Cardiac directed differentiation using small molecule Wnt modulation at single-cell resolution

Clayton E Friedman[1,2,14], Quan Nguyen[1,14], Samuel W Lukowski[1], Abbigail Helfer[1,2], Han Sheng Chiu[1,2], Holly K Voges[2,3], Shengbao Suo Suo[4], Jing-Dong Jackie Han[4], Pierre Osteil[5], Guangdun Peng[6], Naihe Jing[6,7], Greg J Baillie[1], Anne Senabouth[1], Angelika N Christ[1], Timothy J Bruxner[1], Charles E Murry[8], Emily S Wong[3], Jun Ding[9], Yuliang Wang[8,10], James Hudson[2,3], Ziv Bar-Joseph[9], Patrick P L Tam[5,11], Joseph E Powell[1,12,13], and Nathan J Palpant[1,2,13]

[1]Institute for Molecular Bioscience, [2]Centre for Cardiac and Vascular Biology, [3]School of Biomedical Sciences, The University of Queensland, Brisbane, QLD Australia 4072, [4]Key Laboratory of Computational Biology, Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China, [5]Embryology Unit, Children's Medical Research Institute, Westmead, NSW 2145 Australia, [6]State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences; University of Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai 200031, China, [7]School of Life Science and Technology, ShanghaiTech University, 100 Haike Road, Shanghai 201210, China, [8]Departments of Pathology, Bioengineering and Medicine/Cardiology, Institute for Stem Cell and Regenerative Medicine, The University of Washington, Seattle, WA USA, [9]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA, [10]Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA, 98195, [11]School of Medical Sciences, Sydney Medical School, University of Sydney, NSW 2006 Australia, [12]Centre for Neurogenetics and Statistical Genomics, Queensland Brain Institute, University of Queensland, Brisbane, QLD, 4072, Australia, [13] co-senior authors and lead contacts, [14] These authors contributed equally, *Correspondence: n.palpant@uq.edu.au (N.J.P)

Abstract

Differentiation into diverse cell lineages requires orchestration of gene regulatory networks guiding cell fate choices. Here, we present the dissection of cellular composition and gene networks from transcriptomic data of 43,168 cells across five discrete time points during cardiac-directed differentiation. We utilize unsupervised clustering and implement a lineage trajectory prediction algorithm that integrates transcription factor networks to predict cell fate progression of 15 subpopulations that correlate with germ layer and cardiovascular differentiation *in vivo.* These data reveal transcriptional networks underlying lineage derivation of mesoderm, definitive endoderm, vascular endothelium, cardiac precursors, and definitive cell types that comprise cardiomyocytes and a previously unrecognized cardiac outflow tract population. Single cell analysis of genetic regulators governing cardiac fate diversification identified the non-DNA binding homeodomain protein, HOPX, as functionally necessary for endothelial specification. Our findings further implicate dysregulation of HOPX during *in vitro* cardiac-directed differentiation underlying the molecular and physiological immaturity of stem cell-derived cardiomyocytes.

Key words: human pluripotent stem cells, cardiomyocytes, transcriptomic profiling, outflow tract, heart, development, CRISPR, engineered heart tissue, RNA-sequencing, HOPX, *scdiff*

Introduction

Studies of cardiac development at single-cell resolution have provided valuable new insights into cell diversity and genetic regulation of cell types revealing mechanisms underlying cardiovascular differentiation and morphogenesis. Single-cell analysis of *in vivo* mouse heart development have revealed chamber-specific and temporal changes in gene expression underlying embryonic heart development from E9.5 to postnatal day 21 establishing anatomical patterns of gene expression in the heart (Li et al., 2016) and new insights into transcriptional programs underlying cardiac maturation (DeLaughter et al., 2016). These studies provide a valuable resource by which to understand transcriptional mechanisms underlying diverse fate choices involved in cardiac development and morphogenesis *in vivo.* Like many single-cell transcriptomic studies, they further highlight the importance of dissecting cell heterogeneity to understand mechanisms underlying the identity and fate of cells in health and disease.

Human pluripotent stem cells are a key model system to study human cardiovascular developmental biology (Murry and Keller, 2008). However, the fidelity by which cardiac directed differentiation *in vitro* recapitulates the transcriptional programs governing the diversity of cell fates generated *in vivo* is not well understood. Analyzing differentiation efficiency has relied extensively on expression signatures from bulk samples consisting of hundreds of thousands of cells which lack the resolution to dissect gene expression and cell subpopulation heterogeneity. Furthermore, identification of rare populations remains challenging. Cardiac progenitor populations, for example, are difficult to identify but important as they constitute cell states underlying decision points in fate diversification (Qyang et al., 2007). Lastly, modelling development and disease requires an accurate, quantifiable analysis of complex decisions underlying the orchestration of heterogeneous cell types responsible for cell phenotypes, from molecular characterization to physiological function.

In this study, we report RNA-sequencing data captured from more than forty thousand single cells navigating stage-specific transitions through *in vitro* cardiac directed differentiation from pluripotency using an established small molecule Wnt modulation protocol (Burridge et al., 2014; Lian et al., 2012). In coordination with two companion papers (Ding et al., In review; Nguyen et al., in review-a), we present in this study three primary outcomes illustrating the power of this data set for expanding our understanding of stem cell directed differentiation as a platform to study cardiovascular development. First, unsupervised clustering analysis of single cell data identify fifteen transcriptionally distinct cell subpopulations transiting cardiac directed differentiation. *In vitro* gene expression networks underlying these cell subpopulations correlate with mesendoderm fate choices made during gastrulation phases of germ layer specification and progressive developments in cardiovascular differentiation and morphogenesis *in vivo* (Li et al., 2016; Peng et al., 2016). Second, we implement a new lineage trajectory prediction algorithm, *scdiff,* specifically designed for time-course single-cell data (Ding et al., In review). This algorithm utilizes transcription factors and their regulatory networks to provide mechanistic insights into the analysis of coordinated differentiation of diverse cell subpopulations through differentiation from pluripotency into the cardiac lineage. As a result of lineage tracing transcriptional regulatory networks using *scdiff*, we identify cardiomyocytes and outflow tract cells as the dominant terminal cell types derived from *in vitro* differentiation. Lastly, we leverage single-cell transcriptomic data to elucidate new insights into gene regulatory mechanisms controlling cardiovascular fate from pluripotency. We identify the non-DNA binding homeodomain protein HOPX as an early marker of mesoderm and a context-specific regulator of gene networks underlying endothelial and cardiac fate. Using conditional genetic loss of function assays, we determine that HOPX is required for vascular fate specification and is an essential

regulator underlying differentiation of hPSC-derived cardiomyocytes at molecular and physiological levels.

Results

Single-cell RNA-sequencing analysis of cardiac directed differentiation

To gain insights into the genetic regulation of cardiovascular development, we performed single-cell transcriptional profiling of human iPSCs navigating from pluripotency through stage-specific transitions in cardiac differentiation (**Figure 1A**). Small molecule Wnt modulation was used as an efficient method to differentiate pluripotent cells toward the cardiac lineage (Burridge et al., 2014; Lian et al., 2012). WTC-CRISPRi hiPSCs (Mandegar et al., 2016) were chosen as the parental cell line for this study. These cells are genetically engineered with an inducible nuclease-dead Cas9 fused to a KRAB repression domain (**Figure S1A**). Transcriptional inhibition by gRNAs targeted to the transcriptional start site is doxycycline-dependent and can be designed to silence genes in an allele-specific manner. The versatility of this line provides a means to use this scRNA-seq data as a reference point for future studies aiming to assess the transcriptional basis of cardiac differentiation at the single-cell level. Cells were verified to have a normal 46 X,Y male karyotype by Giemsa banding analysis before analysis by scRNA-Seq (**Figure S1B**). Cells were captured at time points corresponding to stage-specific transitions in cell state including pluripotency (day 0), germ layer specification (day 2), and progressing through progenitor (day 5), committed (day 15), and definitive (day 30) cardiac cell states. We harvested a total of 44,020 cells of which 43,168 cells were retained after quality control analysis. In total, we captured expression of 17,718 genes (detected in at least 44 cells and with expression values within the overall expression range of 3 median absolute deviation, as described in our companion paper (Nguyen et al., in review-b)). We used dimensionality reduction approaches to visualize all 43,168 cells in low-dimensional space, in which cell's coordinates were estimated so that they preserve the expression similarity (local and global distance in the original multidimensional space) in *t*-SNE plots (left), and the differentiation pseudotime (transition probability between cells) in diffusion plots (right). These data show distinct transcriptomic clustering and distribution of cells undergoing differentiation (**Figure 1B**).

We generated a time-course gene expression profile using a wide range of known cardiac developmental genes by measuring expression among all cells to reveal the temporally-restricted expression dynamics of stage-specific genes reflecting cardiac fate choices (**Figure 1C**). To confirm that the differentiation follows known developmental trajectories, we identified cells that express known cardiac differentiation genes and observed how they cluster in low-dimensional space (**Figure 1D**) (Coifman et al., 2005; Moignard et al., 2015). Day 0 (pluripotency) cells express POU5F1 (OCT4), a key pluripotency gene that is down-regulated as cells transition to germ layer differentiation at day 2. Day 2 cells are characterised by expression of EOMES, a marker of pan-mesendoderm lineages. TMEM88, previously identified as a regulator of cardiac progenitor fate (Palpant et al., 2013), displays expression primarily in day 5 cells. TNNI1, the foetal isoform of troponin I, marks all cardiac derivatives beginning on day 5 and continuing through to day 30, whereas TTN, a sarcomeric protein found in definitive contractile cardiomyocytes, is expressed predominantly in day 30 cells. Analysis of this time-course data using an unsupervised clustering algorithm (Clustering at Optimal REsolution (CORE) (Nguyen et al., in review-b)) identified fifteen transcriptionally unique subpopulations across all five time points (**Figure 1E**). Using the subpopulation identification, we show that stage-specific genes are consistently expressed

with known developmental state transitions, subpopulations within each time point reveal a diversity of cell states coordinating fate choices during differentiation. Overall, these data show that small molecule-mediated cardiac directed differentiation generates developmentally distinct populations of cells displaying expected temporal-specific transcriptional profiles.

Phenotypic diversity and lineage heterogeneity during differentiation.

With recent developments in high-resolution transcriptomic mapping of mouse *in vivo* development of the cardiovascular system from the earliest stages of gastrulation (DeLaughter et al., 2016; Li et al., 2016; Peng et al., 2016), we set out to map single-cell heterogeneity of human *in vitro* derived subpopulations to cell types against stages of lineage specification *in vivo.* To assist in elucidating the molecular identity of each subpopulation, we analyzed high-resolution spatio-temporal gene expression during mouse *in vivo* gastrulation to identify genes that mark known developmental populations and cell types (**Figure S2**). Using previously published approaches, laser microdissection was used to capture germ layer cells of mid-gastrula stage (E7.0) embryos (Peng et al., 2016), with an expanded analysis to include early- (E6.5) and late-gastrulation (E7.5) mouse embryos (unpublished data). High-throughput RNA-sequencing data were compiled into corn plots, with each plot depicting discrete spatial-temporal patterns of gene expression corresponding to individually sequenced sections. Expression data for selected genes involved in heart development in the epiblast/ectoderm and mesoderm/endoderm for E6.5, E7.0, and E7.5 mouse embryos are shown in **Figure S3**. To determine phenotypic identities based on gene expression networks governing each human *in vitro*-derived subpopulation during differentiation, we visualized the spatio-temporal patterns of gene expression in the gastrulating mouse embryo including: EOMES (pan-mesendoderm), MESP1 and MIXL1 (mesoderm), SOX17 and FOXA2 (endoderm), and NKX2-5 (cardiac lineage transcription factor) (**Figure 2A, Figure S3A-D**). These *in vivo* expression dynamics of mouse gastrulation established spatiotemporal reference points for identifying *in vitro* subpopulations.

Based on these observations, we dissected the transcriptional phenotype of subpopulations identified during human cardiac directed differentiation. From pluripotency **(Figure S4)**, cells navigate through germ layer specification (day 2), comprising three transcriptionally distinct subpopulations that express the pan-mesendoderm gene, EOMES (**Figure 2B-C, Figure S3A**). Specific day 2 subpopulations express genes involved in mesoderm (D2:S2), mesendoderm (D2:S3), and definitive endoderm (D2:S1) differentiation (**Figure 2B-C and Figure S3D and Figure S4C**). Gene ontology (GO) analysis of differentially expressed genes between subpopulations indicated that only D2:S2 showed significant enrichment for cardiogenic gene networks (**Figure 2D, Table S1**). Surprisingly, these data show that only 34% of day 2 cells comprise cardiogenic mesoderm marked by MESP1 with the majority of cells characterized by mesendoderm and definitive endoderm expression patterns. At the progenitor stage (day 5), gene expression network and gene ontology analysis were used to identify subpopulations of cardiac precursors (D5:S1 and D5:S3) (**Figure 2E-G and Figure S4D**), a persistent population of definitive endoderm (D5:S2) (**Figure 2E-F and Figure S4D**), and endothelial cells (D5:S3) (**Figure 2E-G**). At day 15 and day 30 cells split into two distinct cardiac subpopulations (**Figure 2H-M and Figure S4E-F**). NKX2-5, MYH6, TTN and other cardiac structural and regulatory genes were identified in S2 (**Figure 2H-M and Figure S4E-F**). In contrast, S1 was primarily characterized by GO enrichment for genes associated with extracellular matrix deposition, motility, and cell adhesion (**Figure 2J and M**) which was supported by identification of a significant number of fibroblast-like cells marked by THY1 (CD90) in S1 (**Figure 2I and L**). The co-existence of a non-contractile cell population, which is characterized as non-

myocytes, is common in directed cardiac differentiation (Dubois et al., 2011). Taken together, these data show iPSC differentiation into committed (day 15) and definitive (day 30) cardiomyocytes (S2) and non-contractile $THY1^+$ cells (S1).

Overall, these data provide single-cell resolution transcriptomic data of small molecule *in vitro* directed cardiac differentiation. Unsupervised clustering using CORE (Nguyen et al., in review-b) enabled the identification of subpopulations with biologically distinct phenotypes linked to *in vivo* cardiovascular development fate choices (**Figure 2N**). To assess the level of maturity derived from this protocol relative to *in vivo* human development, we compared day 30 clusters against ENCODE RNA-seq data from foetal and adult hearts (**Figure 2O**). Using genes that reflect either early foetal (TNNI1, MYH6) vs late stages of heart development (MYH7, TNNI3, MYL2), the most differentiated *in vitro* derived cardiac population (D30:S2) remains more developmentally immature than even first trimester human hearts.

Lineage predictions based on regulatory gene networks governing differentiation

While these bulk population analyses provided clarity into the diversity of cell types represented in cardiac differentiation, we sought to understand the lineage trajectories and gene regulatory networks governing diversification of cell fates (**Figure S5**). We analyzed the full differentiation data set using two previously published computational packages, Monocle2 and dPath (**Figure S6**) (Gong et al., 2017; Qiu et al., 2017; Trapnell et al., 2014). dPath predicts trajectories on the basis of metagene entropy to rank cells on their differentiation potential and implements a self-organizing map (SOM) and random walk with restart (RWR) algorithm to construct lineage hierarchies in an unbiased manner based on measures of transcriptional entropy. Monocle builds lineage trajectories by aligning cells in pseudotime and establishes branch points based on a manifold tree structure learned from the multidimensional expression data using the discriminative dimensionality reduction via learning tree (DDRTree) algorithm. Analysis of our time-course cardiac differentiation data using Monocle2 or dPath for all 43,168 cells was carried out using default parameters. For gene inputs, we used the 6000 top most variable genes for dPath, and known markers or differentially expressed genes for Monocle2. These analyses did not yield lineage predictions that accurately fit known *in vivo* developmental trajectories possibly due to limitations in linking such transcriptionally disparate populations represented in our time series data set through gene expression entropy or pseudotime analysis (**Figure S6**).

To address these problems, we implemented a probabilistic method for constructing regulatory networks from single-cell time series expression data (*scdiff*: Cell Differentiation Analysis Using Time-series Single-cell RNA-seq Data) (Ding et al., In review) (**Figure S7**). The algorithm utilizes TF-gene databases to model gene regulation relationships based on the directional changes in expression of TFs and target genes at parental and descendant states. These prerequisites impact both cell assignment and model learning since each state is represented by a probabilistic model that takes into account not just the expression but also the regulatory information. As such, *scdiff* does not exclusively rely on expression similarity to connect states allowing it to overcome problems related to sampling since it can still identify descendent states even if they are less similar in terms of their actual expression profiles.

We used *scdiff* to predict lineages underlying the diversity of fates during small molecule-mediated cardiac directed differentiation (**Table S2** and **Figure 3A**). While the algorithm permits movement between days, we did not observe any cells shift from the original day of isolation (**Table S2**), thus highlighting the challenge of connecting lineage trajectories on the basis of transcriptional similarity.

Overall, the algorithm identified three distinct lineages from pluripotency comprising 10 nodes. Since this algorithm reassigns cells based on regulatory networks, we analyzed the distribution of cell subpopulations based on our CORE cluster classifications as outlined in **Figure 2** to establish population identities linking predicted lineages (**Figure 3A**). The nodes classify into known developmental fate trajectories and are anchored by genes that define developmental states and cell populations *in vivo* (**Figure 2A, Figure 3B-C**). The first lineage (N1:N2) diverts from pluripotency into a SOX17/FOXA2/EPCAM^{+} definitive endoderm population that terminates at day 2 and is comprised almost exclusively of D2:S1 and D2:S3 (**Figure 3A-C**). The second lineage, N1:N3:N5, transitions from pluripotency (N1) into node 3 which is primarily comprised of definitive endoderm (D2:S1) and mesendoderm (D2:S3) but includes a larger fraction of MESP1/T^{+} mesoderm (D2:S2).This node is predicted to be the origin of another terminal lineage endpoint, node 5 at day 5, comprising FOXA2/EPCAM^{+} definitive endoderm cells (D5:S2 and D5:S4) (**Figure 3A-C**). The third lineage comprises the longest trajectory through differentiation involving stepwise transitions in cardiac fate (N1:N4:N6-N9 and N6-N10). Pluripotent cells (N1) give rise initially to node 4 mesoderm (D2:S2) and mesendoderm (D2:S3) cells with subsequent progression into TMEM88/TNNI1^{+} cardiac precursor cells (N6: primarily D5:S1 and D5:S3). From day 5 the algorithm predicts a bifurcation of fate giving rise to THY1^{+}/NKX2-5^{-} non-contractile cardiac derivatives (N8-10: D15:S1 and D30:S1) or NKX2-5^{+}/MYH6^{+} committed CM (N7: D15:S2) that progress onto MYH7^{+}/MYL2^{+} definitive CM (N9: D30:S2) (**Figure 3A-C**).

We leveraged the regulatory network predictions to identify key transcription factors and target genes underlying progressive fate changes across all 10 nodes (**Figure 3D and Table S2**). These data reinforce established mechanisms of cardiac lineage specification. In particular, we found evidence for down-regulation of Wnt/β-catenin signaling (LEF1) between N4-N6 which is required to transition from mesoderm into the cardiac progenitor cell (Paige et al., 2010; Ueno et al., 2007). From the progenitor node N6 into contractile cardiomyocytes N7:N9, the data show proper down-regulation of progenitor transcription factors such as YY1 and up-regulation of TFs known to control cardiomyocyte differentiation such as NKX2-5. Downstream target genes involved in governing the transition from progenitor cells to a differentiated cardiac state were expressed concomitantly (**Figure 3D**).

To gain new insights into lineage trajectories derived during differentiation, we sought to understand the gene network underlying specification of THY1^{+} non-contractile cardiac derivatives N8:N10, a population currently not well defined although widely used for tissue engineering applications (Thavandiran et al., 2013). The predicted network underlying this transition showed significant down-regulation of cardiac TFs NKX2-5 and MAZ while other TFs involved in lipid metabolism/sterol regulation (SREBF2) and protein sumoylation (TOPORS) were up-regulated (**Figure 3D**). Of particular note, we observed up-regulation of Pre-B cell leukemia transcription homeobox (PBX1: $P = 1.1e^{-16}$, mean DE target fold change = 2.72), a transcriptional regulator that activates a network of genes associated with cardiac outflow tract (OFT) morphogenesis (Arrington et al., 2012).

To further assess an OFT phenotype, we compared expression of a panel of cardiomyocyte, early developmental vascular endothelial, and OFT development genes across all subpopulations comprising transitions from day 5 to 30 (**Figure 3E**). While early developmental vascular EC differentiation genes (TAL1, CDH5) were expressed in D5:S3, these genes were not expressed in D15 or D30. Furthermore, while D15 and D30 S2 cells expressed cardiac sarcomere genes and transcription factors associated with first heart field specification (IRX4 and HCN4), S1 cells exclusively expressed an extensive network of genes associated with OFT development including PITX2, TBX18, HOXA1-3, FGF10, GJA1, and KDR

(**Figure 3E**). We also performed gene ontology analysis of differentially expressed genes between D30:S1 (N10) and D30:S2 (N9) cells. These data show a significant enrichment for gene networks related to vascular development ($P = 1.1e^{-11}$) and blood vessel morphogenesis ($P = 4.7e^{-9}$) exclusively within node 10 D30:S1 cells. This finding is supported by single-cell visualization showing enrichment of OFT gene expression in S1 vs S2 by *t*-SNE analysis of THY1 (59% D30:S1 vs 2% D30:S2), BMP4 (70% D30:S1 vs 6% D30:S2), and PITX2 (73% D30:S1 vs 17% D30:S2) (**Figure 3F-G**).

Lastly, to anchor this observation to *in vivo* cell types, we used single-cell RNA-seq data of *in vivo* heart development (Li et al., 2016) to identify the top most differentially expressed genes between outflow tract and left ventricle (LV). These data show expression of BMP4, RSPO3, TNC, and COL1A2 in D30:S1 and *in vivo* OFT derivatives and MYL2 and HOPX upregulated in cardiomyocytes (**Figure 3H**). Lastly, to assess cell-type specific transcriptional signatures, we identified differentially expressed genes between D30 S1 vs S2 and performed a Spearman rank correlation analysis against expression profiles of *in vivo* FACS sorted (Quaife-Ryan et al., 2017) or single cell-derived cardiac subtypes (Li et al., 2016). These data show that D30:S1 has a significantly stronger correlation to OFT cells (Spearman's $\rho = 0.442$) than fibroblasts (Spearman's $\rho = 0.219$), endothelium (Spearman's $\rho = 0.175$), or myocardium (Spearman's $\rho = 0.243$) ($P < 2.2x10^{-16}$ for all pairwise comparisons) **(Figure 3I)**. As expected, there was a significantly stronger correlation of D30:S2 to myocardium (Spearman's $\rho = 0.448$) compared to D30:S1 (Spearman's $\rho = 0.243$, $P < 2.2x10^{-16}$) **(Figure 3I)**. Collectively, these data indicate that directed differentiation generates distinctive populations of contractile $MYL2^+$ cardiomyocytes and $THY1^+/PITX2^+$ non-contractile cells that are transcriptionally similar to cardiac outflow tract cells.

HOPX: a context-specific regulator of cardiovascular lineage diversification *in vivo* and *in vitro*

To gain new insights into genetic regulation of cardiac subpopulations not possible without single-cell level resolution, we analyzed a panel of 52 transcription factors and epigenetic regulators known to govern diversification of mesoderm and endoderm lineages represented in this data set (**Table S3**). Expression of these regulatory genes was measured across eleven subpopulations identified between days 2-30 of differentiation (**Table S3**). HOPX, a non-DNA binding homeodomain protein identified in this analysis, has previously been identified as one of the earliest, specific markers of cardiomyocyte development (Jain et al., 2015), and governs cardiac fate by regulating cardiac gene networks through interactions with transcription factors, epigenetic regulators, and signaling molecules (Chen et al., 2002; Jain et al., 2015). We have also recently shown that HOPX functionally regulates blood formation from hemogenic endothelium (Palpant et al., 2017b).

Analysis of single-cell data revealed three unexpected observations about HOPX in the context of cardiac directed differentiation. First, HOPX is expressed as early as day 2 of differentiation during germ layer specification, much earlier than previously detected (**Figure 4A**). Second, HOPX is expressed during cardiomyocyte specification at the progenitor stage of mouse development *in vivo* whereas we detect HOPX only in endothelium (D5:C3) and not in cardiac precursor cells (D5:C1) at an equivalent time point (day 5) of *in vitro* differentiation (**Figure 4A**). Third, in contrast to previous studies *in vivo* where HOPX lineage traces almost all cardiomyocytes of the heart (Jain et al., 2015), HOPX is detected in only 16% of D30:S2 cardiomyocytes (**Figure 4B-C**). To rule out stochastic expression in cardiomyocytes due to low sequencing read depth resulting in dropout, we analyzed expression of a panel of genes known to regulate cardiac lineage specification and differentiation (**Figure 4B-C**). While HOPX is rarely detected, its expression level is equivalent to that of other cardiac TFs detected in a majority of

D30:S2 cardiomyocytes (HAND1: 67%, HAND2: 64%, GATA4: 67%, NKX2-5: 86% vs HOPX: 16%) (**Figure 4B-C**).

To assess the fidelity of HOPX expression during germ layer specification, we analyzed spatio-temporal gene expression throughout mesoderm and endoderm development in the gastrulating mouse *in vivo* and during hPSC directed differentiation *in vitro*. Previous work using MESP1 as a marker of mesoderm specification from pluripotency *in vitro* (MESP1-mCherry reporter hPSCs (Den Hartogh et al., 2015)) shows enrichment of HOPX expression in mCherry positive cells (**Figure S8B**). Similarly, assessment of gene expression during mouse gastrulation *in vivo* shows HOPX expression as early as E6.5 in the proximal portion of the nascent primitive streak (P) (**Figure 4D and Figure S8C-D**) similar to the expression pattern of MESP1 (**Figure S8E-F**). From E7.0 to E7.5, HOPX is increasingly expressed throughout the developing endoderm. By E7.5, HOPX displays residual expression in the remaining distal primitive streak, endoderm (EA to EP), and the anterior mesoderm (MA) in coordination with other cardiogenic genes including NKX2-5 and MESP1 (**Figure 4D and Figure 2A**).

We analyzed expression networks correlated with $HOPX^+$ cells at day 2 of differentiation. These data show that HOPX is most highly expressed and represented in mes-endoderm subpopulations 2 and 3 (**Figure 4A**). At day 2 $HOPX^+$ cells have significantly higher expression of mesoderm genes HAND1, T, MESP1, and MIXL1, no change in anterior primitive streak endodermal genes GSC and NODAL and significantly lower levels of definitive endoderm gene expression (FOXA2, ISL1, and SOX17) compared to $HOPX^-$ cells (**Figure S8G**). These data suggest that $HOPX^+$ cells are associated with cardiogenic mesoderm, consistent with *in vivo* spatio-temporal expression data in the mouse from E6.5-E7.5 (**Figure 4D**).

From E8.5 onward, cardiac development and morphogenesis occurs as chambers, valves, and outflow tract form. We analyzed HOPX expression across diverse cell types contributing to heart development *in vivo* using single-cell transcriptomic analysis of the E9.5 mouse heart (Li et al., 2016). These data indicate that HOPX expression is distributed throughout all chambers and cell types of the heart (**Figure 4E**). While HOPX expression largely coincides with expression of cardiac genes MYH7 and ACTN2, HOPX is also expressed in endothelial cells ($CDH5^+$ and/or $PECAM1^+$), smooth muscle cells ($MYH11^+$ and/or $TAGLIN2^+$), and epicardial cells ($WT1^+$) (**Figure 4D and Table S4**). Consistent with mouse heart development, analysis of human foetal development at each trimester indicate a robust activation of HOPX during heart development *in vivo* (**Figure S8A**). These findings confirm that HOPX is expressed at the early stages of mes-endodermal fate during germ layer specification *in vitro* and *in vivo*, which is much earlier than previously reported, and is expressed in multiple cell lineages that contribute to heart development.

Lineage trajectory of HOPX-expressing cells

We analyzed the lineage trajectory of $HOPX^+$ cells at single-cell resolution during cardiac directed differentiation to determine the core gene networks and transcription factors governing successive fate choices of HOPX expressing cells during cardiac differentiation (**Figure 4F-H, Figure S9, and Table S5**). At day 2 HOPX is expressed predominantly in mes-endoderm (D2:S2 9% and D2:S3 6%) and rarely in definitive endoderm (D2:S1 2%) with the HOPX lineage at this early stage of specification comprising 2 lineages (N2 and N3) enriched for expression of cardiogenic mesoderm genes such as MESP1 (**Figure 4F-H and Figure S9**). From day 2 into day 5, $HOPX^+$ cells remain sparse in progenitor cell populations where day 2 N3 splits into two day 5 lineages N4 and N5 (**Figure 4F-H and Figure S9**). Based on

lineage prediction, an equal proportion of HOPX$^+$ cells give rise to TNNI1$^+$ cardiac precursor cells (N4: 389 cells) or TAL1$^+$ expressing endothelial cells (N5: 381 cells) with both fates governed by established TFs and downstream gene networks required for endothelial (NRP2, KDR) vs cardiac fate specification (TNNI1, TMEM88) (**Figure 4F-H and Table S5**). Progressing to day 15 of differentiation HOPX remains rare (2-4% of cells) and splits into two separate lineages derived from day 5 cardiac precursor cells (N4). Governed in part by increased NKX2-5 and downregulation of the cardiac progenitor TF YY1, HOPX cardiac precursor cells differentiate into MYL2/IRX4$^+$ cardiomyocytes (N6-N7) while a separate branch governed by TFs such as PBX1 differentiate into outflow tract derivatives (N8-N9) (**Figure 4F-H and Figure S9**). Overall, these data show that HOPX is expressed early in germ layer specification and shows spatio-temporal expression broadly across mes-endodermal lineages *in vivo* and *in vitro.* Analysis of HOPX gene expression across populations shows that HOPX is most highly expressed in endothelial cells at day 5 and in committed cardiomyocytes at day 30 of directed differentiation (**Figure 4A**).

Chromatin and expression analysis of HOPX in cardiac vs endothelial lineage specification

We analyzed epigenetic and transcriptional regulation at the HOPX locus in cardiac vs endothelium (Palpant et al., 2017a) (**Figure 5A-B and Figure S10A**). Cells were assayed by chromatin immunoprecipitation for repressive chromatin (H3K27me3), actively transcribed chromatin (H3K4me3), and gene expression by RNA-seq (Palpant et al., 2017b). In the context of cardiac directed differentiation, these data show that the HOPX locus is epigenetically repressed on the basis of abundant H3K27me3 compared to H3K4me3 in day 5 cardiac precursor cells (**Figure 5A-B**). This is consistent with RNA-seq, qRT-PCR data, and expression analysis from HOPX-tdTomato reporter cells showing that HOPX is expressed late during cardiac differentiation, well after sarcomere formation during cardiac directed differentiation (**Figure 5C and Figure S10B-D**). In the context of endothelium, consistent with our previous analysis (Palpant et al., 2017b), the HOPX locus show significant gene expression correlated with abundant H3K4me3 and reduced H3K27me3 in KDR$^+$/CD34$^+$ ECs at day 5 of differentiation (**Figure 5A**). These data show a direct link between chromatin regulation of the HOPX locus and expression of HOPX in cardiac vs endothelial lineage specification.

Functional analysis of HOPX in cardiac and endothelial fate

To dissect the role of HOPX early during differentiation of endothelial fate and late stages of cardiac differentiation, we analyzed single-cell expression data of HOPX$^+$ cells in combination with HOPX conditional loss of function during differentiation. At the progenitor stage (day 5), HOPX is identified predominantly in D5:S3 with 84% of HOPX cells co-expressing the endothelial TF TAL1 and HOPX$^+$ cells express significantly higher levels of endothelial genes (TAL1, CD34) and significantly lower levels of cardiac progenitor genes (TNNI1, NKX2-5) compared to HOPX$^-$ cells (**Figure 5D**). Focusing exclusively on D5:S1 cells which have the most cardiogenic progenitor cell-like transcriptomes (<1% TAL1$^+$, 99% TMEM88$^+$, 96% TNNI1$^+$), we found that HOPX$^+$ cells still represent a population with robust activation of endothelial genes (**Figure S8H**). Analysis of CRISPRi HOPX LOF cells revealed that loss of HOPX blocks expression of endothelial genes including TAL1, CD34, CDH5 but not GATA1 or ETV2. In contrast, HOPX LOF in progenitor cells significantly up-regulates genes involved in cardiac fate (**Figure 5E**). This is consistent with studies showing that regulation of endothelial vs cardiac fate is

430 governed by mutually antagonistic networks (Palpant et al., 2015; Van Handel et al., 2012), and suggests that loss of HOPX blocks endothelial specification and enables cells to acquire a cardiac fate instead.

At later stages of differentiation, $HOPX^+$ cells are identified predominantly with cardiomyocytes (D30:S2) (**Figure 4A**). Analysis of gene networks associated with HOPX expression showed significantly increased expression of nearly all tested genes associated with cardiac development with the exception of
435 MYH6, the foetal myosin heavy chain isoform, which was significantly upregulated in $HOPX^-$ cells (**Figure 5F**). Analysis of LOF cells showed that loss of HOPX specifically blocks expression of late cardiac differentiation gene networks including MYH7, MYL2, and TNNI3 with no effect on early cardiac specification genes (MYL7, TNNI1) (**Figure 5G**). Analysis of single-cell data showed that among cells expressing HOPX at day 30, HOPX expression was significantly higher in subpopulation 2
440 compared to cells expressing HOPX in subpopulation 1 supporting a role in cardiomyocyte specification as opposed to outflow tract differentiation which is consistent with fate mapping studies (Jain et al., 2015) (**Figure 5H**). Collectively these data show a dependency of cardiac differentiation gene networks on expression of HOPX in committed cardiomyocytes.

Based on these findings, we tested whether HOPX is required for cardiac or endothelial fate
445 specification. Analysis of $KDR^+/CD34^+$ endothelial cells at day 5 of differentiation showed a significant loss of $KDR^+/CD34^+$ endothelium in HOPX LOF cells vs controls indicating a direct role in governing specification of the endothelial fate (**Figure 5I**). In contrast, there was no significant effect on generating $cTnT^+$ cardiomyocytes (**Figure 5J**) and no difference in the onset of beating between WT and HOPX LOF cardiomyocytes (**Figure 5K**).

450 Changes in myofilament composition underlie significant changes in functional parameters including calcium sensitivity, force production, and elasticity that govern cardiac contractility during development (Yang et al., 2014). Engineered heart tissues are a well-established model for testing physiological analysis of cardiac differentiation and maturation (Mills et al., 2017). Comparative analysis of hPSC-derived cardiomyocytes in monolayer, engineered heart tissues (EHTs) and adult heart tissue
455 shows that HOPX expression significantly increases in EHTs comparable to monolayer cardiomyocytes cultured for one year (**Figure S8I**) (Mills et al., 2017). Since HOPX directly regulates myofilament genes involved in late stages of differentiation, we tested whether HOPX LOF had an impact on cardiac contractility of heart tissues. EHTs were made comprising HOPX LOF or control cardiomyocytes with 10% HS27a stromal cells and placed on PDMS exercise polls for 10 days (Voges et al., 2017). Analysis
460 of beating rate, a significant determinant of contractility, was not different between groups. However, loss of HOPX significantly reduced the contractile force of EHTs by 65% compared to controls ($P = 0.0009$) (**Figure 5L**). Collectively, these data provide evidence for a role of HOPX in cardiovascular development in which HOPX functions early during progenitor specification in governing differentiation of endothelium. Subsequently HOPX is shown to be essential for cardiomyocyte differentiation and the
465 expression of gene networks required for definitive cardiomyocyte development and physiological function (**Figure 5M**).

Discussion

470 This study provides single-cell resolution RNA-sequencing of human cardiac directed *in vitro* differentiation. Transcriptomic analysis of 43,168 cells traversing stepwise transitions in fate revealed cellular heterogeneity and the underlying gene networks involved in cardiac fate choices from pluripotency. The identification and characterization of *in vitro* derived cell types are supported by spatio-

temporal gene expression of the gastrulating mouse embryo and single-cell analysis of *in vivo* heart development, providing a direct link to the complex restriction of fates underlying cardiovascular lineage specification *in vivo*. We leverage the computational power of single-cell level analysis to identify the framework of transcription factors and gene regulatory networks underlying progressive diversification of fates from pluripotency into mesoderm, definitive endoderm, endothelium, cardiomyocytes, and outflow tract fates. Transcriptional analysis of subpopulations revealed context-specific functions of transcription factors and epigenetic regulators underlying cardiovascular fate choices. Utilizing single-cell computational analysis with CRISPRi hiPSCs for conditional loss of function, we identify HOPX as an early marker of cardiac fate specification during gastrulation and functionally required for endothelial and cardiovascular differentiation from pluripotency. Collectively, we use a widely implemented and efficient protocol using Wnt modulation for cardiac directed differentiation (Burridge et al., 2014; Lian et al., 2012) with CRISPRi hiPSCs (Mandegar et al., 2016) providing a platform to dissect cardiac differentiation at single-cell resolution.

While the progression of heart development and morphogenesis from time points spanning E8.5 to P21 have been analyzed at single-cell resolution (DeLaughter et al., 2016; Li et al., 2016), a comprehensive transcriptomic profiling of the lineages derived by human pluripotent stem cell directed differentiation from pluripotency has not been available. Cardiac directed differentiation protocols using small molecules to modulate Wnt signaling have emerged in recent years as a simple, cost-effective, and reliable method to generate high-purity cardiac derivatives from hPSCs. Stem cell-derived cardiomyocytes generated using this approach have been utilized in translational applications to model patient-specific diseases (Ang et al., 2016; Bayzigitov et al., 2016; Ebert et al., 2014; Smith et al., 2017; Wu et al., 2015), test cardiotoxicity (Maillet et al., 2016), screen novel therapeutic drugs (Casini et al., 2017; Sharma et al., 2014), generate engineered heart tissue constructs to model the 3D environment of the heart (Huebsch et al., 2016; Tzatzalos et al., 2016), and develop cell-based regenerative therapies to repair heart tissue post-infarct (Hartman et al., 2016). The current study provides whole genome-wide analysis of stage-specific changes in gene expression during cardiac differentiation as a resource with which to dissect cell subpopulations at the molecular level. The statistical power of 43,168 transcriptomes establishes a new experimental platform for fundamental and translational applications in cardiovascular biology.

Analysis of subpopulations during early stages of differentiation indicate a surprising contribution of mesendoderm and definitive endoderm coordinately specified with cardiac fates through the progenitor stage of differentiation. In particular, a minority of cells (34%) comprise $MESP1^{+}$ cardiogenic mesoderm at day 2 that ultimately give rise to all cardiac derivatives at day 30. The interaction between endoderm and mesoderm in governing lineage specification *in vivo* is well known, and these data suggest that a critical functional role of induction cues provided by directed differentiation protocols is to establish the necessary population stoichiometry of transiently sustained endoderm required to support mesoderm in the derivation of high purity cardiac fates *in vitro.*

We evaluated lineage trajectories from single-cell data by implementing a lineage prediction algorithm, *scdiff*, specifically designed for learning regulatory networks controlling differentiation from single-cell time series data. This method relies on iterative analysis of known transcription factor-gene interactions to establish a gene regulatory framework that statistically links disparate populations captured across intermittent time intervals through differentiation. Analysis of cardiac directed differentiation using *scdiff* revealed stage-specific genetic regulators underlying diversification of cardiac fate from pluripotency. In particular, these data revealed new insights into the bifurcation of cardiac precursor cells

at day 5 of differentiation into NKX2-5^{+}/MYL2^{+} ventricular cardiomyocytes and NKX2-5^{-}/PITX2^{+} cardiac outflow tract (OFT) cells. While previous studies have routinely described a non-contractile THY1^{+} (CD90^{+}) fibroblast-like cell used commonly for tissue engineering applications (Dubois et al., 2011; Thavandiran et al., 2013), this population remains poorly studied with no strong evidence for an *in vivo* correlate. Using transcriptional fate mapping and gene network analysis, we provide single-cell level transcriptome-wide evidence directly linked to *in vivo* cell cardiac types that non-contractile THY1^{+} cells are a cardiac OFT derivative. Future work will require a more comprehensive analysis of this population in diverse differentiation protocols to determine the reproducibility of this finding across various induction mechanisms, with a key emphasis on single-cell analysis endpoints as a mechanism for phenotyping hPSC-derived cardiac subpopulations. Of importance, congenital heart disease (CHD) is among the most common forms of congenital defects (van der Linde et al., 2011), and OFT anomalies account for roughly 30% of CHD incidences (Thom et al., 2006). While the genetic basis of OFT malformations has been well studied (Arrington et al., 2012), the capacity to study OFT development and disease using hPSC models has not been possible. Our finding of an OFT cell subpopulation derived from *in vitro* cardiac directed differentiation presents new opportunities to develop translational platforms utilizing this cell type for disease modelling or therapies.

It is well-established that *in vitro* cardiac differentiation does not generate cardiomyocytes with the transcriptional profile, cellular diversity, morphometry, or function maturity of adult *in vivo*-derived cardiomyocytes (Yang et al., 2014). This is at least in part a consequence of dysregulation of the stage-specific gene networks that are not properly modelled *in vitro.* In this study, we analyzed a panel of 52 transcription factors and epigenetic regulators across eleven cardiac differentiation-derived cell subpopulations at single-cell resolution. These data provide evidence that HOPX is one of the earliest transcriptional markers of cardiogenic mesoderm during germ layer specification *in vitro* and *in vivo.* Functional studies show that HOPX functionally regulates gene networks required for vascular and cardiac differentiation and function. While HOPX is a key developmental regulator of cardiac myoblasts early in heart development *in vivo* (Jain et al., 2015) with data from this and other studies (Chen et al., 2002; Trivedi et al., 2010) showing a key role in heart maturation, we observe it rarely in *in vitro* derived cardiomyocytes. Engineered heart tissues (EHTs) force significant up-regulation of genes involved in maturation including HOPX (Mills et al., 2017). We used this platform to show that loss of HOPX in EHTs reduces contractile function of hPSC-derived cardiomyocytes implicating it as a key genetic regulator of differentiation. Given the significance of promoting adult-like phenotypes from *in vitro* differentiated cell types for translational applications in disease modelling and therapy, these data provide evidence that HOPX expression is a key transcriptional regulator near absent during cardiac directed differentiation *in vitro* underlying at least in part the transcriptional and functional immaturity of hPSC-derived cardiomyocytes.

Taken together, our study provides transcriptional profiling of human *in vitro* cardiac differentiation from more than forty thousand single-cells revealing cell diversity and genetic networks governing lineage progression of hPSC cardiac directed differentiation by small molecule Wnt modulation. These data provide new insights into the complexity of cell populations represented in stage-specific transitions from pluripotency and, coupled with the use of CRISPRi loss of function, establish a unique reference point for dissecting gene networks involved in human cardiac development and disease.

Acknowledgements. Sequencing was performed by the Institute for Molecular Bioscience Sequencing Facility at the University of Queensland. Assistance with Figure 1A schematic was provided by Suzy

Hur. The WTC CRISPRi hiPSCs and pQM plasmid backbone were kindly provided by the Conklin lab (UCSF, Gladstone Institute). We thank Prof Richard Harvey (Victor Change Cardiac Research Institute) for assistance reviewing the draft manuscript. This work was supported by the Australian Research Council (SR1101002) (NJP), the ARC Discovery Early Career Award (DE160100755) (ESW), National Health and Medical Research Council grants 1107599 and 1083405. ZBJ and JD were supported in part by grant 1R01GM122096 from the National Institute of Health, USA. This work was also supported by a Strategic Priority Research Program of the Chinese Academy of Sciences (XDA01010201 to N.J., XDA01010303 to J.D.J.H), National Key Basic Research and Development Program of China (2014CB964804, 2015CB964500, 2015CB964803), and National Natural Science Foundation of China (91219303, 31430058, 31401261, 91329302, 31210103916, and 91519330). P.P.L.T. is a Senior Principal Research Fellow of the National Health and Medical Research Council of Australia (1110751).

Author contributions:

CEF: Generated cells for single-cell RNA-seq, performed cell-based experiments including HOPX LOF, and wrote the manuscript
QN: Primary lead on computational analysis of single-cell data and wrote the manuscript
SWL: Performed single-cell isolation, barcoding, and sequencing and performed computational analysis of single-cell RNA-seq data and edited the manuscript
AH: Performed computational analysis of single-cell RNA-seq data and edited the manuscript
HSC: Generated cells for single-cell RNA-seq
HKV: Provided bioengineered heart tissue experimental training
SSS: Conceived and generated iTranscriptome data for mouse gastrulation *in vivo*
JDJH: Conceived and generated iTranscriptome data for mouse gastrulation *in vivo*
OP: Assisted with the preparation of iTranscriptome data for publication
GP: Conceived and generated iTranscriptome data for mouse gastrulation *in vivo*
GJB: Assisted with sequencing
AS: Performed computational analysis of single-cell data
ANC: Assisted with sequencing
TJB: Assisted with sequencing
NJ: Conceived and generated iTranscriptome data for mouse gastrulation *in vivo*
CEM: Supervised HOPX LOF experiments and edited the manuscript
ESW: Adapted *scdiff* for large scale single-cell data sets and generated *scdiff* data
JD: Conceived of and developed *scdiff*
ZBJ: Conceived of and developed *scdiff* and wrote the manuscript
YW: Performed computational analysis of single-cell data and edited the manuscript
JH: Assisted with bioengineered heart tissue experiments and edited the manuscript
PPLT: Supervised iTranscriptome analysis, consulted single-cell phenotypes and edited the manuscript
JEP: Conceived and supervised experiments involving computational genomics analysis of single-cell RNA-seq data, wrote manuscript, and generated funding for the work
NJP: Conceived and supervised experiments involving stem cell differentiation, performed HOPX experiments, wrote manuscript, and generated funding for the work.

See Supplemental Information for extended methods.

Figure Legends

Figure 1. Single-cell analysis of cardiac directed differentiation. (A) Schematic of protocol for small molecule directed differentiation from pluripotency into the cardiac lineage. hPSC: human pluripotent stem cell; GLS: germ layer specification; PC: progenitor cell: cCD: committed cardiac derivative; dCD: definitive cardiac derivative. **(B)** 43,168 single-cells transiting cardiac differentiation beginning at pluripotency (day 0) and transitioning through mesoderm (day 2) into progenitor (day 5), committed (day 15), and definitive (day 30) cardiac derivatives. Data are presented using *t*-SNE plot, pseudospacing cells by the nonlinear transformation of similarity in gene expression to preserve the local and global distance of cells in multidimensional space when embedded into two dimensional *t*-SNE space (left), and diffusion plot, pseudospacing cells in a trajectory based on diffusion distance (transition probability) between two cells (right). **(C)** Mean gene expression across all cells at individual time points showing proper temporal expression of stage-specific genes governing differentiation into the cardiac lineage. Shown are pluripotency genes (DNMT3B, POU5F1, NANOG), mes-endoderm genes (EOMES, MIXL1, T, MESP1), and genes governing cardiomyocyte differentiation including signaling regulators (TMEM88), transcription factors (ISL1, HAND1, NKX2-5, TBX5, GATA4), calcium handling genes (ATP2A2, PLN) and sarcomere genes (TNNI1, MYH6, MYH7, MYL7). **(D)** Diffusion plots showing pseudospacing at single-cell resolution for gene expression of stage-specific genes during differentiation based on known genetic regulators of cardiac fate specification including POU5F1 (day 0), EOMES (day 2), TMEM88 (day 5), TNNI1 (day 15), and TTN (day 30). Cells are colored in a binary manner. If the cell expresses the gene it is colored according to the day of isolation (0, 2, 5, 15, or 30). Non-expressing cells are shaded gray. **(E)** Representation of unsupervised clustering analysis (Nguyen et al., in review-b) using *t*-SNE plots to show single-cell level expression of stage-specific gene expression at each day of differentiation based on known genetic regulators of cardiac fate specification including POU5F1 (day 0), EOMES (day 2), ISL1 (day 5), TNNI1 (day 15), and MYL7 (day 30). If the cell expresses the gene it is colored according to subpopulation 1-4 in which the cell is associated. Non-expressing cells are shaded gray. Above each *t*-SNE plot, the percentage of cells expressing the gene in each subpopulation is shown together with the expression histogram and the reference *t*-SNE plot.

Figure 2. Subpopulation identification and characterization. (A) Corn plots showing spatial domains of EOMES, MESP1, SOX17 and NKX2-5 expression in the mesoderm and endoderm of E6.5, E7.0, and E7.5 mouse embryos during gastrulation (unpublished RNA-seq data for E6.5 (n = 6) and E7.5 (n = 6) embryos and published data for E7.0 embryo (Peng et al., 2016)). Positions of the cell populations ("kernels" in the 2D plot of RNA-Seq data) in the germ layers: the proximal-distal location in descending numerical order (1 = most distal site) and in the transverse plane of the mesoderm and endoderm – Anterior half (EA) and Posterior half (EP) of the endoderm, Anterior half (MA) and Posterior half (MP) of the mesoderm, and Posterior epiblast (P) containing the primitive streak. Color scales represent levels of expression as log10 of fragments per kilobase million (FPKM + 1) (see **Figure S3** for schematic of iTranscriptome). **(B-D)** Analysis of day 2 subpopulations represented by **(B)** Reference *t*-SNE (left) and diffusion (right) plots and the percent of cells in each subpopulations (D2:S1-S3), **(C)** analysis of primitive streak genes EOMES (pan-mesendoderm transcription factor), MESP1 (cardiogenic mesoderm transcription factor), and SOX17 (definitive endoderm transcription factor). Below each gene name are shown the following data from left to right: *t*-SNE plot and diffusion plot of cells expressing each gene, percent of cells expressing gene, expression level of gene in each subpopulation. **(D)** Gene ontology

analysis of differentially expressed genes showing enrichment for networks governing cardiac development enriched in subpopulation 2. **(E-G)** Analysis of day 5 progenitor subpopulations represented by **(E)** reference *t*-SNE (left) and diffusion (right) plots and the percent of cells in each subpopulations (D5:S1-S4), **(F)** analysis of progenitor genes TAL1 (endothelial fate transcription factor), TNNI1 (early development sarcomere isoform of TNNI), and SOX17 (definitive endoderm transcription factor). Below each gene name are shown the following data from left to right: *t*-SNE plot and diffusion plot showing cells expressing each gene, percent of cells expressing gene, expression level of gene in each subpopulation. **(G)** Gene ontology analysis of differentially expressed genes showing enrichment for networks governing cardiac development (D5:S1), definitive endoderm (D5:S2), and endothelium (D5:S3). **(H-J)** Analysis of day 15 subpopulations represented by **(H)** reference *t*-SNE (left) and diffusion (right) plots and the percent of cells in each subpopulations (D15:S1-S2), **(I)** analysis of cardiac genes MYL7 (early development sarcomere isoform of MYL), NKX2-5 (cardiac transcription factor), and THY1 (fibroblast marker). Below each gene name are shown the following data from left to right: *t*-SNE plot and diffusion plot showing cells expressing each gene, percent of cells expressing gene, expression level of gene in each subpopulation. **(J)** Gene ontology analysis of differentially expressed genes showing enrichment for networks governing extracellular matrix and cell motility (D15:S1) and cardiac development (D15:S2). **(K-M)** Analysis of day 30 subpopulations represented by **(K)** reference *t*-SNE (left) and diffusion (right) plots and the percent of cells in each subpopulations (D30:S1-S2), **(L)** analysis of cardiac genes TNNI1 (early development sarcomere isoform of TNNI), MYH7 (mature sarcomere isoform of MYH), and THY1 (fibroblast marker). Below each gene name are shown the following data from left to right: *t*-SNE plot and diffusion plot showing cells expressing each gene, percent of cells expressing gene, expression level of gene in each subpopulation. **(M)** Gene ontology analysis of differentially expressed genes showing enrichment for networks governing system development and morphogenesis (D30:S1) and cardiac development (D30:S2). **(N)** Overall phenotypic determinations of subpopulation identity based on *in vivo* anchoring genes outlined through stage-specific transitions in differentiation. CM: cardiomyocyte. **(O)** Expression of cardiac genes in day 30 hPSC-derived cardiomyocytes (all cells vs S1 vs S2) relative to expression levels in human foetal and adult heart samples (ENCODE). Gene expression is measured as counts per million mapped reads and each gene is internally normalized to maximum expression.

Figure 3. Transcription factor regulatory networks predict developmental fate choices during cardiac differentiation. (A) Stepwise transitions from pluripotency into cardiac lineages from pluripotency predicted on the basis of gene regulatory networks (GRN) detected between pairwise changes in cell state during differentiation. Circles indicate distinct nodes governed by a common GRN. Since cells can be re-assigned based on the expression of their genes, the re-distribution of subpopulations established by clustering analysis and phenotyping as outlined in **Figure 2** are represented as pie charts within each circle indicating the percent of cells from each subpopulation contributing to that node. Each node is numbered N1-N10 for reference. **(B)** Expression level within each node for known cardiac and endodermal lineage markers reflecting *in vivo* populations. Gene expression is represented across all nodes: EOMES (pan-mesoderm), MESP1 and T (cardiogenic mesoderm), SOX17, FOXA2, EPCAM (endoderm), TMEM88, GATA4, TNNI1 (cardiac progenitors), THY1 (fibroblasts), MYH6, MYH7, MYL2, TTN, NKX2-5 (late-stage cardiac differentiation markers). **(C)** Phenotypic identity of nodes reflecting stage-specific transitions in cell state through cardiac directed differentiation. **(D)** Analysis of transcription factors (TFs) and genes controlling stage-specific regulatory networks underlying cell fate

transitions. Mean DE target fold change calculates the fold change for the differentially expressed targets of the TF. DE gene fold change shows up or down-regulated fold change of TF target genes. **(E)** Heat map comparing expression across all cells from day 5, 15, and 30 subpopulations for genes involved in progenitor specification, vascular endothelial development, outflow tract development, and primary heart field cardiomyocyte development. **(F)** Gene ontology analysis comparing day 30 S1 vs S2 showing gene networks involved in vascular development enriched in S1 vs cardiac muscle development enriched in S2. **(G)** *t*-SNE and diffusion plots for all cells from days 15 and 30 showing expression distribution of the cardiac gene MYH7 (high in S2 at day 15 and 30) relative to outflow tract development genes THY1, PITX2, and BMP4 (high in S1 at day 15 and 30). **(H)** The top most differentially expressed genes identified by *in vivo* single-cell analysis comparing outflow tract (OFT) vs ventricular cardiomyocyte (Wu Dev Cell) were compared against their expression level in D30:S1 vs D30:S2 *in vitro* derived cardiac derivatives. **(I)** Differentially expressed genes between subpopulations D30:S1 and D30:S2 were used to assess transcriptional similarity to *in vivo* cell types using Spearman's correlation analysis. Values are presented median Spearman's value ρ. Significant differences between pairs of correlation coefficients were calculated using a Fisher Z-transformation. *P*-values for all tests were below the double precision limit of 2.2e-308

Figure 4. HOPX is rarely expressed during *in vitro* cardiac directed differentiation. (A) Analysis of HOPX expression in eleven subpopulations from day 2 to day 5 of differentiation showing expression as early as day 2 mesoderm and highest expression in day 5 endothelial cells (ECs) and day 30 cardiomyocytes (CMs). **(B-C)** Analysis of known genetic regulators of heart development only in subpopulation 2 at day 30 of differentiation. *t*-SNE plots of merged data sets from two continuous days **(B)** for all cells between day 15-30 for each gene showing robust distribution of key cardiac regulatory genes with the exception of HOPX. Mean data **(C)** show expression of gene only in cells in which that gene is expressed and the percent of cells expressing the gene. **(D)** Corn plots showing the spatial domains of HOPX expression in the mesoderm and endoderm of E6.5, E7.0, and E7.5 mouse embryos during gastrulation (unpublished RNA-seq data for E6.5 (n = 6) and E7.5 (n = 6) embryos and published data for E7.0 embryo, (Peng et al., 2016)). Positions of the cell populations ("kernels" in the 2D plot of RNA-Seq data) and in the germ layers: the proximal-distal location in descending numerical order (1 = most distal site) in the transverse plane of the mesoderm and endoderm – Anterior half (EA) and Posterior half (EP) of the endoderm, Anterior half (MA) and Posterior half (MP) of the mesoderm, and Posterior epiblast (P) containing the primitive streak. Color scales represent levels of expression as log10 of fragments per kilobase million (FPKM+1). **(E)** Single-cell expression analysis of E9.5 mouse heart (Li et al., 2016) showing HOPX expression relative to markers of cardiomyocytes (MYH7, ACTN2) and endothelial cells (CDH5, PECAM1) (scale bars are Log2(RPM)). Table (right) shows percent of cardiac (MYH7), endothelial (PECAM1), and smooth muscle (TAGLN2) cells co-expressing HOPX in various regions of the developing mouse heart. **(F)** Expression level of HOPX generated from *in vitro* cardiac differentiation single-cell data as a mean of all cells and subpopulations combined from each time point during the time-course of cardiac differentiation from pluripotency. **(G)** Single-cell expression analysis of HOPX at day 2, 5, 15, and 30. Data presented include *t*-SNE plots indicating distribution and localization of HOPX expressing cells in different subpopulations (bottom), the percentage of HOPX+ cells in each subpopulation (top left), bar graphs showing expression of HOPX in each subpopulation (top middle), and the reference *t*-SNE plot demarcating subpopulations (top right). **(H)** Lineage tracing prediction of only HOPX cells at each stepwise transition from pluripotency into cardiac lineages predicted on the basis

of gene regulatory networks (GRN) detected between pairwise changes in cell state during differentiation. Circles indicate distinct nodes governed by a common GRN. Since cells can be re-assigned based on network stability, the re-distribution of subpopulations established by our initial unsupervised clustering analysis (as outlined in **Figure 2**) are represented as pie charts within each circle indicating the percent of cells from each subpopulation contributing to that node. Nodes are demarcated as N1-N9. **(G)** Expression level within each node for known developmental lineage markers. Gene expression is represented across all nodes: NANOG (pluripotency), MESP1 (cardiac-mesoderm), TAL1 and CDH5 (endothelium) TNNI1, MYH7, NKX2-5, IRX4 (primarily first heart field cardiac genes), THY1, PITX2, HOXA2, and TBX18 (second heart field and outflow tract development).

Figure 5. HOPX functionally governs cardiac and endothelial differentiation. (A-B) ChIP-seq and RNA-seq analysis of the HOPX locus in day 5 cardiac progenitor cells vs endothelial cells showing chromatin levels for H3K4me3 vs H3K27me3 and gene expression by RNA-seq. Quantitative data are shown **(A)** and raw data provided for cardiac differentiation **(B). (C)** Immunohistochemistry of HOPX-tdTomato reporter cells showing co-expression with the progenitor marker ISL1 at day 5 and the mature cardiac marker α-actinin at day 40 of differentiation. **(D-E)** Analysis of gene expression comparing $HOPX^{+}$ vs $HOPX^{-}$ cells for cardiac and endothelial genes in day 5 progenitor cells using single-cell expression data **(D)** and HOPX CRISPRi KD cells (**E**, n = 6-8 group) to determine the functional requirement of HOPX for endothelial and cardiac progenitor gene networks. **(F-G)** Analysis of gene expression comparing $HOPX^{+}$ vs $HOPX^{-}$ cells for genes in day 30 cardiomyocytes using single-cell expression data **(F)** and HOPX CRISPRi KD cells (**G**, n = 6-8 group) to determine the functional requirement of HOPX on cardiac developmental gene networks. **(H)** Analysis of HOPX expression level only in $HOPX^{+}$ cells in subpopulation 1 vs subpopulation 2 cells at day 30. Data are presented as mean ± SEM. * $P < 0.05$. **(I-J)** FACS analysis of control vs HOPX CRISPRi knockdown cells for $KDR^{+}/CD34^{+}$ endothelial cells at day 5 of differentiation **(I)** and $cTnT^{+}$ cardiomyocytes at day 15 of differentiation **(J)** n = 6 per group. **(K)** Analysis of the onset of beating during directed differentiation showing no difference between HOPX CRISPRi knockdown and controls. **(L)** Validation of HOPX KD in engineered heart tissues by qRT-PCR. Functional analysis of engineered heart tissues show significant loss of contractile force with no change in beating rate in HOPX LOF tissues compared to control, n = 4-5 per group. Representative force traces are also shown. **(M)** Schematic lineage tree showing fate choices governed by HOPX during cardiac directed differentiation. Our data provide functional evidence that loss of HOPX is primarily identified within the endothelial fate at the progenitor stage of differentiation and loss of HOPX prevents expression of endothelial cell identity genes. At late stages of cardiac differentiation HOPX is most closely related with progression to a definitive cardiomyocyte and loss of HOPX functionally prevents activation of key cardiomyocyte cell identity genes involved in differentiation. * $P < 0.05$. Data are presented as mean ± SEM.

References

Ang, Y.S., Rivas, R.N., Ribeiro, A.J., Srivas, R., Rivera, J., Stone, N.R., Pratt, K., Mohamed, T.M., Fu, J.D., Spencer, C.I., *et al.* (2016). Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. Cell *167*, 1734-1749.e1722.

Arrington, C.B., Dowse, B.R., Bleyl, S.B., and Bowles, N.E. (2012). Non-synonymous variants in pre-B cell leukemia homeobox (PBX) genes are associated with congenital heart defects. Eur J Med Genet *55*, 235-237.

Bayzigitov, D.R., Medvedev, S.P., Dementyeva, E.V., Bayramova, S.A., Pokushalov, E.A., Karaskov, A.M., and Zakian, S.M. (2016). Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes Afford New Opportunities in Inherited Cardiovascular Disease Modeling. Cardiol Res Pract *2016*, 3582380.

Burridge, P.W., Matsa, E., Shukla, P., Lin, Z.C., Churko, J.M., Ebert, A.D., Lan, F., Diecke, S., Huber, B., Mordwinkin, N.M., *et al.* (2014). Chemically defined generation of human cardiomyocytes. Nat Methods *11*, 855-860.

Casini, S., Verkerk, A.O., and Remme, C.A. (2017). Human iPSC-Derived Cardiomyocytes for Investigation of Disease Mechanisms and Therapeutic Strategies in Inherited Arrhythmia Syndromes: Strengths and Limitations. Cardiovasc Drugs Ther.

Chen, F., Kook, H., Milewski, R., Gitler, A.D., Lu, M.M., Li, J., Nazarian, R., Schnepp, R., Jen, K., Biben, C., *et al.* (2002). Hop is an unusual homeobox gene that modulates cardiac development. Cell *110*, 713-723.

Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Nadler, B., Warner, F., and Zucker, S.W. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: multiscale methods. Proc Natl Acad Sci U S A *102*, 7432-7437.

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

Den Hartogh, S.C., Schreurs, C., Monshouwer-Kloots, J.J., Davis, R.P., Elliott, D.A., Mummery, C.L., and Passier, R. (2015). Dual reporter MESP1 mCherry/w-NKX2-5 eGFP/w hESCs enable studying early human cardiac differentiation. Stem Cells *33*, 56-67.

Ding, J., Aronow, B., Kamminski, N., Kitzmiller, J., Whitsett, J., and Bar-Joseph, Z. (In review). Reconstructing regulatory differentiation networks from time series single cell expression data

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

Ebert, A.D., Kodo, K., Liang, P., Wu, H., Huber, B.C., Riegler, J., Churko, J., Lee, J., de Almeida, P., Lan, F., *et al.* (2014). Characterization of the molecular mechanisms underlying increased ischemic damage in the aldehyde dehydrogenase 2 genetic polymorphism using a human induced pluripotent stem cell model system. Sci Transl Med *6*, 255ra130.

Gong, W., Rasmussen, T.L., Singh, B.N., Koyano-Nakagawa, N., Pan, W., and Garry, D.J. (2017). Dpath software reveals hierarchical haemato-endothelial lineages of Etv2 progenitors based on single-cell transcriptome analysis. Nat Commun *8*, 14362.

Hartman, M.E., Dai, D.F., and Laflamme, M.A. (2016). Human pluripotent stem cells: Prospects and challenges as a source of cardiomyocytes for in vitro modeling and cell-based cardiac repair. Adv Drug Deliv Rev *96*, 3-17.

Huebsch, N., Loskill, P., Deveshwar, N., Spencer, C.I., Judge, L.M., Mandegar, M.A., Fox, C.B., Mohamed, T.M., Ma, Z., Mathur, A.*, et al.* (2016). Miniaturized iPS-Cell-Derived Cardiac Muscles for Physiologically Relevant Drug Response Analyses. Sci Rep *6*, 24726.

Jain, R., Li, D., Gupta, M., Manderfield, L.J., Ifkovits, J.L., Wang, Q., Liu, F., Liu, Y., Poleshko, A., Padmanabhan, A.*, et al.* (2015). HEART DEVELOPMENT. Integration of Bmp and Wnt signaling by Hopx specifies commitment of cardiomyoblasts. Science *348*, aaa6071.

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

Lian, X., Hsiao, C., Wilson, G., Zhu, K., Hazeltine, L.B., Azarin, S.M., Raval, K.K., Zhang, J., Kamp, T.J., and Palecek, S.P. (2012). Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. Proc Natl Acad Sci U S A *109*, E1848-1857.

Maillet, A., Tan, K., Chai, X., Sadananda, S.N., Mehta, A., Ooi, J., Hayden, M.R., Pouladi, M.A., Ghosh, S., Shim, W.*, et al.* (2016). Modeling Doxorubicin-Induced Cardiotoxicity in Human Pluripotent Stem Cell Derived-Cardiomyocytes. Sci Rep *6*, 25333.

Mandegar, M.A., Huebsch, N., Frolov, E.B., Shin, E., Truong, A., Olvera, M.P., Chan, A.H., Miyaoka, Y., Holmes, K., Spencer, C.I.*, et al.* (2016). CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. Cell Stem Cell *18*, 541-553.

Mills, R.J., Titmarsh, D.M., Koenig, X., Parker, B.L., Ryall, J.G., Quaife-Ryan, G.A., Voges, H.K., Hodson, M.P., Ferguson, C., Drowley, L.*, et al.* (2017). Functional screening in human cardiac organoids reveals a metabolic mechanism for cardiomyocyte cell cycle arrest. Proc Natl Acad Sci U S A.

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E.*, et al.* (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat Biotechnol *33*, 269-276.

Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell *132*, 661-680.

Nguyen, Q., Lukowski, S., Chiu, H., Friedman, C., Senabouth, A., Bruxner, T., Christ, A., Palpant, N., and Powell, J. (in review-a). Determining cell fate specification and genetic contribution to cardiac disease risk in hiPSC-derived cardiomyocytes at single cell resolution.

Nguyen, Q., Lukowski, S., Chiu, H., Senabouth, A., Bruxner, T., Christ, A., Palpant, N., and Powell, J. (in review-b). Determining cell fate specification and genetic contribution to cardiac disease risk in hiPSC-derived cardiomyocytes at single cell resolution.

Paige, S.L., Osugi, T., Afanasiev, O.K., Pabon, L., Reinecke, H., and Murry, C.E. (2010). Endogenous Wnt/beta-catenin signaling is required for cardiac differentiation in human embryonic stem cells. PLoS One *5*, e11134.

Palpant, N.J., Pabon, L., Friedman, C.E., Roberts, M., Hadland, B., Zaunbrecher, R.J., Bernstein, I., Zheng, Y., and Murry, C.E. (2017a). Generating high-purity cardiac and endothelial derivatives from patterned mesoderm using human pluripotent stem cells. Nat Protoc *12*, 15-31.

Palpant, N.J., Pabon, L., Rabinowitz, J.S., Hadland, B.K., Stoick-Cooper, C.L., Paige, S.L., Bernstein, I.D., Moon, R.T., and Murry, C.E. (2013). Transmembrane protein 88: A Wnt regulatory protein that specifies cardiomyocyte development. Development (Cambridge) *140*, 3799-3808.

Palpant, N.J., Pabon, L., Roberts, M., Hadland, B., Jones, D., Jones, C., Moon, R.T., Ruzzo, W.L., Bernstein, I., Zheng, Y*., et al.* (2015). Inhibition of β-catenin signaling respecifies anterior-like endothelium into beating human cardiomyocytes. Development *142*, 3198-3209.

Palpant, N.J., Wang, Y., Hadland, B., Zaunbrecher, R.J., Redd, M., Jones, D., Pabon, L., Jain, R., Epstein, J., Ruzzo, W.L*., et al.* (2017b). Chromatin and Transcriptional Analysis of Mesoderm Progenitor Cells Identifies HOPX as a Regulator of Primitive Hematopoiesis. Cell Rep *20*, 1597-1608.

Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G*., et al.* (2016). Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. Dev Cell *36*, 681-697.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat Meth *14*, 979-982.

Quaife-Ryan, G.A., Sim, C.B., Ziemann, M., Kaspi, A., Rafehi, H., Ramialison, M., El-Osta, A., Hudson, J.E., and Porrello, E.R. (2017). Multicellular Transcriptional Analysis of Mammalian Heart Regeneration. Circulation *136*, 1123-1139.

Qyang, Y., Martin-Puig, S., Chiravuri, M., Chen, S., Xu, H., Bu, L., Jiang, X., Lin, L., Granger, A., Moretti, A*., et al.* (2007). The renewal and differentiation of Isl1+ cardiovascular progenitors are controlled by a Wnt/beta-catenin pathway. Cell Stem Cell *1*, 165-179.

Sharma, A., Marceau, C., Hamaguchi, R., Burridge, P.W., Rajarajan, K., Churko, J.M., Wu, H., Sallam, K.I., Matsa, E., Sturzu, A.C*., et al.* (2014). Human induced pluripotent stem cell-derived cardiomyocytes as an in vitro model for coxsackievirus B3-induced myocarditis and antiviral drug screening platform. Circ Res *115*, 556-566.

Smith, A.S., Macadangdang, J., Leung, W., Laflamme, M.A., and Kim, D.H. (2017). Human iPSC-derived cardiomyocytes and tissue engineering strategies for disease modeling and drug screening. Biotechnol Adv *35*, 77-94.

Thavandiran, N., Dubois, N., Mikryukov, A., Massé, S., Beca, B., Simmons, C.A., Deshpande, V.S., McGarry, J.P., Chen, C.S., Nanthakumar, K*., et al.* (2013). Design and formulation of functional pluripotent stem cell-derived cardiac microtissues. Proc Natl Acad Sci U S A *110*, E4698-4707.

Thom, T., Haase, N., Rosamond, W., Howard, V.J., Rumsfeld, J., Manolio, T., Zheng, Z.J., Flegal, K., O'Donnell, C., Kittner, S*., et al.* (2006). Heart disease and stroke statistics--2006 update: a report from the

American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation *113*, e85-151.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol *32*, 381-386.

Trivedi, C.M., Zhu, W., Wang, Q., Jia, C., Kee, H.J., Li, L., Hannenhalli, S., and Epstein, J.A. (2010). Hopx and Hdac2 interact to modulate Gata4 acetylation and embryonic cardiac myocyte proliferation. Dev Cell *19*, 450-459.

Tzatzalos, E., Abilez, O.J., Shukla, P., and Wu, J.C. (2016). Engineered heart tissues and induced pluripotent stem cells: Macro- and microstructures for disease modeling, drug screening, and translational studies. Adv Drug Deliv Rev *96*, 234-244.

Ueno, S., Weidinger, G., Osugi, T., Kohn, A.D., Golob, J.L., Pabon, L., Reinecke, H., Moon, R.T., and Murry, C.E. (2007). Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. Proc Natl Acad Sci U S A *104*, 9685-9690.

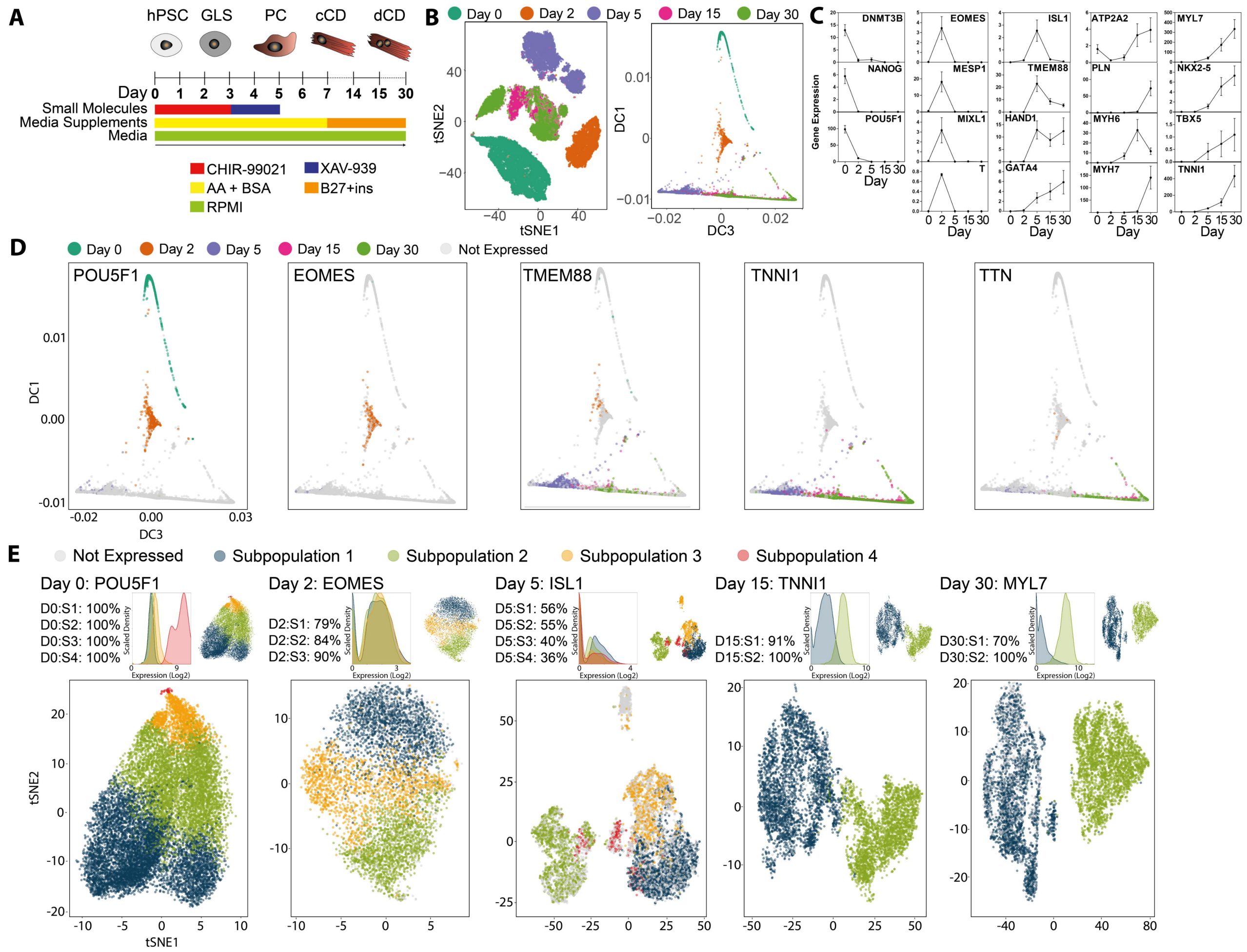
van der Linde, D., Konings, E.E., Slager, M.A., Witsenburg, M., Helbing, W.A., Takkenberg, J.J., and Roos-Hesselink, J.W. (2011). Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. J Am Coll Cardiol *58*, 2241-2247.

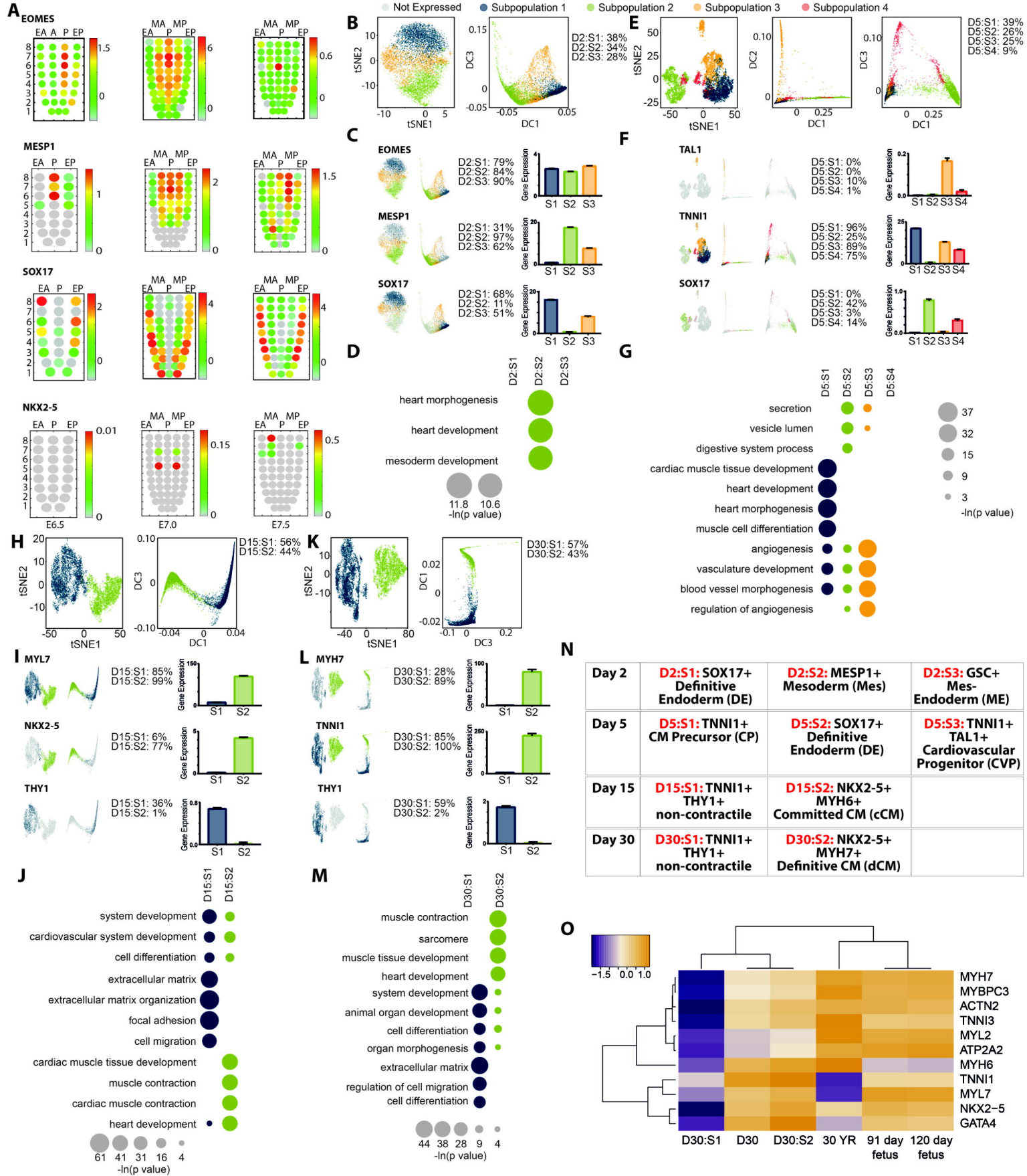
Van Handel, B., Montel-Hagen, A., Sasidharan, R., Nakano, H., Ferrari, R., Boogerd, C.J., Schredelseker, J., Wang, Y., Hunter, S., Org, T.*, et al.* (2012). Scl represses cardiomyogenesis in prospective hemogenic endothelium and endocardium. Cell *150*, 590-605.

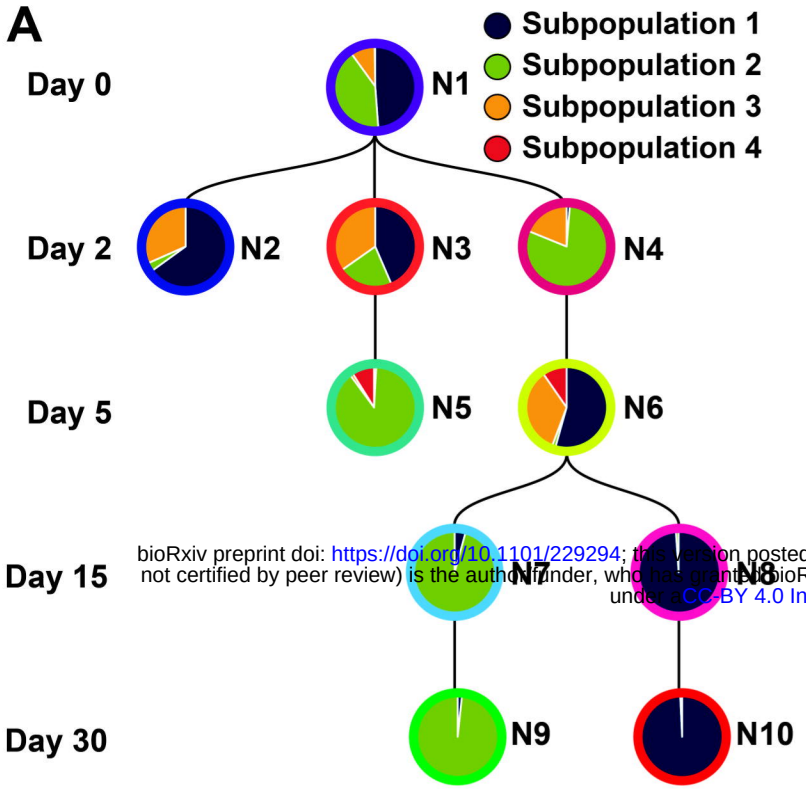
Voges, H.K., Mills, R.J., Elliott, D.A., Parton, R.G., Porrello, E.R., and Hudson, J.E. (2017). Development of a human cardiac organoid injury model reveals innate regenerative potential. Development *144*, 1118-1127.

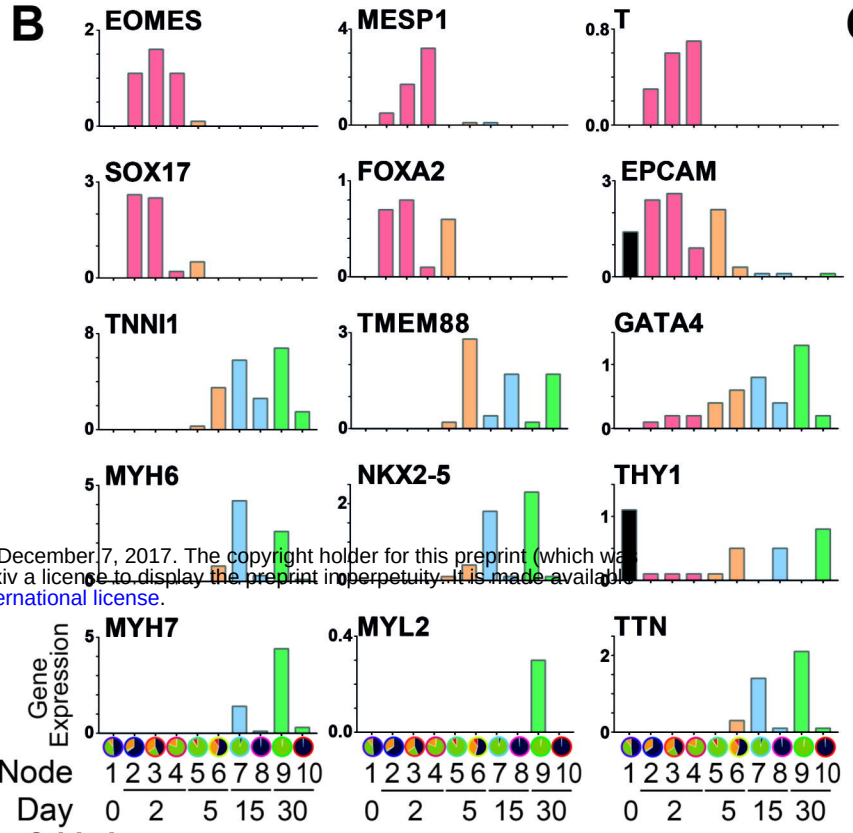
Wu, H., Lee, J., Vincent, L.G., Wang, Q., Gu, M., Lan, F., Churko, J.M., Sallam, K.I., Matsa, E., Sharma, A.*, et al.* (2015). Epigenetic Regulation of Phosphodiesterases 2A and 3A Underlies Compromised β-Adrenergic Signaling in an iPSC Model of Dilated Cardiomyopathy. Cell Stem Cell *17*, 89-100.

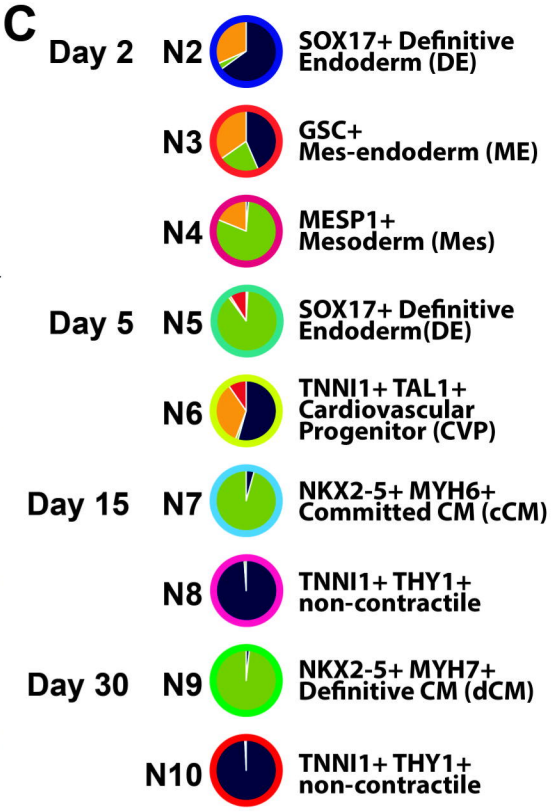
Yang, X., Pabon, L., and Murry, C.E. (2014). Engineering adolescence: maturation of human pluripotent stem cell-derived cardiomyocytes. Circ Res *114*, 511-523.

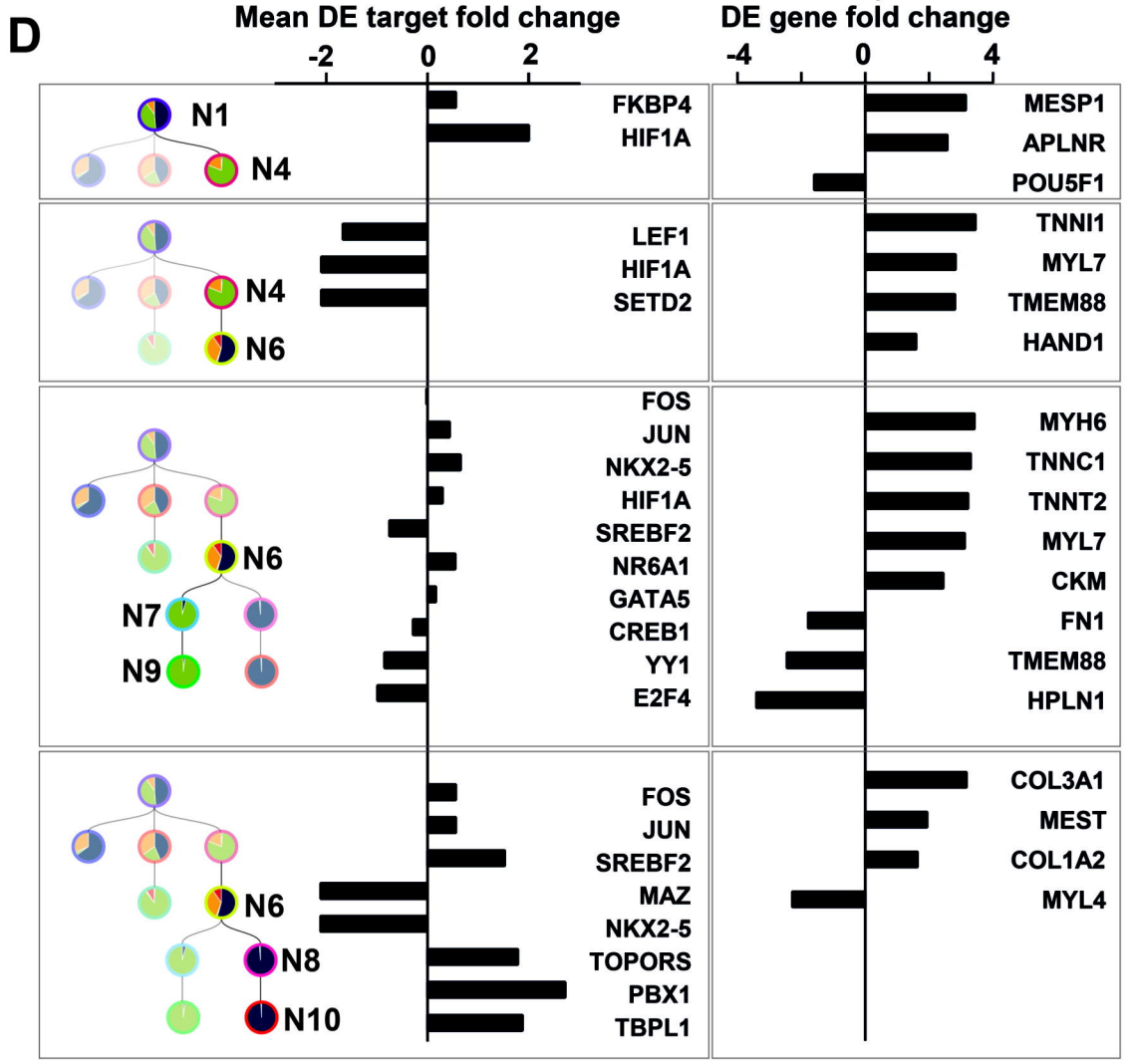

A
hPSC
GLS
PC
cCD
dCD
Day 0 1 2 3 4 5 6 7 14 15 30
Small Molecules
Media Supplements
Media
CHIR-99021
XAV-939
AA + BSA
B27+ins
RPMI
B
Day 0
Day 2
Day 5
Day 15
Day 30
tSNE1
tSNE2
DC3
DC1
C
Gene Expression
DNMT3B
NANOG
POU5F1
EOMES
MESP1
MIXL1
T
ISL1
TMEM88
HAND1
GATA4
ATP2A2
PLN
MYH6
MYH7
MYL7
NKX2-5
TBX5
TNNI1
Day
D
Day 0
Day 2
Day 5
Day 15
Day 30
Not Expressed
POU5F1
EOMES
TMEM88
TNNI1
TTN
DC1
DC3
E
Not Expressed
Subpopulation 1
Subpopulation 2
Subpopulation 3
Subpopulation 4
Day 0: POU5F1
D0:S1: 100%
D0:S2: 100%
D0:S3: 100%
D0:S4: 100%
Day 2: EOMES
D2:S1: 79%
D2:S2: 84%
D2:S3: 90%
Day 5: ISL1
D5:S1: 56%
D5:S2: 55%
D5:S3: 40%
D5:S4: 36%
Day 15: TNNI1
D15:S1: 91%
D15:S2: 100%
Day 30: MYL7
D30:S1: 70%
D30:S2: 100%
Scaled Density
Expression (Log2)
tSNE1
tSNE2

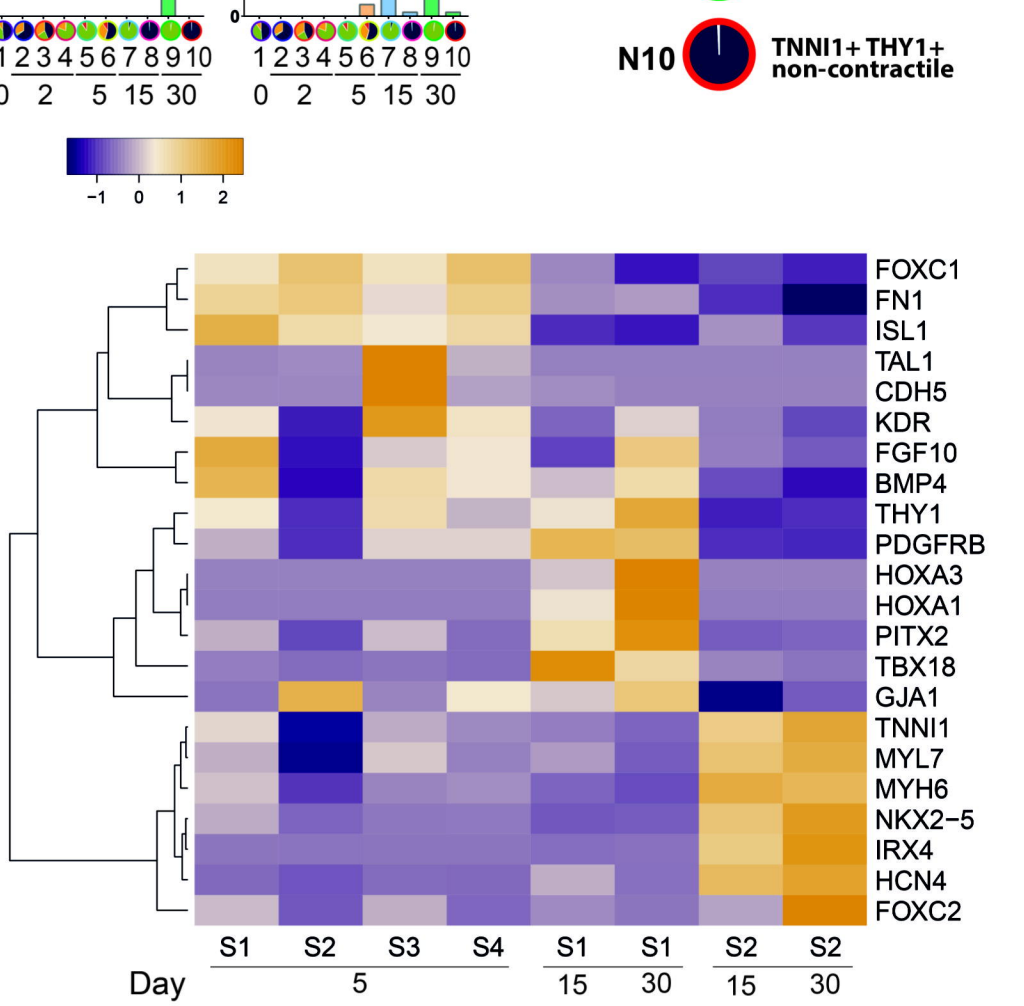

A
EOMES
MESP1
SOX17
NKX2-5
EA A P EP
MA MP
E6.5
E7.0
E7.5
B
Not Expressed
Subpopulation 1
Subpopulation 2
Subpopulation 3
Subpopulation 4
tSNE1
tSNE2
DC1
DC3
D2:S1: 38%
D2:S2: 34%
D2:S3: 28%
E
D5:S1: 39%
D5:S2: 26%
D5:S3: 25%
D5:S4: 9%
DC2
C
EOMES
D2:S1: 79%
D2:S2: 84%
D2:S3: 90%
MESP1
D2:S1: 31%
D2:S2: 97%
D2:S3: 62%
SOX17
D2:S1: 68%
D2:S2: 11%
D2:S3: 51%
Gene Expression
S1 S2 S3
F
TAL1
D5:S1: 0%
D5:S2: 0%
D5:S3: 10%
D5:S4: 1%
TNNI1
D5:S1: 96%
D5:S2: 25%
D5:S3: 89%
D5:S4: 75%
SOX17
D5:S1: 0%
D5:S2: 42%
D5:S3: 3%
D5:S4: 14%
S1 S2 S3 S4
D
D2:S1
D2:S2
D2:S3
heart morphogenesis
heart development
mesoderm development
11.8
10.6
-ln(p value)
G
D5:S1
D5:S2
D5:S3
D5:S4
secretion
vesicle lumen
digestive system process
cardiac muscle tissue development
heart development
heart morphogenesis
muscle cell differentiation
angiogenesis
vasculature development
blood vessel morphogenesis
regulation of angiogenesis
37
32
15
9
3
-ln(p value)
H
D15:S1: 56%
D15:S2: 44%
K
D30:S1: 57%
D30:S2: 43%
I
MYL7
D15:S1: 85%
D15:S2: 99%
NKX2-5
D15:S1: 6%
D15:S2: 77%
THY1
D15:S1: 36%
D15:S2: 1%
L
MYH7
D30:S1: 28%
D30:S2: 89%
TNNI1
D30:S1: 85%
D30:S2: 100%
THY1
D30:S1: 59%
D30:S2: 2%
N
Day 2
D2:S1: SOX17+ Definitive Endoderm (DE)
D2:S2: MESP1+ Mesoderm (Mes)
D2:S3: GSC+ Mes-Endoderm (ME)
Day 5
D5:S1: TNNI1+ CM Precursor (CP)
D5:S2: SOX17+ Definitive Endoderm (DE)
D5:S3: TNNI1+ TAL1+ Cardiovascular Progenitor (CVP)
Day 15
D15:S1: TNNI1+ THY1+ non-contractile
D15:S2: NKX2-5+ MYH6+ Committed CM (cCM)
Day 30
D30:S1: TNNI1+ THY1+ non-contractile
D30:S2: NKX2-5+ MYH7+ Definitive CM (dCM)
J
D15:S1
D15:S2
system development
cardiovascular system development
cell differentiation
extracellular matrix
extracellular matrix organization
focal adhesion
cell migration
cardiac muscle tissue development
muscle contraction
cardiac muscle contraction
heart development
61 41 31 16 4
-ln(p value)
M
D30:S1
D30:S2
muscle contraction
sarcomere
muscle tissue development
heart development
system development
animal organ development
cell differentiation
organ morphogenesis
extracellular matrix
regulation of cell migration
cell differentiation
44 38 28 9 4
-ln(p value)
O
−1.5 0.0 1.0
MYH7
MYBPC3
ACTN2
TNNI3
MYL2
ATP2A2
MYH6
TNNI1
MYL7
NKX2−5
GATA4
D30:S1
D30
D30:S2
30 YR
91 day fetus
120 day fetus

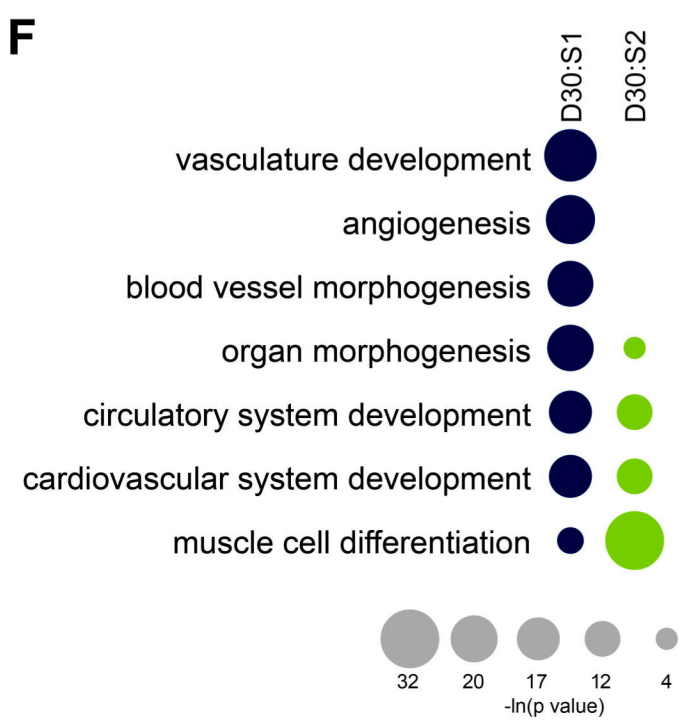


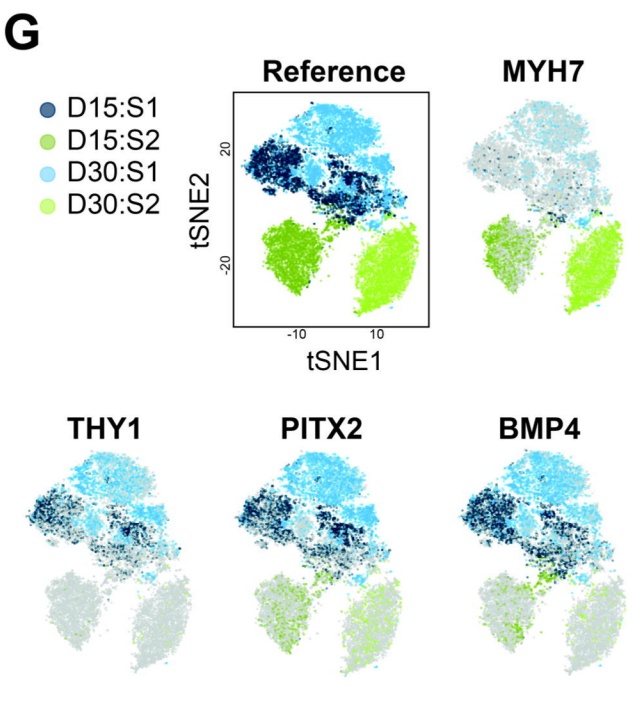


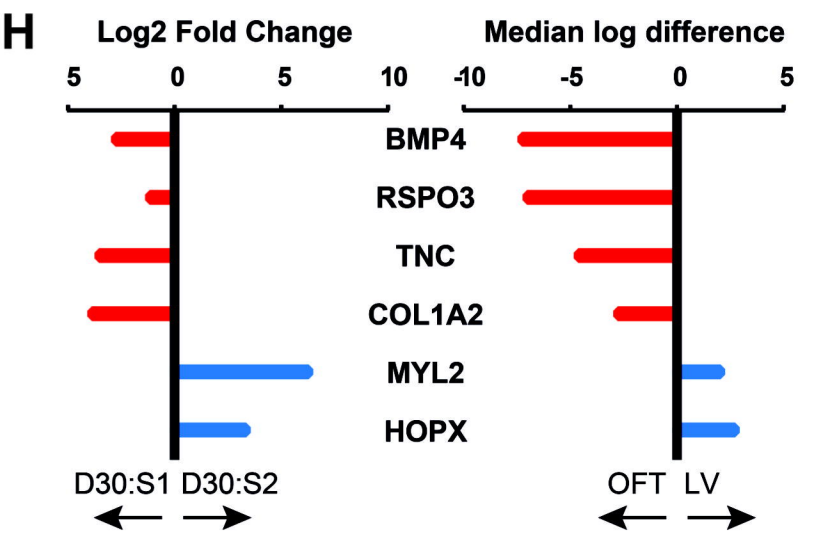








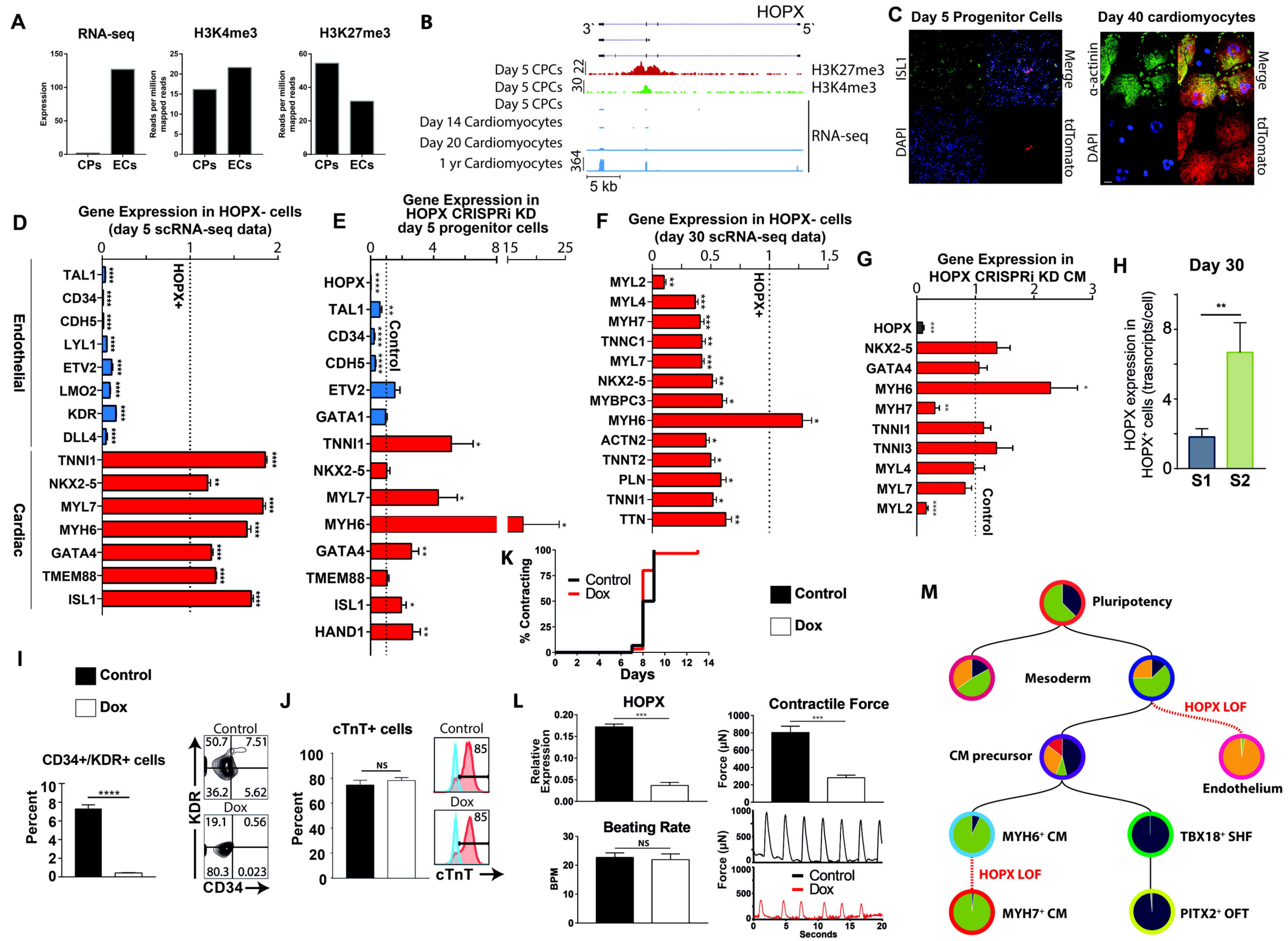




I

Correlation	ρ		Correlation	ρ
D30:S1 vs OFT	0.442	:	D30:S1 vs myocardium	0.243*
		:	D30:S1 vs fibroblasts	0.219*
		:	D30:S1 vs endothelium	0.175*
D30:S1 vs myocardium	0.243	:	D30:S2 vs myocardium	0.448*

	MYH7	PECAM1	TAGLN2
e9.5 left atrium	66	14	37
e9.5 left ventricle	65	14	47
e9.5 left ventricular septum	87	14	41
e9.5 outflow tract	61	7	52
e9.5 right atrium	71	10	26
e9.5 right ventricle	83	6	75
e9.5 right ventricular septum	91	5	58


A
RNA-seq
H3K4me3
H3K27me3
Expression
Reads per million mapped reads
CPs
ECs
B
HOPX
3`
5`
Day 5 CPCs
Day 5 CPCs
Day 5 CPCs
Day 14 Cardiomyocytes
Day 20 Cardiomyocytes
1 yr Cardiomyocytes
H3K27me3
H3K4me3
RNA-seq
5 kb
C
Day 5 Progenitor Cells
Day 40 cardiomyocytes
ISL1
DAPI
Merge
tdTomato
α-actinin
D
Gene Expression in HOPX- cells
(day 5 scRNA-seq data)
Endothelial
Cardiac
HOPX+
TAL1
CD34
CDH5
LYL1
ETV2
LMO2
KDR
DLL4
TNNI1
NKX2-5
MYL7
MYH6
GATA4
TMEM88
ISL1
E
Gene Expression in
HOPX CRISPRi KD
day 5 progenitor cells
Control
HOPX
TAL1
CD34
CDH5
ETV2
GATA1
TNNI1
NKX2-5
MYL7
MYH6
GATA4
TMEM88
ISL1
HAND1
F
Gene Expression in HOPX- cells
(day 30 scRNA-seq data)
HOPX+
MYL2
MYL4
MYH7
TNNC1
MYL7
NKX2-5
MYBPC3
MYH6
ACTN2
TNNT2
PLN
TNNI1
TTN
G
Gene Expression in
HOPX CRISPRi KD CM
HOPX
NKX2-5
GATA4
MYH6
MYH7
TNNI1
TNNI3
MYL4
MYL7
MYL2
Control
H
Day 30
HOPX expression in
HOPX+ cells (trasncripts/cell)
S1
S2
K
% Contracting
Days
Control
Dox
I
Control
Dox
CD34+/KDR+ cells
Percent
Control
50.7
7.51
36.2
5.62
Dox
19.1
0.56
80.3
0.023
KDR
CD34
J
cTnT+ cells
Percent
NS
Control
Dox
85
cTnT
L
HOPX
Relative Expression
Beating Rate
BPM
Contractile Force
Force (μN)
Control
Dox
Seconds
M
Pluripotency
Mesoderm
HOPX LOF
CM precursor
Endothelium
MYH6+ CM
TBX18+ SHF
HOPX LOF
MYH7+ CM
PITX2+ OFT