Combined Lineage Tracing and scRNA-seq Reveals Unexpected First Heart Field Predominance of Human iPSC Differentiation

Francisco X. Galdos[1,2], Carissa Lee[1], Soah Lee[8], William Goodyer[1,3], Sharon Paige[1,3], Gabriela V. Escobar[1], Adrija Darsha[7], Aimee Beck[1], Sidra Xu[1] Rasmus O. Bak[6], Matthew Porteus[1,5], Sean M. Wu[1,2,4,*]

[1]Stanford Cardiovascular Institute, [2] Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, [3]Division of Pediatric Cardiology, Department of Pediatrics [4] Division of Cardiovascular of Medicine, Department of Medicine, [5]Department of Pediatrics, Stanford University School of Medicine, Stanford CA [6]Department of Biomedicine, Aarhus University, Aarhus, Denmark, [7] University of California San Diego School of Medicine, [8]Biopharmaceutical Convergence, Department of Pharmacy, Sungkyunkwan University

*Correspondance:
Sean M. Wu, MD, PhD
G1120 Lokey Stem Cell Building
265 Campus Drive
Stanford, CA 94305
smwu@stanford.edu

Keywords
Cardiac Development, transcription factor, cardiomyocyte differentiation, pluripotent stem cells, lineage tracing, genetic engineering

ABSTRACT

During mammalian development, the left and right ventricles arise from early populations of cardiac progenitors known as the first and second heart fields, respectively. While these populations have been extensively studied in non-human model systems, their identification and study *in vivo* human tissues have been limited due to the ethical and technical limitations of accessing gastrulation stage human embryos. Human induced pluripotent stem cells (hiPSCs) present an exciting alternative for modeling early human embryogenesis due to their well-established ability to differentiate into all embryonic germ layers. Here, we describe the development of a TBX5/MYL2 lineage tracing reporter system that allows for the identification of FHF-progenitors and their descendants including left ventricular cardiomyocytes. Furthermore, using single cell RNA sequencing (scRNA-seq) with oligonucleotide-based sample multiplexing, we extensively profiled differentiating hiPSCs across 12 timepoints in two independent iPSC lines. Surprisingly, our reporter system and scRNA-seq analysis revealed a predominance of FHF differentiation using the small molecule Wnt-based 2D differentiation protocol. We compared this data with existing murine and 3D cardiac organoid scRNA-seq data and confirmed the dominance of left ventricular cardiomyocytes (>90%) in our hiPSC-derived progeny. Together, our work provides the scientific community with a powerful new genetic lineage tracing approach as well as a single cell transcriptomic atlas of hiPSCs undergoing cardiac differentiation.

INTRODUCTION

The human heart is one of the first organs to develop during embryogenesis with critical events in progenitor specification and differentiation occurring during the first 3 weeks of human gestation (Buckingham et al., 2005; Cui et al., 2019; Hikspoors et al., 2022; Meilhac and Buckingham, 2018; Tan and Lewandowski, 2020). Due to ethical and technical limitations in the study of human embryogenesis prior to 5 weeks gestation, developmental biologists have largely relied upon animal models to study cardiac development (Hyun et al., 2021; Meilhac and Buckingham, 2018). Early studies in mammalian cardiac progenitor biology identified the presence of two definitive multipotent progenitor populations known as the first (FHF) and second (SHF) heart fields which give rise to the left and right ventricles, respectively (Cai et al., 2003; Dyer and Kirby, 2009; Meilhac et al., 2004; Mjaatvedt et al., 2001; Moretti et al., 2006; Waldo et al., 2001). Furthermore, early lineage tracing studies using the mesodermal progenitor marker, *Mesp1*, have revealed that the early specification of these lineages likely occurs during the earliest stages of gastrulation, with the FHF emerging as the first wave of cardiac progenitors, followed by the SHF (Lescroart et al., 2014; Saga et al., 1999; Scialdone et al., 2016). With the advent of single cell RNA sequencing (scRNA-seq), these progenitor populations have been extensively characterized and shown to exhibit unique transcriptional expression profiles (de Soysa et al., 2019; Hill et al., 2019; Xiong et al., 2019). Moreover, scRNA-seq profiling of murine left and right ventricles during early cardiac development have shown that transcriptional differences can be detected up to E10.5 of murine development, suggesting that early left and right

ventricular development is characterized by unique transcriptional regulatory networks (DeLaughter et al., 2016; Li et al., 2019, 2016).

While the distinct identities of the first and second heart field progenitors have been well established in the murine system, the identification of these progenitor populations within a human model has been severely limited by lack of access to human embryonic tissues. Over the past decade, the advent of human induced pluripotent stem cells (hiPSCs) has allowed for the developmental modeling of multiple different embryonic lineages *in vitro* (Holloway et al., 2020; Kanton et al., 2019; Karagiannis et al., 2019; Lian et al., 2013; Takahashi and Yamanaka, 2006; Yamanaka, 2008). In the cardiac field, small molecule-based protocols modulating WNT signaling have become standard due to their remarkable efficiency at generating large numbers of beating cardiomyocytes that can be utilized for disease modeling, drug discovery, and the study of cellular functions (Burridge et al., 2015; Chen et al., 2016; Feyen et al., 2020; Lian et al., 2013; Sacchetto et al., 2020). Several questions remain as to whether hiPSC cardiac differentiations are capable of modeling early cardiac progenitor biology as seen during *in vivo* mouse development (Protze et al., 2019a). Moreover, evidence is lacking as to whether current hiPSC differentiation protocols give rise to FHF- and SHF-derived LV and RV cardiomyocytes, respectively (Protze et al., 2019a).

A major bottleneck in the identification of these cell types during hiPSC differentiation is the lack of lineage tracing tools that have been extensively used in murine models to understand the developmental lineage contributions of progenitor populations (Barnes et al., 2010; Cai et al., 2003; Meilhac et al., 2004; Moretti et al., 2006; Tyser et al., 2020; Vincentz et al., 2017; Zhang et al., 2021). Early studies

profiling the expression of the T-box transcription factor, *Tbx5,* identified its specific expression at the cardiac crescent and its role as a marker of early FHF progenitors (Bruneau et al., 2001, 1999). More recently, studies using inducible CreER/LoxP lineage tracing have shown the exquisite specificity of *Tbx5* to label ventricular cardiomyocytes on the left but not the right, demonstrating a clear boundary between cell origins of left and right ventricular cardiomyocytes during embryogenesis (Devine et al., 2014). While lineage tracing tools have provided insight into the cellular contributions of the FHF in mice, no lineage tracing tool is currently available for tracing LV and RV cardiomyocytes in a human model system.

To identify FHF-progenitors and their derived cell types during hiPSC differentiation, we used a CRISPR-Cas9 targeting platform to engineer a TBX5 expression-driven, highly-sensitive, Cre/LoxP lineage tracing system in hiPSCs that contain a ventricular cardiomyocyte-specific myosin light chain-2 (*MYL2)*-tdTomato fluorescent protein. By conducting a time course analysis of cardiomyocyte differentiation, we identified a left ventricular cardiomyocyte predominant differentiation across two distinct cell lines based on the high percentage of TBX5-lineage positive ventricular cardiomyocytes (>90%). Using chemically modified lipid-oligonucleotides (CMOs), we conducted multiplexed scRNA-seq assays on 12 different timepoints across two independent hiPSC lines. Using differentiation trajectory analysis, we compared our scRNA-seq data with murine heart field development scRNA-seq data and validate the FHF origin and LV identity of cardiomyocytes generated. Finally, we conduct a comparison of our scRNA-seq data with a recently published 3D cardiac organoid differentiation (Drakhlis et al., 2021) and identify the greater potential of a 3D

system to generate SHF derived cell types. Together, our findings provide a powerful new tool for human *in vitro* cardiac development studies and a validated single cell expression atlas for identifying the human FHF lineage during *in vitro* hiPSC differentiation.

RESULTS

Generation of a TBX5-Lineage Tracing and Ventricular Reporter Line By CRISPR/Cas9 Genome Editing

Given the well-established role of the T-box transcription factor, *Tbx5,* as a specific marker of the early FHF and left ventricular lineage (Bruneau et al., 2001, 1999; DeLaughter et al., 2016; Devine et al., 2014), we engineered a fluorescent lineage tracing system that would allow for the determination of whether *TBX5* lineage tracing could correctly identify left ventricular cardiomyocytes using a human iPSC model of cardiac differentiation. Previously, our laboratory developed an *MYL2*-tdTomato construct targeting a P2A-TdTomato to the stop codon of the *MYL2* gene that was validated to specifically isolate ventricular cardiomyocytes during hiPSC differentiations (Chirikian et al., 2021).To construct a reporter system that could isolate left ventricular cardiomyocytes, we employed a triple construct system that would allow for the identification of *MYL2*-positive ventricular cardiomyocytes and the identification of *TBX5*-lineage-positive left ventricular cardiomyocytes **(Figure 1A).** To lineage trace *TBX5* expressing cells during hiPSC differentiation, we developed two new genetic constructs based on a P2A self-cleaving peptide system that allows the tethering of genetic construct expression with a gene of interest (Liu et al., 2017). The first construct consists of tandem P2A-Cre Recombinase genes that is targeted to replace the stop

codon of *TBX5* (**Figure 1B)**. The second construct consists of a constitutively active CMV promoter followed by a floxed stop cassette and a downstream TurboGFP with the goal to only allow for TurboGFP expression after the excision of the stop cassette by Cre (**Figure 1B**). Using CRISPR/Cas9 genome editing, we first targeted the MYL2-tdTomato construct into two hiPSC lines derived from healthy donor patients **(Figure 1B)**. Using an inside-out PCR strategy (Galdos et al., 2021; Ran et al., 2013), we confirmed the successful integration of the MYL2 construct based on the integration of the 5' and 3' ends of the construct and selected a heterozygous integrated clone **(Figure 1C)**. We subsequently integrated the CMV-Lox-STOP-Lox-TurboGFP construct into the CCR5 safe harbor site and confirmed successful integration via inside-out PCR **(Figure 1C)**. Next, we integrated the P2A-Cre-P2A-Cre construct into the *TBX5* locus by replacing the stop codon of the gene. To ensure maximal sensitivity of our lineage tracing system, we integrated the *P2A-Cre* construct in a homozygous manner to ensure high expression of Cre recombinase upon expression of TBX5 **(Figure 1C)**. Importantly, the expression of TBX5 is preserved with this approach since the Cre recombinase is inserted after the TBX5 coding sequence and the fusion protein product undergoes self-cleaving at the P2A sequence (Liu et al., 2017). Using sanger sequencing, we validated the in-frame integration of the P2A sequences of both the MYL2 and TBX5 constructs (**Figure 1D)**. Lastly, we confirmed the maintenance of pluripotency after three rounds of genome editing by immunostaining of pluripotency marker *OCT4* (**Figure 1E),** thus demonstrating the successful genome editing of three independent genetic constructs into two different hiPSC lines.

TBX5-Lineage/MYL2 Reporter System Reveals Predominance of Left Ventricular Differentiation Using Small Molecule Wnt Protocol

To determine the proportion of hiPSC-derived cardiomyocytes that exhibit a TBX5-lineage positive phenotype, we conducted cardiac differentiations using a widely published differentiation protocol consisting of biphasic activation and subsequent inhibition of WNT signaling using small molecules (**Figure 1A, see Methods**) (Lian et al., 2013). We employed a strategy where we conducted a high throughput flow cytometry analysis of cardiac troponin, TurboGFP, and tdTomato expression across multiple timepoints during cardiac differentiation and across the two independent cell lines containing our reporter system **(Figure 2A, 2B, and 2C)**. Analysis of *TNNT2* expression from day 3 to day 30 of differentiation revealed a gradual upregulation of *TNNT2* expression starting at day 7 of differentiation **(Figure 2B and 2D)**, with the greatest increase in *TNNT2*+ cardiomyocytes being reported between day 7 and 11 of differentiation. Overall cardiac differentiation at day 30 across both reporter lines (WTC and SCVI-111) averaged 93.2%±0.80% and 92.1%± 1.60% of *TNNT2*+ cells out of total cells analyzed, respectively (**Figure 2D**). We further we analyzed the proportion of cells that were positive for TurboGFP+ between day 3 and day 30 of differentiation **(Figure 2B and 2E)** and found a large increase in TurboGFP+ cells between days 7 and 11 with a continuous increase in the level of GFP signal as the differentiation proceeded. By day 30, both WTC and SCVI-111 lines exhibited 99.0%±0.21% and 93.2±0.84% of total TNNT2+ cardiomyocytes expressing TurboGFP, respectively, indicating a predominance of cardiomyocytes from the TBX5 lineage (**Figure 2E).**

Since TBX5 is known to be expressed in both atrial and ventricular cardiomyocytes, we next determined the percentage of ventricular cardiomyocytes that are within the TBX5-lineage by analyzing the proportion of MYL2-tdTomato+ cardiomyocytes that express TurboGFP. As *MYL2* expression gradually increases over time during both hiSPC cardiac differentiation and *in vivo* development and is highly tied to the overall maturational status of hiPSC-derived cardiomyocytes (Bizy et al., 2013; Chirikian et al., 2021; DeLaughter et al., 2016; Li et al., 2016; O'Brien et al., 1993), we show, consistent with these previous studies, that the percentage of MYL2-tdTomato+ cardiomyocytes increases between days 15 and 30 of cardiomyocyte differentiation with some line to line variability likely tied to variation in hiPSC-CM maturation rate (**Figure 2C and 2F)**. Across day 15, 20, and 30 we observed that for both cell lines the proportion of ventricular cardiomyocytes marked by the tdTomato reporter were more than 95% for TurboGFP indicating that nearly all ventricular cardiomyocytes were within the TBX5-lineage (**Figure 2G, Figure 2F, Figure 3A)**.

We further validated the expression kinetics of our reporter system by conducting bulk gene expression analyses using RT-qPCR across multiple timepoints during hiPSC differentiation (**Figure 3B-3E)**. We evaluated the expression of known markers of early FHF progenitors and left ventricular cardiomyocytes, *HAND1* and *TBX5* (Barnes et al., 2010; de Soysa et al., 2019; Devine et al., 2014; Vincentz et al., 2017). We also evaluated the expression of Cre recombinase throughout differentiation. Relative to day 0 we observed that all three markers exhibited high expression values with *HAND1* exhibiting more than 50,000 fold upregulation relative to day 0 by day 7 of differentiation across both lines **(Figure 3B)**. Similarly, by day 30 of differentiation *TBX5* exhibited

nearly 3,000-fold upregulation relative to day 0 (**Figure 3B)**. The expression of Cre recombinase incrased over time and was consistent with the expected increase in TurboGFP expression observed in the flow cytometry data (**Figure 3B)**.

In addition to analyzing FHF marker expression, we also examined the expression of SHF markers such as ISL1, FGF8, and TBX1 (**Figure 3C)** (Cai et al., 2003; Park et al., 2008; Rana et al., 2014). While ISL1 has been reported to be expressed in the early FHF lineage (Ma et al., 2008), a well-established observation is that ISL1 expression is sustained during the emergence of the SHF (Cai et al., 2003). Interestingly, we observed in both the WTC and SCVI-111 that ISL1 expression peaked at day 5 of differentiation, which is indicative of an early cardiac progenitor population at that timepoint. We did not observe a sustained expression of ISL1 and rather observed its downregulation over time. Similarly, FGF8 has been reported to be important for early cardiomyocyte differentiation in both the FHF and SHF, however, its expression is maintained within SHF progenitors during cardiogenesis. We observed that while FGF8 expression was present at the mesodermal stage of differentiation (**Figure 3C**), a significant drop in expression was observed after day 7. *TBX1*, a pharyngeal endoderm and mesoderm marker (Chapman et al., 1996; Mesbah et al., 2012; Rana et al., 2014; Vitelli et al., 2002), also known as marker of the anterior second heart field (Liao et al., 2008; Meilhac and Buckingham, 2018; Nevis et al., 2013), was unexpectedly upregulated during gastrulation and early cardiac progenitor stages of differentiation but declined after day 7 of differentiation with some line-to-line and batch-to-batch variability (**Figure 3C**). This lack of SHF markers upregulation after day 7 contrasts with the continued or even increased expression of FHF markers *HAND1* and *TBX5*, thus

indicating the predominance of FHF-derived cells at later stages of differentiation. Lastly, we also observed a continued increase in cardiomyocyte (*TNNT2)* and ventricular cardiomyocyte (*MYL2)* (**Figure 3E**) markers consistent with our flow cytometry data (**Figure 2B and 2E)**.

scRNA-seq Time Course Reveals Three Major Developmental Trajectories During hiPSC Differentiation

While our Cre/LoxP-based fluorescent reporter system showed a predominance of *TBX5*-lineage cardiomyocytes (**Figure 2D**) and our qPCR data showed an upregulation of FHF, but not SHF, gene markers at late stages of differentiation (**Figure 3B, 3C**), we asked whether scRNA-seq may help to pinpoint the developmental trajectories that bifurcates between FHF and SHF during hiPSC differentiation. Using sample multiplexing with CMO in our scRNA-seq experiment where we captured cells from both the WTC and SCVI-111 lines and at 12 different timepoints hiPSC cardiac differentiation (Days 0 to 7 and 11, 13, 15, and 30) for a total of 27,595 cells after sample demultiplexing and quality control **(Figure 4A, Supplementary Figure 1, 2, and 3**).

Using a well published batch correction method known as the mutual nearest neighbor algorithm (Haghverdi et al., 2018), we batch corrected the effect of the scRNA-seq runs and conducted downstream dimensionality reduction and unsupervised clustering **(Supplementary Figure 4A)**. Annotation of unsupervised clusters revealed 13 major populations during cardiac differentiation **(Figure 4 and Supplementary Figure 4A)**. During the early days of differentiation we identified cell populations

consistent with pluripotent stem cells, primitive streak, and definitive endoderm populations marked by the expression of POU5F1, MIXL1, and SOX17, respectively **(Supplementary Figure 4B)** (Mead et al., 1996; Pijuan-Sala et al., 2019; Takahashi and Yamanaka, 2006; Tyser et al., 2021).. By day 3 of differentiation, we observed the emergence of mesodermal progenitors, early cardiac progenitors, late cardiac progenitors, cardiomyocytes, and epicardial populations marked by the expression of MESP1, ISL1, NKX2-5, TNNT2, and WT1, respectively **(Supplementary Figure 4B)** (Barnes et al., 2010; Christoffels et al., 2009; Rudat and Kispert, 2012; Zeng et al., 2011)..

Given the identification of distinct cell types within our single cell data, we asked whether we could further identify developmental trajectories during hiPSC cardiac differentiation. We used a Python based bioinformatic pipeline known as STREAM to automatically identify and visualize differentiation trajectories within our scRNA-seq data (Chen et al., 2019). STREAM uses a low dimensional manifold such as a UMAP plot and calculates a principal graph that identifies differentiation paths throughout the dataset. Intriguingly, the STREAM algorithm fit a principal graph that identified two major bifurcations during hiPSC differentiation **(Figure 4D and Supplementary Figure 5C).** Along with the fitting of a principal graph, we calculated STREAM pseudotime by setting the pluripotent stem cell cluster as the root of the differentiation. We then reordered and projected cells along the principal graph according to increasing pseudotime to visualize our annotated cell types as they progressed during differentiation (**Figure 4D-F and Supplementary Figure 5C).** To further characterize the distinct trajectories identified by STREAM, we correlated the expression of gene

expression with the cell pseudotime along each unique branch (**Table 6)**. This analysis recovered multiple gene markers known to be expressed during the course gastrulation and cardiac development. The first bifurcation identified occurred at the late primitive streak stage and represents the bifurcation into mesodermal and endodermal cell lineages. Gene markers identified for the endodermal lineage includes markers such as *EPCAM and FOXA2,* as well as hepatic-like endodermal markers such as *APOA1* and *AFP* (**Figure 4G)** (Hurrell et al., 2019; Pijuan-Sala et al., 2019; Sarrach et al., 2018)**.** Along the mesodermal differentiation path, we observed the upregulation of *MESP1* followed by ISL1 expression as cells became specified along the cardiac progenitor lineage. ISL1 expression preceded the expression of NKX2-5 and encompasses both endodermal precursors as well as cardiac progenitors, supporting the earlier but less cardiac specific expression of ISL1. At the second bifurcation, we identified an NKX2-5 population that bifurcated into myocardial and epicardial lineages. Top ranking markers for the epicardial lineage included known epicardial marker *TBX18, WT1, TCF21,* and *IGF2*, while the myocardial lineage was characterized by the elevated expression of sarcomeric genes **(Figure 4G and Supplementary Figure 5D)** (Christoffels et al., 2009; Hu et al., 2020; Li et al., 2011; Rudat and Kispert, 2012; Tandon et al., 2013). Within the epicardial lineage, we observed a high expression of *HAND1* along with the enrichment of extracellular matrix genes such as *COL3A1, VIM,* and *COL1A1* at the epicardial progenitor population, suggesting the emergence of *WT1* positive epicardial cells from a *HAND1* expressing precursor (**Table 6).** Together, STREAM revealed the emergence of an epicardial and myocardial lineage from a common cardiac progenitor during hiPSC differentiation.

Predominance of FHF Cardiomyocyte Differentiation by hiPSCs Confirmed by Comparison with scRNA-seq Data from Murine Heart Field Development

To confirm the FHF cardiomyocyte-predominant differentiation of hiPSCs that we observed in our TBX5 lineage tracing (**Figure 2D**) and qPCR (**Figure 3B, 3C**) data, we conducted a comparison between previously published murine scRNA-seq heart field data (de Soysa et al., 2019; Hill et al., 2019; Pijuan-Sala et al., 2019) and our hiPSC cardiac differentiation data. We clustered data from the murine datasets representing 7 major cell types of interest including nascent mesoderm, heart field progenitors, epicardial cells, left and right ventricular cardiomyocytes, and outflow tract cardiomyocytes (**Figure 5A)**. As previously reported (de Soysa et al., 2019; Hill et al., 2019), we observed a bifurcation of the FHF and SHF cells from the nascent mesoderm and observed a clear contribution of both heart field progenitors to the development of left and right ventricular cardiomyocytes with FHF cells contributing to the LV and SHF cells contributing to the RV/OFT. Intriguingly, we observed the epicardial lineage branched off from FHF progenitor cells (**Figure 5B**). This observation is consistent with the recent lineage tracing literature that indicated the contribution of a subset of FHF progenitors to both left ventricular cardiomyocytes as well as epicardial cells (Tyser et al., 2020; Zhang et al., 2021).

To further dissect the gene expression changes that occur during FHF and SHF development, we replotted each FHF and SHF cell population using STREAM and displayed the expression of known FHF and SHF progenitors markers during murine heart field development and hiPSC cardiac differentiation (**Figure 5B**). Consistent with the literature, Tbx5 and Hcn4 were upregulated in FHF cells during mouse development

in vivo and are completely absent in the aSHF lineage (**Figure 5C)** (Andersen et al., 2018; Bruneau et al., 2001; Devine et al., 2014; Später et al., 2013). Of note, Tbx5 appeared to gradually increase in expression during the transition from FHF progenitors to LV CMs with a gradual downregulation as development progresses, indicating a dynamic expression pattern through the course of development. In contrast to the FHF markers, we observed a clear upregulation of aSHF markers Tbx1 and Fgf8 during early aSHF progenitor development in mice with Fgf8 exhibiting its highest expression pattern at prior to the bifurcation between OFT CM and RV CM **(Figure 5C)** (Nevis et al., 2013; Park et al., 2008; Vitelli et al., 2002). As expected, the expression of Tbx1 and Fgf8 were absent in FHF progenitors, albeit a low Fgf8 expression was found in early LV CM, which is expected given the role of Fgf signaling during early cardiomyocyte differentiation (Khosravi et al., 2021; Reifers et al., 2000). We next plotted the expression of these markers during our hiPSC cardiac differentiations and observed a striking consistency in marker expression with the FHF lineage in the mouse (**Figure 5D**). Importantly, we observed the upregulation of TBX5 starting at the late progenitor stage and increasing during the myocardial branch of the differentiation, like the kinetics observed in the FHF trajectory of the mouse (**Figure 5D**). Similarly, HCN4 expression remained high during cardiomyocyte differentiation further supporting the left ventricular identity of the myocardial branch given reported role of HCN4 as an early FHF and LV marker (Später et al., 2013).

Having observed the similarities between murine FHF development and our hiPSC differentiations, we asked whether our hiPSC differentiations exhibited a ventricular specific differentiation trajectory. We confirmed the ventricular specific

trajectory of our hiPSC differentiation given the gradual upregulation of the ventricular specific Iroquois transcription factor, IRX4 (**Figure 5E)** (Nelson et al., 2016, 2014). Previous studies have shown IRX4 to mark early ventricular specific cardiomyocytes, which we effectively observed in both the left and right ventricular differentiation lineages in the murine data (Nelson et al., 2014). Lastly, consistent with the MYL2-tdtomato expression pattern observed with our reporter lines (**Figure 2E)**, we observed the gradual upregulation of MYL2 in both the human and murine datasets. Together, our analysis provides evidence for the predominance of FHF ventricular cardiomyocyte development during hiPSC differentiation.

Comparison of 2D and 3D Cardiac Differentiation Uncovers Potential of Organoid System for SHF Generation

Recently, multiple groups have proposed the use of 3D differentiation to better model chamber morphogenesis and potentially model both first and second heart field development *in vitro* (Andersen et al., 2018; Drakhlis et al., 2021; Protze et al., 2019b; Rossi et al., 2021). Interestingly, murine gastruloid and precardiac organoids have been shown to exhibit aspects of first and second heart field development (Andersen et al., 2018; Rossi et al., 2021); however, data on the ability to generate both heart fields in human iPSCs are lacking. Given the predominance of FHF progenitors and LV cardiomyocytes made from our 2D hiPSC differentiation platform, we assessed whether a greater repertoire of cardiac cells can be generated from a 3D hiPSC differentiation platform by analyzing scRNA-seq data from a recently published cardiac organoid study (Drakhlis et al., 2021). This paper demonstrated the close relationship

between anterior endoderm lineages and anterior second heart field cells during hiPSC differentiation by showing that anterior foregut endoderm can be generated alongside cardiac lineage cell types (Kelly et al., 2001; Rochais et al., 2009). We compared scRNA-seq data from our hiPSC-derived cardiac cells with those generated by the Drakhlis etal group using their 3D differentiation protocol (Drakhlis et al., 2021). We first conducted a cross dataset comparison of FHF and SHF marker analysis where we focused exclusively on cell types composing the myocardial lineages of both datasets (**Figure 6A)**. In our 2D cardiac differentiation we observed a clear increased expression of FHF markers TBX5, HCN4 and HAND1 during differentiation with the absence of SHF markers FGF8 and TBX1 (**Figure 6B and 6C)**. Interestingly, the day 13 data from Drakhlis etal showed two clusters exhibiting distinct transcriptional expression patterns suggestive of FHF and SHF progenitors (**Figure 6D)**, with both clusters appearing to give rise to TNNI1 and NKX2-5 positive cardiomyocytes. Cluster 2 and 8 of the cardiac organoid data indicated a high expression of TBX5, HAND1, along with the upregulation HCN4 during differentiation which was consistent with the FHF trajectory we observed in our 2D data (**Figure 6E)**. Interestingly, we found that cells in cluster 4, 7, 10 of Drakhlis et al exhibited a high expression of TBX1, FGF8, and ISL1 all of which were consistent with a SHF identity (Cai et al., 2003; Mesbah et al., 2012; Park et al., 2008). Moreover, we also observed a cardiomyocyte population emerging from the TBX1+ population that was negative for FHF markers TBX5 and HCN4, but highly expressing ISL1 and HAND1, which is suggestive of an OFT CM population that is known to express HAND1 and emerge from ISL1 expressing SHF progenitors.

To determine whether this TBX5-ISL1+HAND1+ cardiomyocyte population was indeed OFT CMs, we conducted a joint analysis of the cardiomyocytes from our 2D differentiations and the Drakhlis etal cardiac organoid. Using unsupervised clustering, we observed a cardiomyocyte population emerge from this study that displayed the absence of TBX5. Interestingly, we observed that the rest of the organoid cardiomyocytes co-clustered with the cells from our 2D differentiations indicating shared gene expression profiles (**Figure 6F)**. We conducted differential gene expression analysis of the putative OFT CM cluster and the rest cardiomyocytes and found a statistically significant enrichment of markers associated with OFT development including HAND1, BMP2, WNT5A, and PITX2 (Délot et al., 2003; Li et al., 2016; Ma et al., 2013; Schleiffarth et al., 2007), while the rest of the cardiomyocytes exhibited high expression of markers associated with early LV development such as *TBX5* and *NPPA* (**Figure 6G)** (Li et al., 2016). Overall, these data thus provide strong evidence for the emergence of a SHF derived cell type within a 3D organoid differentiation protocol and reinforce the left ventricular identity that we identified using a standard 2D small molecule differentiation protocol.

DISCUSSION

Over the past decade, the development of highly efficient cardiac directed differentiation protocols have significantly advanced efforts to model cardiovascular diseases *in vitro* (Burridge et al., 2015; Lian et al., 2013). While non-human model systems have provided significant insight into the developmental lineages that contribute to cardiac development, the inaccessibility of early human embryonic tissue

has significantly limited the creation of an *in vivo* reference atlas of human heart field development. Importantly, questions remain as to whether the human iPSC system can be used to efficiently generate cell types representing distinct chambers of the heart such as left/right ventricular cardiomyocytes, outflow tract cardiomyocytes, or atrioventricular canal cells. Currently, a major gap exists in the ability to model the development of specific structures of the human heart in part due to the lack of genetic tools to mark distinct cell lineage *in vitro.* Importantly, the ability to identify chamber specific cardiomyocyte is vitally important to modeling the early developmental mechanisms that give rise to structural congenital heart defects (Doyle et al., 2015; Reller et al., 2008; van der Linde et al., 2011).

In this paper, we sought to build a novel genetic lineage tracing tool to elucidate the identities of cardiomyocytes generated using a well cited and standard differentiation protocol. We successfully implemented a *TBX5* lineage tracing scheme into the hiPSC system by targeting a highly sensitive Cre recombinase to the 3' end of the endogenous TBX5 locus. Moreover, by including an *MYL2*-TdTomato direct reporter into our hiPSC lines, we evaluated the percentage of ventricular cardiomyocytes that were descended from *TBX5* expressing precursors. By implementing a lineage tracing method, our reporter system provides several advantages for studying the descendants of FHF progenitor cell types compared with previous approaches. While previous studies have shown the isolation of TBX5-postive cell types using direct reporter schemes (Zhang et al., 2019), a major advantage of a lineage tracing approach is the permanent and robust labeling of descendants from progenitor populations or cardiomyocyte populations that express *TBX5,* thus allowing

for the evaluation of cell type identity at later stages of differentiation when TBX5 expression is downregulated. Moreover, the combination of our lineage tracing and a ventricular reporter system allowed us to evaluate the proportion of definitive ventricular cardiomyocytes from the TBX5-lineage during iPSC differentiation that generates atrial cardiomyocytes as well.

Surprisingly, our data indicates that in two distinct hiPSC lines, TBX5-lineage-positive cardiomyocytes represent more than 95% of all cardiomyocytes generated. Furthermore, our scRNA-seq time course data further revealed the emergence of a core cardiac progenitor differentiation trajectory that displayed a gradual upregulation of FHF markers and no expression of known SHF genes. The unexpected finding of FHF predominance raises important questions on whether the widely used small molecule WNT modulation protocol used for the generation of hiPSC cardiomyocytes is biased towards the generation of FHF lineage cardiomyocytes. An additional finding from our scRNA-seq analysis that was quite intriguing was the bifurcation of cardiac progenitors into epicardial and myocardial cells. Importantly, the epicardial lineage arose from a cell cluster that exhibited high expression of HAND1 representing early FHF progenitor cells. Recently, two groups have published lineage tracing results in mice that suggest a subset of FHF progenitor cells exhibit contributions to both the proepicardial and myocardial lineages (Tyser et al., 2020; Zhang et al., 2021). Consistent with these studies, our data provides evidence of this bifurcation *in vitro* and demonstrates the high expression of left ventricular markers HAND1, TBX5, and HCN4 along the myocardial lineage. The high expression of HAND1 during the earliest stages of mesoderm differentiation further suggests that even at the earliest stages of mesoderm

specification the progenitors in our hiPSC differentiations were already bound for a FHF fate given the known restriction of HAND1 to give rise to FHF derived cell types during early embryonic development (Barnes et al., 2010; Vincentz et al., 2017). This observation is consistent with a growing body of literature suggesting that the bifurcation of first and second heart field occurs during the earliest patterning of mesoderm during its emergence from the primitive streak (Lescroart et al., 2014).

Having established the FHF identity of our cardiac differentiation conducted in 2D, we used our dataset as a reference point for determining whether a recently published cardiac organoid protocol could give rise to a greater diversity of cardiac progenitor cell types. While multiple 3D protocols have recently been published, the study by Drakhlis etal is the only protocol to date to demonstrate the co-emergence of anterior endoderm cell types that are closely related the emergence of an anterior cell type such as the anterior second heart field during embryonic development. Interestingly, our comparison revealed the presence of *bona fide* second heart field progenitors within the organoids generated in the Drakhlis etal protocol. Importantly, we were able to combine cardiomyocytes generated from our study with those from the Drakhlis et al 3D protocol and reveal the identity of a true OFT cell type present in the Drakhlis et al dataset (**Figure 6F**). This analysis thus reveals the promising application of 3D differentiation protocols for generating a greater diversity of cardiac cell types.

Overall, here we provide a novel reporter system that allows for the identification of left ventricular cardiomyocytes during the course hiPSC differentiation. By allowing for the permanent labeling of *TBX5* descended cell types, we envision our system being used to conduct more complex studies to study chamber specific cardiomyocytes in the

context of congenital heart diseases as well as for the development of novel hiPSC differentiation protocols for generating both left and right ventricular cardiomyocytes. Moreover, by generating a scRNA-seq dataset profiling multiple consecutive day of hiPSC cardiac differentiation we provide here reference atlas of the differentiation events that occur during human *in vitro* cardiac development. Together, our study provides extensive evidence of the identification of the FHF lineage in a human system and reveals the early bifurcation of this lineage into an epicardial and myocardial lineage.

MATERIALS AND METHODS

Cell Lines:

Human induced pluripotent stem cell lines used in this study were provided by the Stanford Cardiovascular Institute Biobank (SCVI-111, Sendai virus reprogrammed peripheral blood mononuclear cells, healthy male with normal karyotype, 46, XY). The WTC-11 (reprogrammed from healthy males with normal karyotype, 46, XY) hiPSC line was graciously provided by Bruce Conklin's laboratory at the University of California, San Francisco. Studies involved human iPSC approved under protocol #460 of the Stanford Stem Cell Research Oversight (SCRO) committee.

Cardiac Differentiation:

HiPSCs were maintained in DMEM/F12 (ThermoFisher) supplemented with essential eight (E8) (henceforth referred to as E8 media) that is prepared in-house as previously

described (Burridge et al., 2015) and cultured on growth factor reduced Matrigel (Corning) coated plates at a 1:350 dilution. Upon reaching 75-80% confluency, hiPSCs were passaged using 0.5 mM EDTA in PBS for 8 minutes at 37°C. Passaging was conducted with gentile dissociation of cell clusters and plated in E8 media supplemented 10µM ROCK inhibitor. Passaging was performed using 1:12 splitting ratio to achieve approximately 10,000 cells per cm^2. 24 hours after passaging media was changed to E8 media. Daily media changes were conducted until cells reached 90-95% confluency at which point media was change to RPMI-1640 containing 6µM CHIR99021 (Selleckchem) and 2% B27 minus insulin supplement (Invitrogen). Two days after initial treatment with CHIR, media was changed to RPMI-1640 with 2% B27 minus insulin for 24 hrs. Between day 3-5, media was changed to 2µM C59 in RPMI-1640 media with 2% B27 minus insulin. On day 5 of differentiation, media was changed for RPMI-1640 with 2% B27 minus insulin for 48 hrs and was subsequently changed to RPMI-1640 with 2% B27 Plus Insulin for another 48 hrs. On day 9, cells underwent glucose deprivation for 48 hrs to purify cardiomyocytes by changing media to RPMI-1640 minus glucose with 2% B27 Plus insulin. Cardiomyocytes were subsequently maintained in RPMI-1640 with glucose with 2% B27 Plus Insulin.

<u>Donor Construct Plasmids</u>

Cre recombinase gene sequence was provided by Connie Cepko lab (Addgene plasmid # 13775) (Matsuda and Cepko, 2007). TurboGFP gene was obtained from the pMaxGFP plasmid obtained from the Lonza P3 Primary Cell 4D-Nucleofector X Kit L.

Plasmids were constructed using plasmid construct service by Genscript Biotech. All donor plasmids are available upon request.

CRISPR/Cas9 Genome Editing:

Genetic constructs were targeted to hiPSCs following the schematic presented in Figure 1. TurboGFP sequence was cloned from the pMaxGFP plasmid from Lonza P3 Primary Cell Nucleofection kit. Protocol for targeting of genetic constructs was followed as previously described (Galdos et al., 2021). Briefly, hiPSCs were dissociated into a single cell suspension at 75% confluency using an Accutase-EDTA solution (Millipore) containing a 0.02% blebbistatin (Sigma-Aldrich). Dissociation reaction was quenched using a solution of E8 media with 10µM ROCK inhibitor an 0.02% blebbistatin. Cells were then pelleted by centrifugation at 200g for 3 minutes. We subsequently conducted nucleofection of the dissociated cells using the Lonza P3 Primary Cell 4D-Nucleofector X Kit L. A transfection mix was prepared containing the Lonza P3 Solution, Supplement, single guide RNA/Cas9 expressing plasmid (1µg), and the donor template plasmid (3µg). Cells were electroporated using a Lonza 4D Nucleofector machine using protocol number “CM150”. After electroporation, 1 mL of E8 media supplemented with 10µM ROCK inhibitor was gently added to the cuvette containing the cells. Cells were allowed to rest for 10 minutes after which they were plated onto two wells of a 6 well plate. 24 hrs after plating, media was changed to regular E8 media, and regular maintenance until cells reach 50% confluency. At the 50% confluency mark, cells were dissociated into single cell suspensions and passaged onto 6 well plates at 1000 cells per well of 6-well. We subsequently maintained the transfected hiPSCs in E8 media supplemented

with appropriate antibiotic for selection of successfully targeted cells. Concentrations were as follows for the constructs targeted: TBX5-Cre (Hygromycin 150 µg/mL), CCR5-CLSL-TurboGFP (G418 150 µg/mL), and MYL2-Tdtomato (Puromycin 0.2 µg/mL). After 5 days of treatment with antibiotic, cells were switched to regular E8 to allow for colony expansion derived from single cells plated. After 4 days of colony expansion, colonies were picked into 24-well plates containing E8 plus 10 µM ROCK inhibitor and were expanded for downstream DNA extraction using Qiagen DNeasy Kit and for cell freezing using Bambanker.

PCR validation of construct integration was conducted using the “Inside-Outside” approach where one primer was designed outside of the homology arms of the donor template and one primer was designed insight of the construct to be integrated. PCRs were conducted using the GoTaq Master Mix (Promega Corporation), and products were run on a 1% agarose gel in 1X TAE Buffer.

Flow Cytometry:

A Beckman Coulter CytoFLEX flow cytometer was used for high throughput analysis of TNNT2, TurboGFP, and TdTomato expression across time of hiPSC cardiac differentiation. On day of timepoint collection, cells were dissociated into single cell suspensions by incubating in 10X TrypLE Select (ThermoFisher) for 5 minutes at 37 degrees Celsius. For later stage cardiomyocytes (day 15 and onwards) incubation time was extended to 10 minutes to achieve single cell dissociation. Cells were subsequently pelleted by centrifugation at 200g for 5 minutes. Cell pellets were resuspended in 4% PFA for 10 minutes and were rinsed with a 5% Fetal Bovine Serum solution in 1x PBS.

Cells were permeabilized in a 0.5% Saponin solution containing 5% FBS in 1X PBS (hereafter referred to as Saponin Solution). After permeabilization cells were incubated for 45 minutes in a monoclonal mouse anti-Troponin primary antibody (ThermoFisher Cat**.** MA5-12960) at a 1:200 dilution in 0.5% saponin solution. Cells were rinsed twice in saponin solution and then incubated in secondary antibody AlexaFluor 647 goat anti-mouse (ThermoFisher) at a 1:1000 dilution in 0.5% saponin solution. Cells were subsequently rinsed in 1x PBS twice and analyzed using CytoFLEX flow cytometer.

RNA Extraction and Quantitative RT-PCR

Cells were collected for RNA extraction by dissociation with 10X TrypLE Select and pelleted as described in the Flow Cytometry section. Cell pellets were then resuspended in 300 µL of TRIZOL reagent (Invitrogen) at room temperature for 3 minutes. After complete resuspension of the cells, RNA was extracted using the Zymo DIRECT-Zol extraction kit per manufacuter's protocol. Purified RNA was reverse transcribed into cDNA using the High-Capacity RNA-to-cDNA kit (ThermoFisher). Quantitative PCR was subsequently run on a Biorad qPCR 384-well machine using the Biorad SYBR qPCR master mix.

Sample Preparation for Multiplexed Single Cell RNA Sequencing

To prepare cells for single cell RNA sequencing, we dissociated cells at desired timepoints by incubating in 10X TRYPLE Select for 5 minutes at 37 degrees Celsius. Cells were gently dissocated by repeated pipetting. For later timepoints (Days 15 and 30), we extended the TRYPLE Select incubation by 5 minutes in order to ensure cells

were fully dissociated. After single cell suspensions were obtained, TRYPLE Select was neutralized with an equal volume of replating media consisting of 10% Knockout Serum Replacement, 2% B27 Plus Insulin Supplement, and RPMI-1640 with glucose. Cells were centrifuged at 200g for 5 minutes to obtain cell pellets. For days 0 to 6 we resuspended cells in BamBanker freezing medium and control rate freezed vials of cells at each timepoint collected to obtain 2 million cells per freezing vial. For days 7 onwards we froze down cells in a cardiomyocyte freezing medium consisting of 90% Knockout Serum Replacement and 10% cell culture grade DMSO.

For running the scRNA-seq experiment, we thawed the desired timepoints for each experimental run by thawing vials of cell at 37 degrees Celsius and adding replating media dropwise to each vial. Cells were then centrifuged at 200g for 5 minutes to obtain cell pellets. We conducted two washes in sterile 1X DPBS for each sample to wash away any remaining Knockout Serum that could interfere with the chemically modified-oligonucleotide (CMO) staining. Using the 10X Genomics CellPlex kit, we added 100 µL of CMO to each cell pellet and resuspended the cells to allow the lipid conjugated oligonucleotides to bind to each of our samples. Each CellPlex CMO consists of a unique barcode that is used for identifying individual samples that are run within a single channel of a 10X Genomics ship. After incubating for 5 minutes, 1000 µL of sterile 1X DPBS was added to each sample, and all samples were centrifuged for 5 minutes as 200g. Cells were transferred to 5 mL eppendorfs which allowed for two more 1x DPBS washes with a total of 5 mL DPBS for each sample. Washing of unbound CMO was critical to ensuring minimal cross contamination of CMOs when combining samples. After washing, we proceeded to follow the 10X Genomics 3’ 3.1 with CellPlex

protocol for cell capture and combined samples as listed in **Table 4**. We aimed to capture a total of 30,000 cells per well of a 10X Genomics GEM Chip. After cell capture, we proceeded with preparing gene expression libraries and CellPlex libraries by following 10X Genomic manufacturer's protocol. Libraries were sequenced using an Illumina NovaSeq 6000 with S4 v1.5 flowcell reagents. We sequenced the gene expression libraries at a depth of 25,000 paired-end reads per cell and the CellPlex libraries at 5,000 paired-end reads per cell. Base calling during sequencing was performed using Real Time Analysis Version 3 software.

Bioinformatic Analysis of HiPSC scRNA-seq Time Course

Raw FASTQ files were obtained for gene expression and CellPlex libraries and were aligned using CellRanger-6.0.0 using the count function. We aligned the gene expression libraries to prebuilt GRCh38 Human genome reference provided by 10X Genomics at: https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest. We aligned the CellPlex libraries using a list of CMO barcodes as a reference as found in **Table 4.** After alignment, we obtained gene by cell expression matrices containing individual counts for each gene detected for each individual cell captured. We also obtained matrices containing the counts for each CMO detected per cell.

Gene expression matrices for each single cell run were corrected for ambient RNA contamination using the SoupX package v1.5.2, which detects levels of ambient RNA contamination using empty droplets processed during the single cell capture. Following ambient RNA correction, we then imported the CMO expression matrix into

the Seurat R package version 4.1.0 and conducted log-ratio normalization using the NormalizeData function. To demultiplex each sample according to the CMO used for labeling, we ran the HTODemux function using default parameters. This function assigns cell labels according to the amount of CMO counts detected per cell. It also identifies cells that can be classified as signlets, doublets, or negative (did not stain for any of the CMO labels).

After sample demultiplexing, we removed doublet and negative cells from the dataset as part of our quality control pipeline. We then calculated the percent of all RNA counts belonging to ribosomal and mitochondrial genes. To further remove low quality, dead, or doublet cells, we calculated the median percentage of mitochondrial and ribosomal RNA counts detected as well as the total RNA counts per cell and genes detected per cell. We calculated an upper and lower cutoff for elimintation of low quality cells by calculating the threshold at three times the median absolute deviation above and below the median value of each of these quality control metrics. Cells above and below these cutoffs were eliminated and the cells passing quality control were used for subsequent analyses.

After quality control analyses, we proceeded to merge data from all three single cell runs conducted. We normalized RNA counts per cell using the NormalizeData function in Seurat and using default parameters. We then proceeded to integrate the three single cell runs by using mutual nearest neighbor batch correction algorithm and using the individual as the batch correction variable (Haghverdi et al., 2018). This was done in order to correct for the technical variation that occurs from running single cell samples in individual capture runs. To conduct the batch correction, we first identified a

common set of integration features across the three runs integrated by running the SelectIntegrationFeatures function in Seurat. After identifying highly variable features commonly found between the datasets, we further filtered these features by removing features associated with cell cycle in order to remove the effect of cell cycle from downstream analyses and clustering. We then ran the RunFastMNN function from the SeuratWrappers version 0.3.0 package. After integration we then proceeded with constructing the nearest neighbors graph by using the FindNeighbors function and used 15 principal components that were identified using the ElbowPlot method in Seurat. We then proceeded to conduct non-linear dimensionality reduction by running the RunUMAP function using 15 principal components. Lastly, we conducted unsupervised clustering using the FindClusters function. We then conducted differential gene expression analysis to annotate clusters based on literature reported cell markers.

Lineage Trajectory Analysis Using STREAM

To conduct lineage trajectory analysis of hiPSC cardiac differentiations, we imported our integrated Seurat object into an AnnData object using ScanPy v1.8.2 package in Python. After obtaining an AnnData object, we then fit a principal graph to the UMAP plot calculated using Seurat using the seed_elastic_principal_graph and elastic_principal_graph functions in STREAM version 1.1. In addition to fitting the principal graphs, we also calculated STREAM pseudotime and ordered cells along this pseudotime using the principal graph calculation functions. After calculating pseudotime, we projected cells across distinct differentiation trajectories using the plot_flat_tree function and also plotted cells in a subway plot using plot_stream_sc.

These functions allowed for the visualization of cellular differentiation along distinct differentiation paths identified by the STREAM algorithm and for downstream identification of gene markers that are expressed during each trajectory identified. After obtaining distinct differentiation trajectory branches across pseudotime, we calculated the top differentially expressed genes across each major branch identified by STREAM by using the detect_transition_markers function.

Comparison of hiPSC and Murine Cardiac Development scRNA-seq

To compare our hiPSC differentiation scRNA-seq data with murine data, we downloaded data from GEO database from three previously published datasets. We extracted cells of interest from these datasets representing nascent mesoderm, heart field progenitors, left and right ventricular cardiomyocytes, outflow tract cardiomyocytes, and epicardial cells. In order to jointly plot these cells across multiple different datasets, we used the Fast Mutual Nearest Neighbor algorithm and integrated the cells for each dataset together. We then proceeded to construct the shared nearest neighbor graph and conduct dimensionality reduction using the RunUMAP function by using the corrected principal components derived using the FastMNN algorithm.

To analyze the development of FHF and aSHF differentiation lineages, respectively, we subsetted cells that are known from literature to fall along each lineage. For example, for the FHF differentiation trajectory, we reclustered cells identified as FHF Progenitors, epicardial cells, and left ventricular cardiomyocytes. For the anterior second heart field, we clustered SHF progenitors along with right ventricular and outflow tract cardiomyocytes. After subsetting the cells, we reran the FastMNN algorithm to

recalculate the corrected PCA space for the subsetted cells and continued with deriving UMAP plots. These plots were then used to calculate principal graphs using STREAM and to identify the differentiation lineages that arise during the progenitor's differentiation trajectories. STREAM then allowed for the plotting of STREAM plots along the differentiation trajectories identified, where we probed for the expression of well-established FHF and aSHF genes. These genes were also evaluated using the the human iPSC cardiac differentiations using STREAM plots.

<u>scRNA-seq Comparison of 2D and previously published 3D hiPSC Cardiac Differentiation</u>

To compare our 2D hiPSC cardiac differentiations to a previously published cardiac organoid protocol, we downloaded data from Drakhlis et al from GSE150202 and subsetted the dataset for putative cardiomyocytes and cardiac progenitors. The Drakhlis et al dataset was generated from two individual heart forming organoids, therefore, to correct for the batch effect from these two independent organoids, we ran the FastMNN algorithm to conduct dimensionality reduction using the FindNeighbors and RunUMAP functions of Seurat. Moreover, we conducted unsupervised clustering of the Drakhlis et al dataset where we identified major clusters of cardiac progenitors and cardiomyocytes. To compare between our hiPSC 2D data and the Drakhlis et al data, we plotted the expression of multiple FHF and SHF markers using features plots. Moreover, we focused on the myocardial lineage on our hiPSC 2D differentiations to conduct a direct comparison of the cardiomyocyte differentiation lineages in the datasets.

To conduct unsupervised clustering of both the Drakhlis et al cardiomyocytes and those from our 2D differentiations, we merged only the cardiomyocytes from the datasets and batch corrected using FastMNN. We subsequently conducted dimensionality reduction using FindNeighbors and RunUMAP in Seurat and conducted unsupervised clustering of the cardiomyocyte populations. We then conducted differential gene expression analysis by using the FindMarkers function in Seurat and compared the expression of the putative outflow tract cluster with the rest of the cardiomyocyte clusters. We then plotted a volcano plot using the EnhancedVolcano package which allowed for the visualization of statistically significant upregulated and downregulated markers in the OFT cluster relative to the other cardiomyocyte populations. Adjusted P-values for the differential expression analysis were calculated using Bonferroni correction for multiple comparisons.

Statistics:

Data presented in bar graphs for flow cytometry and RT-qPCR are presented as a mean ± standard error of the mean. Two-way ANOVA was conducted for statistical analysis of flow cytometry data with first independent variable analyzed being time and second variable being cell line. One way ANOVA with Dunnet's multiple comparisons correction was conducted for statistical analysis of RT-qPCR data.

For differential gene expression analysis of single cell data, we ran FindMarkers or FindAllMarkers function in Seurat which allows for the evaluation of differentially expressed genes between cell populations of interest. For each gene evaluated, the log base 2 of the fold change between the population of interest and the

comparator population was calculated along with the adjusted p-values based on Bonferroni correction using all the genes in the dataset.

To identify differential expression of genes along distinct differentiation branches during the STREAM analysis, we conducted the leaf gene detection which involves calculating the average gene expression of genes along the leaves of the developmental trajectory. Detailed explanation for the statistical calculations conducted by the STREAM package to find differentially expressed genes along differentiation branches can be found in the original STREAM publication (Chen et al., 2019).

Data Availability Statement

All raw data for single cell RNA-sequencing has been deposited in the GEO repository as reported in **Table 1.**

ACKNOWLEDGEMENTS

We would like to thank all members of the Sean Wu laboratory for their input and feedback in the work presented in this manuscript. A special thanks to the members of the Stanford Stem Cell Institute Flow Cytometry Core facility for their training and experimental advice. We would also like to thank Sneha Venkatramen and Daniel Lee for their assistance in supply management in the laboratory. We thank Vittorio Sebastiano's laboratory for providing us access to the Stem Cell Core electroporation equipment for genome editing experiments.

FUNDING

This work was supported by the NIH/NIGMS grant 1RM1 GM131981-02, American Heart Association - Established Investigator Award, Hoffmann/Schroepfer Foundation, Additional Venture Foundation, Joan and Sanford I. Weill Scholar Fund, and NIHLBI grant R01HL13483004 (S.M.W). This work was also supported by the NIH K08 Mentored Clinical Scientist Research Career Development Award (NHLBI) (1K08HL15378501) (W.R.G.) and NIH/NHLBI F30Hl149152 grant, the Dorothy Dee and Marjorie Boring Trust Award, and the Stanford Medical Scientist Training Program (F.X.G).

REFERENCES

Andersen, P., Tampakakis, E., Jimenez, D.V., Kannan, S., Miyamoto, M., Shin, H.K., Saberi, A., Murphy, S., Sulistio, E., Chelko, S.P., Kwon, C., 2018. Precardiac organoids form two heart fields via Bmp/Wnt signaling. Nat Commun 9, 3140. https://doi.org/10.1038/s41467-018-05604-8

Barnes, R.M., Firulli, B.A., Conway, S.J., Vincentz, J.W., Firulli, A.B., 2010. Analysis of the Hand1 Cell Lineage Reveals Novel Contributions to Cardiovascular, Neural Crest, Extra-Embryonic, and Lateral Mesoderm Derivatives. Dev Dyn 239, 3086–3097. https://doi.org/10.1002/dvdy.22428

Bizy, A., Guerrero-Serna, G., Hu, B., Ponce-Balbuena, D., Willis, B.C., Zarzoso, M., Ramirez, R.J., Sener, M.F., Mundada, L.V., Klos, M., Devaney, E.J., Vikstrom, K.L., Herron, T.J., Jalife, J., 2013. Myosin light chain 2-based selection of human iPSC-derived early ventricular cardiac myocytes. Stem Cell Res 11, 1335–1347. https://doi.org/10.1016/j.scr.2013.09.003

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. Dev Biol 211, 100–108. https://doi.org/10.1006/dbio.1999.9298

Bruneau, B.G., Nemer, G., Schmitt, J.P., Charron, F., Robitaille, L., Caron, S., Conner, D.A., Gessler, M., Nemer, M., Seidman, C.E., Seidman, J.G., 2001. A Murine Model of Holt-Oram Syndrome Defines Roles of the T-Box Transcription Factor Tbx5 in Cardiogenesis and Disease. Cell 106, 709–721. https://doi.org/10.1016/S0092-8674(01)00493-7

Buckingham, M., Meilhac, S., Zaffran, S., 2005. Building the mammalian heart from two sources of myocardial cells. Nat Rev Genet 6, 826–835. https://doi.org/10.1038/nrg1710

Burridge, P.W., Holmström, A., Wu, J.C., 2015. Chemically Defined Culture and Cardiomyocyte Differentiation of Human Pluripotent Stem Cells. Curr Protoc Hum Genet 87, 21.3.1-21.3.15. https://doi.org/10.1002/0471142905.hg2103s87

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. Dev Cell 5, 877–889. https://doi.org/10.1016/s1534-5807(03)00363-0

Chapman, D.L., Garvey, N., Hancock, S., Alexiou, M., Agulnik, S.I., Gibson-Brown, J.J., Cebra-Thomas, J., Bollag, R.J., Silver, L.M., Papaioannou, V.E., 1996. Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. Dev Dyn 206, 379–390. https://doi.org/10.1002/(SICI)1097-0177(199608)206:4<379::AID-AJA4>3.0.CO;2-F

Chen, H., Albergante, L., Hsu, J.Y., Lareau, C.A., Lo Bosco, G., Guan, J., Zhou, S., Gorban, A.N., Bauer, D.E., Aryee, M.J., Langenau, D.M., Zinovyev, A., Buenrostro, J.D., Yuan, G.-C., Pinello, L., 2019. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat Commun 10, 1903. https://doi.org/10.1038/s41467-019-09670-4

Chen, I.Y., Matsa, E., Wu, J.C., 2016. Induced Pluripotent Stem Cells: at the Heart of Cardiovascular Precision Medicine. Nat Rev Cardiol 13, 333–349. https://doi.org/10.1038/nrcardio.2016.36

Chirikian, O., Goodyer, W.R., Dzilic, E., Serpooshan, V., Buikema, J.W., McKeithan, W., Wu, H., Li, G., Lee, S., Merk, M., Galdos, F., Beck, A., Ribeiro, A.J.S., Paige, S., Mercola, M., Wu, J.C., Pruitt, B.L., Wu, S.M., 2021. CRISPR/Cas9-based targeting of fluorescent reporters to human iPSCs to isolate atrial and ventricular-specific cardiomyocytes. Sci Rep 11, 3026. https://doi.org/10.1038/s41598-021-81860-x

Christoffels, V.M., Grieskamp, T., Norden, J., Mommersteeg, M.T.M., Rudat, C., Kispert, A., 2009. Tbx18 and the fate of epicardial progenitors. Nature 458, E8–E9. https://doi.org/10.1038/nature07916

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 Rep 26, 1934-1950.e5. https://doi.org/10.1016/j.celrep.2019.01.079

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, 120–124. https://doi.org/10.1038/s41586-019-1414-x

DeLaughter, D.M., Bick, A.G., Wakimoto, H., McKean, D., Gorham, J.M., Kathiriya, I.S., Hinson, J.T., Homsy, J., Gray, J., Pu, W., Bruneau, B.G., Seidman, J.G., Seidman, C.E., 2016. Single-Cell Resolution of Temporal Gene Expression during Heart Development. Developmental Cell 39, 480–490. https://doi.org/10.1016/j.devcel.2016.10.001

Délot, E.C., Bahamonde, M.E., Zhao, M., Lyons, K.M., 2003. BMP signaling is required for septation of the outflow tract of the mammalian heart. Development 130, 209–220. https://doi.org/10.1242/dev.00181

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. https://doi.org/10.7554/eLife.03848

Doyle, M.J., Lohr, J.L., Chapman, C.S., Koyano-Nakagawa, N., Garry, M.G., Garry, D.J., 2015. Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes as a Model for Heart Development and Congenital Heart Disease. Stem Cell Rev and Rep 11, 710–727. https://doi.org/10.1007/s12015-015-9596-6

Drakhlis, L., Biswanath, S., Farr, C.-M., Lupanow, V., Teske, J., Ritzenhoff, K., Franke, A., Manstein, F., Bolesani, E., Kempf, H., Liebscher, S., Schenke-Layland, K., Hegermann, J., Nolte, L., Meyer, H., de la Roche, J., Thiemann, S., Wahl-Schott, C., Martin, U., Zweigerdt, R., 2021. Human heart-forming organoids recapitulate early heart and foregut development. Nat Biotechnol 39, 737–746. https://doi.org/10.1038/s41587-021-00815-9

Dyer, L.A., Kirby, M.L., 2009. The role of secondary heart field in cardiac development. Developmental Biology 336, 137–144. https://doi.org/10.1016/j.ydbio.2009.10.009

Feyen, D.A.M., McKeithan, W.L., Bruyneel, A.A.N., Spiering, S., Hörmann, L., Ulmer, B., Zhang, H., Briganti, F., Schweizer, M., Hegyi, B., Liao, Z., Pölönen, R.-P., Ginsburg, K.S., Lam, C.K., Serrano, R., Wahlquist, C., Kreymerman, A., Vu, M., Amatya, P.L., Behrens, C.S., Ranjbarvaziri, S., Maas, R.G.C., Greenhaw, M., Bernstein, D., Wu, J.C., Bers, D.M., Eschenhagen, T., Metallo, C.M., Mercola, M., 2020. Metabolic Maturation Media Improve Physiological Function of Human iPSC-Derived Cardiomyocytes. Cell Rep 32, 107925. https://doi.org/10.1016/j.celrep.2020.107925

Galdos, F.X., Darsha, A.K., Paige, S.L., Wu, S.M., 2021. Purification of Pluripotent Stem Cell-Derived Cardiomyocytes Using CRISPR/Cas9-Mediated Integration of Fluorescent Reporters. Methods Mol Biol 2158, 223–240. https://doi.org/10.1007/978-1-0716-0668-1_17

Haghverdi, L., Lun, A.T.L., Morgan, M.D., Marioni, J.C., 2018. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 36, 421–427. https://doi.org/10.1038/nbt.4091

Hikspoors, J.P.J.M., Kruepunga, N., Mommen, G.M.C., Köhler, S.E., Anderson, R.H., Lamers, W.H., 2022. A pictorial account of the human embryonic heart between 3.5 and 8 weeks of development. Commun Biol 5, 1–22. https://doi.org/10.1038/s42003-022-03153-x

Hill, M.C., Kadow, Z.A., Li, L., Tran, T.T., Wythe, J.D., Martin, J.F., 2019. A cellular atlas of Pitx2-dependent cardiac development. Development 146, dev180398. https://doi.org/10.1242/dev.180398

Holloway, E.M., Wu, J.H., Czerwinski, M., Sweet, C.W., Wu, A., Tsai, Y.-H., Huang, S., Stoddard, A.E., Capeling, M.M., Glass, I., Spence, J.R., 2020. Differentiation of Human Intestinal Organoids with Endogenous Vascular Endothelial Cells. Developmental Cell 54, 516-528.e7. https://doi.org/10.1016/j.devcel.2020.07.023

Hu, H., Lin, S., Wang, S., Chen, X., 2020. The Role of Transcription Factor 21 in Epicardial Cell Differentiation and the Development of Coronary Heart Disease. Front Cell Dev Biol 8, 457. https://doi.org/10.3389/fcell.2020.00457

Hurrell, T., Segeritz, C.-P., Vallier, L., Lilley, K.S., Cromarty, A.D., 2019. A proteomic time course through the differentiation of human induced pluripotent stem cells into hepatocyte-like cells. Sci Rep 9, 3270. https://doi.org/10.1038/s41598-019-39400-1

Hyun, I., Bredenoord, A.L., Briscoe, J., Klipstein, S., Tan, T., 2021. Human embryo research beyond the primitive streak. Science 371, 998–1000. https://doi.org/10.1126/science.abf3751

Kanton, S., Boyle, M.J., He, Z., Santel, M., Weigert, A., Sanchís-Calleja, F., Guijarro, P., Sidow, L., Fleck, J.S., Han, D., Qian, Z., Heide, M., Huttner, W.B., Khaitovich, P., Pääbo, S.,

Treutlein, B., Camp, J.G., 2019. Organoid single-cell genomic atlas uncovers human-specific features of brain development. Nature 574, 418–422. https://doi.org/10.1038/s41586-019-1654-9

Karagiannis, P., Takahashi, K., Saito, M., Yoshida, Y., Okita, K., Watanabe, A., Inoue, H., Yamashita, J.K., Todani, M., Nakagawa, M., Osawa, M., Yashiro, Y., Yamanaka, S., Osafune, K., 2019. Induced Pluripotent Stem Cells and Their Use in Human Models of Disease and Development. Physiological Reviews 99, 79–114. https://doi.org/10.1152/physrev.00039.2017

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, 435–440. https://doi.org/10.1016/S1534-5807(01)00040-5

Khosravi, F., Ahmadvand, N., Bellusci, S., Sauer, H., 2021. The Multifunctional Contribution of FGF Signaling to Cardiac Development, Homeostasis, Disease and Repair. Frontiers in Cell and Developmental Biology 9.

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. Nat Cell Biol 16, 829–840. https://doi.org/10.1038/ncb3024

Li, G., Tian, L., Goodyer, W., Kort, E.J., Buikema, J.W., Xu, A., Wu, J.C., Jovinge, S., Wu, S.M., 2019. Single cell expression analysis reveals anatomical and cell cycle-dependent transcriptional shifts during heart development. Development 146. https://doi.org/10.1242/dev.173476

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, 491–507. https://doi.org/10.1016/j.devcel.2016.10.014

Li, P., Cavallero, S., Gu, Y., Chen, T.H.P., Hughes, J., Hassan, A.B., Brüning, J.C., Pashmforoush, M., Sucov, H.M., 2011. IGF signaling directs ventricular cardiomyocyte proliferation during embryonic heart development. Development 138, 1795–1805. https://doi.org/10.1242/dev.054338

Lian, X., Zhang, J., Azarin, S.M., Zhu, K., Hazeltine, L.B., Bao, X., Hsiao, C., Kamp, T.J., Palecek, S.P., 2013. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. Nat Protoc 8, 162–175. https://doi.org/10.1038/nprot.2012.150

Liao, J., Aggarwal, V.S., Nowotschin, S., Bondarev, A., Lipner, S., Morrow, B.E., 2008. Identification of downstream genetic pathways of Tbx1 in the second heart field. Developmental Biology 316, 524–537. https://doi.org/10.1016/j.ydbio.2008.01.037

Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., Liu, J., 2017. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci Rep 7, 2193. https://doi.org/10.1038/s41598-017-02460-2

Ma, H.-Y., Xu, J., Eng, D., Gross, M.K., Kioussi, C., 2013. Pitx2-mediated cardiac outflow tract remodeling. Dev Dyn 242, 456–468. https://doi.org/10.1002/dvdy.23934

Ma, Q., Zhou, B., Pu, W.T., 2008. Reassessment of Isl1 and Nkx2-5 cardiac fate maps using a Gata4-based reporter of Cre activity. Dev Biol 323, 98–104. https://doi.org/10.1016/j.ydbio.2008.08.013

Mead, P.E., Brivanlou, I.H., Kelley, C.M., Zon, L.I., 1996. BMP-4-responsive regulation of dorsal-ventral patterning by the homeobox protein Mix.1. Nature 382, 357–360. https://doi.org/10.1038/382357a0

Meilhac, S.M., Buckingham, M.E., 2018. The deployment of cell lineages that form the mammalian heart. Nat Rev Cardiol 15, 705–724. https://doi.org/10.1038/s41569-018-0086-9

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. Dev Cell 6, 685–698. https://doi.org/10.1016/s1534-5807(04)00133-9

Mesbah, K., Rana, M.S., Francou, A., van Duijvenboden, K., Papaioannou, V.E., Moorman, A.F., Kelly, R.G., Christoffels, V.M., 2012. Identification of a Tbx1/Tbx2/Tbx3 genetic pathway governing pharyngeal and arterial pole morphogenesis. Hum Mol Genet 21, 1217–1229. https://doi.org/10.1093/hmg/ddr553

Mjaatvedt, C.H., Nakaoka, T., Moreno-Rodriguez, R., Norris, R.A., Kern, M.J., Eisenberg, C.A., Turner, D., Markwald, R.R., 2001. The Outflow Tract of the Heart Is Recruited from a Novel Heart-Forming Field. Developmental Biology 238, 97–109. https://doi.org/10.1006/dbio.2001.0409

Moretti, A., Caron, L., Nakano, A., Lam, J.T., Bernshausen, A., Chen, Y., Qyang, Y., Bu, L., Sasaki, M., Martin-Puig, S., Sun, Y., Evans, S.M., Laugwitz, K.-L., Chien, K.R., 2006. Multipotent Embryonic Isl1+ Progenitor Cells Lead to Cardiac, Smooth Muscle, and Endothelial Cell Diversification. Cell 127, 1151–1165. https://doi.org/10.1016/j.cell.2006.10.029

Nelson, D.O., Jin, D.X., Downs, K.M., Kamp, T.J., Lyons, G.E., 2014. Irx4 identifies a chamber-specific cell population that contributes to ventricular myocardium development. Dev Dyn 243, 381–392. https://doi.org/10.1002/dvdy.24078

Nelson, D.O., Lalit, P.A., Biermann, M., Markandeya, Y.S., Capes, D.L., Addesso, L., Patel, G., Han, T., John, M.C., Powers, P.A., Downs, K.M., Kamp, T.J., Lyons, G.E., 2016. Irx4 marks a multipotent, ventricular-specific progenitor cell. Stem Cells 34, 2875–2888. https://doi.org/10.1002/stem.2486

Nevis, K., Obregon, P., Walsh, C., Guner-Ataman, B., Burns, C.G., Burns, C.E., 2013. Tbx1 is required for second heart field proliferation in zebrafish. Dev Dyn 242, 550–559. https://doi.org/10.1002/dvdy.23928

O'Brien, T.X., Lee, K.J., Chien, K.R., 1993. Positional specification of ventricular myosin light chain 2 expression in the primitive murine heart tube. PNAS 90, 5157–5161. https://doi.org/10.1073/pnas.90.11.5157

Park, E.J., Watanabe, Y., Smyth, G., Miyagawa-Tomita, S., Meyers, E., Klingensmith, J., Camenisch, T., Buckingham, M., Moon, A.M., 2008. An FGF autocrine loop initiated in second heart field mesoderm regulates morphogenesis at the arterial pole of the heart. Development 135, 3599–3610. https://doi.org/10.1242/dev.025437

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, 490–495. https://doi.org/10.1038/s41586-019-0933-9

Protze, S.I., Lee, J.H., Keller, G.M., 2019a. Human Pluripotent Stem Cell-Derived Cardiovascular Cells: From Developmental Biology to Therapeutic Applications. Cell Stem Cell 25, 311–327. https://doi.org/10.1016/j.stem.2019.07.010

Protze, S.I., Lee, J.H., Keller, G.M., 2019b. Human Pluripotent Stem Cell-Derived Cardiovascular Cells: From Developmental Biology to Therapeutic Applications. Cell Stem Cell 25, 311–327. https://doi.org/10.1016/j.stem.2019.07.010

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8, 2281–2308. https://doi.org/10.1038/nprot.2013.143

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. Circ Res 115, 790–799. https://doi.org/10.1161/CIRCRESAHA.115.305020

Reifers, F., Walsh, E.C., Léger, S., Stainier, D.Y., Brand, M., 2000. Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (fgf8/acerebellar). Development 127, 225–235. https://doi.org/10.1242/dev.127.2.225

Reller, M.D., Strickland, M.J., Riehle-Colarusso, T., Mahle, W.T., Correa, A., 2008. Prevalence of Congenital Heart Defects in Metropolitan Atlanta, 1998-2005. The Journal of Pediatrics 153, 807–813. https://doi.org/10.1016/j.jpeds.2008.05.059

Rochais, F., Mesbah, K., Kelly, R.G., 2009. Signaling Pathways Controlling Second Heart Field Development. Circulation Research 104, 933–942. https://doi.org/10.1161/CIRCRESAHA.109.194464

Rossi, G., Broguiere, N., Miyamoto, M., Boni, A., Guiet, R., Girgin, M., Kelly, R.G., Kwon, C., Lutolf, M.P., 2021. Capturing Cardiogenesis in Gastruloids. Cell Stem Cell 28, 230-240.e6. https://doi.org/10.1016/j.stem.2020.10.013

Rudat, C., Kispert, A., 2012. Wt1 and Epicardial Fate Mapping. Circulation Research 111, 165–169. https://doi.org/10.1161/CIRCRESAHA.112.273946

Sacchetto, C., Vitiello, L., de Windt, L.J., Rampazzo, A., Calore, M., 2020. Modeling Cardiovascular Diseases with hiPSC-Derived Cardiomyocytes in 2D and 3D Cultures. Int J Mol Sci 21, 3404. https://doi.org/10.3390/ijms21093404

Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J. i, Inoue, T., 1999. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. Development 126, 3437–3447.

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., Paoluzzi Tomada, E., Kieslinger, M., Jeremias, I., Scialdone, A., Gires, O., 2018. Spatiotemporal patterning of EpCAM is important for murine embryonic endo- and mesodermal differentiation. Sci Rep 8, 1801. https://doi.org/10.1038/s41598-018-20131-8

Schleiffarth, J.R., Person, A.D., Martinsen, B.J., Sukovich, D.J., Neumann, A., Baker, C.V.H., Lohr, J.L., Cornfield, D.N., Ekker, S.C., Petryk, A., 2007. Wnt5a is required for cardiac outflow tract septation in mice. Pediatr Res 61, 386–391. https://doi.org/10.1203/pdr.0b013e3180323810

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N.K., Macaulay, I.C., Marioni, J.C., Göttgens, B., 2016. Resolving Early Mesoderm Diversification through Single Cell Expression Profiling. Nature 535, 289–293. https://doi.org/10.1038/nature18633

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. Nat Cell Biol 15, 1098–1106. https://doi.org/10.1038/ncb2824

Takahashi, K., Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126, 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Tan, C.M.J., Lewandowski, A.J., 2020. The Transitional Heart: From Early Embryonic and Fetal Development to Neonatal Life. FDT 47, 373–386. https://doi.org/10.1159/000501906

Tandon, P., Miteva, Y.V., Kuchenbrod, L.M., Cristea, I.M., Conlon, F.L., 2013. Tcf21 regulates the specification and maturation of proepicardial cells. Development 140, 2409–2421. https://doi.org/10.1242/dev.093385

Tyser, R.C.V., Ibarra-Soria, X., McDole, K., Jayaram, S.A., Godwin, J., Brand, T.A.H. van den, Miranda, A.M.A., Scialdone, A., Keller, P.J., Marioni, J.C., Srinivas, S., 2020. Characterization of a common progenitor pool of the epicardium and myocardium. Science. https://doi.org/10.1126/science.abb2986

Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., Srinivas, S., 2021. Single-cell transcriptomic characterization of a gastrulating human embryo. Nature 600, 285–289. https://doi.org/10.1038/s41586-021-04158-y

van der Linde, D., Konings, E.E.M., Slager, M.A., Witsenburg, M., Helbing, W.A., Takkenberg, J.J.M., Roos-Hesselink, J.W., 2011. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. J Am Coll Cardiol 58, 2241–2247. https://doi.org/10.1016/j.jacc.2011.08.025

Vincentz, J.W., Toolan, K.P., Zhang, W., Firulli, A.B., 2017. Hand factor ablation causes defective left ventricular chamber development and compromised adult cardiac function. PLOS Genetics 13, e1006922. https://doi.org/10.1371/journal.pgen.1006922

Vitelli, F., Morishima, M., Taddei, I., Lindsay, E.A., Baldini, A., 2002. Tbx1 mutation causes multiple cardiovascular defects and disrupts neural crest and cranial nerve migratory pathways. Human Molecular Genetics 11, 915–922. https://doi.org/10.1093/hmg/11.8.915

Waldo, K.L., Kumiski, D.H., Wallis, K.T., Stadt, H.A., Hutson, M.R., Platt, D.H., Kirby, M.L., 2001. Conotruncal myocardium arises from a secondary heart field. Development 128, 3179–3188. https://doi.org/10.1242/dev.128.16.3179

Xiong, H., Luo, Y., Yue, Y., Zhang, J., Ai, S., Li, X., Wang, X., Zhang, Y.-L., Wei, Y., Li, H.-H., Hu, X., Li, C., He, A., 2019. Single-Cell Transcriptomics Reveals Chemotaxis-Mediated Intraorgan Crosstalk During Cardiogenesis. Circ Res 125, 398–410. https://doi.org/10.1161/CIRCRESAHA.119.315243

Yamanaka, S., 2008. Induction of pluripotent stem cells from mouse fibroblasts by four transcription factors. Cell Proliferation 41, 51–56. https://doi.org/10.1111/j.1365-2184.2008.00493.x

Zeng, B., Ren, X., Cao, F., Zhou, X., Zhang, J., 2011. Developmental patterns and characteristics of epicardial cell markers Tbx18 and Wt1 in murine embryonic heart. Journal of Biomedical Science 18, 67. https://doi.org/10.1186/1423-0127-18-67

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, 802-811.e5. https://doi.org/10.1016/j.stem.2019.02.015

Zhang, Q., Carlin, D., Zhu, F., Cattaneo, P., Ideker, T., Evans, S.M., Bloomekatz, J., Chi, N.C., 2021. Unveiling Complexity and Multipotentiality of Early Heart Fields. Circulation Research 129, 474–487. https://doi.org/10.1161/CIRCRESAHA.121.318943

Main Figure Legends

Figure 1: Integration of TBX5/MYL2 Lineage Tracing Reporter System into Human Induced Pluripotent Stem Cells

(A) Schematic of lineage tracing strategy for identifying left ventricular cardiomyocytes *in vitro*.
(B) CRISPR/Cas9 gene targeting strategy of genetic constructs for TBX5 lineage tracing and MYL2 direct reporter. MYL2, CCR5, and TBX5 constructs contain Puromycin (PuroR), Neomycin (NeoR), and Hygromycin (HygroR) resistance cassettes for selection of hiPSC after targeting of genetic constructs. Blue arrows indicate location of PCR primer binding sites for confirmation of construct integration.
(C) Inside-Outside (I/O) and Outside-Outside (O/O) PCR DNA agarose gels for confirmation of integration of genetic constructs into MYL2, CCR5, and TBX5 genetic loci.
(D) Sanger sequencing traces for C-terminal regions of MYL2 and TBX5 genes indicating in-frame integration of P2A site.
(E) Bright field and immunofluorescence images of pluripotency marker expression in hiPSC lines after integration of all three genetic constructs.

Figure 2: TBX5-Lineage Tracing Reveals a Predominance of Lineage Positive Cardiomyocytes Through the Course of HiPSC Cardiac Differentiation

(A) Schematic of analysis approach of reporter system expression through the course of cardiac differentiation.
(B) Representative flow cytometry plots for the expression of TurboGFP and cardiac troponin T staining between Day 3 to Day 30 of differentiation for WTC and SCVI-111 reporter lines. Gating set based on Day 3 of differentiation.
(C) Representative flow cytometry for the expression MYL2-tdTomato and TurboGFP between Day 3 an Day 30. Gating set based on Day 3 of differentiation.
(D) Quantification of percentage cardiac troponin T (TNNT2) expressing cells across all cells sampled.
(E) Quantification of the percentage of TurboGFP positive cells out of cardiac troponin positive cardiomyocytes.
(F) Quantification of tdTomato expressing cells across total cells sampled. Quantification of TurboGFP-positive cells out of MYL2-Tdtomato cardiomyocytes

Minimum of 4 independent differentiations collected per sample. *pvalue<0.05, **pvalue<0.01, ***pvalue<0.001, ****pvalue<0.0001

Figure 3: Immunofluorescence Imaging of Day 30 Reporter Cardiomyocytes and RT-qPCR Profiling of FHF and SHF Markers Across Differentiation

(A) Immunofluoresence images of TurboGFP, TdTomato, and TNNT2 expression at Day 30 of cardiac differentiation for WTC and SCVI-111 reporter lines.
(B) RT-qPCR profiling of FHF markers HAND1 and TBX5 along with Cre Recombinase between Day 0 and Day 30 of differentiation.
(C) RT-qPCR profiling of pan-cardiac progenitor marker ISL1, and SHF markers TBX1 and FGF8 between Day 0 and Day 30 of differentiation.
(D) RT-qPCR profiling of pan-cardiomyocyte marker TNNT2.
(E) RT-qPCR profiling of ventricular marker MYL2.
N=3-4 biological replicates per timepoint. *pvalue<0.05, **pvalue<0.01, ***pvalue<0.001, ****pvalue<0.0001

Figure 4: scRNA-seq Profiling and Trajectory Inference Reveals Emergence of Myocardial and Epicardial Lineages During hiPSC Cardiac Differentiation

(A) Diagram of scRNA-seq multiplexing for profiling of 12 timepoints across two lines during hiPSC cardiac differentiation.
(B) Left, UMAP plot with identification of 13 cell populations over the course cardiac differentiation. Right-Top, Plot indicating labeling of cells captured from WTC and SCVI-111 line. Right-Bottom, Plot indicating timepoints of differentiation at which cells were captured for scRNA-seq.
(C) Dotplot presenting expression of top markers for major cell populations identified during hiPSC cardiac differentiation.
(D) Subway map plot showing projection of cells along cell lineages detected using STREAM analysis.
(E) Stream plot indicating the relative cell type composition along each branch of differentiation identified by STREAM.
(F) Graph indicating pseudotime values calculated by STREAM for ordering cells along a continuous developmental projection axis.
(G) Feature plots for top gene markers identified during each major developmental phase of hiPSC cardiac differentiation.

Figure 5: Comparison of Heart Field Development Between Murine and hiPSC Cardiac Differentiation Reveals FHF Identity of hiPSC Cardiac Lineages

(A) Top, UMAP plot showing clustering of murine cell types encompassing early mesodermal progenitors, FHF/SHF progenitors, LV/RV/OFT cardiomyocytes, and epicardial cells. Bottom-left, labeling of cell types by timepoint analyzed. Bottom-right, annotation by reference from which data was obtained.
(B) UMAP embeddings of heart field development split by FHF and SHF lineages.
(C) FHF and aSHF trajectory analysis and plotting of expression FHF (Tbx5 and Hcn4) and anterior SHF (Tbx1 and Fgf8) markers.
(D) Analysis of FHF and aSHF marker expression during hiPSC cardiac differentiation.

(E) Gene expression analysis of ventricular markers IRX4 and MYL2 during hiPSC cardiac differentiation.

Figure 6: Comparison of 2D and 3D Cardiac Differentiation Uncovers Potential of Organoid System for SHF Generation

(A) Schematic of Comparison Strategy of scRNA-seq data from 2D hiPSC cardiac differentiation generated using our 2D protocol and a previously published organoid protocol.
(B) UMAP plot of the myocardial lineage identified during 2D hiPSC cardiac differentiation.
(C) Feature plots of FHF (TBX5, HCN4, HAND1), pan-cardiac (ISL1, NKX2-5, TNNI1), and SHF (FGF8, TBX1) during 2D hiPSC differentiation.
(D) UMAP plot of Drakhlis etal cardiac lineage cells with annotation of unsupervised clusters.
(E) Feature plots of FHF (TBX5, HCN4, HAND1), pan-cardiac (ISL1, NKX2-5, TNNI1), and SHF (FGF8, TBX1) during organoid differentiation
(F) Co-clustering of 2D and 3D cardiomyocytes with annotation of source dataset (right-top) and expression of TBX5 (right-bottom).
(G) Volcano plot for top differentially expressed genes between OFT identified cluster and remainder of identified cardiomyocytes identified as left ventricular.

Table Legends

Table 1: GEO Accession Numbers for Datasets

Table 2: Single Guide RNA Sequences Used for Gene Targeting

Table 3: qPCR Primer Sequences

Table 4: Description of scRNA-seq Run and Hashtag Oligo Sequences Used for Sample Multiplexing

Table 5: Differentially Expressed Genes Identified For Annotated Cell Types

Table 6: Genes Correlated with Differentiation Trajectories Identified During hiPSC Cardiac Differentiation

Supplementary Figure Legends

Supplementary Figure 1: Quality Control and Hashtag Oligo Labeling of Samples for scRNA-seq Sample Galdos_Seq_Run1

(A) Cutoff values (dotted red lines) calculated for quality control filtering of four metrics evaluated.

(B) Ridge plots demonstrating highly specific labeling of chemically modified oligos for multiplexing of day 1-6 samples across two hiPSC lines.

Supplementary Figure 2: Quality Control and Hashtag Oligo Labeling of Samples for scRNA-seq Sample Galdos_Seq_Run2

(A) Cutoff values (dotted red lines) calculated for quality control filtering of four metrics evaluated.

(B) Ridge plots demonstrating highly specific labeling of chemically modified oligos for multiplexing of day 0, 7, and 11 samples across two hiPSC lines.

Supplementary Figure 3: Quality Control and Hashtag Oligo Labeling of Samples for scRNA-seq Sample Galdos_Seq_Run3

(A) Cutoff values (dotted red lines) calculated for quality control filtering of four metrics evaluated.

(B) Ridge plots demonstrating highly specific labeling of chemically modified oligos for multiplexing of day 13, 15, and 30 samples across two hiPSC lines.

Supplementary Figure 4: Unsupervised Clustering and Marker Expression of Combined hiPSC scRNA-seq Data from WTC and SCVI-111 lines.

(A) UMAP embedding for combined scRNA-seq data from WTC and SCVI-111 lines, along with unsupervised clustered identified.

(B) Stacked violin plots for distinct cell markers enriched across all clusters identified.

Supplementary Figure 5: Trajectory Inference Analysis of hiPSC Cardiac Differentiation Across WTC and SCVI-111 Lines

(A) UMAP embedding reproduced from Figure 4A demonstrating annotation of 12 distinct cell types.

(B) Compositional analysis showing proportion of each cell type identified across all times points for both WTC and SCVI-111 lines.

(C) STREAM algorithm fitting of principal graphs and trajectory inference detection. Cells were ordered along principal graphs and calculated pseudotime values.

(D) Expression of statistically significant gene markers along differentiation trajectories identified during hiPSC cardiac differentiation.

FIGURE 1:

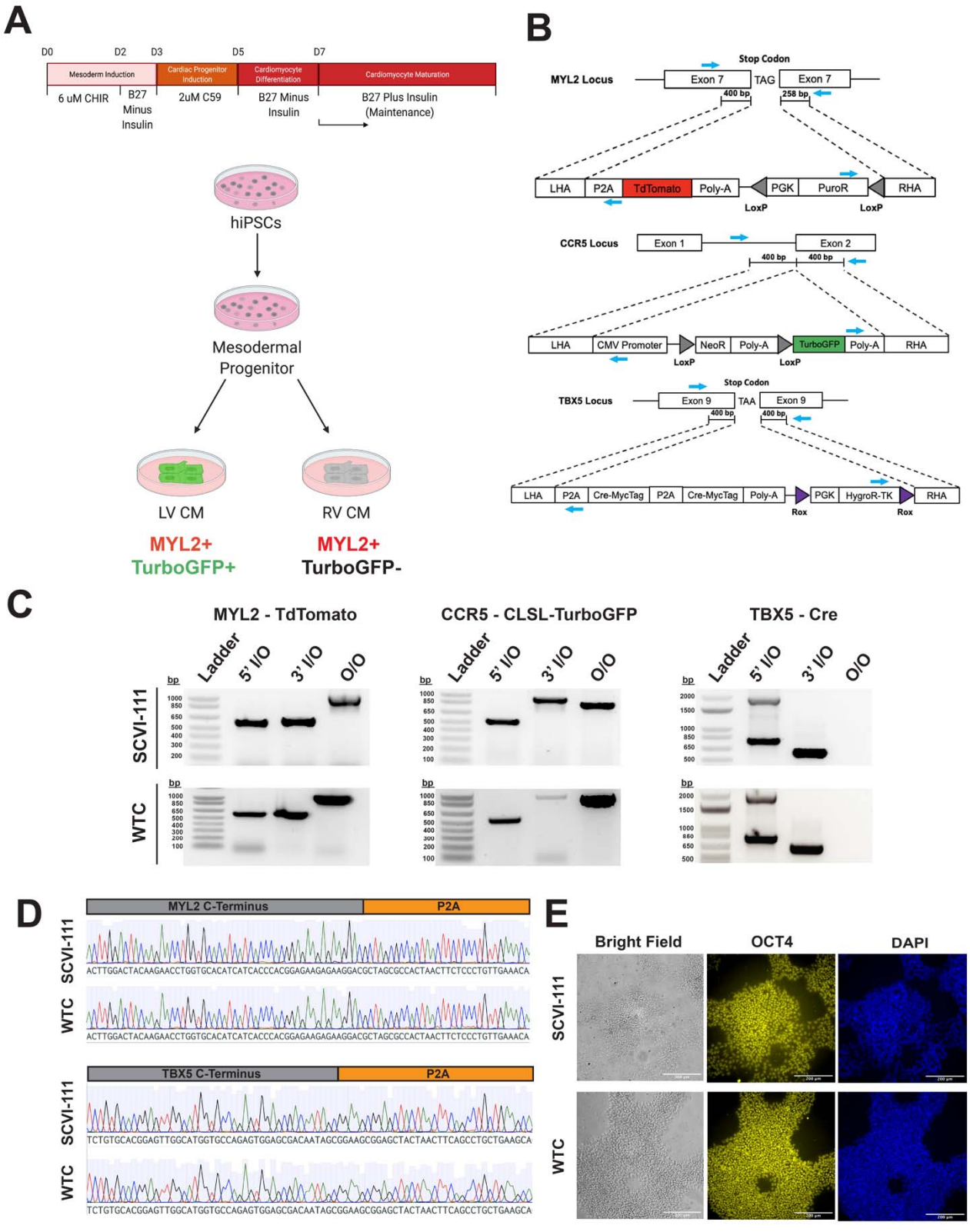

A
D0
D2
D3
D5
D7
Mesoderm Induction
Cardiac Progenitor Induction
Cardiomyocyte Differentiation
Cardiomyocyte Maturation
6 uM CHIR
B27 Minus Insulin
2uM C59
B27 Minus Insulin
B27 Plus Insulin (Maintenance)
hiPSCs
Mesodermal Progenitor
LV CM
RV CM
MYL2+ TurboGFP+
MYL2+ TurboGFP-
B
MYL2 Locus
Stop Codon
Exon 7
TAG
Exon 7
400 bp
258 bp
LHA
P2A
TdTomato
Poly-A
PGK
PuroR
RHA
LoxP
LoxP
CCR5 Locus
Exon 1
Exon 2
400 bp
400 bp
LHA
CMV Promoter
NeoR
Poly-A
TurboGFP
Poly-A
RHA
LoxP
LoxP
TBX5 Locus
Stop Codon
Exon 9
TAA
Exon 9
400 bp
400 bp
LHA
P2A
Cre-MycTag
P2A
Cre-MycTag
Poly-A
PGK
HygroR-TK
RHA
Rox
Rox
C
MYL2 - TdTomato
CCR5 - CLSL-TurboGFP
TBX5 - Cre
Ladder
5' I/O
3' I/O
O/O
bp
SCVI-111
WTC
D
MYL2 C-Terminus
P2A
SCVI-111
WTC
ACTTGGACTACAAGAACCTGGTGCACATCATCACCCACGGAGAAGAGAAGGACGCTAGCGCCACTAACTTCTCCCTGTTGAAACA
TBX5 C-Terminus
P2A
SCVI-111
WTC
TCTGTGCACGGAGTTGGCATGGTGCCAGAGTGGAGCGACAATAGCGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCA
E
Bright Field
OCT4
DAPI
SCVI-111
WTC

FIGURE 2:

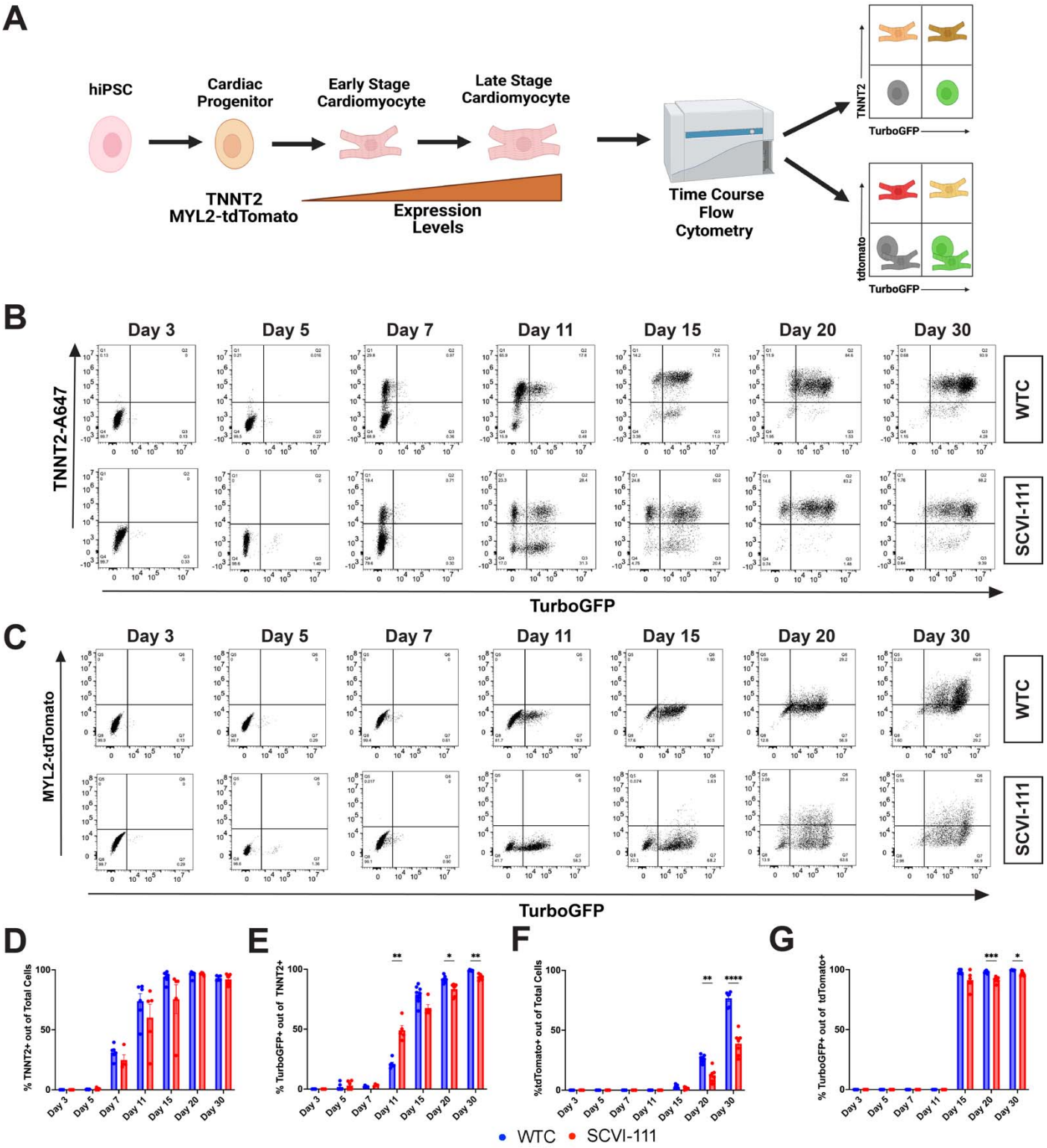

A
hiPSC
Cardiac Progenitor
Early Stage Cardiomyocyte
Late Stage Cardiomyocyte
TNNT2
MYL2-tdTomato
Expression Levels
Time Course Flow Cytometry
TNNT2
TurboGFP
tdtomato
TurboGFP
B
Day 3
Day 5
Day 7
Day 11
Day 15
Day 20
Day 30
TNNT2-A647
WTC
SCVI-111
TurboGFP
C
Day 3
Day 5
Day 7
Day 11
Day 15
Day 20
Day 30
MYL2-tdTomato
WTC
SCVI-111
TurboGFP
D
% TNNT2+ out of Total Cells
E
% TurboGFP+ out of TNNT2+
F
%tdTomato+ out of Total Cells
G
% TurboGFP+ out of tdTomato+
Day 3
Day 5
Day 7
Day 11
Day 15
Day 20
Day 30
WTC
SCVI-111

FIGURE 3:

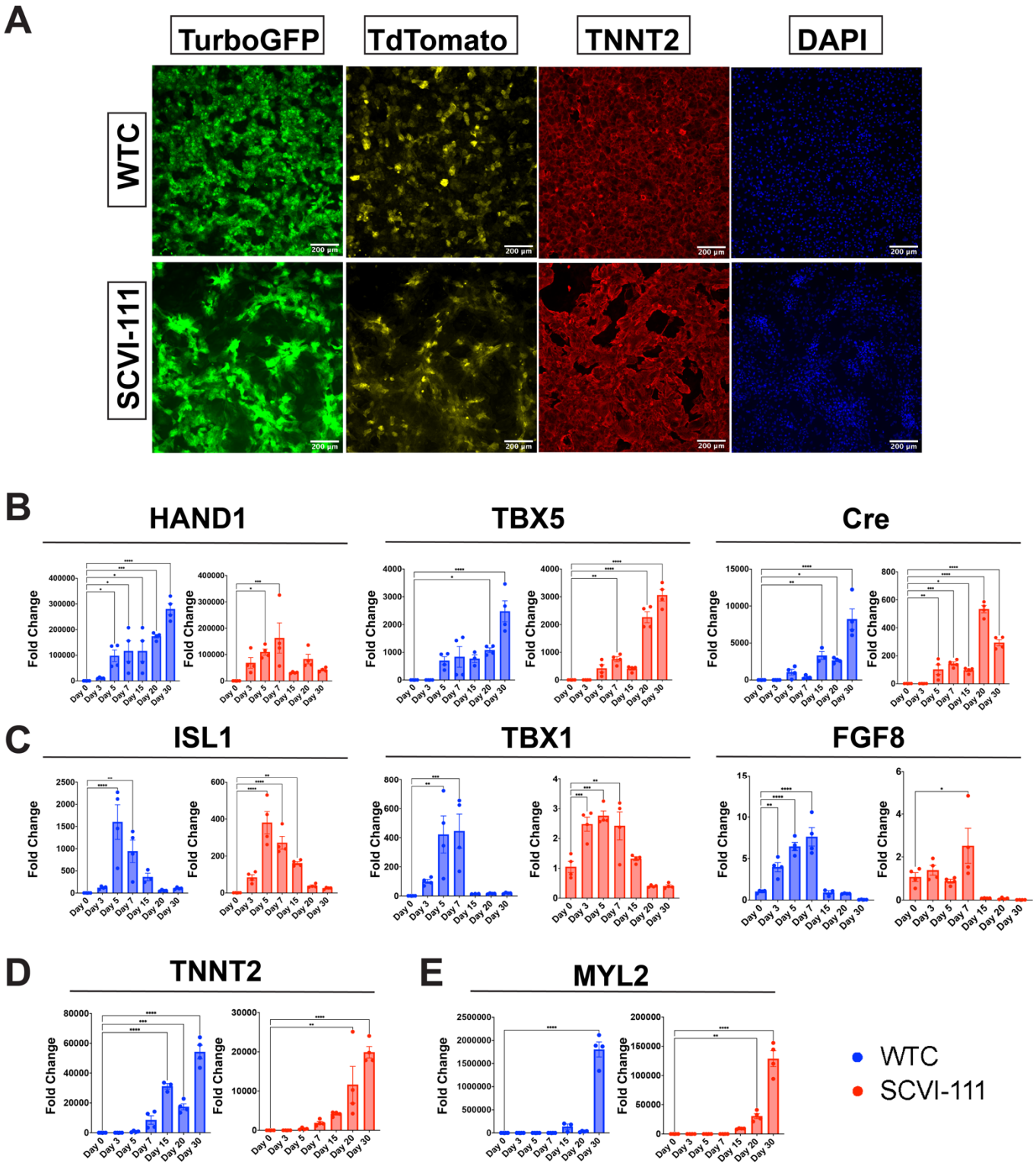

A
TurboGFP
TdTomato
TNNT2
DAPI
WTC
SCVI-111
200 µm
B
HAND1
TBX5
Cre
C
ISL1
TBX1
FGF8
D
TNNT2
E
MYL2
Fold Change
Day 0
Day 3
Day 5
Day 7
Day 15
Day 20
Day 30
WTC
SCVI-111

FIGURE 4:

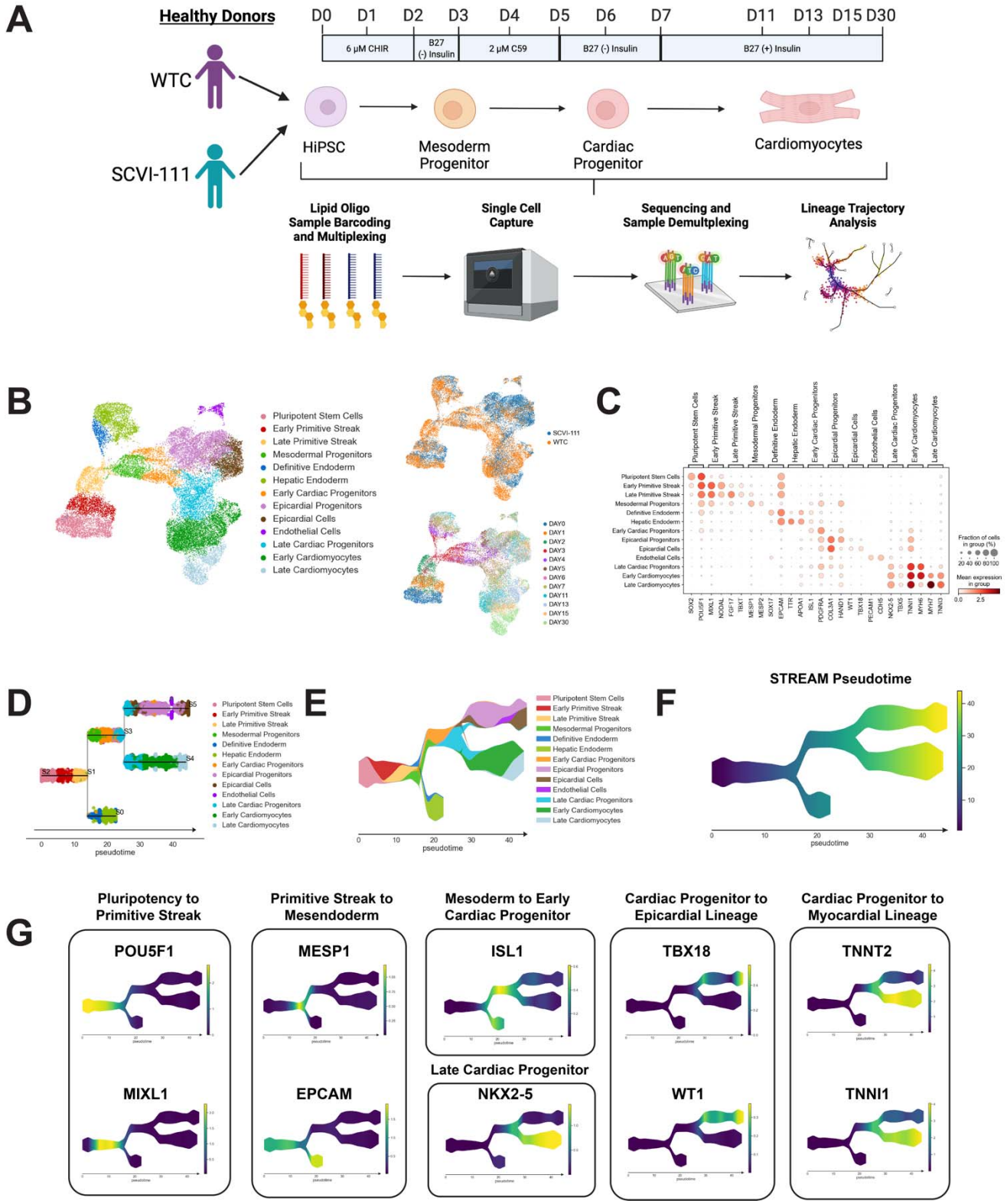

A
Healthy Donors
D0 D1 D2 D3 D4 D5 D6 D7 D11 D13 D15 D30
6 μM CHIR
B27 (-) Insulin
2 μM C59
B27 (-) Insulin
B27 (+) Insulin
WTC
SCVI-111
HiPSC
Mesoderm Progenitor
Cardiac Progenitor
Cardiomyocytes
Lipid Oligo Sample Barcoding and Multiplexing
Single Cell Capture
Sequencing and Sample Demultplexing
Lineage Trajectory Analysis
B
Pluripotent Stem Cells
Early Primitive Streak
Late Primitive Streak
Mesodermal Progenitors
Definitive Endoderm
Hepatic Endoderm
Early Cardiac Progenitors
Epicardial Progenitors
Epicardial Cells
Endothelial Cells
Late Cardiac Progenitors
Early Cardiomyocytes
Late Cardiomyocytes
SCVI-111
WTC
DAY0
DAY1
DAY2
DAY3
DAY4
DAY5
DAY6
DAY7
DAY11
DAY13
DAY15
DAY30
C
Fraction of cells in group (%)
Mean expression in group
D
pseudotime
E
pseudotime
F
STREAM Pseudotime
pseudotime
G
Pluripotency to Primitive Streak
POU5F1
MIXL1
Primitive Streak to Mesendoderm
MESP1
EPCAM
Mesoderm to Early Cardiac Progenitor
ISL1
Late Cardiac Progenitor
NKX2-5
Cardiac Progenitor to Epicardial Lineage
TBX18
WT1
Cardiac Progenitor to Myocardial Lineage
TNNT2
TNNI1

FIGURE 5:

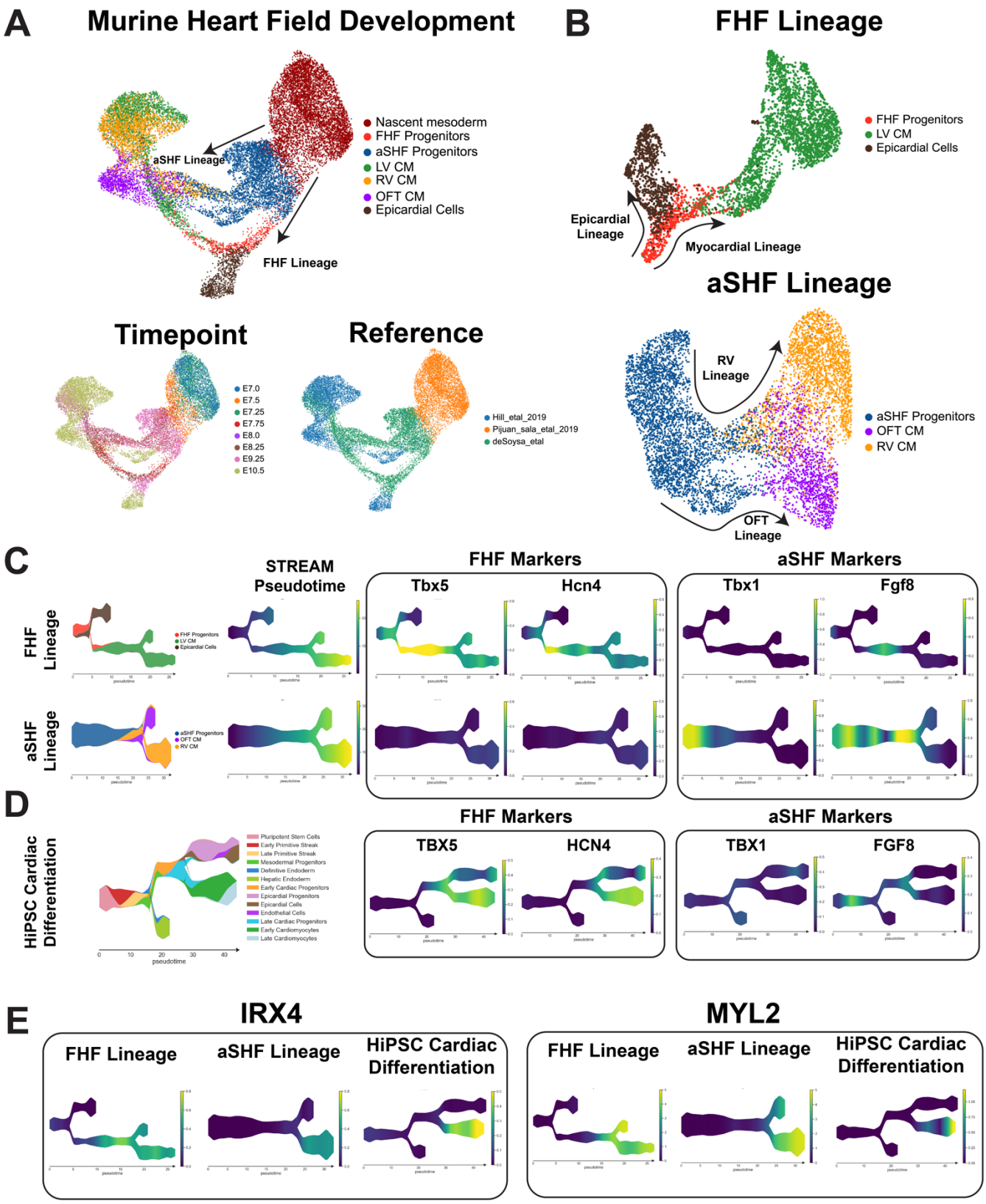
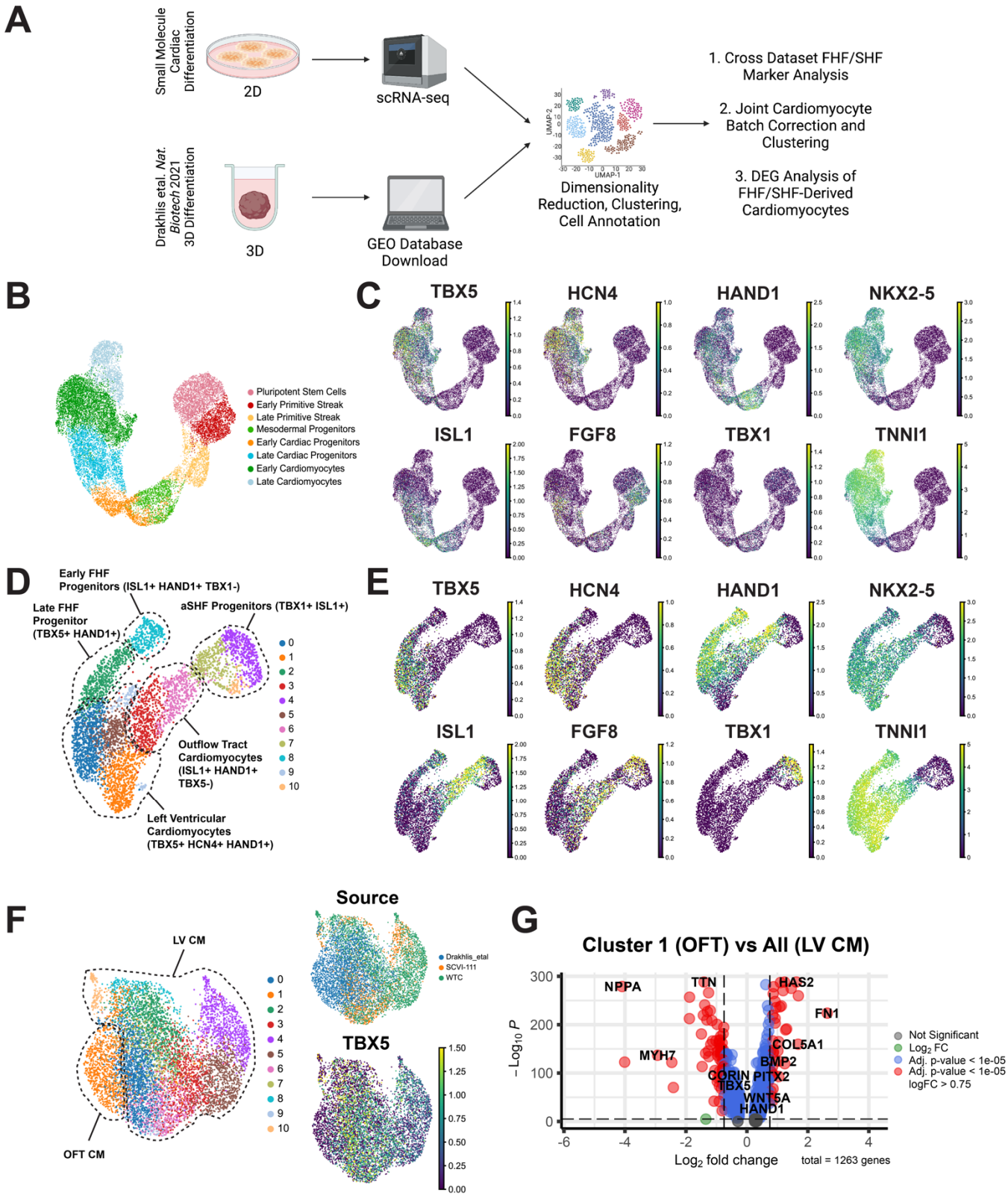
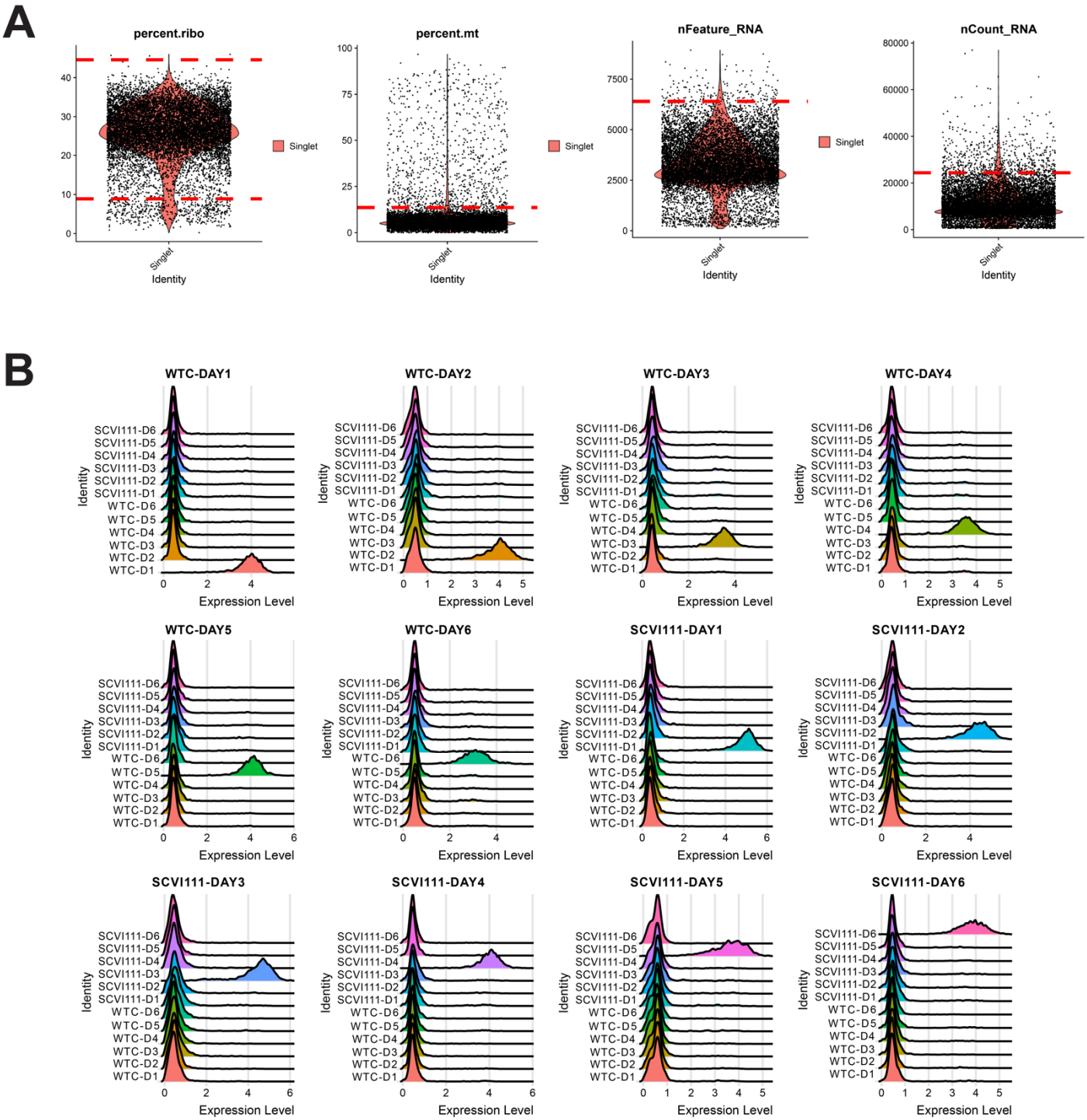

A
Murine Heart Field Development
aSHF Lineage
FHF Lineage
Nascent mesoderm
FHF Progenitors
aSHF Progenitors
LV CM
RV CM
OFT CM
Epicardial Cells
B
FHF Lineage
FHF Progenitors
LV CM
Epicardial Cells
Epicardial Lineage
Myocardial Lineage
aSHF Lineage
RV Lineage
OFT Lineage
aSHF Progenitors
OFT CM
RV CM
Timepoint
E7.0
E7.5
E7.25
E7.75
E8.0
E8.25
E9.25
E10.5
Reference
Hill_etal_2019
Pijuan_sala_etal_2019
deSoysa_etal
C
STREAM Pseudotime
FHF Markers
aSHF Markers
Tbx5
Hcn4
Tbx1
Fgf8
FHF Lineage
aSHF Lineage
D
HiPSC Cardiac Differentiation
FHF Markers
aSHF Markers
TBX5
HCN4
TBX1
FGF8
E
IRX4
MYL2
FHF Lineage
aSHF Lineage
HiPSC Cardiac Differentiation

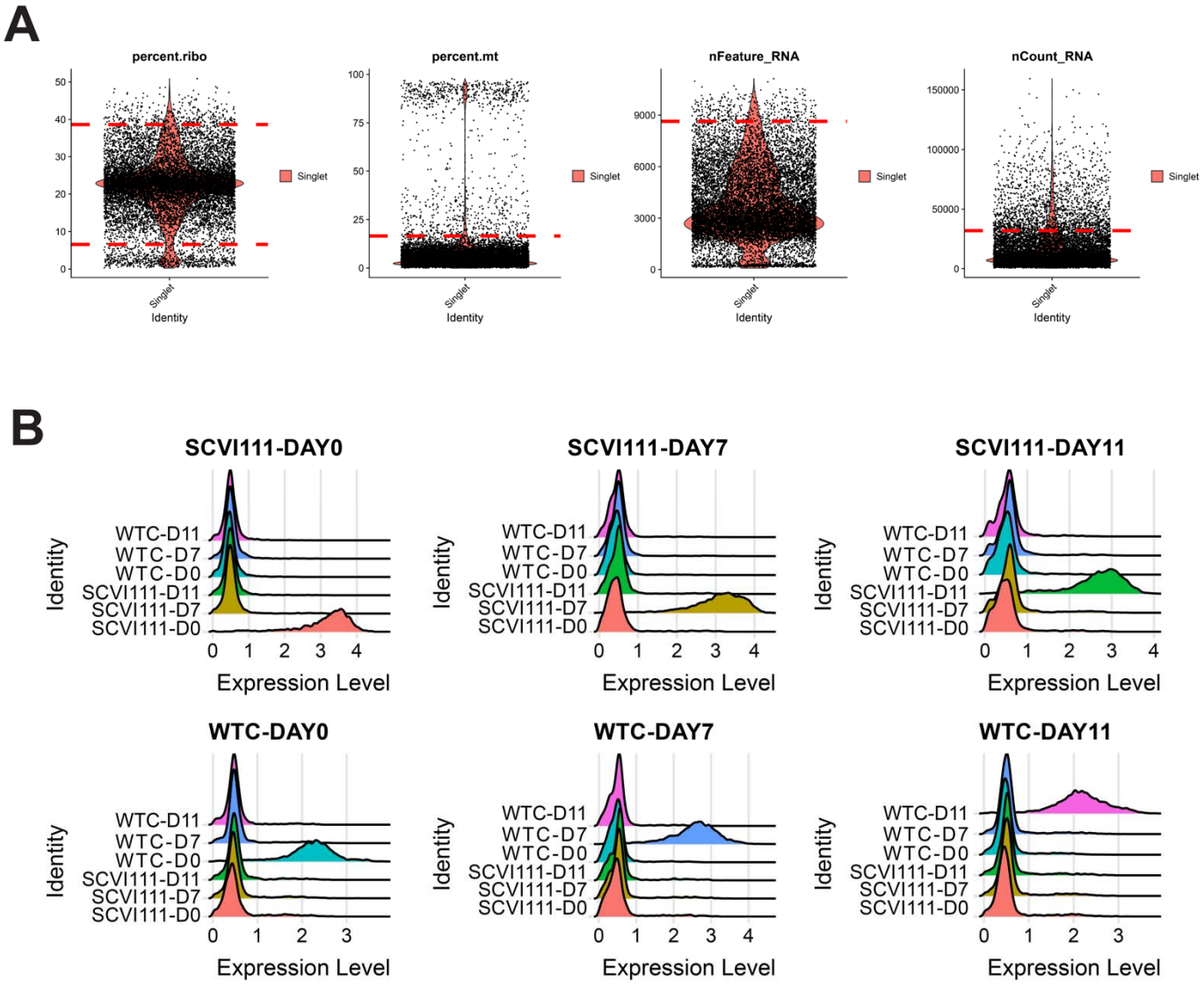
FIGURE 6:


A
Small Molecule Cardiac Differentiation
2D
scRNA-seq
Drakhlis etal. Nat. Biotech 2021 3D Differentiation
3D
GEO Database Download
Dimensionality Reduction, Clustering, Cell Annotation
1. Cross Dataset FHF/SHF Marker Analysis
2. Joint Cardiomyocyte Batch Correction and Clustering
3. DEG Analysis of FHF/SHF-Derived Cardiomyocytes
B
Pluripotent Stem Cells
Early Primitive Streak
Late Primitive Streak
Mesodermal Progenitors
Early Cardiac Progenitors
Late Cardiac Progenitors
Early Cardiomyocytes
Late Cardiomyocytes
C
TBX5
HCN4
HAND1
NKX2-5
ISL1
FGF8
TBX1
TNNI1
D
Early FHF Progenitors (ISL1+ HAND1+ TBX1-)
Late FHF Progenitor (TBX5+ HAND1+)
aSHF Progenitors (TBX1+ ISL1+)
Outflow Tract Cardiomyocytes (ISL1+ HAND1+ TBX5-)
Left Ventricular Cardiomyocytes (TBX5+ HCN4+ HAND1+)
E
TBX5
HCN4
HAND1
NKX2-5
ISL1
FGF8
TBX1
TNNI1
F
LV CM
OFT CM
Source
Drakhlis_etal
SCVI-111
WTC
TBX5
G
Cluster 1 (OFT) vs All (LV CM)
NPPA
TTN
HAS2
FN1
COL5A1
MYH7
BMP2
CORIN
PITX2
TBX5
WNT5A
HAND1
– Log10 P
Log2 fold change
total = 1263 genes
Not Significant
Log2 FC
Adj. p-value < 1e-05
Adj. p-value < 1e-05 logFC > 0.75

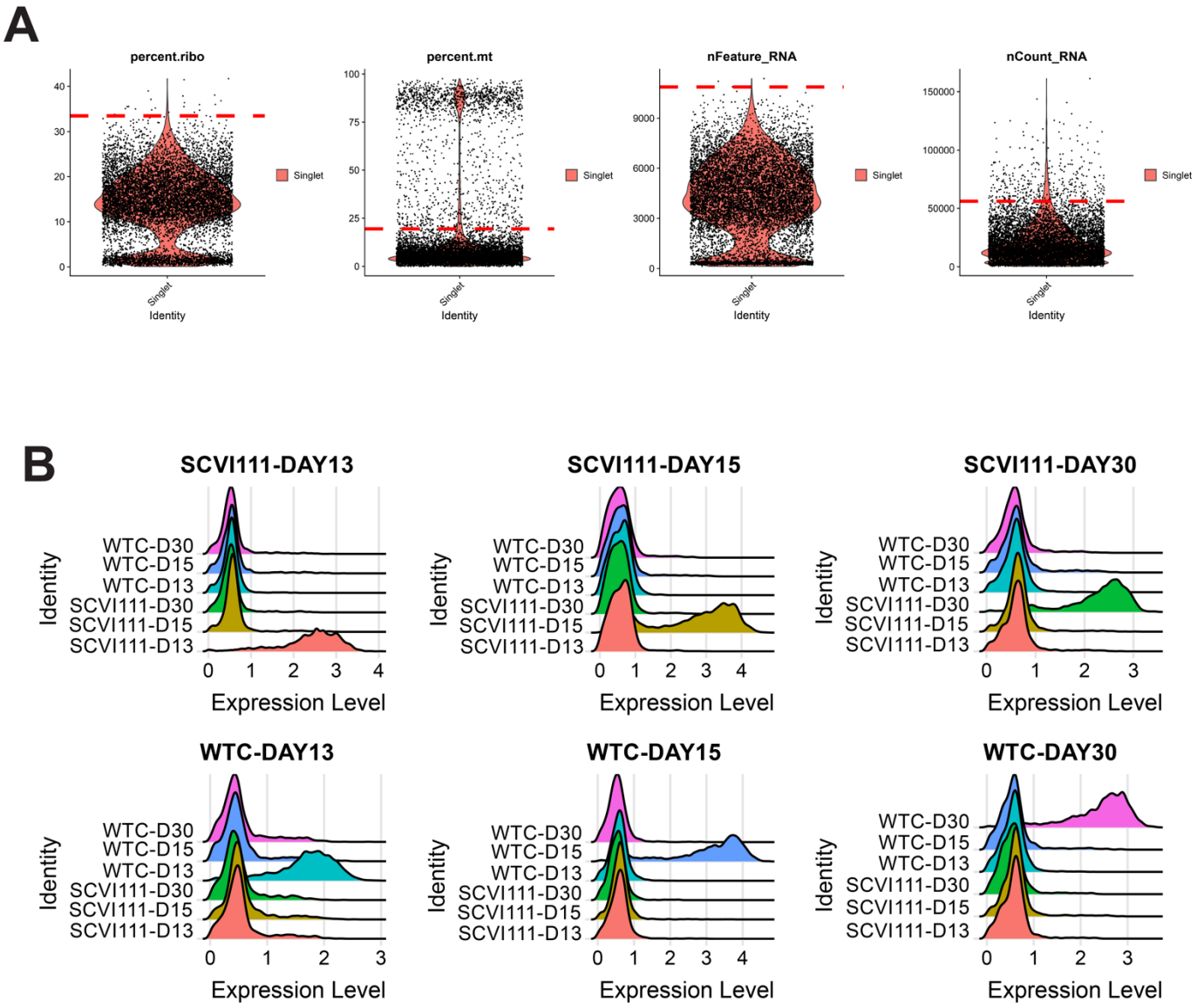
SUPPLEMENTARY FIGURE 1:

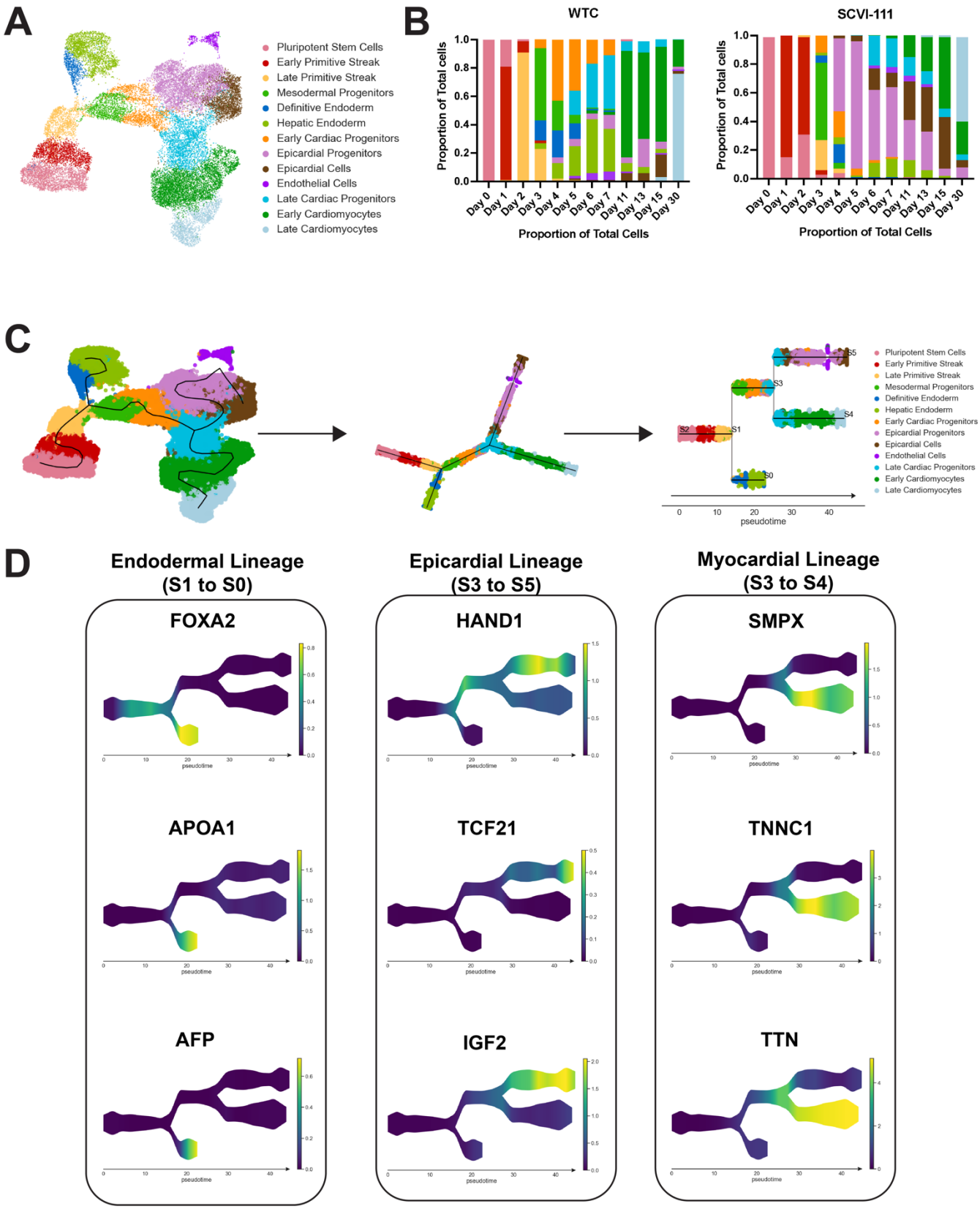


SUPPLEMENTARY FIGURE 2:



SUPPLEMENTARY FIGURE 3:



SUPPLEMENTARY FIGURE 4:

SUPPLEMENTARY FIGURE 5: