both of RA differentiation and eGFP (NKX2.5) at day 4.5 of noRA differentiation were aggregated at analysed at day 30. **(d)** Dot plot showing selected differentially expressed genes for clusters. ECs: endothelial cells; FBs: fibroblasts; VCs: valvular cells; SMCs: smooth muscle cells. T = TBX5; N = NKX.2.5. Adapted from Zawada et al., 2022 (preprint).

Overall, based on scRNA-Seq analysis, aggregates derived from noRA-N+ sample constituted of approximately 15.5% non-myocytic cell types whereas in aggregates formed from sorted populations exposed to RA, non-myocytic populations oscillated between 4 and 12% (Fig. 35).
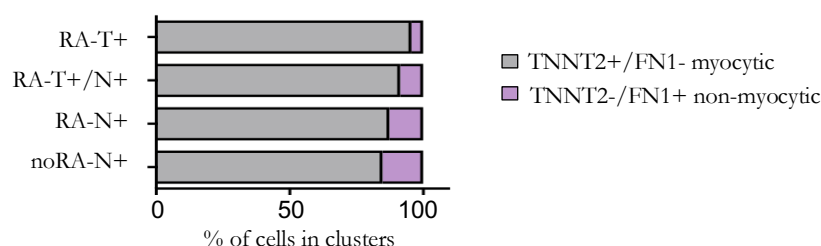
Figure 35: Human heart field-like progenitors differentially contribute to myocytic and non-myocytic clusters.

Percentage of cells from d30 aggregates derived from FACS sorted cells at day 4.5 expressing mCherry (TBX5), eGFP (NKX2.5) of RA differentiation and eGFP (NKX2.5) of noRA differentiation contributing to myocytic and non-myocytic clusters. T = TBX5; N = NKX2.5.

Immunofluorescence staining confirmed that aggregates of all samples were formed mainly from cardiomyocytes (cTnT+) with a small proportion of endoderm-derived cells (E-cadherin+) and endothelial cells (CD31+; Fig. 36a'-", 36a""). Larger clusters of endothelial cells were only found in aggregates derived from noRA-N+ sorted progenitors. While some cells in

noRA-N+ derived aggregates were positive for mCherry+ (TBX5+), substantial number of cells remained eGFP+ (NKX2.5+) only (Fig. 36a''').
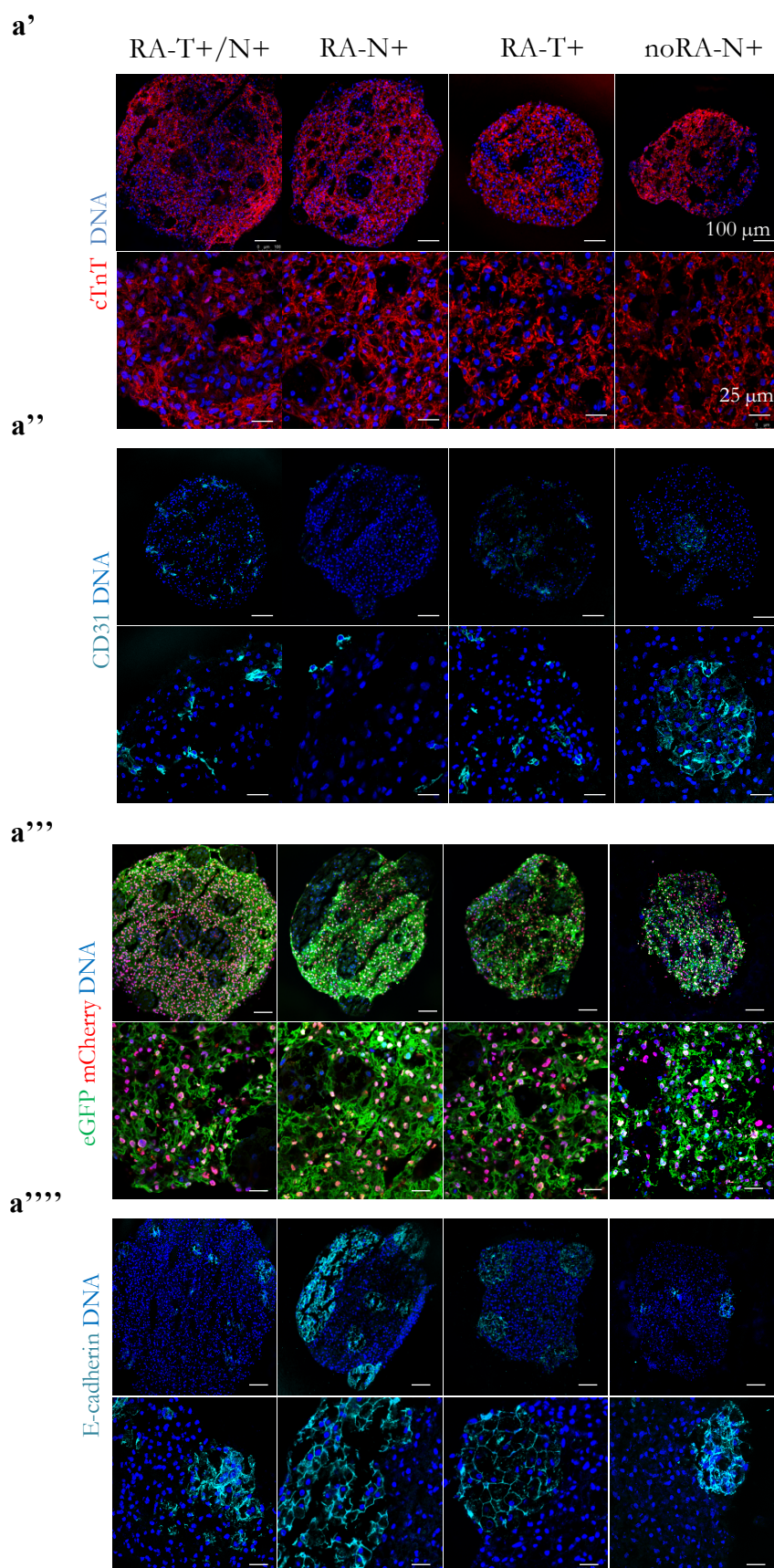
Figure 36: Human heart field-like progenitors give rise to distinct cell types according to their predicted lineage potential.

(a) Representative immunofluorescence images of day 30 aggregates, 25 days after reaggregation of cells FACS-sorted for TBX5 (mCherry) or NKX2.5 (eGFP) or both at day 4.5 during RA differentiation and of cells FACS-sorted for NKX2.5 (eGFP) at d4.5 during noRA differentiation; using antibodies against **(a')** cTnT (red); **(a")** mCherry (red), eGFP (green); **(a''')** CD31 (light blue); **(a'''')** E-cadherin (light blue). Nuclei were counterstained with Hoechst (blue). Size bars: upper panels a', a", a''', a'''': 100 μM; bottom panels of a', a", a''', a'''': 25 μM.

scRNA-Seq, immunofluorescence and flow cytometry analyses suggested that sorted populations of cells are likely representing different stages of progression of the same differentiation trajectory. Specifically, it showed that RA-T+ and RA-N+ cells during differentiation converge into RA-T+/N+ population. Therefore, we hypothesized that a similar cell composition could be obtained without sorting and re-aggregation in a simple 2D monolayer cell culture system. Flow cytometry analysis of various pluripotent stem cell lines differentiated as a monolayer till d30 showed that in the absence of RA, on average, more than 80% of cells were cTnT+ cardiomyocytes and approximately 10% to 20% were FN1+ non-myocytic cells, while RA exposure resulted in more than 90% cTNT+ cells at day 30 and less than 10% FN1+ (Fig. 37).
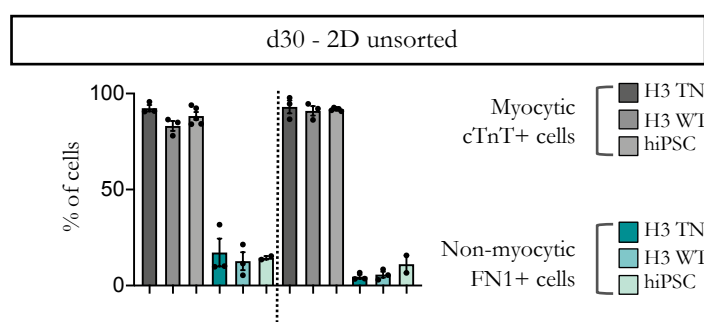
Figure 37: RA/noRA monolayer differentiation approach applied to various human pluripotent cell lines shows similar efficiency in generating myocytic and non-myocytic cell types.

Summary of flow cytometry analysis of cells expressing cTnT and Fibronectin at day 30 of noRA and RA differentiation of various cell lines. Data are mean ± SEM; n = 3 differentiations. Adapted from Zawada et al., 2022 (preprint).

Across multiple cell lines, unsorted cells differentiated as monolayer till d30 showed similar patterns of cardiomyocyte marker expression, with enrichment of *RSPO3* in the absence of RA and *TBX5* enrichment upon RA exposure (Fig. 38).
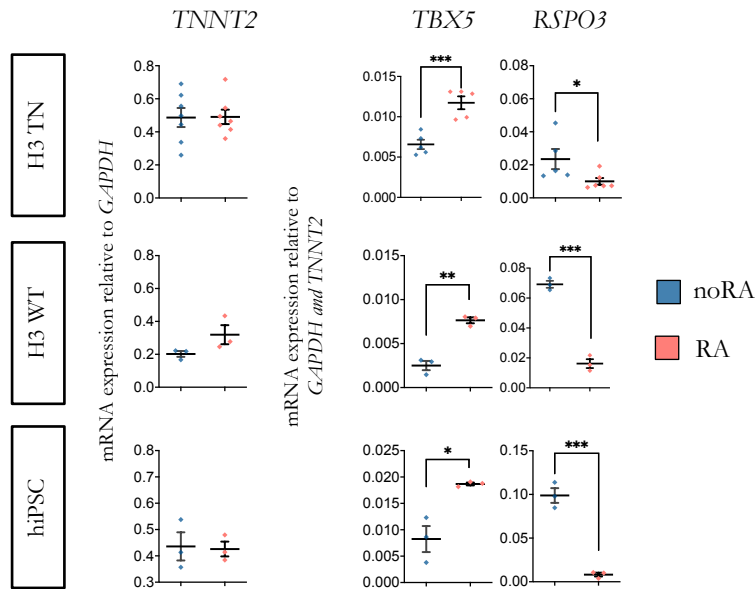

Figure 38: Application of RA/noRA differentiation approaches to various human pluripotent cell lines shows similar expression patterns of key markers of cardiomyocyte subtypes.

Analysis of mRNA expression *TNNT2, SOX4, RSPO3, TBX5* at d30 of noRA and RA differentiation. mRNA expression relative to *GAPDH* or *GAPDH* and *TNNT2*. Data are mean ± SEM; n = 3 differentiations/time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed *t*-test between RA and noRA). Adapted from Zawada et al., 2022 (preprint).

Taken together, the differentiation efficiency into different cardiac lineages was similar between sorted and unsorted cells. This highlights that cell sorting is not necessary to allow the emergence of defined cardiomyocyte subtypes; the modulation of RA is sufficient to observe differences.

3.3 Unravelling a novel progenitor population resembling murine juxta-cardiac field progenitors

3.3.1 ITGA8 allows for isolation of progenitors resembling juxta-cardiac field cells

As mentioned above, our analysis of scRNA-Seq of day 4.5 cells, captured a population similar to the murine JCF, which is the first time this population has been described in a human system. To isolate this population for investigations, screening for possible cell surface markers was conducted. ITGA8 was identified as a candidate based on its co-expression with key JCF markers (*MAB21L2*, *HOXB6*, *HOXB5*, *HAND1*, *BNC2*; Fig. 39a-b). Adapted from Zawada et al., 2022 (preprint).
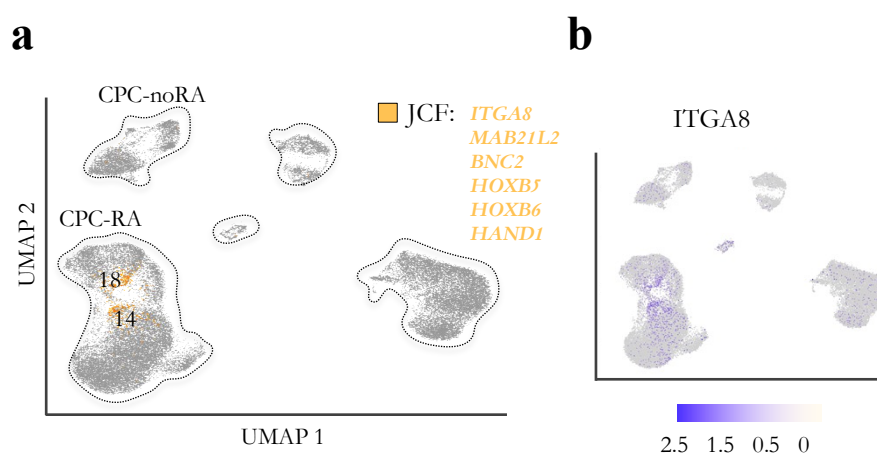
Figure 39: A cell population corresponding to mouse JCF is found within human FHF-like progenitors.

(a) Feature plot showing the co-expression of key genes characteristic for the JCF cluster Me5 (*MAB21L2*, *HAND1*, *HOXB6*, *HOXB5*, *BNC2*, *ITGA8)* defined by Tyser et al. (2021). **(b)** Feature plot showing expression of *ITGA8*. Adapted from Zawada et al., 2022 (preprint).

Analysis of mRNA expression showed a dynamic expression of JCF markers during differentiation in presence of RA (Fig. 40). Expression of *HOXB6* and *MAB21L2* peaked on

day 3, followed by *BNC2* and *ITGA8* on day 4: all markers then decreased on day 5 to become virtually absent by day 7.
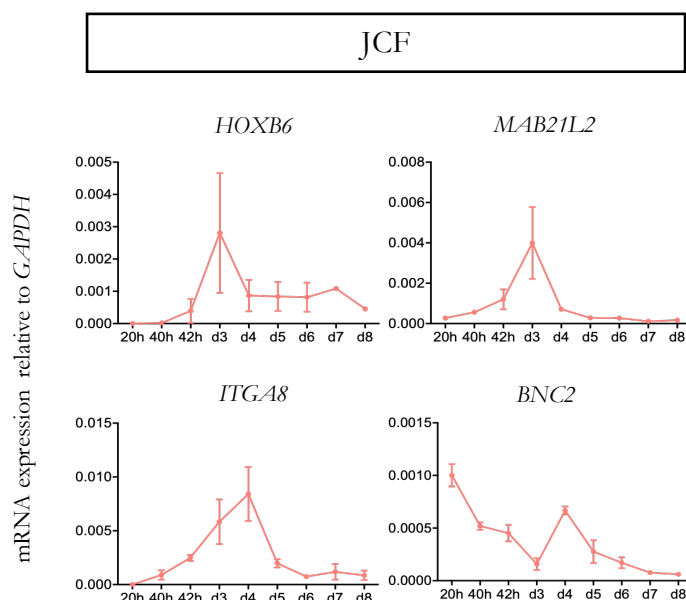
Figure 40: Sequential expression of JCF markers during cardiac differentiation of hPSC.

Analysis of mRNA expression of key JCF genes as defined by Tyser et al. (2021) between 20h – d8 of RA differentiation. mRNA expression relative to *GAPDH*. Data are mean ± SEM; n = 3 differentiations/time point.

Flow cytometry analysis revealed a transient population of ITGA8+ cells during cardiac differentiation. Consistent with the gene expression dynamics described above, a small proportion of cells expressed ITGA8+ on day 4, that population peaked on day 5 and went down on day 6 (Fig. 41a-b). In line with mouse data, the majority of ITGA8+ cells were TBX5+ (mCherry+)/NKX2.5- (eGFP+) early on, and quickly upregulated NKX2.5 (fig. 41c-d; Tyser et al., 2021). RT-qPCR analysis of JCF markers demonstrated a significant enrichment in *BNC2, MAB21L2* and *HAND1* expression in ITGA8+ cells, while on day 5 the expression of these markers decreased to the level of their unsorted counterparts (Fig. 41e). The levels of expression were not significantly different at day 5 between sorted and unsorted counterparts most probably due to timing of sorting which happen one day after peak expression of those genes.
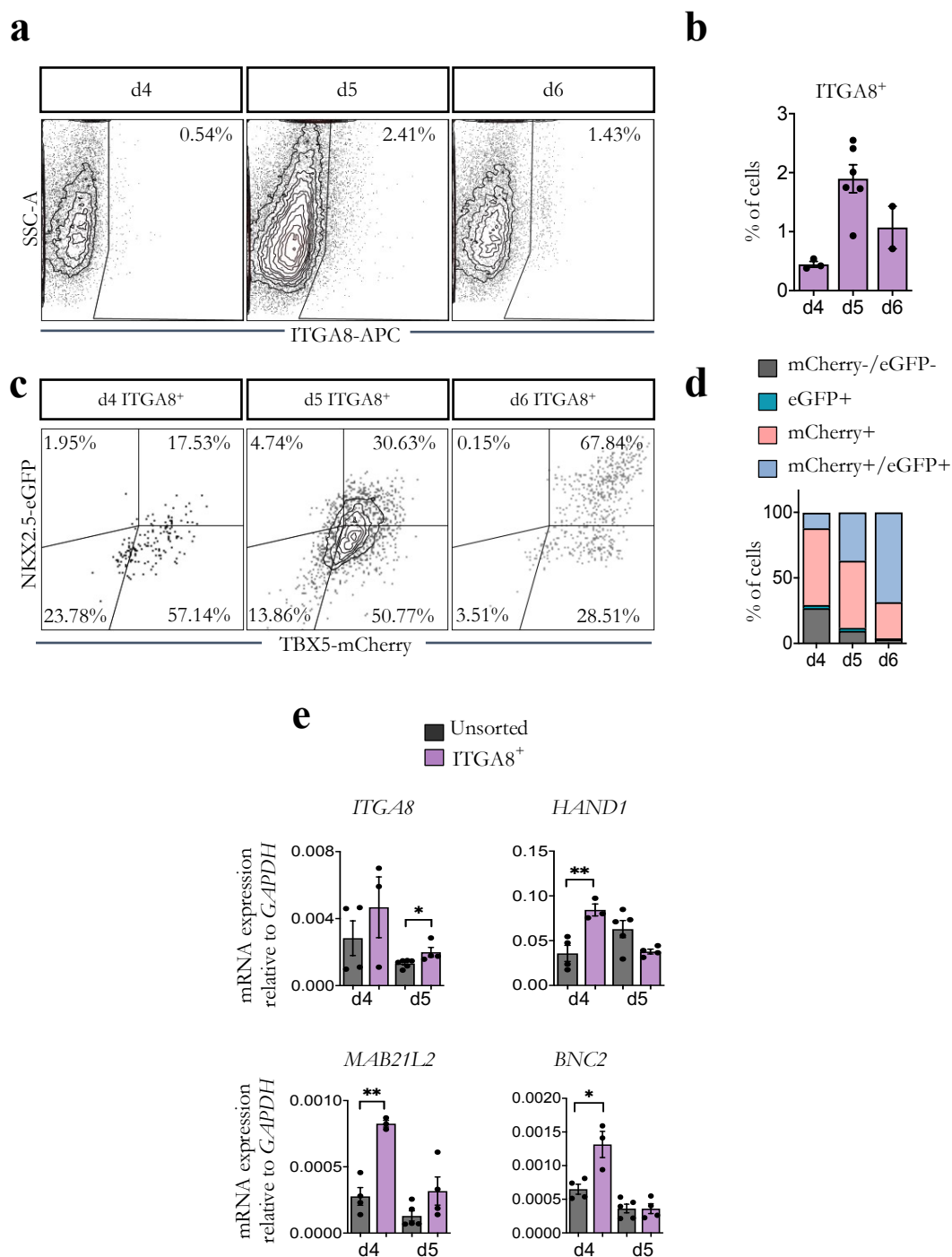
Figure 41: ITGA8 allows for isolation of a subset of progenitors resembling mouse JCF.

(a) Representative plots of live flow cytometry analysis of cells expressing ITGA8 (APC) between day 3.5 and day 5.5 of RA differentiation. **(b)** Summary of live flow cytometry analysis of ITGA8 (APC) expressing cells between days 4 and 6 of RA differentiation. Data are mean ± SEM; n = 2-6 differentiations. **(c)** Representative plots of live flow cytometry analysis of cells expressing mCherry

(TBX5) and (eGFP) NKX2.5 within ITGA8 (APC) population between days 4 and 6 of RA differentiation. **(d)** Summary of live flow cytometry analysis of cells expressing mCherry (TBX5) and (eGFP) NKX2.5 within ITGA8 (APC) population between days 4 and 6 of RA differentiation. Data are mean ± SEM; n = 2-5 differentiations. **(e)** Analysis of mRNA expression of JCF markers in day 4 and day 5 FACS sorted cells for ITGA8 (APC) and unsorted during RA differentiation. mRNA expression relative to *GAPDH*. Data are mean ± SEM; n = 3 - 6 differentitions/time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed *t*-test). Adapted from Zawada et al., 2022 (preprint).

In mice, the JCF population was reported to give rise to both myocardial and epicardial lineages. To explore the lineage potential of our ITGA8+ population, the isolated cells were subjected to an epicardial differentiation protocol developed by Bao et al. (2017) or to previously described myocardial culture conditions (Fig. 42a). Immunostaining analysis showed that ITGA8+ cells differentiated into both cTnT expressing cardiomyocytes and epicardial cells forming a tight epithelial layer marked by ZO1, CK18 and TCF21, validating their dual potentiality (Fig. 42b).
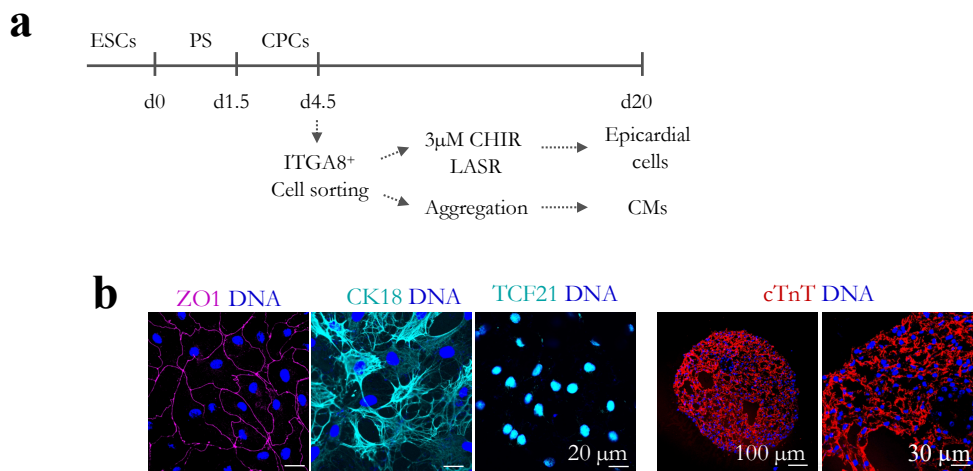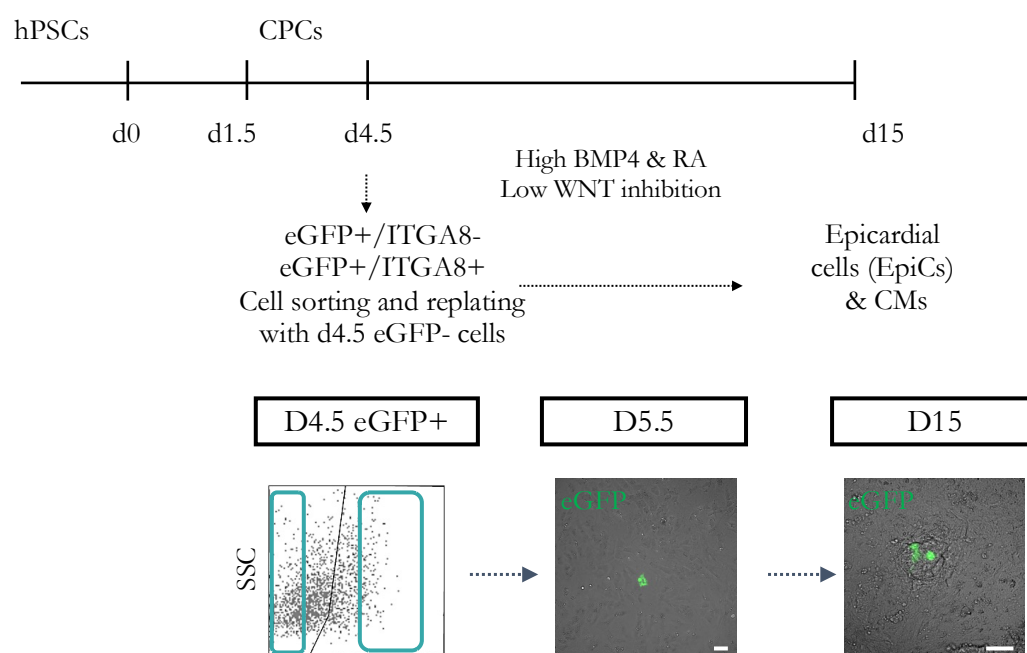
Figure 42: ITGA8 expressing cells differentiate towards epicardial and myocardial lineages.

(a) Schematic representation of FACS sorting followed by epicardial and myocardial differentiation. **(b)** Representative images of cells expressing epicardial (ZO1, CK18 and TCF21) and myocardial (TNNT2) markers 15 days after replating / reaggregation of FACS sorted progenitors for ITGA8 at day 4.5 of RA differentiation. Adapted from Zawada et al., 2022 (preprint).

Genetic clonal analysis performed in mouse by Zhang et al. (2021) revealed the multipotentiality of Hand1+ cardiac progenitors which likely overlap with JCF progenitors

described by Tyser et al., (2021). To explore the possible bipotentiality of human JCF-like cells, hiPSCs constitutively expressing *eGFP* and unlabeled hiPSCs were differentiated in the presence of RA. At day 4.5, eGFP+ hiPSC-derived progenitors were sorted for ITGA8, resulting in eGFP+ITGA8+ and eGFP+ITGA8- populations that were subsequently mixed with eGFP- cells. Approximately 5 eGFP+ cells of each fraction were re-seeded per well of a 12 well chamber slide, mixed with 20,000 eGFP- cells acting as a support for survival and differentiation. Cells were then differentiated for another 10 days in culture conditions that allow for simultaneous generation of cardiomyocytes and epicardial cells (Fig. 43). Modulations of BMP, RA and WNT signaling cues were described previously to favor either myocardial or epicardial lineages (Wiesinger et al., 2022), therefore we decided to use low levels of WNT signaling inhibition with concomitant addition of BMP4 and RA for our experiment.

Figure 43: ITGA8 expressing cells differentiate towards epicardial and myocardial lineages.

Schematic representation of FACS sorting of eGFP+/ITGA8+ and eGFP+/ITGA8- followed by simultaneous differentiation towards epicardial and myocardial cells. Scale bar = 50 μm. hPSCs: human pluripotent stem cells; CPCs: cardiac progenitor cells; CMs: cardiomyocytes;. Adapted from Zawada et al., 2022 (preprint).

Analysis of eGFP+ clusters revealed that eGFP+/ITGA8+ cells were about twice as prone to give rise to clusters built only from epicardial cells as eGFP+/ITGA8- cells, though they also maintained the ability to differentiate towards cardiomyocytes (Fig. 44a-b). Most of the eGFP+/ITGA8+ or eGFP+/ITGA8- cells differentiated into a single lineage. However, around 10% of eGFP+/ITGA8+ cells gave rise to clusters consisting of both epicardial cells and cardiomyocytes (Fig. 44c). This data provides further proof that JCF-like cells of bipotential capacity also exist during human cardiogenesis *in vitro* and can be enriched using ITGA8 as a cell surface marker.
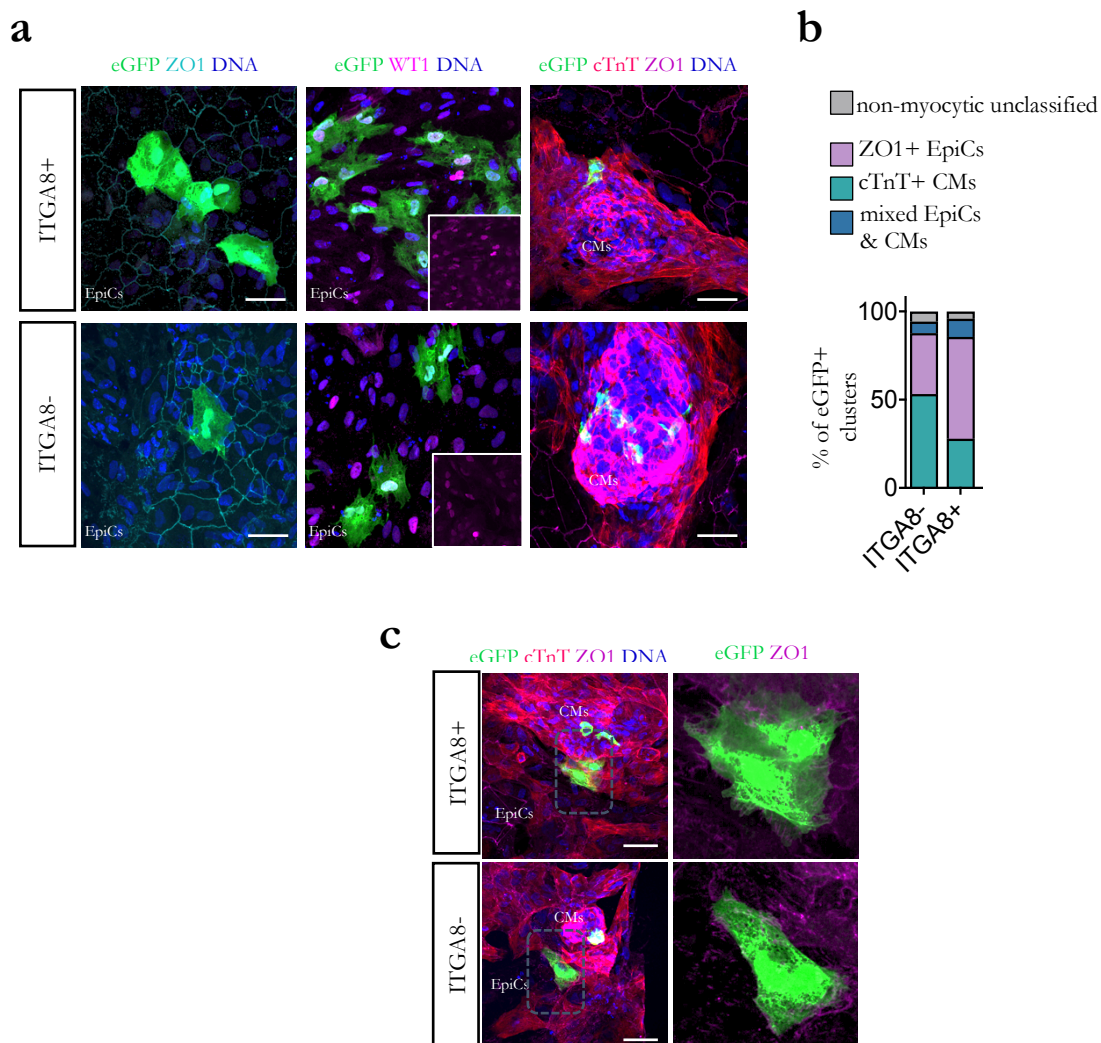
Figure 44: Human JCF-like cells form uni- and bi-lineage clusters of epicardial cells and cardiomyocytes.

(a) Representative images of cells stained for epicardial (ZO1, WT1) and myocardial (cTnT) markers 10 days after replating of FACS sorted progenitors for eGFP+/ITGA8+ and eGFP+/ITGA8- at day

4.5 of RA differentiation with unsorted eGFP- cells. EpiCs: epicardial cells; CMs: cardiomyocytes. Size bar = 50 μm. **(b)** Quantification of eGFP+ cell clusters stained for epicardial marker ZO1, myocardial marker cTnT, both or none of those markers 10 days after replating eGFP+/ITGA8+ and eGFP+/ITGA8- cells. n = 3 differentiations, for eGFP+/ITGA8+ N = 87 clusters; for eGFP+/ITGA8- N = 130 clusters. EpiCs: epicardial cells; CMs: cardiomyocytes. **(c)** EpiCs: epicardial cells; CMs: cardiomyocytes **(c)** Representative images of clusters of cells built from cells stained for epicardial (ZO1) and myocardial (cTnT) markers 10 days after replating of FACS sorted progenitors for eGFP+/ITGA8+ and eGFP+/ITGA8- at day 4.5 of RA differentiation with unsorted eGFP- cells. EpiCs: epicardial cells; CMs: cardiomyocytes. Size bar = 50 μm. Adapted from Zawada et al., 2022 (preprint).

3.3.2 The key JCF marker *MAB21L2* as a putative direct target of RA signalling.

To explore whether RA can directly induce the expression of JCF-associated genes, promoters of several well-known markers of the JCF (*HAND1*, *MAB21L2*, *HOXB6*, *BNC2*) were analysed for the presence of sequences that could be bound by RA receptors (RAR/RXR), known as RA response elements (RAREs). Specifically, sequences located ±5 kb from transcription start sites (TSS) were bioinformatically investigated for the presence of direct repeats (DR). Ten different variations of sequences were considered. This uncovered putative repeats in the promoter sequence of all tested genes (Fig. 45a).
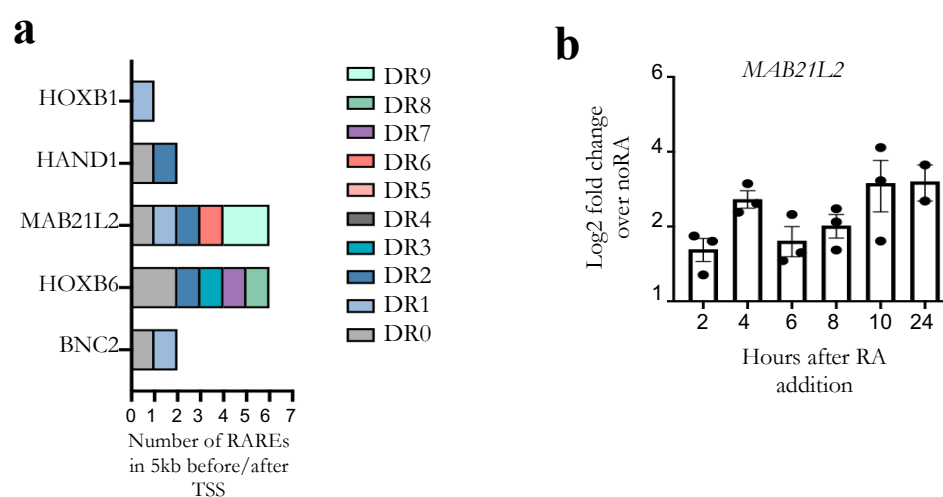

Figure 45: *MAB21L2* **is a putative direct target of RA signalling.**

(a) Number of putative RA response elements binding sequences repeats within 5kb before transcription start site (TSS) of tested genes. DR: direct repeats; RAREs: retinoic acid response elements. **(b)** Analysis of mRNA expression of *MAB21L2* between 42h and 64h during RA differentiation. mRNA expression relative to *GAPDH,* represent fold change over noRA sample collected at the same timepoint and displayed as log2. Data are mean ± SEM; n = 3 differentiations/time point.

One of the genes with the biggest number of putative RA receptors binding sites, *MAB21L2* (6 predicted RAREs), which was described as the key marker of the murine JCF, was chosen to perform an analysis of the functionality of the uncovered elements. Validation of *MAB21L2* regulation by RA was performed by collecting RNA between 2 and 24 hours after RA exposure for RT-qPCR. Analysis revealed upregulation of *MAB21L2* within 24 hours of RA exposure when compared with noRA treated samples, suggesting that *MAB21L2* could be a potential direct target of RA signalling (Fig. 45b). Additional analyses would be necessary to determine whether RA is indeed bound to RAREs sequence associated with *MAB21L2* promoter at this time of differentiation, but these data provide a potential explanation for the emergence of JCF progenitors in the differentiation with RA.

3.4 Utility of *in vitro*-obtained defined progenitors and their derivatives in studying developmental defects

As CHDs often affect derivatives of a specific pool of progenitors, *in vitro* modelling using a differentiation protocol that does not generate defined precursors can produce confounding results of limited translational value. To test the pathological relevance of our approach, the two versions of our protocol were applied to a hiPSC line derived from a patient with hypoplastic left heart syndrome (HLHS), one of the most complex CHDs characterised predominantly by an underdeveloped left ventricle (Krane et al., 2021).

HLHS hiPSC-derived cardiomyocytes obtained using the protocol that enriches for FHF-derivatives (RA treatment) displayed lower levels of transcripts related to the ventricular cardiomyocyte fate (*MYH7, MYL2, FHL2*) compared to healthy hiPSC-derived cardiomyocytes on day 30 (Fig. 46).
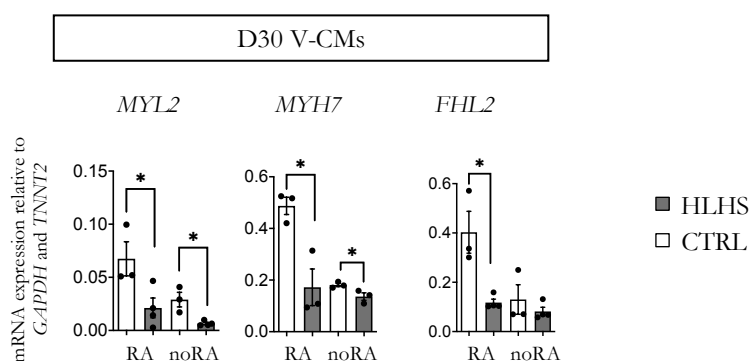
Figure 46: HLHS-CMs show reduced expression of ventricular cardiomyocyte markers.

Analysis of mRNA expression of ventricular CM markers at day 30 of RA differentiation of HLHS-patient-specific hiPSCs (HLHS) and healthy control hiPSCs. mRNA expression relative to *GAPDH* and *TNNT2*. Data are mean ± SEM; n = 3 differentiation per time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed t-test).

Furthermore, they exhibited significantly higher expression levels of genes that are specific for cardiomyocytes of aSHF-origin (*SOX4*, *RSPO3*; Fig. 47), which are normally derived in the absence of RA. This could suggest that HLHS cells fail to respond to cues that guide FHF specification during *in vitro* differentiation.
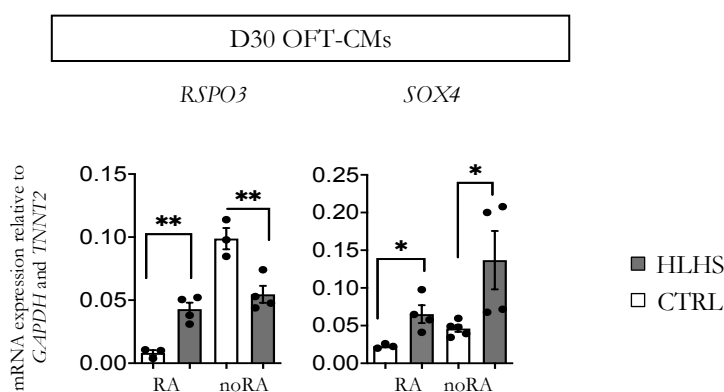
Figure 47: HLHS-CMs show aberrant expression of aSHF-CMs markers.

Analysis of mRNA expression of aSHF-CMs-enriched markers at day 30 of RA and noRA differentiation of HLHS-patient-specific hiPSCs (HLHS) and healthy hiPSCs. mRNA expression

relative to *GAPDH* and *TNNT2*. Data are mean ± SEM; n = 3 - 5 differentiations per time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed t-test).

In addition, in line with previous studies, the gene expression profile of HLHS hiPSC-derived cardiomyocytes exhibited a more immature phenotype as shown by expression ratio of genes prevalent at fetal stages of cardiac development *(TNNI1* and *MYL7)* and in adult hearts *(TNNI3* and *MYL2)*, regardless of the protocol version (Fig. 48; Krane et al., 2021).
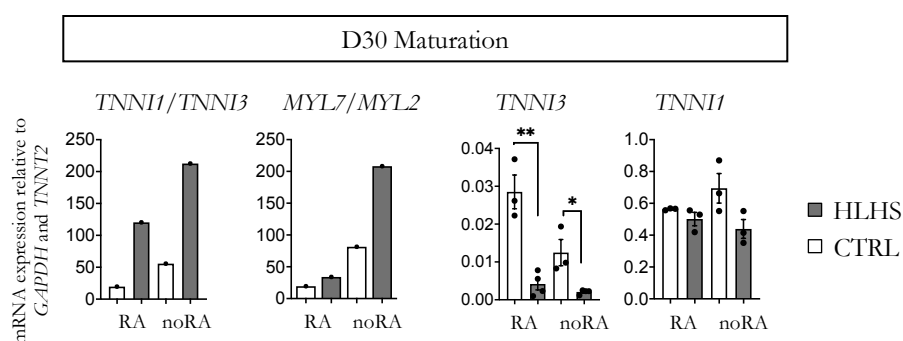
Figure 48: HLHS-CMs show a maturation defect.

Ratio between mRNA expression levels of *TNNI1* and *TNNI3* as well as *MYL7* and *MYL2* relative to *GAPDH* and *TNNT2* and analysis of mRNA expression of maturation markers at day 30 of RA and noRA differentiation of HLHS-patient-specific hiPSC line (HLHS) and healthy person derived hiPSC line (previously referred to as hiPSC and here as CTRL). mRNA expression relative to *GAPDH* and *TNNT2*. Data are mean ± SEM; n = 3 differentiation per time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed t-test).

As FHF and aSHF derivatives are differently affected in HLHS, the approach presented here shows how these differences can be better appreciated when applying differentiation protocols that enrich for either FHF or aSHF derivatives. Though limited in scope, this analysis underlines the importance of applying a well-characterised differentiation protocol that ensures generation of specific cell subtypes to correctly assess the mechanism of disease.

3.5 lncRNAs in human cardiac differentiation

3.5.1 Distinct lncRNAs characterise human heart field-like progenitors

Mounting evidence shows that lncRNAs are more specific for particular cell types and states than protein-coding genes. Thus, lncRNAs profiling provides a novel opportunity for better capturing elusive cell types. In order to discover human heart field-specific lncRNAs, different progenitor populations were sorted and subjected to deep RNA-seq profiling. Based on previous analyses on heart field identity, noRA-N+ and RA-T+/N+ populations were chosen and sorting was performed on days 5.5 and 6.5.

Differential expression gene (DEG) analysis confirmed the aSHF and FHF nature of sorted progenitors (noRA-N+ and RA-T+/N+, respectively). This analysis recovered 3213 genes on day 5.5 (Fig. 49a) and 593 on day 6.5 (Fig. 50a), fitting the notion that cardiac progenitors differ more at early stages and, upon myocardial commitment, the gene profile of the two populations becomes more similar (Tyser et al., 2021).

Gene Ontology analysis of the DEGs of the noRA-N+ sample collected on day 5.5 produced categories in line with their aSHF-like characteristics, such as *Extracellular matrix regulation of vasculature development*, *Response to BMP*, *Muscle cell proliferation*, *Cardiac chamber development*, *Dorsal-ventral pattern formation*, *Heart valve development*, *Neural crest cell differentiation* and *Endocardial cushion morphogenesis* (Fig. 49b). Some of these categories persisted into myocardial specification stage - *Heart morphogenesis*, *Response to BMP*, *Muscle cell proliferation* and *Heart valve development*. On the other hand, gene ontology analysis of DEGs associated with the RA-T+/N+ population sorted at day 5.5 resulted in categories indicative of their FHF-like identity, including *Cardiac muscle contraction*, *Anterior-posterior pattern specification*, *Regulation of heart rate*, *Cardiac ventricle morphogenesis*, *Heart growth*, *Retinoic acid receptor signalling pathway*, *Cardiac atrium development*, and *Cardiac left ventricle morphogenesiss* (Fig. 50b).

Analysis of differentially expressed lncRNAs on day 5.5 recovered 427 lncRNAs that differed between noRA-N+ and RA-T+/N+ populations (of which 234 were downregulated in noRA N+ versus RA-T+/N+ sample while 193 were upregulated, and of which 272 were still unannotated; Fig. 49c). Among lncRNAs specific for noRA-N+, a number of lncRNAs paired with genes that are important for aSHF specification were discovered, such as *ISL1-DT* (ISL1-divergent transcript); *FGF10-AS1* (*FGF10-antisense*), *GATA3-AS1*, *NKX2.1-AS1* as well as *EVX1-AS,* which mediates mesendodermal lineage progression and has been reported to

be induced by BMP4 (Bell et al., 2016; Luo et al., 2016). On the other hand, among lncRNAs that were downregulated in the noRA-N+ sample vs RA-T+/N+, lncRNAs associated with genes important for the FHF lineage (*THBS4-AS1*, *TBX5-AS1*) as well as anterior-posterior patterning (*HOXB-AS1*) and RA signalling (*RARA-AS1*) were detected. Correlated expression patterns of detected lncRNAs with their corresponding coding genes suggested potential regulation in *cis*, i.e., regulation the transcription of nearby genes (Anderson et al., 2016, Daneshvar et al., 2016, Engreitz et al., 2016).
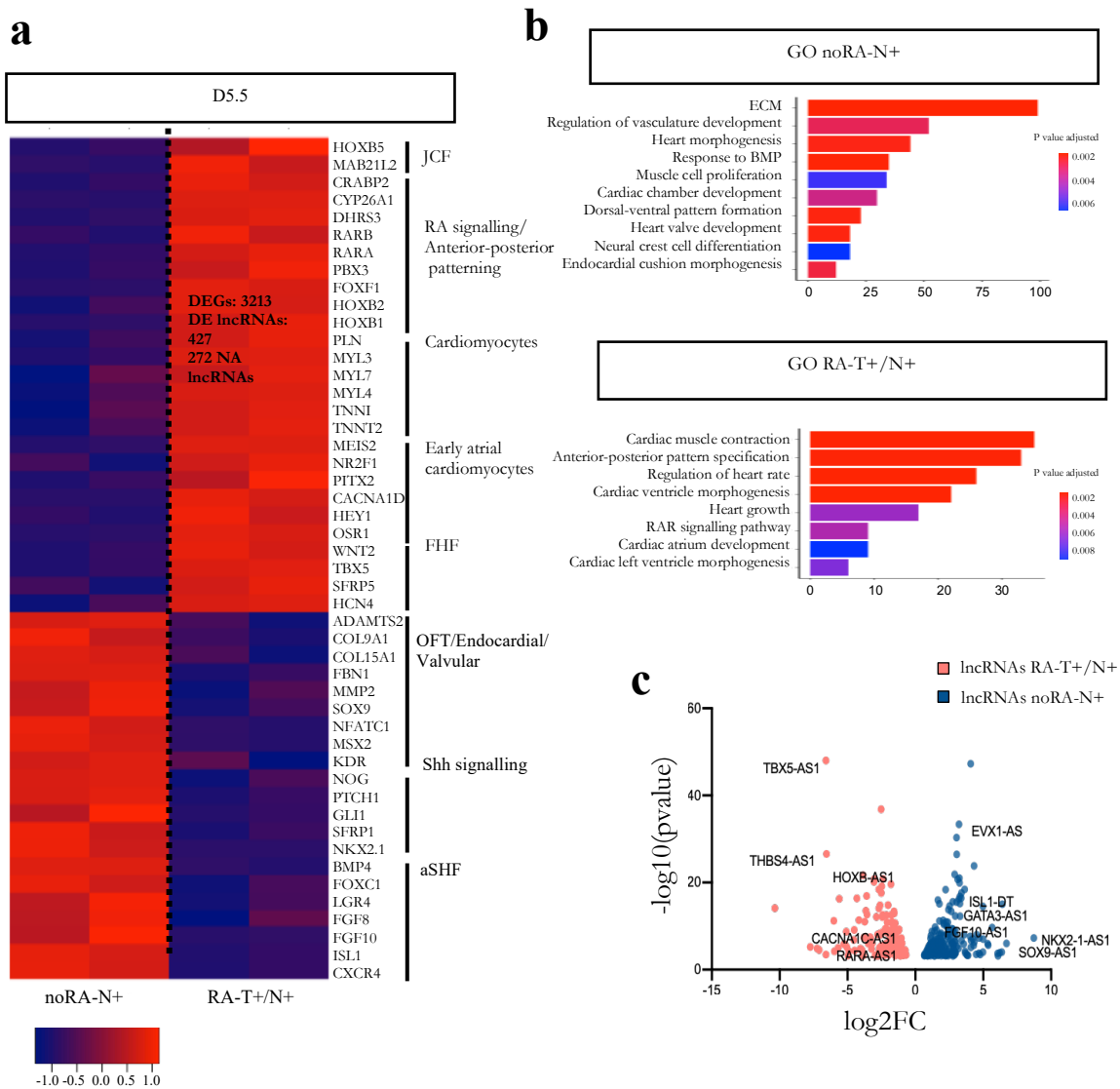
Figure 49: Deep RNA-Seq at early progenitor stage (day 5.5) allows for detection of heart-field-like specific lncRNAs.

(a) Heatmap of key markers selected from differentially expressed genes of FACS-sorted cells expressing eGFP (NKX2.5) of noRA differentiation (noRA-N+) and mCherry (TBX5) eGFP

(NKX2.5) of RA differentiation (RA-T+/N+) both captured at day 5.5. DEGs analysis of noRA-N+ vs RA-T+/N+ - values represent log fold change. JCF: juxtacardiac field; FHF: first heart field; pSHF: posterior second heart field; aSHF: anterior second heart field; Shh: sonic hedgehog; OFT: outflow-tract. **(b)** Selected Gene Ontology categories among top 100 of FACS-sorted cells expressing eGFP (NKX2.5) of noRA differentiation (noRA-N+) and mCherry (TBX5)/eGFP (NKX2.5) of RA differentiation (RA-T+/N+) both captured at day 5.5 - values represent log fold change. ECM: extracellular matrix. RAR: retinoic acid receptor **(c)** Volcano plot showing differentially expressed lncRNAs between noRA-N+ and RA-T+/N+. N = NKX2.5; T = TBX5.

Upon entering myocardial commitment on day 6.5, a much smaller number of differentially expressed lncRNAs were recovered (294, among which 193 were unannotated; Fig. 50c). Within upregulated annotated lncRNAs of the noRA-N+ population, 41 were maintained from day 5.5, while 8 emerged on day 6.5; and 48 were not differentially upregulated anymore, including *LHX1-DT*. LHX1 is transiently expressed in the nascent mesoderm and progenies of LHX1+ contribute exclusively to the cardiac and head mesoderm (Costello et al. 2011). Thus, downregulation of *LHX1-DT* from day 5.5 to day 6.5 is indicative of differentiation progression. Within downregulated lncRNAs, 40 were maintained while 26 were not expressed anymore and 12 appeared at day 6.5, including *IGF2-AS*, associated with a gene playing an important role in the thickening of the ventricular wall and cardiomyocyte proliferation (Liu et al., 1996).
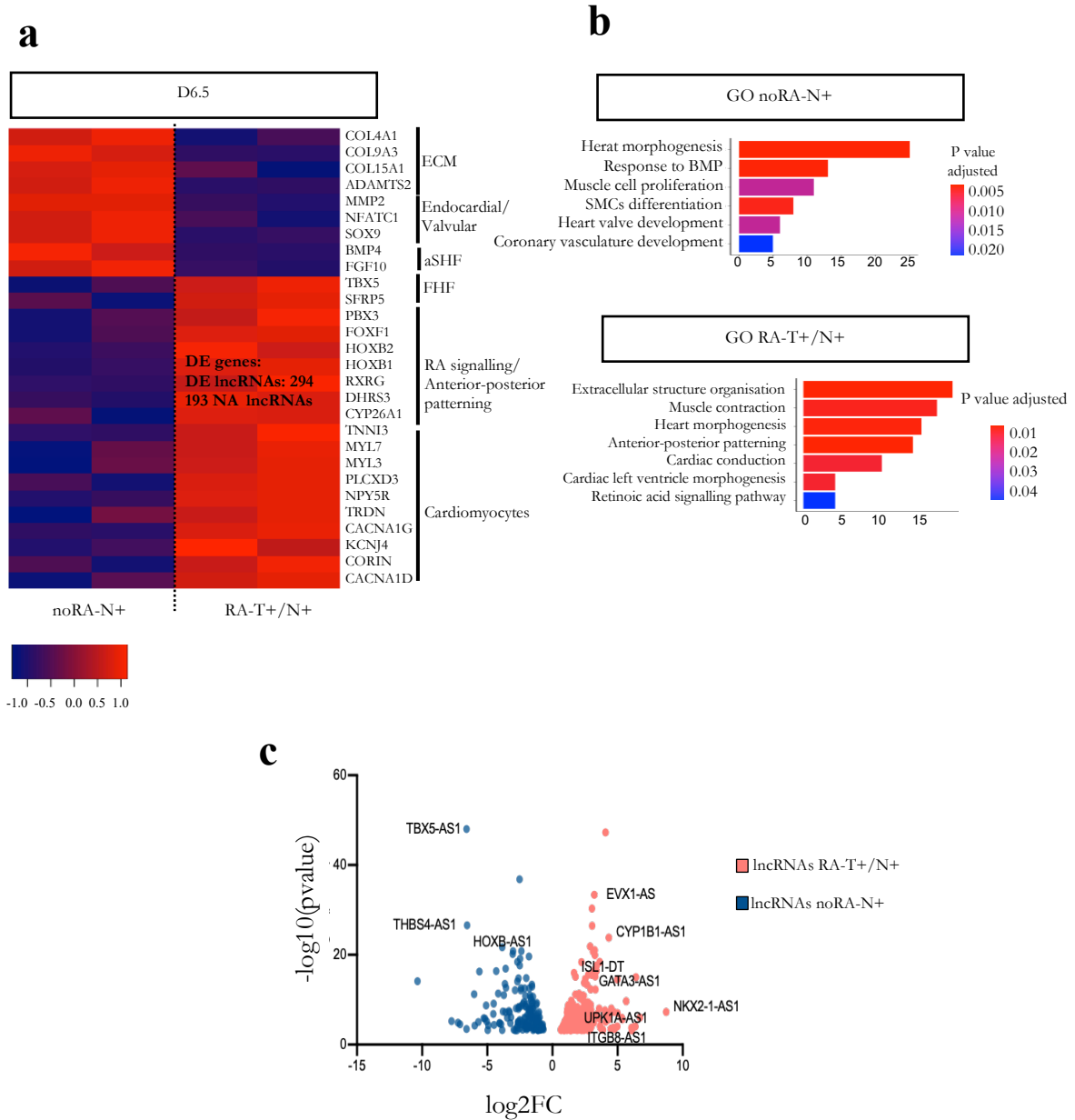
a

b

c

Figure 50: Deep RNA-Seq at late cardiac progenitor/early cardiomyocyte stage (d6.5) allows for detection of heart-field-like specific lncRNAs.

(a) Heatmap of key markers selected from differentially expressed genes of FACS-sorted cells expressing eGFP (NKX2.5) of noRA differentiation (noRA-N+) and mCherry (TBX5) eGFP (NKX2.5) of RA differentiation (RA-T+/N+) both captured at day 6.5. DEGs analysis of noRA-N+ vs RA-T+/N+ - values represent log fold change. ECM: extracellular matrix; FHF: first heart field; aSHF: anterior second heart field; OFT: outflow-tract. **(b)** Selected Gene Ontology categories among top 100 of FACS-sorted cells expressing eGFP (NKX2.5) of noRA differentiation (noRA-N+) and mCherry (TBX5) eGFP (NKX2.5) of RA differentiation (RA-T+/N+) both captured at day 6.5 - values

represent log fold change. SMCs: smooth muscle cells. **(c)** Volcano plot showing differentially expressed lncRNAs between noRA-N+ and RA-T+/N+. N = NK2.5; T = TBX5.

3.5.2 Relevance of the lncRNA Handsdown during cardiac differentiation

The lncRNA Handsdown (Hdn) has been reported to be essential for mouse development (Ritter et al., 2019). Specifically, homozygous *Hdn* mutants die embryonically, while heterozygous mutants develop into adults but exhibit hyperplasia of the right ventricular wall, a derivative of the aSHF. Therefore, it could be expected that FHF- and aSHF-related cardiac programs are also dysregulated in human *HDN* mutant cardiac cells during development. So far, the *HDN* locus has not been described to be functionally relevant in humans, though we identified *HDN* as a conserved lncRNA locus between mice and humans (Fig. 51).

To investigate whether the HDN locus plays a similar role in humans as in mice, a part of the HDN promoter was deleted from a hiPSC line using CRISPR/Cas9 genome editing. Specifically, a ~5,6 kb fragment was deleted which encompasses a putative transcriptional start site (TSS) of HDN, located approximately 6,6 bp downstream of the HAND2 mRNA sequence and around 700 bp upstream of the putative HDN sequence (Fig. 52a, 52c). Correctly targeted clones carrying the deletion were identified by PCR and validated by sequencing (Fig. 52d). The validation primers either encompassed the deleted fragment, resulting in an amplicon smaller than 600bp in the case of correct removal of part of the HDN promoter (PCR 3), or were nested inside of the deleted fragment and just upstream of the guide, resulting in an approximately 1kb amplicon in the case of WT alleles (PCR 1). Karyotyping of the generated knock-out cell line confirmed the absence of chromosomal abnormalities (Fig. 52b).
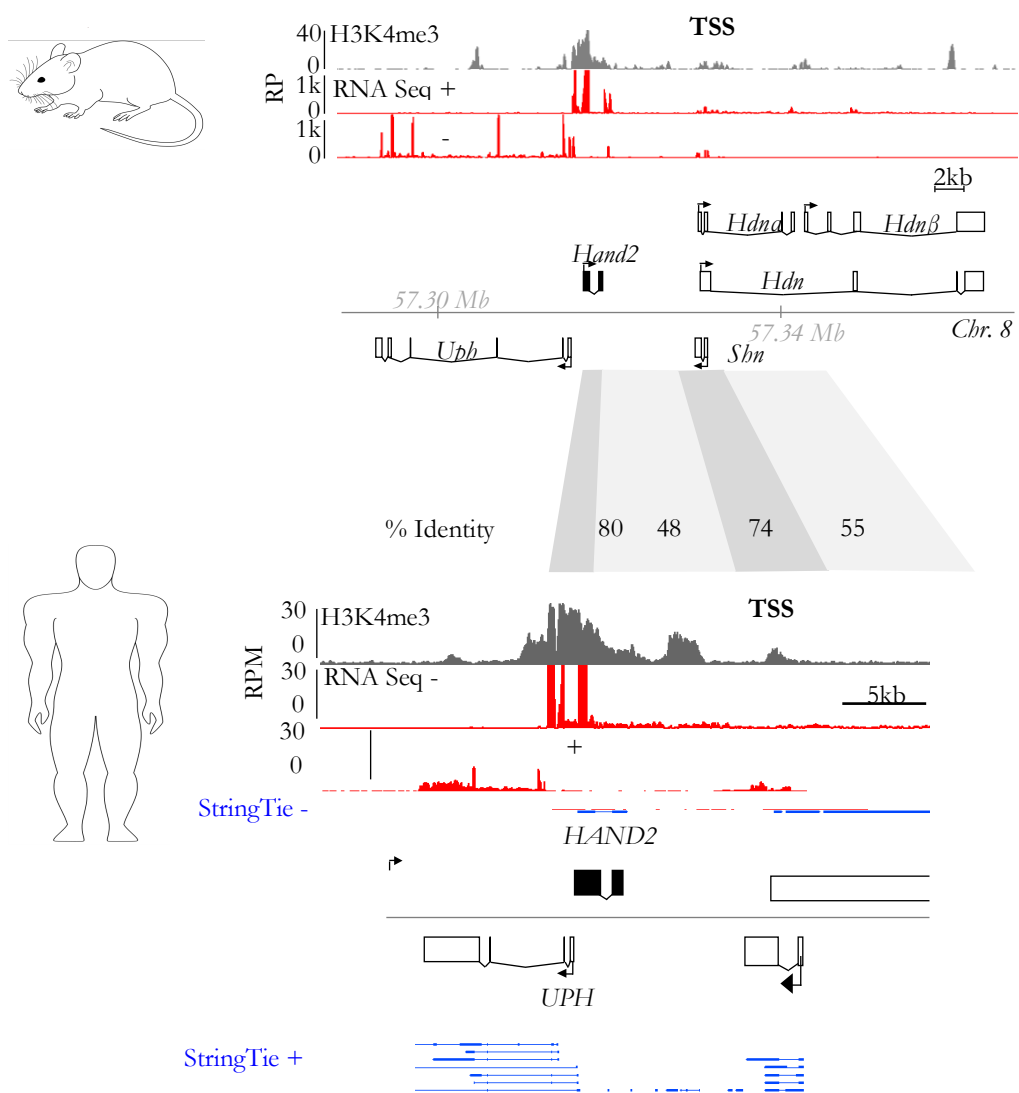
Figure 51: The Handsdown Transcriptional Start Site is highly conserved between mice and humans.

Comparison of the mouse and human Handsdown loci. Hdn = Handsdown, Uph = Upperhand, Shn = Sweetheart, TSS = Transcriptional Start Site. Schematics prepared in collaboration with Dr. Philip Grote.
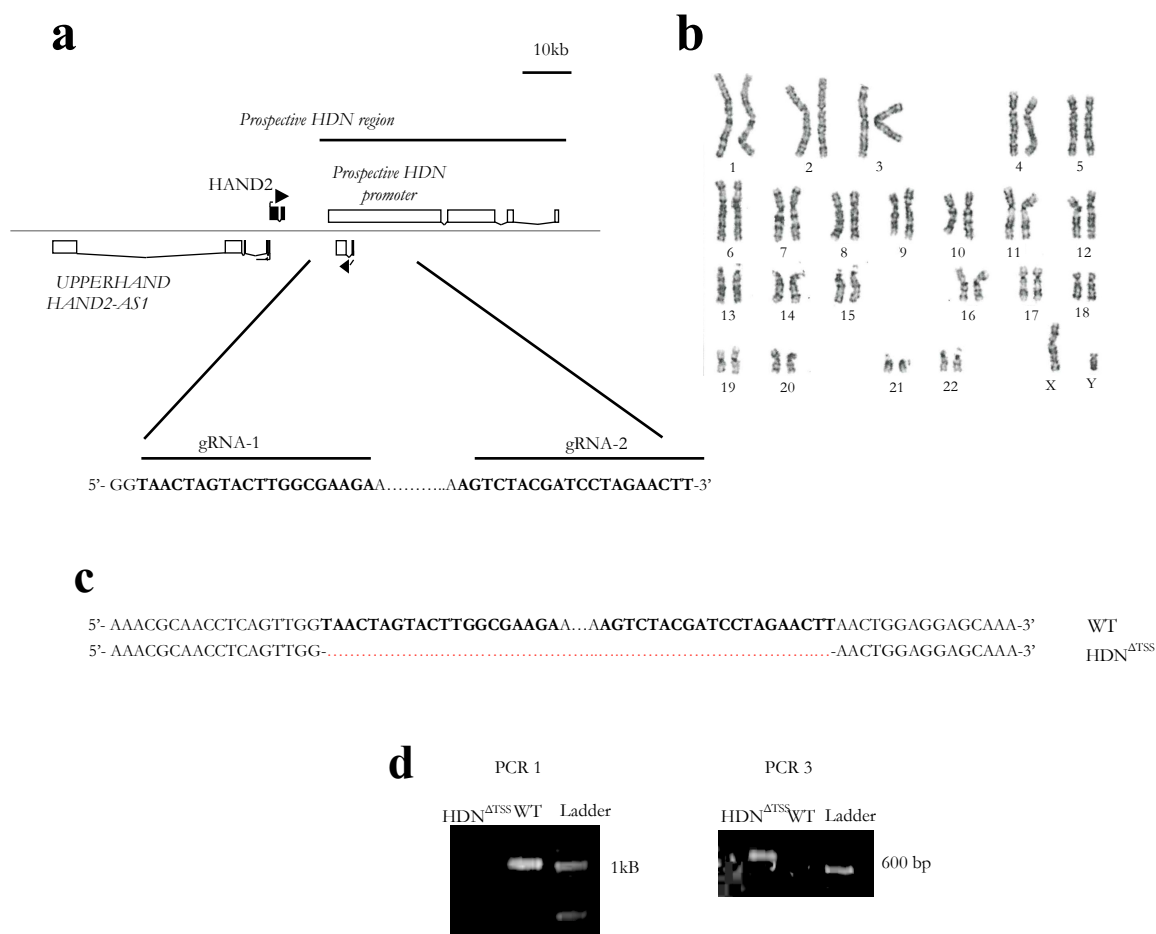
a

10kb

Prospective HDN region

HAND2

Prospective HDN promoter

UPPERHAND HAND2-AS1

gRNA-1

gRNA-2

5'- GG**TAACTAGTACTTGGCGAAGA**A...........A**AGTCTACGATCCTAGAACTT**-3'

b

c

5'- AAACGCAACCTCAGTTGG**TAACTAGTACTTGGCGAAGA**A...A**AGTCTACGATCCTAGAACTT**AACTGGAGGAGCAAA-3'   WT

5'- AAACGCAACCTCAGTTGG-...-AACTGGAGGAGCAAA-3'   HDN$^{\Delta TSS}$

d   PCR 1   PCR 3

HDN$^{\Delta TSS}$ WT   Ladder   HDN$^{\Delta TSS}$ WT   Ladder

1kB   600 bp

Figure 52: Deletion of the Transcriptional Start Site (TSS) of the lncRNA Handsdown using genome editing by CRISPR/Cas9.

(a) Schematic diagram of Handsdown coding region and the targeting loci of gRNA. **(b)** Correct karyotype of HDN$^{\Delta TSS}$ cell line. **(c)** Sequencing results of wild type (WT) and HDN$^{\Delta TSS}$ cell lines. **(d)** DNA electrophoresis of PCR products confirming successful deletion of HDN TSS. TSS: transcriptional start site; WT: wild type; HDN: *HANDSDOWN*; gRNA: guide RNA.

Then, to determine the extent and nature of dysregulated genes the edited and isogenic control lines were subjected to the two versions of the differentiation protocol previously described to derive early human heart field-specific cells. In mice, loss of *Hdn* causes an upregulation of its *cis*-located gene *Hand2* and other core cardiac transcription factors (*NKX2.5*, *HAND1*) (Ritter et al., 2019). These findings are in line with the notion that the *Hdn* region is a negative regulator of *Hand2* and the other genes. However, in our human system, RT-qPCR analysis did not reveal any significant differences in expression of the genes affected in mice (*NKX2.5*, *HAND1*, *HAND2*, *TNNT2*) between control and *HDN* mutant lines, neither at

early stages of differentiation between days 1 and 8 nor on day 30 (Fig. 53a-b). Analysis of day 30 cardiomyocytes did not detect any significant differences in markers of the ventricles (*MYL2*) or OFT (*RSPO3*) – structures affected in *Hdn* mutant mice. Though the importance of the HDN TSS in the regulation of human HDN and on human cardiogenesis cannot be excluded by the presented analysis, the identified TSS of HDN or HDN itself might play different roles in humans and mice. These results highlight the complexity of mammalian development, specifically, differences between species (rodents vs primates).

a [a]

b
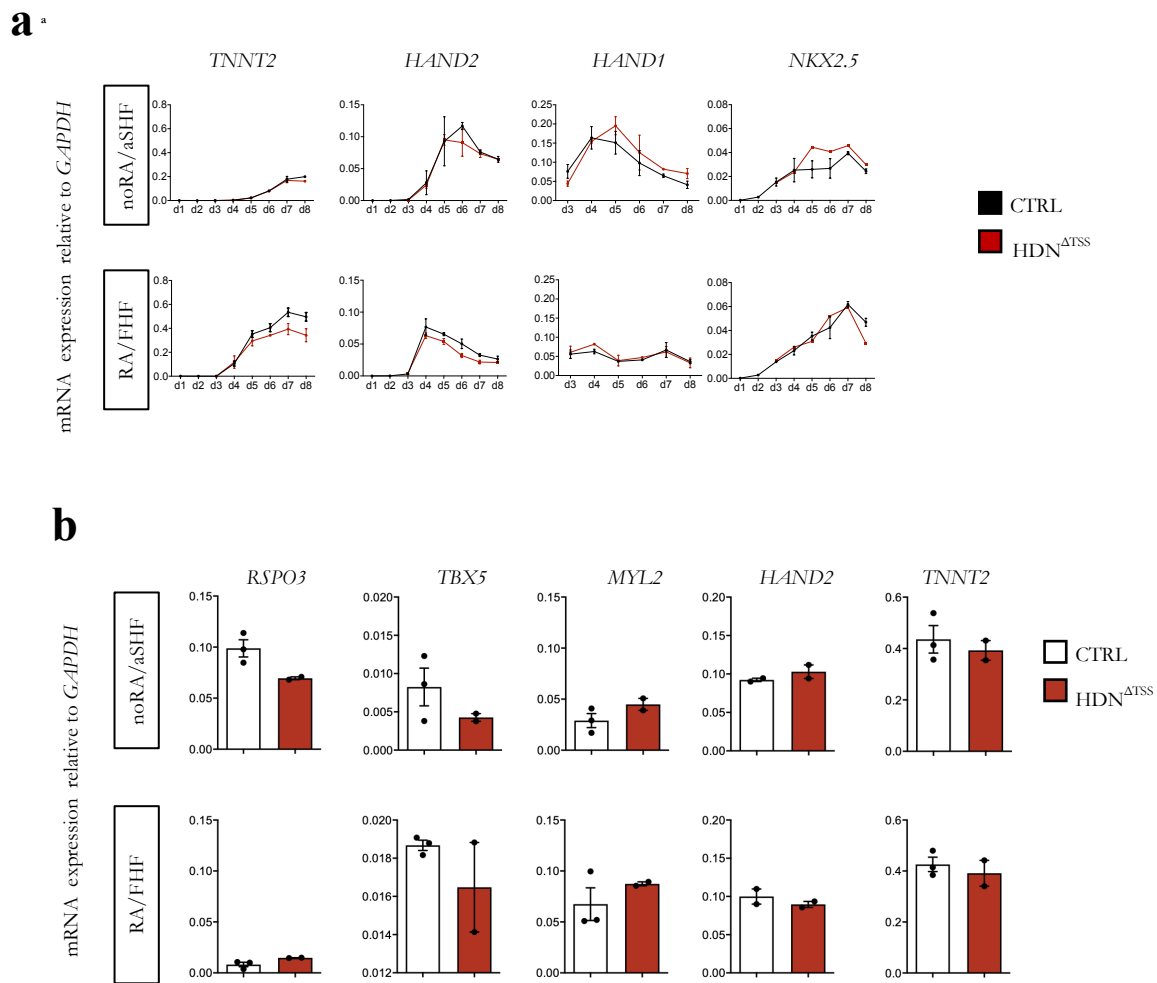
Figure 53: Comparison of the extent and nature of dysregulated genes between human wild type and HDN$^{\Delta TSS}$.

(a) Time course of mRNA expression of genes affected in mice by Hdn knockout between d1 and d8 during RA and noRA differentiation of HDN$^{\Delta TSS}$ cell line and parental cell line. mRNA expression relative to *GAPDH*. Data are mean ± SEM; n =2 differentiations/time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed *t*-test). **(b)** RT-qPCR analysis of d30 of RA and noRA differentiation

of HDN$^{\Delta TSS}$ cell line and parental cell line. mRNA expression relative to *TNNT2* and/or *GAPDH*. Data are mean ± SEM; n = 2 - 3 differentiations/time point; *p<0.05, **p<0.005, ***p<0.001 (unpaired two-tailed *t*-test).

4 Discussion

CHDs are one of the predominant causes of infant mortality in developed countries (reviewed by Roth et al., 2020; Hoffman & Kaplan, 2002). In the last years, advances in pediatric cardiology and heart surgery have resulted in increased survival of the children with CHDs, which led to the increase in the number of adults suffering from cardiac birth defects (Knowles et al., 2012). This has driven research to elucidate the mechanisms of CHDs and discover novel therapeutic strategies. As certain congenital heart malformations affect heart structures that originate from specific pools of cardiac progenitors (Ward et al., 2005; Crucean et al., 2017; Krane et al., 2021), a comprehensive understanding of the molecular mechanisms governing early human cardiac progenitor specification could shed light on the causes of these diseases. The difficulty to access human foetal tissue at the onset of cardiogenesis has shifted emphasis to other human-derived models, mainly *in vitro* differentiated cells.

Here, we presented protocols to direct hPSCs towards the cardiac lineage to obtain various populations of mesodermal heart field-like progenitors and their definitive progenies. Using genetic labelling, cell sorting as well as bulk and single-cell transcriptional analysis, we thoroughly characterized the transcriptional profile and potentiality of these populations. Our analyses allowed us to explore the heterogeneity of human heart field-like progenitors and identify human progenitors resembling the JCF for the first time. We were able to isolate JCF-like cells from hPSCs-based differentiation and showed that, similarly to their mouse counterpart, they can give rise to cardiomyocytes as well epicardial cells. Additionally, we showed that key markers of JCF could be potentially directly regulated by RA signalling through RA response elements. To explore the utility of our differentiation approaches in studying developmental defects, we applied the established protocols to hiPSCs derived from a patient with HLHS. Though limited in scope, this analysis showcased how fate-committed cardiac progenitors can increase the resolution of *in vitro* disease models and provide additional insights into human CHDs. Finally, we also analysed lncRNAs signatures of the different heart field-like progenitors. This showed that the deletion of a promoter region consisting of the transcriptional start site of the lncRNA HDN – essential for heart development in mice – does not affect cardiomyocyte formation *in vitro,* highlighting the importance of human-derived models.

4.1 Application of fluorescent reporters to capture human heart-field-like cell types *in vitro*

HiPSCs- and hESCs-derived cell types are accessible sources to explore early human development and disease *in vitro*. To allow for targeted mechanistic studies, specific hPSC-derived populations can be obtained through time-specific exposure to signalling cues, followed by sorting cells based on a specific marker – either a surface marker or one driving the expression of a fluorescent protein. Fluorescent reporters are also useful tools to identify and track specific cell populations during *in vitro* differentiation of hPSCs into cardiomyocytes (Elliott et al., 2011; Dubois et al., 2011); to obtain pure populations of cardiomyocyte subtypes for drug testing (Gunawan et al., 2021; Lauschke et al., 2021) or for tracking cardiomyocytes after transplantation (Poch et al., 2022). In this study, we used a double reporter cell line engineered through CRISPR/Cas9-mediated homologous recombination resulting in one allele of TBX5 and NKX2.5 being replaced with the sequence encoding for the fluorescent proteins mCherry and eGFP, respectively. This allowed us to observe the endogenous gene reporter expression. There are several limitations of this strategy, including reporting efficiency dependent on a locus, and disruption of one gene allele, possibly affecting differentiation and cellular functions. However, in the studied experimental conditions, we did not observe major differences between this haploinsufficient line and several other hPSCs lines.

Recently, Zhang et al., (2019) and Pezhouman et al. (2021) employed a similar hPSC double reporter system consisting of fluorescent proteins under the control of *TBX5* and *NKX2.5* promoters to isolate early cardiomyocytes of both FHF- and SHF-like. Both above-mentioned works followed one simple protocol that rely on temporal activation and inhibition of WNT signalling (Lian et al., 2012). Specifically, Zhang et al., (2019), sorted cells at days 10 and 35 of differentiation, and Pezhouman et al. (2021) at days 10 and 20, and assigned to four discrete populations identities of FHF-ventricular cardiomyocytes (NKX2.5+/TBX5+), SHF-atrial cardiomyocytes (NKX2.5+), epicardial / nodal cells (TBX5+) and other cell types, including endothelial cells (NKX2.5-/TBX5-). Contrary to their work, we observed that various populations expressing TBX5 and/or NKX2.5 can be dissected much earlier, before the onset of beating, between days 3 and 5 of RA/noRA differentiations. This is important, as during later stages of human development SHF-cardiomyocytes (RV-, atrial cardiomyocytes) also upregulate TBX5 (Hatcher et al., 2020). Many markers of cellular states and types are context and time-specific, and thus relying on a small set of markers for cell identification and isolation

may be misleading. Ultimately, though reporter cell lines are useful tools to track and isolate FHF- and SHF-like cardiomyocytes, caution should be taken when classifying the obtained pools of cells. This is why, to study human development *in vitro*, fluorescent reporter cell lines should be combined with protocols that direct differentiation through precisely defined stages into specific cell types as showed in here presented work.

4.2 Retinoic acid signalling as an important cue driving early cardiac progenitor lineage segregation *in vitro*

RA signalling is one of the major developmental cues by which cells determine their positional identity within various structures during embryogenesis. It plays a pivotal role in the development of many organs including the brain, limbs, eye, spinal cord, pancreas, and heart (reviewed by Niederreither and Dolle, 2008 and Duester, 2008). Misdistribution or lack of enzymes involved in both RA synthesis and degradation have deleterious effects on developing embryos (Niederreither et al., 1999; Abu-Abed et al., 2001). During development, RA signalling influences cardiogenesis at many different stages, from the heart tube through looping, heart chambers specification, septation of chambers and the OFT, to the trabeculation of ventricles (reviewed by Stefanovic and Zaffran, 2016). As studies in animal models reported the importance of RA in maintaining the pSHF domain giving rise to the atria (Hochgreb et al., 2003), many groups tried to utilize different RA dosage regimes to facilitate atrial cardiomyocytes differentiation from hPSCs *in vitro*. The first report on RA usage comes from Zhang et al. (2011). They showed that administration of a high dosage of RA (1 μM) between days 6 and 8 of *in vitro* cardiac differentiation enhances the generation of atrial cardiomyocytes. In line with this, Devalla et al. (2015) reported that high RA concentrations (1 μM) between days 4 to 7 of cardiac differentiation led to upregulation of COUP-TFI *(NR2F1)* and COUP-TFII *(NR2F2)*, which are expressed in atrial cardiomyocytes. Interestingly, they also showed that RA at lower concentrations (1 – 10 nM) applied in the same time window enhances cardiac differentiation in general. The analysis presented in this thesis parallels these findings: in our system, higher concentration of RA (1 μM) resulted in the slight upregulation of atrial-associated markers, while low/middle concentrations of RA led to the generation of mainly ventricular cardiomyocytes with high efficiency. However, it is worth noting that the timing of

RA administration in the current study differs substantially, as in this work RA was applied much earlier than in previous studies, at the mid-anterior primitive streak/emergent mesoderm stage at day 1.5 of differentiation.

In mice, mesodermal progenitors that give rise to ventricular cardiomyocytes arise first, followed by progenitors that differentiate into atrial cardiomyocytes (Lescroart et al., 2014; Ivanovitch et al., 2021). Following the assumption that *in vitro* differentiations of hPSC progresses through the stages described *in vivo* in mice, Lee et al. (2017) explored how RA application could influence the early induction of different human mesodermal cells, and ultimately atrial and ventricular lineage segregation. Specifically, they used different ratios of Activin A/Nodal and BMP signals to enrich for distinct atrial and ventricular mesoderm populations. In their system, a relatively lower concentration of Activin A induced RALDH2-expressing mesodermal cells that were able to generate atrial cells upon RA or retinol (RAOH) addition, while a higher Activin A to BMP4 ratio supported the generation of CYP26A1/CD235A-expressing cells that differentiated towards ventricular cardiomyocytes. However, they showed that atrial cardiomyocytes could be also obtained from CYP26A1/CD235A-expressing cells by adding RA to culture media, though with suboptimal efficiency. We exposed cells to a similar high ratio of Activin A to BMP4, however with an additional addition of CHIR and bFGF as well as longer BMP4 exposure. We observed high expression of *CYP26A1* in day 1.5 early mesodermal cells, suggesting that we may have generated the mesoderm subtype that mainly gives rise to ventricular cardiomyocytes. In line with previous work, the application of RA induced some atrial cardiomyocytes generation, even from mesodermal cells patterned with a high Activin A to BMP4 ratio, but to a much smaller extent.

So far, RA has not been reported to directly play a role in FHF development in animal models. Studies in mouse embryos indicated that RA represses FGF signalling and activates expression of TBX5 (Sirbu et al., 2008), which leads to direct and indirect inhibition of some aSHF-associated markers (Sirbu et al., 2008; De Bono et al., 2018; Rankin et al., 2021). This provides a permissive environment for pSHF emergence, but potentially also FHF, both marked by expression of TBX5. Recently, RA was also used to obtain cardiac organoids of FHF origin (Hofbauer et al., 2021). They described the generation of cardiac organoids ("cardioids") with early left ventricular chamber-like identity using a differentiation protocol similar to the one applied here. Our study showed that in an *in vitro* system, RA influences the expression of many genes that are associated with the FHF lineage (*TBX5, WNT2, THBS4*) at

the expense of aSHF-related markers (*BMP4*, *ISL1*, *FGF10*, *LGR5*), which ultimately leads to robust ventricular differentiation. Surprisingly, we detected only a minor induction of pSHF and atrial cardiomyocyte markers. It would be of great interest to further dissect the mechanisms of RA driving lineage segregation of human heart field-like progenitors.

4.3 Translational relevance of the generation of human aSHF-like progenitors and their progenies

The vast majority of work related to modelling cardiovascular development *in vitro* focuses on ventricular or atrial cardiomyocyte subtypes or non-myocytic cells including cardiac fibroblasts and endothelial/endocardial cells. Little focus so far was drawn to OFT development and cell types building the OFT. Given that differences between human and murine OFT development exist (Sahara et al., 2019), human-derived models are a prerequisite to deepen our understanding of the molecular mechanisms underlying OFT morphogenesis. This is crucial as approximately 30% of congenital heart malformations are caused by defects during OFT development, specifically patterning of the anterior heart region (Thom et al. 2006). The first suggestion that human OFT-like cells can be generated *in vitro* came from a scRNA-Seq analysis performed by Friedman et al. (2018) of cells collected during directed cardiac differentiation based on a WNT modulation protocol. They found a population of *BMP4+/RSPO3+/COL1A2+/TNC+* non-contractile cells that correlated with the transcriptional signature of mouse OFT cells (Li et al. 2016). Interestingly, we identified human OFT-like cells derived from aSHF-like progenitors (noRA-N+) with a similar transcriptional profile. However, this OFT-like population does not correspond to fibroblast-like cells of OFT, but rather OFT-like cardiomyocytes. These OFT-like cardiomyocytes are characterised by myocardial and SMC markers, similarly to mouse OFT cardiomyocytes (Liu et al., 2019). One of the few analyses of the human OFT region comes from Sahara et al. (2019). They uncovered early OFT progenitors in human foetal tissues marked by expression of *LGR5*, *ISL1*, *RSPO3*, and *BMP4,* termed conoventricular progenitors (CVPs). This gene profile is also characteristic of the aSHF-like progenitors we describe.

The formation of valves is a complex process that occurs concomitantly with the morphogenesis of the myocardium and changing haemodynamics within the growing heart. Briefly, endocardial cells lining the myocardium undergo a process of activation, migration and invasion into the cardiac jelly, known as an endothelial-to-mesenchymal transition, to initially form poorly organized swellings – primitive cardiac cushions – that ultimately become well-developed valves (Puceat et al., 2013). A fundamental understanding of the pathways governing valvulogenesis and valve maturation remains elusive.

Here, we described a cell population with a transcriptomic signature corresponding to OFT valvular cells. This population was almost exclusively derived from human aSHF-like cells, which is consistent with SHF-derived endocardial progenitors' contribution to valve formation in mice (Verzi et al., 2005; Crucean et al., 2017). The analysis focused on three time points spanning only 30 days, precluding the deconstruction of meaningful lineage trajectories. However, it would be of great interest to explore further whether valvular-like cell types obtained with this protocol can transition through all key developmental stages from aSHF cells into cells forming valves. Such research could benefit the advancement of therapeutic approaches aiming at valve replacement.

4.4 Detection of a new population of progenitors in human *in vitro* differentiation

scRNA-Seq allows for the detection of rare transient populations that would be missed in bulk transcriptomic analyses. Tyser et al., 2021 and Zhang et al., 2021 uncovered an unexpected complexity of cardiovascular lineages by combining traditional lineage tracing with scRNA-Seq-based differentiation trajectory analyses. This allowed them to discover a new population during mouse cardiogenesis, termed the juxta-cardiac field. We captured a small subset of JCF-like cells during hPSCs differentiation *in vitro*, providing the first evidence that an equivalent population exists in humans. To further explore the JCF population, we looked for surface markers that would allow for its isolation. Using the scRNA-seq dataset of Tyser et al. (2021) we identified several candidates showing co-expression with other key JCF markers through comparative differential gene expression analysis. We ultimately selected *ITGA8* as the most promising surface marker due to its high specificity for JCF cells.

ITGA8 is the subunit a8 of the integrin family of cell surface receptors that constitute a link between the extracellular matrix and cytoskeleton within a cell. In mice, *Itga8* is one of the genes upregulated in early *Mesp1+* cardiac precursors (Lescroart et al., 2014). It was also detected in mesenchymal cells of the septum transversum (E9.5), a subset of hepatic stellate cells and perivascular mesenchymal cells in the developing liver (from E11.5) and at day 12.5 in the epicardium (Ogawa et al., 2018). In chick embryos, itga8 was detected in the proepicardial organ, epicardial mesothelial cells and at HH19 in both the epicardium and the myocardium (So et al., 2008). The reported localisation of ITGA8 in animal models largely correlates with the bipotentiality of JCF to differentiate towards myocardial and epicardial lineages.

RA signalling plays an important role during many stages of epicardium development. One of the enzymes involved in RA synthesis, RALDH2, is present in the proepicardial organ as well as epicardial cells (Moss et al., 1998; Perez-Pomares et al., 2002; Xavier-Neto et al., 2000; Jenkins et al., 2005). These observations have led to the utilization of RA in protocols aiming to generate epicardial cells *in vitro* (Iyer et al., 2015; Guadix et al., 2017). Within this work, we showed that RA potentially also influences the emergence of JCF-like cells by uncovering RAREs within ±5 kb from the transcription start site of one of the key JCF markers – *MAB21L2*.

4.5 lncRNAs as important gene regulatory elements modulating the specification of progenitors during cardiogenesis

lncRNAs are increasingly recognized as important regulatory elements that can modulate gene expression at multiple levels: at the chromatin level – by influencing its structure through interacting with DNA, RNA and proteins, at the transcriptional level – by modulating expression of neighbouring and distant genes, and at the translational level – by affecting splicing and RNA stability (reviewed by Statello et al., 2021). Several lncRNAs essential for specific stages of heart development have been reported (Klattenhoff et al., 2013; Grote et al., 2013; Ounzain et al., 2015; Ritter et al., 2019). However, so far, human heart-field-like specific lncRNAs have not been described. In this work, we used deep RNA-seq profiling to provide a comprehensive list of hundreds of lncRNAs differentially expressed between human aSHF and

FHF-like cells. It would be of great interest to further explore whether these lncRNAs are essential for heart field-specific differentiation.

Interestingly, we identified several human heart field-like specific lncRNAs that seem to mirror expression of their protein-coding counterparts, such as *THBS4-AS* and *THBS4*; *ISL1* and *ISL1-DT*, *FGF10-AS* and *FGF10,* and others, suggesting their potential relevance. Recent work by Frank et al., (2018) identified a class of divergent lncRNAs mirroring expression of their protein-coding counterparts in a cell-specific manner, termed yin yang lncRNAs (yylncRNAs). Using an *in vitro* cardiac differentiation model, they found 781 yylncRNAs. Further, they focused on paired with *BRACHYURY (T)* - yylncT that they found to be required for transcription of *BRACHYURY (T)* and the subsequent progression of hESCs into a mesodermal stage. Further studies would be required to confirm whether the lncRNAs described here have similar functions.

Using mouse models to study the mechanism of diseases has been very informative. However, understanding the similarities and differences between both species is crucial. The work presented here provides an insight into how deletion of the same loci can lead to different results in mouse and human models. Ritter et al. (2019) described a conserved lncRNA locus downstream of Hand2, termed Handsdown (Hdn), which negatively regulates transcription of Hand2 and subsequently other core cardiac transcription factors (*NKX2.5*, *TNNT2*). The presence and transcriptional activity of Hdn are essential for a proper expression level of *Hand2*. However, we showed here that the deletion of the part of *HDN* promoter that abolishes Hdn activity in mice does not affect the transcription of human *HAND2* or cardiac differentiation. It should be noted that the presented analysis focused only on few transcripts that were observed to be dysregulated in mice, and that results may be different in an *in vivo* context. However, Guo et al (2020) recently showed that a lncRNA conserved between mice and humans (mFast and hFAST) can play different roles in the regulation of mouse embryonic stem cells (mESCs) and hESCs pluripotency. Differing RNA processing and subcellular localization of lncRNAs between the two species can lead to functional divergence, it is therefore possible that HDN indeed affects human heart development differently than in mice.

Overall, the work presented here emphasises the importance of human-derived models and the detailed characterisation of cell types, states and gene regulatory networks within them to gain a better understanding of human development and disease.

4.6 Limitations & Outlook

The work presented here showed the *in vitro* derivation of mesodermal cardiac progenitor populations that contribute to the heart formation, including aSHF, FHF, as well as the recently discovered JCF. Though the well-characterized protocol described here has many important applications, as a 2D cell culture protocol, it also has certain limitations. Cells grown in 2D do not reproduce many features of their *in vivo* counterparts, including the recapitulation of the spatiotemporal self-organization of cell types as well as certain morphological and functional characteristics. Hence, many researchers have turned to 3D models to produce more complex heart-like structures (Drakhlis et al., 2021, preprint: Silva et al, 2020; Lewis-Israeli et al., 2021; Hofbauer et al., 2021). However, these cardiac organoids still lack the self-organization of all three layers of the heart – myocardium, endocardium, and epicardium – as well as many structures important for the function of the heart, including valves or septa. Being able to introduce them would be of great benefit for modelling CHDs that are characterized by defects in those heart regions. A major obstacle to develop functional, properly organized microtissues is a failure of differentiation protocols to fully recapitulate the complexity of signalling driving organogenesis *in vivo*. Here, we provided a thorough description of differentiation protocols guiding cells through certain developmental stages upon tight modulation of signalling cues. Our work in combination with current progress in the generation of 3D models could further advance the development of complex cardiac organoids.

Within this work, we also showed that the comparison between scRNA-Seq datasets of *in vitro* cells and their *in vivo* counterparts can greatly support our understanding of human development and disease and uncover novel populations of cells. Single-cell transcriptomics offers exceptional insights into the complexity of studied systems and allows us to predict possible cell differentiation trajectories which deepen our understanding of lineage segregation and differentiation progression. It can be expected that knowledge gained from the growing number of single-cell multi-omics datasets from human tissues will even further advance 2D and 3D protocols for the *in vitro* derivation of therapeutically important cell types.

Current models to study CHDs fail to recapitulate many aspects of those diseases. A basic understanding of processes during healthy development is a prerequisite to grasp cellular and molecular changes of diseased states. Here we expanded knowledge on early human cardiogenesis by generating various human cardiac cell types and states that resemble the *in vivo*

counterparts, and explored some elements of gene regulatory networks that characterize them. We showed the utility of generating lineage-specific progenitors in studying one of the most severe CHDs – HLHS. Overall, this work contributed to the ongoing effort to advance pre-clinical models for better and faster development of therapeutic interventions.

References

1. Abu-Abed, S., Dollé, P., Metzger, D., Beckett, B., Chambon, P., Petkovich, M. (2001). The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes and Development,* 1*5*(2), 226–240.

2. Aggarwal, V. S., Liao, J., Bondarev, A., Schimmang, T., Lewandoski, M., Locker, J., Shanske, A., Campione, M., Morrow, B. E. (2006). Dissection of Tbx1 and Fgf interactions in mouse models of 22q11DS suggests functional redundancy. *Human Molecular Genetics*, *15*(21), 3219–3228.

3. Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J. C., Geurts, P., Aerts, J., Van Den Oord, J., Atak, Z. K., Wouters, J., Aerts, S. (2017). SCENIC: Single-cell regulatory network inference and clustering. *Nature Methods*, *14*(11), 1083–1086.

4. Alexanian, M., Ounzain, S. (2020). Long Noncoding RNAs in Cardiac Development. *Cold Spring Harbour Perspectives Biology, 12*(11), a037374.

5. Anderson, K. M., Anderson, D. M., McAnally, J. R., Shelton, J. M., Bassel-Duby, R., Olson, E. N. (2016). Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature*, *539*(7629), 433–436.

6. Ang, Y. S., Rivas, R. N., Ribeiro, A. J. S., Srivas, R., Rivera, J., Stone, N. R., Pratt, K., Mohamed, T. M. A., Fu, J. D., Spencer, C. I., Tippens, N. D., Li, M., Narasimha, A., Radzinsky, E., Moon-Grady, A. J., Yu, H., Pruitt, B. L., Snyder, M. P., Srivastava, D. (2016). Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell*, *167*(7), 1734-1749.e22.

7. Arai, A., Yamamoto, K., Toyama, J. (1997). Murine Cardiac Progenitor Cells Require Visceral Embryonic Endoderm and Primitive Streak for Terminal Differentiation. *Developmental Dynamics* 210(3), 344-53.

8. Arnold, S. J., Robertson, E. J. (2009). Making a commitment: Cell lineage allocation and axis patterning in the early mouse embryo. In *Nature Reviews Molecular Cell Biology*, *10*(2), 91–103.

9. Asp, M., Giacomello, S., Larsson, L., Wu, C., Fürth, D., Qian, X., Wärdell, E., Custodio, J., Reimegård, J., Salmén, F., Österholm, C., Ståhl, P. L., Sundström, E., Åkesson, E., Bergmann, O., Bienko, M., Månsson-Broberg, A., Nilsson, M., Sylvén, C., Lundeberg, J. (2019). A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell*, *179*(7), 1647-1660.e19.

10. Bao, X., Lian, X., Qian, T., Bhute, V. J., Han, T., Palecek, S. P. (2017). Directed differentiation and long-term maintenance of epicardial cells derived from human pluripotent stem cells under fully defined conditions. *Nature Protocols*, *12*(9), 1890–1900.

11. Basson, C.T., Bachinsky, D.R., Lin, R.C., Levi, T., Elkins, J.A., Soults, J. et al. (1997). Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat. Genet.* 15, 30–35.

12. Bell, C. C., Amaral, P. P., Kalsbeek, A., Magor, G. W., Gillinder, K. R., Tangermann, P., Di Lisio, L., Cheetham, S. W., Gruhl, F., Frith, J., Tallack, M. R., Ru, K. L., Crawford, J., Mattick, J. S., Dinger, M. E., Perkins, A. C. (2016). The Evx1/Evx1as gene locus regulates anterior-posterior patterning during gastrulation. *Scientific Reports*, *6*, 26657.

13. Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E. J. J., Constam, D. B. (2006). The Nodal Precursor Acting via Activin Receptors Induces Mesoderm by Maintaining a Source of Its Convertases and BMP4. *Developmental Cell*, *11*(3), 313–323.

14. Benson, D.W., Silberbach, G.M., Kavanaugh-Mchugh, A., Cottrill, C., Zhang, Y., Riggs, S. et al. (1999). Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J. Clin. Invest.* 104, 1567–1573.

15. Bernheim, S., Meilhac, S. M. (2020). Mesoderm patterning by a dynamic gradient of retinoic acid signalling. *Philosophical Transactions of the Royal Society B: Biological Sciences, 375*(1809), 20190556.

16. Bondue, A., Lapouge, G., Paulissen, C., Semeraro, C., Iacovino, M., Kyba, M., Blanpain, C. (2008). Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. *Cell Stem Cell*, *3*(1), 69–84.

17. Bruneau, B. G., Logan, M., Davis, N., Levi, T., Tabin, C. J., Seidman, J. G., Seidman, C. E. (1999). Chamber-Specific Cardiac Expression of Tbx5 and Heart Defects in Holt-Oram Syndrome. *Developmental Biology, 211*(1), 100-108.

18. Cai, C. L., Liang, X., Shi, Y., Chu, P.-H., Pfaff, S. L., Chen, J., Evans, S. (2003). Isl1 Identifies a Cardiac Progenitor Population that Proliferates Prior to Differentiation and Contributes a Majority of Cells to the Heart. *Developmental Cell 5(*6), 877–889.

19. Chen, L., Fulcoli, F. G., Tang, S., Baldini, A. (2009). Tbx1 regulates proliferation and differentiation of multipotent heart progenitors. *Circulation Research*, *105*(9), 842–851.

20. Cheng, L. X., Xie, M. H., Qiao, W. H., Song, Y., Zhang, Y. Y., Geng, Y. C., Xu, W. L., Wang, L., Wang, Z., Huang, K., Dong, N. G., Sun, Y. H. (2021). Generation and characterization of cardiac valve endothelial-like cells from human pluripotent stem cells. *Communications Biology*, *4*(1039).

21. Ciruna, B., Rossant, J. (2001). FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Developmental cell*, *1*(1), 37–49.

22. Cohen, E. D., Miller, M. F., Wang, Z., Moon, R. T., Morrisey, E. E. (2012). Wnt5a and wnt11 are essential for second heart field progenitor development. *Development*, *139*(11), 1931–1940.

23. Cohen, E. D., Wang, Z., Lepore, J. J., Min, M. L., Taketo, M. M., Epstein, D. J., Morrisey, E. E. (2007). Wnt/β-catenin signaling promotes expansion of Isl-1-positive cardiac progenitor cells through regulation of FGF signaling. *Journal of Clinical Investigation*, *117*(7), 1794–1804.

24. Corsten-Janssen, N. et al. (2013). The cardiac phenotype in patients with a CHD7 mutation. *Circ. Cardiovasc. Genet.* 6, 248–254.

25. Costello, I., Pimeisl, I. M., Dräger, S., Bikoff, E. K., Robertson, E. J., Arnold, S. J. (2011). The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nature Cell Biology*, *13*(9), 1084–1092.

26. Crucean, A., Alqahtani, A., Barron, D. J., Brawn, W. J., Richardson, R. V., O'Sullivan, J., Anderson, R. H., Henderson, D. J., Chaudhry, B. (2017). Re-evaluation of hypoplastic left heart syndrome from a developmental and morphological perspective. *Orphanet Journal of Rare Diseases*, *12*(1).

27. Cui, Y., Zheng, Y., Liu, X., Yan, L., Fan, X., Yong, J., Hu, Y., Dong, J., Li, Q., Wu, X., Gao, S., Li, J., Wen, L., Qiao, J., Tang, F. (2019). Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. *Cell Reports*, *26*(7), 1934-1950.e5.

28. Cunningham, F. *et al.* (2022). Ensembl 2022. *Nucleic Acids Res. 50*, D988–D995.

29. Cunningham, T. J., Duester, G. (2015). Mechanisms of retinoic acid signalling and its roles in organ and limb development. In *Nature Reviews Molecular Cell Biology*, 16(2), 110–123.

30. Cyganek, L., Tiburcy, M., Sekeres, K., Gerstenberg, K., Bohnenberger, H., Lenz, C., Henze, S., Stauske, M., Salinas, G., Zimmermann, W. H., Hasenfuss, G., Guan, K. (2018). Deep phenotyping of human induced pluripotent stem cell-derived atrial and ventricular cardiomyocytes. *JCI Insight*, *3*(12).

31. D'Aniello, E., Rydeen, A. B., Anderson, J. L., Mandal, A., Waxman, J. S. (2013). Depletion of Retinoic Acid Receptors Initiates a Novel Positive Feedback Mechanism that Promotes Teratogenic Increases in Retinoic Acid. *PLoS Genetics*, *9*(8).

32. Daneshvar, K., Pondick, J. V., Kim, B. M., Zhou, C., York, S. R., Macklin, J. A., Abualteen, A., Tan, B., Sigova, A. A., Marcho, C., Tremblay, K. D., Mager, J., Choi, M. Y., Mullen, A. C. (2016). DIGIT Is a Conserved Long Noncoding RNA that Regulates GSC Expression to Control Definitive Endoderm Differentiation of Embryonic Stem Cells. *Cell Reports*, *17*(2), 353–365.

33. De Bono, C., Thellier, C., Bertrand, N., Sturny, R., Jullian, E., Cortes, C., Stefanovic, S., Zaffran, S., Théveniau-Ruissy, M., & Kelly, R. G. (2018). T-box genes and retinoic acid signaling regulate the segregation of arterial and venous pole progenitor cells in the murine second heart field. *Human molecular genetics*, *27*(21), 3747–3760.

34. de Soysa, T. Y., Ranade, S. S., Okawa, S., Ravichandran, S., Huang, Y., Salunga, H. T., Schricker, A., del Sol, A., Gifford, C. A., Srivastava, D. (2019). Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. *Nature*, *572*(7767), 120–124.

35. Devalla, H. D., Schwach, V., Ford, J. W., Milnes, J. T., El-Haou, S., Jackson, C., Gkatzis, K., Elliott, D. A., Chuva de Sousa Lopes, S. M., Mummery, C. L., Verkerk, A. O., Passier, R. (2015). Atrial-like cardiomyocytes from human pluripotent stem cells are a robust preclinical model for assessing atrial-selective pharmacology. *EMBO Molecular Medicine*, *7*(4), 394–410.

36. Devine, W. P., Wythe, J. D., George, M., Koshiba-Takeuchi, K., Bruneau, B. G. (2014). Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *ELife*, *3* (e03848).

37. Digilio, M.C., Gnazzo. M., Lepri, F., Dentici, M.L., Pisaneschi, E., Baban, A., Passarelli, C., Capolino, R., Angioni, A., Novelli, A., Marino, B., Dallapiccola, B. (2017).

Congenital heart defects in molecularly proven Kabuki syndrome patients. *Am J Med Genet A.*, 173(11), 2912-2922.

38. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21.

39. Dubois, N.C. , Craft, A.M., Sharma, P. , Elliott, D.A., Stanley, E.G., Elefanty, A.G. (2011). SIRPA is a specific cell-surface marker for isolating cardiomyocytes derived from human pluripotent stem cells, *Nat Biotechnol*, 29:1011-1018.

40. Dyer, L. A., Kirby, M. L. (2009). Sonic hedgehog maintains proliferation in secondary heart field progenitors and is required for normal arterial pole formation. *Developmental Biology*, *330*(2), 305–317.

41. Elliott, D.A., Braam, S.R., Koutsis, K., Ng, E.S., Jenny, R., Lagerqvist, E.L. *et al.* (2011) NKX2-5(eGFP/w) hESCs for isolation of human cardiac progenitors and cardiomyocytes *Nat Methods*, 8:1037-1040.

42. Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., McDonel, P. E., Guttman, M., Lander, E. S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, *539*(7629), 452–455.

43. Fischer, B. *et al.* (2018). A complete workflow for the differentiation and the dissociation of hiPSC-derived cardiospheres. *Stem Cell Res. 32.*

44. Frank, S., Ahuja, G., Bartsch, D., Russ, N., Yao, W., Kuo, J. C. C., Derks, J. P., Akhade, V. S., Kargapolova, Y., Georgomanolis, T., Messling, J. E., Gramm, M., Brant, L., Rehimi, R., Vargas, N. E., Kuroczik, A., Yang, T. P., Sahito, R. G. A., Franzen, J., (…) Kurian, L. (2019). yylncT Defines a Class of Divergently Transcribed lncRNAs and Safeguards the T-mediated Mesodermal Commitment of Human PSCs. *Cell Stem Cell*, *24*(2), 318-327.e8.

45. Friedman, C. E., Nguyen, Q., Lukowski, S. W., Helfer, A., Chiu, H. S., Miklas, J., Levy, S., Suo, S., Han, J. D. J., Osteil, P., Peng, G., Jing, N., Baillie, G. J., Senabouth, A., Christ, A. N., Bruxner, T. J., Murry, C. E., Wong, E. S., Ding, J., (…) Palpant, N. J. (2018). Single-Cell Transcriptomic Analysis of Cardiac Differentiation from Human PSCs Reveals HOPX-Dependent Cardiomyocyte Maturation. *Cell Stem Cell*, *23*(4), 586-598.e8.

46. Funa, N. S., Schachter, K. A., Lerdrup, M., Ekberg, J., Hess, K., Dietrich, N., Honoré, C., Hansen, K., Semb, H. (2015). β-Catenin Regulates Primitive Streak Induction

through Collaborative Interactions with SMAD2/SMAD3 and OCT4. *Cell Stem Cell*, *16*(6), 639–652.

47. Gadue, P., Huber, T. L., Nostro, M. C., Kattman, S., & Keller, G. M. (2005). Germ layer induction from embryonic stem cells. *Experimental hematology*, *33*(9), 955–964.

48. Galli, D., Domínguez, J. N., Zaffran, S., Munk, A., Brown, N. A., Buckingham, M. E. (2008). Atrial myocardium derives from the posterior region of the second heart field, which acquires left-right identity as Pitx2c is expressed. *Development*, *135*(6), 1157–1167.

49. Garg, V., Kathiriya, I.S., Barnes, R., Schluterman, M.K., King, I.N., Butler, C.A. et al. (2003). GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* 424, 443–447.

50. Ge, X., Ren, Y., Bartulos, O., Lee, M.Y., Yue, Z., Kim, K.Y. et al. (2012). Modeling supravalvular aortic stenosis syndrome with human induced pluripotent stem cells. *Circulation* 126, 1695–1704.

51. Gessert, S., Kühl, M. (2010). The multiple phases and faces of Wnt signaling during cardiac differentiation and development. In *Circulation Research,* 107(2), 186–199.

52. Gittenberger-de Groot, A. C., Winter, E. M., & Poelmann, R. E. (2010). Epicardium-derived cells (EPDCs) in development, cardiac disease and repair of ischemia. *Journal of cellular and molecular medicine*, *14*(5), 1056–1060.

53. Gouon-Evans, V., Boussemart, L., Gadue, P., Nierhoff, D., Koehler, C. I., Kubo, A., Shafritz, D. A., Keller, G. (2006). BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nature Biotechnology*, *24*(11), 1402–1411.

54. Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., Herrmann, B. G. (2013). The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Developmental Cell*, *24*(2), 206–214.

55. Guadix, J.A., Orlova, V.V., Giacomelli, E., Bellin, M., Ribeiro, M.C., Mummery, C.L., Pérez-Pomares, J.M., Passier, R. (2017). Human Pluripotent Stem Cell Differentiation into Functional Epicardial Progenitor Cells. *Stem Cell Reports*, 9(6):1754-1764.

56. Gunawan, M.G., Sangha, S.S., Shafaattalab, S., Lin, E., Heims-Waldron, D.A., Bezzerides, V.J., Laksman, Z., Tibbits, G.F. (2021). Drug screening platform using human induced pluripotent stem cell-derived atrial cardiomyocytes and optical mapping. *Stem Cells Transl Med.*, 10(1):68-82.

57. Guo, C.J., Ma, X.K., Xing, Y.H., Zheng, C.C., Xu, Y.F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael, G.G., Yang, L., Chen, L.L. (2020). Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell*, 181(3):621-636.e22.

58. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., (…) Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell, 184*(13), 3573-3587.e29.

59. Hatcher, C. J., Goldstein, M. M., Mah, C. S., Delia, C. S., Basson, C. T. (2000). Identification and Localization of TBX5 Transcription Factor During Human Cardiac Morphogenesis. *Developmental Dynamics, 219*(1), 90-5.

60. High, F. A., Jain, R., Stoller, J. Z., Antonucci, N. B., Min, M. L., Loomes, K. M., Kaestner, K. H., Pear, W. S., Epstein, J. A. (2009). Murine Jagged1/Notch signaling in the second heart field orchestrates Fgf8 expression and tissue-tissue interactions during outflow tract development. *Journal of Clinical Investigation*, 119(7), 1986–1996.

61. Hochgreb, T., Linhares, V. L., Menezes, D. C., Sampaio, A. C., Yan, C. Y. I., Cardoso, W. V., Rosenthal, N., Xavier-Neto, J. (2003). A caudorostral wave of RALDH2 conveys anteroposterior information to the cardiac field. *Development*, 130(22), 5363–5374.

62. Hofbauer, P., Jahnel, S. M., Papai, N., Giesshammer, M., Deyett, A., Schmidt, C., Penc, M., Tavernini, K., Grdseloff, N., Meledeth, C., Ginistrelli L. C., Ctortecka, C., Šalic, Š., Novatchkova, M., Mendjan, S. (2021). Cardioids reveal self-organizing principles of human cardiogenesis. *Cell, 184*(12), 3299-3317.e22.

63. Hoffman, J.I., and Kaplan, S. (2002). The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* 39, 1890–1900.

64. Hong, N., Zhang, E., Xie, H., Jin, L., Zhang, Q., Lu, Y., Chen, A. F., Yu, Y., Zhou, B., Chen, S., Yu, Y., Sun, K. (2021). The transcription factor Sox7 modulates endocardiac cushion formation contributed to atrioventricular septal defect through Wnt4/Bmp2 signaling. *Cell Death and Disease*, 12, 393.

65. Hrstka, S.C., Li, X., Nelson, T.J., and Wanek Program Genetics Pipeline Group. (2017). NOTCH1-dependent nitric oxide signaling deficiency in hypoplastic left heart syndrome revealed through patient-specific phenotypes detected in bioengineered cardiogenesis. *Stem Cells* 35, 1106–1119.

66. Hu, T., Yamagishi, H., Maeda, J., McAnally, J., Yamagishi, C., Srivastava, D. (2004). Tbx1 regulates fibroblast growth factors in the anterior heart field through reinforcing autoregulatory loop involving forkhead transcription factors. *Development*, *131*(21), 5491–5502.

67. Ivanovitch, K., Soro-Barrio, P., Chakravarty, P., Jones, R. A., Bell, D. M., Neda Mousavy Gharavy, S., Stamataki, D., Delile, J., Smith, J. C., Briscoe, J. (2021). Ventricular, atrial, and outflow tract heart progenitors arise from spatially and molecularly distinct regions of the primitive streak. *PLoS Biology*, *19*(5).

68. Iyer, D., Gambardella, L., Bernard, W.G., Serrano, F., Mascetti, V.L., Pedersen, R.A., Talasila, A., Sinha, S. (2015). Robust derivation of epicardium and its differentiated smooth muscle cell progeny from human pluripotent stem cells. *Development*, 142(8):1528-41.

69. Jenkins, S.J., Hutson, D.R., Kubalak, S.W. (2005). Analysis of the proepicardium-epicardium transition during the malformation of the RXRalpha(-/-) epicardium. *Dev Dyn.,* 233:1091–1101.

70. Jiang, Y., Habibollah, S., Tilgner, K., Collin, J., Barta, T., Al-Aama, J.Y., Tesarov, L., Hussain, R., Trafford, A.W., Kirkwood, G., Sernagor, E., Eleftheriou, C.G., Przyborski, S., Stojković, M., Lako, M., Keavney, B., Armstrong, L. (2014) An induced pluripotent stem cell model of hypoplastic left heart syndrome (HLHS) reveals multiple expression and functional differences in HLHS-derived cardiac myocytes. *Stem Cells Transl Med.*, 3(4):416-23.

71. Keegan, B. R., Feldman, J. L., Begemann, G., Ingham, P. W., & Yelon, D. (2005). Retinoic acid signaling restricts the cardiac progenitor pool. *Science, 307*(5707), 247–249.

72. Kelly, R. G., Brown, N. A., Buckingham, M. E. (2001). The Arterial Pole of the Mouse Heart Forms from Fgf10-Expressing Cells in Pharyngeal Mesoderm. *Developmental Cell* 1(3), 435-440.

73. Kinder, S. J., Tsang, T. E., Quinlan, G. A., Hadjantonakis, A. K., Nagy, A., Tam, P. P. (1999). The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development*, *126*(21), 4691–4701.

74. Kitajima, S., Miyagawa-Tomita, S., Inoue, T., Kanno, J., Saga, Y. (2006). Mesp1-nonexpressing cells contribute to the ventricular cardiac conduction system. *Developmental dynamics*, *235*(2), 395–402.

75. Klattenhoff, C. A., Scheuermann, J. C., Surface, L. E., Bradley, R. K., Fields, P. A., Steinhauser, M. L., Ding, H., Butty, V. L., Torrey, L., Haas, S., Abo, R., Tabebordbar, M., Lee, R. T., Burge, C. B., Boyer, L. A. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*, *152*(3), 570–583.

76. Klaus, A., Saga, Y., Taketo, M. M., Tzahor, E., Birchmeier, W. (2007). Distinct roles of Wnt/-catenin and Bmp signaling during early cardiogenesis. *Proceedings of the National Academy of Sciences of the United States of America, 104*(47), 18531-6.

77. Knowles, R.L., Bull, C., Wren, C., Dezateux, C. (2012). Mortality with congenital heart defects in England and Wales, 1959-2009: exploring technological change through period and birth cohort analysis. *Arch Dis Child.*, 97(10):861-5.

78. Kobayashi, J., Yoshida, M., Tarui, S., Hirata, M., Nagai, Y., Kasahara, S., Naruse, K., Ito, H., Sano, S., Oh, H. (2014). Directed differentiation of patient-specific induced pluripotent stem cells identifies the transcriptional repression and epigenetic modification of NKX2-5, HAND1, and NOTCH1 in hypoplastic left heart syndrome. *PLoS One*, 9(7):e102796.

79. Koyano-Nakagawa, N., Garry, D. J. (2017). Etv2 as an essential regulator of mesodermal lineage development. *Cardiovascular Research*, 113(11), 1294–1306.

80. Krane, M., Dreßen, M., Santamaria, G., My, I., Schneider, C. M., Dorn, T., Laue, S., Mastantuono, E., Berutti, R., Rawat, H., Gilsbach, R., Schneider, P., Lahm, H., Schwarz, S., Doppler, S. A., Paige, S., Puluca, N., Doll, S., Neb, I., (…) Moretti, A. (2021). Sequential Defects in Cardiac Lineage Commitment and Maturation Cause Hypoplastic Left Heart Syndrome. *Circulation*, *144*(17), 1409–1428.

81. Krishnan, A., Samtani, R., Dhanantwari, P., Lee, E., Yamada, S., Shiota, K., Donofrio, M. T., Leatherbury, L., Lo, C. W. (2014). A detailed comparison of mouse and human cardiac development. *Pediatric Research*, *76*(6), 500–507.

82. Kwon, C., Qian, L., Cheng, P., Nigam, V., Arnold, J., Srivastava, D. (2009). A regulatory pathway involving Notch1/β-catenin/Isl1 determines cardiac progenitor cell fate. *Nature Cell Biology*, *11*(8), 951–957.

83. Lancaster, M. A., Knoblich, J. A. (2014). Generation of cerebral organoids from human pluripotent stem cells. *Nature Protocols*, *9*(10), 2329–2340.

84. Langston, A. W., Thompson, J. R., Gudas, L. J. (1997). Retinoic acid-responsive enhancers located 3' of the Hox A and Hox B homeobox gene clusters. Functional analysis. *Journal of Biological Chemistry*, *272*(4), 2167–2175.

85. Lauschke, K., Treschow, A.F., Rasmussen, M.A., Davidsen, N., Holst, B., Emnéus, J., Taxvig, C., Vinggaard, A.M. (2021). Creating a human-induced pluripotent stem cell-based NKX2.5 reporter gene assay for developmental toxicity testing. *Arch Toxicol*, 95(5):1659-1670.

86. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*, *9*(8).

87. Lee, J. H., Protze, S. I., Laksman, Z., Backx, P. H., Keller, G. M. (2017). Human Pluripotent Stem Cell-Derived Atrial and Ventricular Cardiomyocytes Develop from Distinct Mesoderm Populations. *Cell Stem Cell*, *21*(2), 179-194.e4.

88. Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolosse, A., Auer, H., Achouri, Y., Dubois, C., Bondue, A., Simons, B. D., Blanpain, C. (2014). Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nature Cell Biology*, *16*(9), 829–840.

89. Lescroart, F., Zaffran, S. (2018). Hox and tale transcription factors in heart development and disease. *International Journal of Developmental Biology*, *62*(11–12), 837–846.

90. Lewis-Israeli, Y.R., Wasserman, A.H., Gabalski, M.A., Volmert, B.D., Ming, Y., Ball, K.A., Yang, W., Zou, J., Ni, G., Pajares, N., Chatzistavrou, X., Li, W., Zhou, C., Aguirre, A. (2021) Self-assembling human heart organoids for the modeling of cardiac development and congenital heart disease. *Nat Commun*. 2021, 12(1), 5142.

91. Li, G., Xu, A., Sim, S., Priest, J. R., Tian, X., Khan, T., Quertermous, T., Zhou, B., Tsao, P. S., Quake, S. R., Wu, S. M. (2016). Transcriptomic Profiling Maps Anatomically Patterned Subpopulations among Single Embryonic Cardiac Cells. *Developmental Cell*, *39*(4), 491–507.

92. Lian, X., Hsiao, C., Wilson, G., Zhu, K., Hazeltine, L. B., Azarin, S. M., Raval, K. K., Zhang, J., Kamp, T. J., Palecek, S. P. (2012). Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(27).

93. Lien, C. L., McAnally, J., Richardson, J. A., Olson, E. N. (2002). Cardiac-specific activity of an Nkx2-5 enhancer requires an evolutionarily conserved Smad binding site. *Developmental Biology, 244*(2), 257–266.

94. Lin, H., McBride, K.L., Garg, V., Zhao, M.T. (2021). Decoding Genetics of Congenital Heart Disease Using Patient-Derived Induced Pluripotent Stem Cells (iPSCs). *Front Cell Dev Biol.*, 9:630069.

95. Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C. L., Lindberg, E. L., Kanda, M., Polanski, K., Heinig, M., Lee, M., Nadelmann, E. R., Roberts, K., Tuck, L., Fasouli, E. S., DeLaughter, D. M., McDonough, B., Wakimoto, H., Gorham, J. M., Samari, S., (…) Teichmann, S. A. (2020). Cells of the adult human heart. *Nature, 588*(7838), 466–472.

96. Liu, Q., Yan, H., Dawes, N. J., Mottino, G. A., Frank, J. S., Zhu, H. (1996). Insulin-like Growth Factor II Induces DNA Synthesis in Fetal Ventricular Myocytes In Vitro. *Circulation Research, 79*(4), 716-726.

97. Liu, X., Chen, W., Li, W., Li, Y., Priest, J. R., Zhou, B., Wang, J., Zhou, Z. (2019). Single-Cell RNA-Seq of the Developing Cardiac Outflow Tract Reveals Convergent Development of the Vascular Smooth Muscle Cells. *Cell Reports, 28*(5), 1346-1361.e4.

98. Love, M. I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12).

99. Luna-Zurita, L., Stirnimann, C. U., Glatt, S., Kaynak, B. L., Thomas, S., Baudin, F., Samee, M. A. H., He, D., Small, E. M., Mileikovsky, M., Nagy, A., Holloway, A. K., Pollard, K. S., Müller, C. W., Bruneau, B. G. (2016). Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. *Cell, 164*(5), 999–1014.

100. Luo, S., Lu, J. Y., Liu, L., Yin, Y., Chen, C., Han, X., Wu, B., Xu, R., Liu, W., Yan, P., Shao, W., Lu, Z., Li, H., Na, J., Tang, F., Wang, J., Zhang, Y. E., Shen, X. (2016). Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell, 18*(5), 637–652. .

101. Martínez-Morales, P. L., del Corral, R. D., Olivera-Martínez, I., Quiroga, A. C., Das, R. M., Barbas, J. A., Storey, K. G., Morales, A. V. (2011). FGF and retinoic acid activity gradients control the timing of neural crest cell emigration in the trunk. *Journal of Cell Biology, 194*(3), 489–503.

102.   Mayanagi, T., & Sobue, K. (2011). Diversification of caldesmon-linked actin cytoskeleton in cell motility. *Cell adhesion & migration*, *5*(2), 150–159.

103.   Meilhac, S. M., Esner, M., Kelly, R. G., Nicolas, J.-F., Buckingham, M. E. (2004). The Clonal Origin of Myocardial Cells in Different Regions of the Embryonic Mouse Heart. *Developmental Cell*, *6(*5), 685-698.

104.   Meilhac, S.M. Buckingham, M.E. (2018). The deployment of cell lineages that form the mammalian heart Nature Reviews Cardiology 15(11), 705-724.

105.   Mendjan, S., Mascetti, V. L., Ortmann, D., Ortiz, M., Karjosukarso, D. W., Ng, Y., Moreau, T., Pedersen, R. A. (2014). NANOG and CDX2 pattern distinct subtypes of human mesoderm during exit from pluripotency. *Cell Stem Cell*, *15*(3), 310–325.

106.   Miao, S., Zhao, D., Wang, X., Ni, X., Fang, X., Yu, M., Ye, L., Yang, J., Wu, H., Han, X., Qu, L., Li, L., Lan, F., Shen, Z., Lei, W., Zhao, Z. A., Hu, S. (2020). Retinoic acid promotes metabolic maturation of human embryonic stem cell-derived cardiomyocytes. *Theranostics*, *10*(21), 9686–9701.

107.   Miao, Y., Tian, L., Martin, M., Paige, S.L., Galdos, F.X., Li, J. et al. (2020b). Intrinsic endocardial defects contribute to hypoplastic left heart syndrome. *Cell Stem Cell* 27, 574–589.

108.   Mikryukov, A. A., Mazine, A., Wei, B., Yang, D., Miao, Y., Gu, M., Keller, G. M. (2021). BMP10 Signaling Promotes the Development of Endocardial Cells from Human Pluripotent Stem Cell-Derived Cardiovascular Progenitors. *Cell Stem Cell*, *28*(1), 96-111.e7.

109.   Miyamoto, M., Kannan, S., Uosaki, H., Kakani, T., Murphy, S., Andersen, P., Kwon, C. (2021). Cardiac progenitors auto-regulate second heart field cell fate via Wnt secretion. *Biorxiv*, *2021.01.31.428968*, https://doi.org/10.1101/2021.01.31.428968.

110.   Moss, J.B., Xavier-Neto, J., Shapiro, M.D., Nayeem, S.M., McCaffery, P., Drager, U.C., Rosenthal, N. (1998). Dynamic patterns of retinoic acid synthesis and response in the developing mammalian heart. *Dev Biol.,* 199:55–71.

111.   Murtagh, F., Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J. Classif. 31*, 274–295.

112.   Neri, T., Hiriart, E., van Vliet, P. P., Faure, E., Norris, R. A., Farhat, B., Jagla, B., Lefrancois, J., Sugi, Y., Moore-Morris, T., Zaffran, S., Faustino, R. S., Zambon, A. C., Desvignes, J. P., Salgado, D., Levine, R. A., de la Pompa, J. L., Terzic, A., Evans, S. M., (…) Pucéat, M. (2019). Human pre-valvular endocardial cells derived from

pluripotent stem cells recapitulate cardiac pathophysiological valvulogenesis. *Nature Communications*, *10,* 1929.

113.   Niederreither, K., Dollé, P. (2008). Retinoic acid in development: towards an integrated view. *Nature Reviews Genetics, 9*(7), 541-553.

114.   Niederreither, K., Subbarayan, V., Dollé, P., Chambon, P. (1999). Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nature genetics, 21,* 444-448.

115.   Nowotschin, S., Setty, M., Kuo, Y. Y., Liu, V., Garg, V., Sharma, R., Simon, C. S., Saiz, N., Gardner, R., Boutet, S. C., Church, D. M., Hoodless, P. A., Hadjantonakis, A. K., Pe'er, D. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature*, *569*(7756), 361–367.

116.   Ogawa, T., Li, Y., Lua, I., Hartner, A., Asahina, K. (2018). Isolation of a Unique Hepatic Stellate Cell Population Expressing Integrin a8 from Embryonic Mouse Livers. *Developmental Dynamics, 247*(6), 867-881.

117.   Ounzain, S., Micheletti, R., Arnan, C., Plaisance, I., Cecchi, D., Schroen, B., Reverter, F., Alexanian, M., Gonzales, C., Ng, S. Y., Bussotti, G., Pezzuto, I., Notredame, C., Heymans, S., Guigó, R., Johnson, R., Pedrazzini, T. (2015). CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. *Journal of Molecular and Cellular Cardiology*, *89*, 98–112.

118.   Parameswaran, M., Tam, P. P. L. (1995). Regionalisation of Cell Fate and Morphogenetic Movement of the Mesoderm During Mouse Gastrulation. *Developmental Genetics, 17*(1), 16-28.

119.   Parikh, A., Wu, J., Blanton, R. M., Tzanakakis, E. S. (2015). Signaling Pathways and Gene Regulatory Networks in Cardiomyocyte Differentiation. *Tissue Engineering - Part B: Reviews 21*(4), 377–392.

120.   Pediatric Cardiac Genomics Consortium., Gelb, B., Brueckner, M., Chung, W., Goldmuntz, E., Kaltman, J. et al. (2013). The congenital heart disease genetic network study: rationale, design, and early results. *Circ. Res.* 112, 698–706.

121.   Perez-Pomares, J.M., Phelps, A., Sedmerova, M., Carmona, R., Gonzalez-Iriarte, M., Munoz-Chapuli, R., Wessels, A. (2002). Experimental studies on the spatiotemporal expression of WT1 and RALDH2 in the embryonic avian heart: a model

for the regulation of myocardial and valvuloseptal development by epicardially derived cells (EPDCs). *Dev Biol.*, 247:307–326.

122. Pezhouman, A., Engel, J. L., Nguyen, N. B., Skelton, R. J. P., Gilmore, W. B., Qiao, R., Sahoo, D., Zhao, P., Elliott, D. A., Ardehali, R. (2021). Isolation and characterization of human embryonic stem cell-derived heart field-specific cardiomyocytes unravels new insights into their transcriptional and electrophysiological profiles. *Cardiovascular Research*, cvab102.

123. Pierpont, M.E., Brueckner, M., Chung, W.K., Garg, V., Lacro, R.V., Mcguire, A.L. et al. (2018). Genetic basis for congenital heart disease: revisited: a scientific statement from the American heart association. *Circulation* 138, e653–e711.

124. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., Reik, W., Srinivas, S., Simons, B.D., Nichols, J., Marioni, J.C., Göttgens, B. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*, *566*(7745):490-495.

125. Poch, C., Foo, K. S., De Angelis, M. T., Jennbacken, K., Santamaria, G., Bähr, A., Wang, Q., Reiter, F., Hornaschewitz, N., Zawada, D., Bozoglu, T., My, I., Meier, A., Dorn, T., LehtieN, M., Tsoi, Y. L., Hovdal, D., Hyllner, J., Schwarz, S., Jurisch, V., Sini, M., Fellows, M. D., Cummings, M., Clarke, J., Baptista, R., Eroglu, E., Tomasi, R., Dendorfer, A., Gaspari, M., Cuda, G., Krane, M., Sinnecker, D., Hoppmann, P., Kupatt, C., Fritsche-Danielson, R., Moretti, A., Chien, K. R., Laugwitz, K. L. (2022) Migratory and anti-fibrotic programmes define the regenerative potential of human cardiac progenitors. *Nature Cell Biology*, 24, 659–671.

126. Pucéat, M. (2013). Embryological origin of the endocardium and derived valve progenitor cells: From developmental biology to stem cell-based valve repair. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1833*(4), 917–922.

127. Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–2308.

128. Rana, M. S., Théveniau-Ruissy, M., De Bono, C., Mesbah, K., Francou, A., Rammah, M., Domínguez, J. N., Roux, M., Laforest, B., Anderson, R. H., Mohun, T., Zaffran, S., Christoffels, V. M., Kelly, R. G. (2014). Tbx1 coordinates addition of posterior second heart field progenitor cells to the arterial and venous poles of the heart. *Circulation Research*, *115*(9), 790–799.

129. Rankin, S. A., Steimle, J. D., Yang, X. H., Rydeen, A. B., Agarwal, K., Chaturvedi, P., Ikegami, K., Herriges, M. J., Moskowitz, I. P., Zorn, A. M. (2021). Tbx5 drives Aldh1a2 expression to regulate a RA-Hedgehog-Wnt gene regulatory network coordinating cardiopulmonary development. *ELife, 10*, e69288.

130. Ritter, N., Ali, T., Kopitchinski, N., Schuster, P., Beisaw, A., Hendrix, D. A., Schulz, M. H., Müller-McNicoll, M., Dimmeler, S., Grote, P. (2019). The lncRNA Locus Handsdown Regulates Cardiac Gene Programs and Is Essential for Early Mouse Development. *Developmental Cell, 50*(5), 644-657.e8.

131. Rivera-Pérez, J. A., Hadjantonakis, A. K. (2015). The dynamics of morphogenesis in the early mouse embryo. *Cold Spring Harbor Perspectives in Biology, 7*(11).

132. Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J. R., Catapano, A. L., Chugh, S., Cooper, L. T., Coresh, J., (…) Fuster, V. (2020). Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. In *Journal of the American College of Cardiology, 76*(25), 2982–3021.

133. Sahara, M., Santoro, F., Sohlmér, J., Zhou, C., Witman, N., Leung, C. Y., Mononen, M., Bylund, K., Gruber, P., Chien, K. R. (2019). Population and Single-Cell Analysis of Human Cardiogenesis Reveals Unique LGR5 Ventricular Progenitors in Embryonic Outflow Tract. *Developmental Cell, 48*(4), 475-490.e7.

134. Santini, M. P., Forte, E., Harvey, R. P., & Kovacic, J. C. (2016). Developmental origin and lineage plasticity of endogenous cardiac stem cells. *Development, 143*(8), 1242–1258.

135. Sarrach, S., Huang, Y., Niedermeyer, S., Hachmeister, M., Fischer, L., Gille, S., Pan, M., Mack, B., Kranz, G., Libl, D., Merl-Pham, J., Hauck, S. M., Tomada, E. P., Kieslinger, M., Jeremias, I., Scialdone, A., Gires, O. (2018). Spatiotemporal patterning of EpCAM is important for murine embryonic endo-And mesodermal differentiation. *Scientific Reports, 8, 1801.*

136. Saunders, V., Dewing, J. M., Sanchez-Elsner, T., Wilson, D. I. (2018). Expression and localisation of thymosin beta-4 in the developing human early fetal heart. *PLoS ONE, 13*(11).

137.   Schuhheiss, T. M., Burch, J. B., Lassar, A. B. (1997). A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes and Development, 11*(4), 451-462.

138.   Sendra, M., Domínguez, J.N., Torres, M., Ocaña, O.H. (2022). Dissecting the Complexity of Early Heart Progenitor Cells. *J. Cardiovasc. Dev. Dis.* 9(1), 5.

139.   Sifrim, A., Hitz, M.P., Wilsdon, A., Breckpot, J., Turki, S.H., Thienpont, B., McRae, J., Fitzgerald, T.W., Singh, T., Swaminathan, G.J., Prigmore, E., Rajan, D., Abdul-Khaliq, H., Banka, S., Bauer, U.M., Bentham, J., Berger, F., Bhattacharya, S., Bu'Lock, F., Canham, N., Colgiu, I.G., Cosgrove, C., Cox, H., Daehnert, I., Daly, A., Danesh, J., Fryer, A., Gewillig, M., Hobson, E., Hoff, K., Homfray, T. INTERVAL Study, Kahlert, A.K., Ketley, A., Kramer, H.H., Lachlan, K., Lampe, A.K., Louw, J.J., Manickara, A.K., Manase, D., McCarthy, K.P., Metcalfe, K., Moore, C., Newbury-Ecob, R., Omer, S.O., Ouwehand, W.H., Park, S.M., Parker, M.J., Pickardt, T., Pollard, M.O., Robert, L., Roberts, D.J., Sambrook, J., Setchfield, K., Stiller, B., Thornborough, C., Toka, O., Watkins, H., Williams, D., Wright, M., Mital, S., Daubeney, P.E., Keavney, B., Goodship, J., UK10K Consortium, Abu-Sulaiman, R.M., Klaassen, S., Wright, C.F., Firth, H.V., Barrett, J.C., Devriendt, K., FitzPatrick, D.R., Brook, J.D., Deciphering Developmental Disorders Study, Hurles, M.E. (2016). Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet.*, 48(9):1060-5.

140.   Silva, A.C., Matthys, O.B., Joy, D.A., Kauss, M.A., Natarajan, V. , Lai, M.H., Turaga, D., Blair, A.P., Alexanian, M., Bruneau, B.G., McDevitt, T.C. Developmental co-emergence of cardiac and gut tissues modeled by human iPSC-derived organoids. Preprint at: bioRxiv 2020.04.30.071472 (2020).

141.   Sirbu, I. O., Zhao, X., Duester, G. (2008). Retinoic acid controls heart anteroposterior patterning by down-regulating Isl1 through the Fgf8 pathway. *Developmental Dynamics*, *237*(6), 1627–1635.

142.   So, H. P., Dokic, D., Dettman, R. W. (2008). Communication between integrin receptors facilitates epicardial cell adhesion and matrix organization. *Developmental Dynamics*, *237*(4), 962–978.

143.   Später, D., Abramczuk, M. K., Buac, K., Zangi, L., Stachel, M. W., Clarke, J., Sahara, M., Ludwig, A., Chien, K. R. (2013). A HCN4+ cardiomyogenic progenitor

derived from the first heart field and human pluripotent stem cells. *Nature Cell Biology*, *15*(9), 1098–1106.

144.   Statello, L., Guo, C. J., Chen, L. L., Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. In *Nature Reviews Molecular Cell Biology, 22*(2), 96–118.

145.   Stefanovic, S., Laforest, B., Desvignes, J. P., Lescroart, F., Argiro, L., Maurel-Zaffran, C., Salgado, D., Plaindoux, E., De Bono, C., Pazur, K., Théveniau-Ruissy, M., Béroud, C., Puceat, M., Gavalas, A., Kelly, R. G., Zaffran, S. (2020). Hox-dependent coordination of mouse cardiac progenitor cell patterning and differentiation. *ELife*, *9*, 1–32.

146.   Stefanovic, S., Zaffran, S. (2017). Mechanisms of retinoic acid signaling during cardiogenesis. *Mechanisms of Development*, *143*, 9–19.

147.   Steimle, J. D., Rankin, S. A., Slagle, C. E., Bekeny, J., Rydeen, A. B., Chan, S. S. K., Kweon, J., Yang, X. H., Ikegami, K., Nadadur, R. D., Rowton, M., Hoffmann, A. D., Lazarevic, S., Thomas, W., Boyle Anderson, E. A. T., Horb, M. E., Luna-Zurita, L., Ho, R. K., Kyba, M., … Moskowitz, I. P. (2018). Evolutionarily conserved Tbx5-Wnt2/2b pathway orchestrates cardiopulmonary development. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), E10615–E10624.

148.   Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7), 1888-1902.e21.

149.   Sun, W., Zhao, R., Yang, Y., Wang, H., Shao, Y., Kong, X. (2013). Comparative study of human aortic and mitral valve interstitial cell gene expression and cellular function. *Genomics*, *101*(6), 326–335.

150.   Theodoris, C.V., Zhou, P., Liu, L., Zhang, Y., Nishino, T., Huang, Y. et al. (2020). Network-based screen in iPSC-derived cells reveals therapeutic candidate for heart valve disease. *Science* 2020:eabd0724.

151.   Thom, T., Haase, N., Rosamond, W., Howard, V. J., Rumsfeld, J., Manolio, T., Zheng, Z. J., Flegal, K., O'Donnell, C., Kittner, S., Lloyd-Jones, D., Goff, D. C., Hong, Y. (2006). Heart disease and stroke statistics - 2006 Update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*, *113*(6).

152. Tsuji-Tamura, K., Morino-Koga, S., Suzuki, S., & Ogawa, M. (2021). The canonical smooth muscle cell marker TAGLN is present in endothelial cells and is involved in angiogenesis. *Journal of cell science*, *134*(15), jcs254920.

153. Tyser, R. C. V., Ibarra-Soria, X., McDole, K., Jayaram, S. A., Godwin, J., Brand, T. A. H. V. Den, Miranda, A. M. A., Scialdone, A., Keller, P. J., Marioni, J. C., Srinivas, S. (2021). Characterization of a common progenitor pool of the epicardium and myocardium. *Science*, *371*(6533).

154. Tyser, R. C. V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., Srinivas, S. (2021b). Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature, 600*, 285-289.

155. Van Den Berg, G., Abu-Issa, R., De Boer, B. A., Hutson, M. R., De Boer, P. A. J., Soufan, A. T., Ruijter, J. M., Kirby, M. L., Van Den Hoff, M. J. B., Moorman, A. F. M. (2009). A caudal proliferating growth center contributes to both poles of the forming heart tube. *Circulation Research*, *104*(2), 179–188.

156. Varona, S., Orriols, M., Galán, M., Guadall, A., Cañes, L., Aguiló, S., Sirvent, M., Martínez-González, J., Rodríguez, C. (2018). Lysyl oxidase (LOX) limits VSMC proliferation and neointimal thickening through its extracellular enzymatic activity. *Scientific Reports*, *8*, 13258.

157. Verzi, M. P., McCulley, D. J., De Val, S., Dodou, E., Black, B. L. (2005). The right ventricle, outflow tract, and ventricular septum comprise a restricted expression domain within the secondary/anterior heart field. *Developmental Biology*, *287*(1), 134–145.

158. Wang, G., Mccain, M.L., Yang, L., He, A., Pasqualini, F.S., Agarwal, A. et al. (2014). Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nat. Med.* 20, 616–623.

159. Wang, J., Greene, S. B., Bonilla-Claudio, M., Tao, Y., Zhang, J., Bai, Y., Huang, Z., Black, B. L., Wang, F., Martin, J. F. (2010). Bmp Signaling Regulates Myocardial Differentiation from Cardiac Progenitors Through a MicroRNA-Mediated Mechanism. *Developmental Cell*, *19*(6), 903–912.

160. Warren, I., Maloy, M., Guiggey, D., Ogoke, O., Groth, T., Mon, T., Meamardoost, S., Liu, X., Szeglowski, A., Thompson, R., Chen, P., Paulmurugan, R., Parashurama, N. (2020). Foxa1 and Foxa2 together control developmental gene regulatory networks, and differentiation genes, in both human stem-cell derived liver progenitors and in

a human liver cell line: evidence of a collapse of human liver differentiation. *BioRxiv*, 2020.06.01.128108, https://doi.org/10.1101/2020.06.01.128108.

161.   Waxman, J. S., Keegan, B. R., Roberts, R. W., Poss, K. D., Yelon, D. (2008). Hoxb5b Acts Downstream of Retinoic Acid Signaling in the Forelimb Field to Restrict Heart Field Potential in Zebrafish. *Developmental Cell*, *15*(6), 923–934.

162.   Wessels, A., Sedmera, D. (2003). Developmental anatomy of the heart: a tale of mice and man. *Physiological Genomics, 15*(3), 165-176.

163.   Wiesinger, A., Li, J., Fokkert, L., Bakker, P., Verkerk, A. O., Christoffels, V. M., Boink, G. J. J., Devalla, H. D. (2022). A single cell transcriptional roadmap of human pacemaker cell differentiation. *eLife*, *11*, e76781.

164.   Witman, N., Zhou, C., Grote Beverborg, N., Sahara, M., Chien, K. R. (2020). Cardiac progenitors and paracrine mediators in cardiogenesis and heart regeneration. *Seminars in Cell and Developmental Biology, 100*, 29–51.

165.   Wobus, A. M., Kaomei, G., Shan, J., Wellner, M.-C., Rgen Rohwedel, J., Guanju, J., Fleischmann, B., Katus, H. A., Rgen Hescheler, J., Franz, W. M. (1997). Retinoic acid accelerates embryonic stem cell-derived cardiac differentiation and enhances development of ventricular cardiomyocytes. *Journal of Molecular and Cellular Cardiology*, *29*(6), 1525-1539.

166.   Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (China)*, *2*(3).

167.   Xavier-Neto, J., Shapiro, M.D., Houghton, L., Rosenthal, N. (2000). Sequential programs of retinoic acid synthesis in the myocardial and epicardial layers of the developing avian heart. *Dev Biol.,* 219:129–141.

168.   Yagi, H., Liu, X., Gabriel, G.C., Wu, Y., Peterson, K., Murray, S.A., Aronow, B.J., Martin, L.J., Benson, D.W., Lo, C.W. (2018) The Genetic Landscape of Hypoplastic Left Heart Syndrome. *Pediatr Cardiol.*, 39(6):1069-1081.

169.   Yamagishi, H., Maeda, J., Hu, T., McAnally, J., Conway, S. J., Kume, T., Meyers, E. N., Yamagishi, C., Srivastava, D. (2003). Tbx1 is regulated by tissue-specific forkhead proteins through a common Sonic hedgehog-responsive enhancer. *Genes and Development*, *17*(2), 269–281.

170. Yang, C., Xu, Y., Yu, M., Lee, D., Alharti, S., Hellen, N. et al. (2017). Induced pluripotent stem cell modelling of HLHS underlines the contribution of dysfunctional NOTCH signalling to impaired cardiogenesis. *Hum. Mol. Genet.* 26, 3031–3045.

171. Zaffran, S., Kelly, R. G., Meilhac, S. M., Buckingham, M. E., Brown, N. A. (2004). Right ventricular myocardium derives from the anterior heart field. *Circulation Research*, *95*(3), 261–268.

172. Zawada, D., Kornherr, J., Meier, A.B., Santamaria, G., Dorn, T., Ortmann, D., Lachmann, M., Ortiz, M., Harmer, S.C., Nobles, M., Tinker, A., Pedersen, R.A., Grote, P., Laugwitz, K-L., Moretti, A., Goedel, A. (2022). Retinoic acid signaling modulation guides *in vitro* specification of human heart field-specific progenitor pools. *bioRxiv* 2022.05.30.494027.

173. Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., To, J. K., (…) Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research*, *46*(D1), D754–D761.

174. Zhang, H., Lui, K. O., Zhou, B. (2018). Endocardial cell plasticity in cardiac development, diseases and regeneration. *Circulation Research*, *122*(5), 774–789.

175. Zhang, J. Z., Termglinchan, V., Shao, N. Y., Itzhaki, I., Liu, C., Ma, N., Tian, L., Wang, V. Y., Chang, A. C. Y., Guo, H., Kitani, T., Wu, H., Lam, C. K., Kodo, K., Sayed, N., Blau, H. M., Wu, J. C. (2019). A Human iPSC Double-Reporter System Enables Purification of Cardiac Lineage Subpopulations with Distinct Function and Drug Response Profiles. *Cell Stem Cell*, *24*(5), 802-811.e5.

176. Zhang, Qiangzhe, Jiang, J., Han, P., Yuan, Q., Zhang, J., Zhang, X., Xu, Y., Cao, H., Meng, Q., Chen, L., Tian, T., Wang, X., Li, P., Hescheler, J., Ji, G., Ma, Y. (2011). Direct differentiation of atrial and ventricular myocytes from human embryonic stem cells by alternating retinoid signals. *Cell Research*, *21*(4), 579–587.

177. Zhang, Q., Carlin, D., Zhu, F., Cattaneo, P., Ideker, T., Bloomekatz, J., Chi, N. C. (2021). Unveiling unexpected complexity and multipotentiality of early heart fields. *Circulation Research, 129*(4), 474-487.

Appendix

List of figures

List of tables

List of publications

1. **Zawada, D.,** Kornherr, J., Meier, A.B., Santamaria, G., Dorn, T., Ortmann, D., Lachmann, M., Ortiz, M., Harmer, S.C., Nobles, M., Tinker, A., Pedersen, R.A., Grote, P., Laugwitz, K-L., Moretti, A., Goedel, A. (2022). Retinoic acid signaling modulation guides *in vitro* specification of human heart field-specific progenitor pools. *bioRxiv* 2022.05.30.494027. (*Nature Communications*, in press)

2. Goedel, A.*, **Zawada, D.*,** Zhang, F., Chen, Z., Moretti, A., Sinnecker, D. (2018). Subtype-specific Optical Action Potential Recordings in Human Induced Pluripotent Stem Cell-derived Ventricular Cardiomyocytes. *J. Vis. Exp.*, 2018(139), e58134. *Equal contribution

3. Meier, A.*, **Zawada, D.*,** De Angelis, M.T.*, Martens, L.*, Santamaria, G. *, Zengerle, S., Nowak-Imialek, M., Kornherr, J., Zhang, F., Tian, Q., Wolf, C., Kupatt, C., Sahara, M., Lipp, P., Theis, F., Gagneur, J., Goedel, A., Laugwitz, K. L., Dorn, T., Moretti, A. Epicardioid single-cell genomics uncover principles of human epicardium biology in heart development and disease. (*Nature Biotechnology;* in press) *Equal contribution

4. Poch, C., Foo, K. S., De Angelis, M. T., Jennbacken, K., Santamaria, G., Bähr, A., Wang, Q., Reiter, F., Hornaschewitz, N., **Zawada, D.,** Bozoglu, T., My, I., Meier, A., Dorn, T., LehtieN, M., Tsoi, Y. L., Hovdal, D., Hyllner, J., Schwarz, S., Jurisch, V., Sini, M., Fellows, M. D., Cummings, M., Clarke, J., Baptista, R., Eroglu, E., Tomasi, R., Dendorfer, A., Gaspari, M., Cuda, G., Krane, M., Sinnecker, D., Hoppmann, P., Kupatt, C., Fritsche-Danielson, R., Moretti, A., Chien, K. R., Laugwitz, K. L. (2022) Migratory and anti-fibrotic programmes define the regenerative potential of human cardiac progenitors. *Nature Cell Biology*, 24, 659–671.

5. Dorn, T., Kornherr, J., Parrotta, E. I, **Zawada, D.,** Ayetey, H., Santamaria, G., Iop, L., Mastantuono, E., Sinncker, D., Goedel, A., Dirschinger, R. J., My, I., Laue, S., Bozoglu, T., Baarlink, C., Ziegler, T., Graf, E., Hinkel, R., Cuda, G., Kääb, S., Grace, A. A., Grosse, R., Kupatt, C., Meitinger, T., Smith, A. G., Laugwitz, K. L., Moretti, A. (2018). Interplay of cell-cell contacts and RhoA / MRTF-A signaling regulates cardiomyocyte identity. *EMBO J.*, 15;37(12).

6. Rawat, H.*, Kornherr, J.*, **Zawada, D.,** Bakhshiyeva, S., Kupatt, C., Laugwitz, K-L., Bähr, A., Dorn, T., Moretti, A., Nowak-Imialek, M. (2023) Recapitulating porcine cardiac development in vitro: from embryo culture to expanded potential stem cell models, Frontiers in Cell and Developmental Biology, 11:1111684.

7. Jaudas, F., Bartenschlager, F., Shashikadze, B., Santamaria, G., Schnell, A., Gräber, S., Bähr, A., Cambra-Bort, M., Krebs, S., Schulz, C., **Zawada, D**., Janda, M., Caballero-Posadas, I., Kunzelmann, K., Morretti, A., Laugwitz, K-L., Kupatt, C., Saalmüller, A., Fröhlich, T., Wolf, E., Mall, M., Mundhenk, L., Gerner, W., Klymiuk, N. Increased pulmonary infiltration and attenuated phagocytosis defines perinatal dysfunction of innate immunity in Cystic Fibrosis. *bioRxiv* 2023.09.11.556517 (in revision)

8. Bolesani, E., Bornhorst, D., Iyer, L., **Zawada, D.,** Friese, N., Morgan, M., Franke, A., Schambach, A., Goedel, A., Moretti, A., Zelarayan, L., Seyfried, S., Zweigerdt, R. Synergistic WNT pathway and chromatin modulation stablizes cardiac prohgenitors from hPSC in vitro, and zebrafish heart development in vivo. (in revision)

Acknowledgements