

RESEARCH ARTICLE

Single cell expression analysis reveals anatomical and cell cycle-dependent transcriptional shifts during heart development

Guang Li^{1,2,*}, Lei Tian¹, William Goodyer¹, Eric J. Kort^{3,4}, Jan W. Buikema⁵, Adele Xu⁶, Joseph C. Wu^{1,7,8,9}, Stefan Jovinge^{1,3,4,*} and Sean M. Wu^{1,7,8,*}

ABSTRACT

The heart is a complex organ composed of multiple cell and tissue types. Cardiac cells from different regions of the growing embryonic heart exhibit distinct patterns of gene expression, which are thought to contribute to heart development and morphogenesis. Single cell RNA sequencing allows genome-wide analysis of gene expression at the single cell level. Here, we have analyzed cardiac cells derived from early stage developing hearts by single cell RNA-seq and identified cell cycle gene expression as a major determinant of transcriptional variation. Within cell cycle stage-matched CMs from a given heart chamber, we found that CMs in the G2/M phase downregulated sarcomeric and cytoskeletal markers. We also identified cell location-specific signaling molecules that may influence the proliferation of other nearby cell types. Our data highlight how variations in cell cycle activity selectively promote cardiac chamber growth during development, reveal profound chamber-specific cell cycle-linked transcriptional shifts, and open the way to deeper understanding of pathogenesis of congenital heart disease.

KEY WORDS: Cell cycle, Compact myocardium, Embryonic heart, Ligand receptor interaction, Single cell RNA sequencing, Trabecular myocardium

INTRODUCTION

The mammalian heart develops from precursors in the first and second heart fields (Buckingham et al., 2005). By embryonic (E) day 10.5 in mice, a four-chambered heart consisting of the left and right atria and ventricles can be observed. The first heart field precursors give rise to the left ventricle (LV), atrial ventricular canal (AVC) and parts of the left and right atria, while the second heart field gives rise to the right ventricle (RV), the outflow tract (OFT) and some cells in the atria (Black, 2007; Buckingham et al., 2005; Devine et al., 2014;

Evans et al., 2010). Beyond the first and second heart fields, the epicardium originates from the pro-epicardial organ, and covers the outer surface of all four chambers (von Gise and Pu, 2012). In addition, the endocardium, which originates from anterolateral plate mesodermal precursors, lines the inner surfaces of all four chambers. Within the ventricular myocardium, compact and trabecular myocardium can be distinguished on histological sections (Christoffels et al., 2000; Paige et al., 2015).

During early development, cells within the heart grow rapidly and are active in distinct phases of the cell cycle (e.g. G1, when RNA and proteins are synthesized; S, when DNA replication occurs; G2/M, when active nuclear division and cytokinesis take place) (Bruneau, 2002; Harvey, 2002). Given the complex steps involved in forming different parts of the heart to achieve proper heart morphology, cell division in different regions of the heart must be precisely controlled. To precisely define the cell cycle state of single cardiac cells and further understand how cell cycle activity contributes to heart development, a technology is needed that can capture the genome-wide expression of gene transcripts at the single cell level.

Single cell RNA sequencing (scRNAseq) has emerged as a powerful tool for analyzing genome-wide expression of genes in single cells. Recently, single cell platforms have been used to identify rare cell populations in addition to analyzing their cellular heterogeneity and precise anatomical location (Durruthy-Durruthy et al., 2014; Li et al., 2017; Satija et al., 2015). Among the many single cell capture and sequencing platforms, the C1 platform from Fluidigm captures each single cell separately and performs full transcript profiling (Wu et al., 2014). By contrast, the Chromium platform from 10X Genomics uses droplets to capture single cells and selectively profiles 5' or 3' ends of transcripts (Zheng et al., 2017).

Owing to the inherent advantages and disadvantages to each approach, we profiled gene expression of mouse cardiac cells isolated at an early developmental stage (E10.5) using both techniques. Unsupervised bioinformatics analysis found that the cell cycle acted as a major source of transcriptional variation in all cardiac cell types. As C1-profiled cells were pre-isolated by cardiac zones and the 10X-profiled cells were labeled by heart field-specific lineage-tracing reporters, we were able to capitalize on these two approaches in order to separate cardiac cells based on their anatomical locations. From single cell gene expression data, we identified that the cell cycle phase drives a transcriptional shift in each cardiac cell type. We further revealed the presence of differential cell cycle activity in different cardiac regions. We then performed a ligand-receptor interaction analysis to identify distinct epicardium- and endocardium-secreted signals that regulate the growth of compact or trabecular myocardium. Together, these data highlight how variations in cell cycle activity selectively promote cardiac chamber growth during development and highlight the profound transcriptional shifts that occur within the single cells at different phases of the cell cycle. Identifying alterations in the

¹Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. ²Department of Developmental Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15201, USA. ³DeVos Cardiovascular Research Program of Spectrum Health and Van Andel Research Institute, 100 Michigan Street NE, Grand Rapids, MI 49503, USA. ⁴Michigan State University, College of Human Medicine, 15 Michigan Street NE, Grand Rapids, MI 49503, USA. ⁵Department of Cardiology, Utrecht Regenerative Medicine Center, University Medical Center Utrecht, Utrecht University, 3508 GA Utrecht, The Netherlands. ⁶Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA. ⁷Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. ⁸Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. ⁹Department of Radiology, Stanford University School of Medicine, Stanford, CA 94305, USA.

*Authors for correspondence (smwu@stanford.edu; stefan.jovinge@vai.org; guangli@pitt.edu)

© G.L., 0000-0002-8546-2364; E.J.K., 0000-0002-9966-7483; S.J., 0000-0002-8446-0228; S.M.W., 0000-0002-0000-3821

expression of key signaling molecules that drive CM proliferation may lead to greater understanding of the onset or progression of congenital heart disease.

RESULTS

Transcriptional analysis of single mouse cardiac cells at E10.5

We have previously used the C1 platform to profile single cardiac cells from heart chambers at E10.5 stage where the chambers were first manually dissected into five anatomic zones: left ventricle, left septum, right septum, right ventricle and atrial ventricular canal (Fig. 1A) (Li et al., 2016). Viable single cells were lysed, reverse transcribed and pre-amplified for library preparation (Fig. 1A).

As the number of cardiac cells captured using the C1 platform was limited, we also employed a droplet-based system ('10xX') to capture and profile more than 10,000 cells isolated from E10.5 ventricular chambers. Using this approach, each single cell is captured by an aqueous droplet containing reverse transcriptase reagents and a bead with unique barcoding oligos. Lysis and reverse transcription then occurs in these droplets while they are suspended in oil. Subsequently, the drops containing single cells were fused and the barcoded DNA was pooled followed by amplification and library preparation (Fig. 1A). To include the cardiac zone information in the experiment, we employed a well-described cardiac reporter mouse model. Specifically, we crossed the second heart-field reporter strain *Isl1-Cre*, expressing the Cre recombinase under the *Isl1* promoter, with the *Rosa26-mTomato/mGFP* reporter mice to generate two-colored heart cells to partially preserve their anatomical information (Yang et al., 2006). In this model, the first heart field lineage descendants, including LV and AVC cells, will remain tdTomato positive, and second heart field lineage descendants (i.e. *Isl1* positive descendants), including

RV and OFT cells, will be labeled with GFP expression (Cai et al., 2003; Yang et al., 2006).

The libraries from both systems were sequenced, de-multiplexed and mapped to the mouse genome. The C1 platform captured ~500 single cells, each yielding more than two million raw reads, 80% of which uniquely mapped to the mouse genome and covering ~8000 genes. In contrast, the droplet platform captured more than 10,000 cells and reports on the expression of about 2000 genes per cell at two different sequencing depths (Fig. 1B, Fig. S1).

Annotation of cell types in the single cell RNA sequencing data

Modeling our analytic approach on the previously published work of Zheng et al. (2017), we first identified those genes with greatest normalized dispersion – i.e. those genes exhibiting the largest variation between cells normalized by expression level. The resulting 1583 genes were carried forward in our analysis. Correlation between the cells in terms of expression of these genes was then visualized by means of t-Distributed Stochastic Neighbor Embedding (tSNE) (Butler et al., 2018; Maaten and Hinton, 2008; Macosko et al., 2015) based on the principle components of the gene expression (see Materials and Methods for details). The resulting visualization reveals several distinct clusters of cells based on the C1 data (Fig. 2A). Based on expression level of previously identified lineage genes (Li et al., 2016), it is apparent that each cluster represents a specific cell type. Cluster 0, 3 and 4 represent cardiomyocytes (CMs), which highly express CM marker genes such as *Tnni3*; cluster 1 cells specifically express mesenchyme cell (MC) marker genes such as *Fbln2*; cluster 2 represents endothelial cells (EDCs), which express lineage gene *Pecam1*; cluster 5 is defined as epicardial cells (EPs) by high expression of *Wt1*. A few red blood cells (RBCs) were also

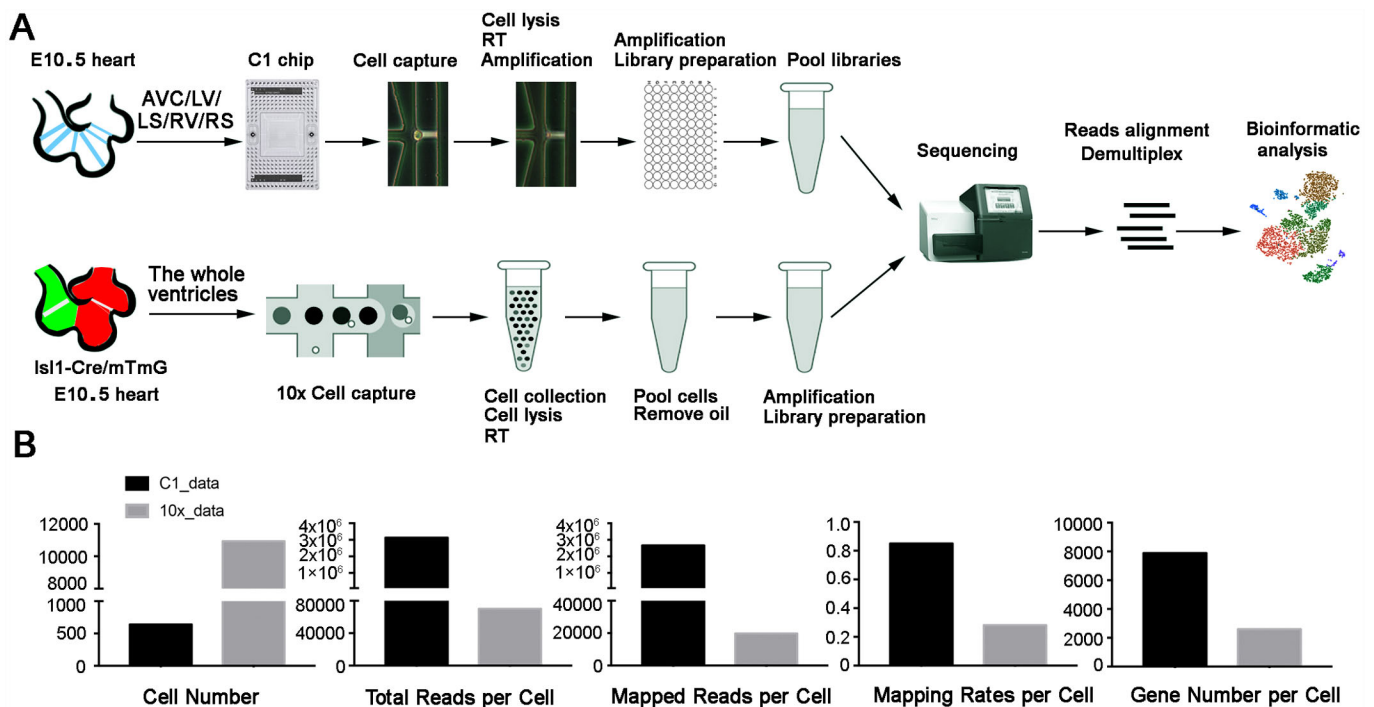


Fig. 1. Experimental design and workflow to capture and profile single embryonic cardiac cells. Single cells from embryonic day (E) 10.5 stage hearts were analyzed using both droplet-based Chromium (10X Genomics) and microfluidic C1 (Fluidigm) platforms. (A) The workflow of both cell capture systems to obtain single cell transcriptional profile. The embryonic heart ventricles were processed separately for the cell capture experiments. (B) Comparison of baseline characteristics of single cell RNA sequencing data from each cell capture platform.

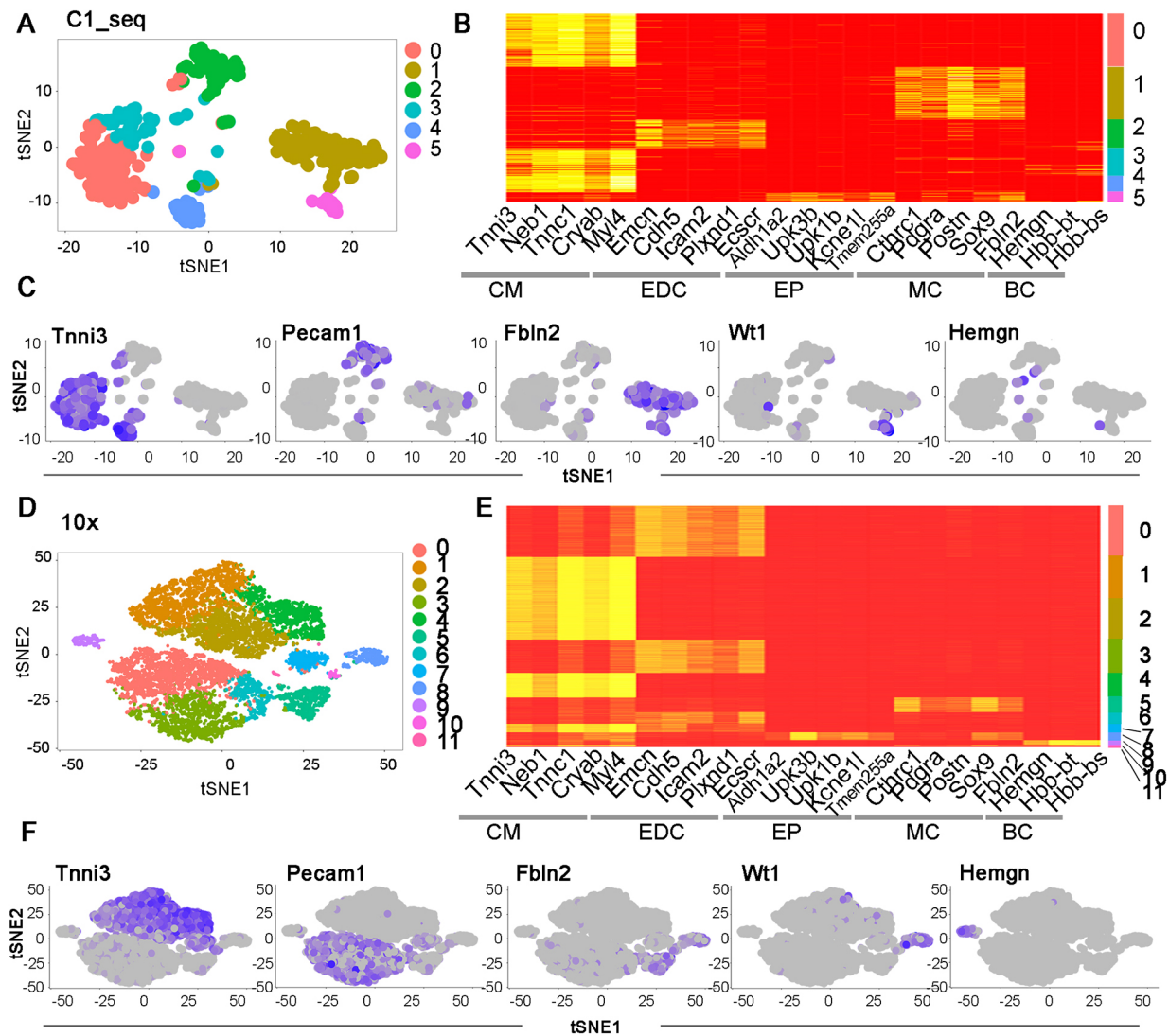


Fig. 2. Cell type analysis of single cardiac cells. (A) Unsupervised analysis of the whole transcriptional expression data revealed the C1 profiled single cells grouped into six cell clusters. (B) Lineage-specific genes were used to assign the cell type for each cell cluster. (C) The expression pattern of representative lineage genes in the C1 profiled cells. (D,E) The droplet-platform profiled cells were grouped into 12 clusters by unsupervised analysis of the whole transcriptional expression data, and the cell type of each cluster was defined by a panel of lineage genes. (F) The expression pattern of representative lineage genes in the profiled cells.

identified with high expression of *Hemgn*; however, they did not form a distinct cluster likely owing to the small number of RBCs within our preparation (Fig. 2B,C).

In contrast, the droplet-based single cell data (~10,000 cells total) revealed twelve tSNE clusters (Fig. 2D). Clusters 1, 2 and 4 represent CMs; clusters 0, 3 and 6 are EDCs; cluster 5 is MCs; cluster 8 is EPs; cluster 9 represents red blood cells; cluster 7 expresses marker genes from cardiomyocytes and endothelial cells; and cluster 10 expresses marker genes from cardiomyocytes and mesenchymal cells [based on their cell percentage and the UMI count of lineage genes expression (data not shown), clusters 7 and 10 likely reflect sequencing reads of two independent cells]; and cluster 11 did not express any of the selected lineage genes (Fig. 2E,F, Fig. S2A,B). Through differential gene expression analysis, we found that cluster 11 highly expressed a platelet marker gene named platelet factor 4 (*Pf4*), and gene ontology analysis of enriched genes within cluster 11 revealed several significant platelet-related pathways, consistent with this cluster representing a platelet and/or megakaryocyte population (Fig. 2D, Fig. S2C).

Finally, through integration analysis of the C1 data and droplet data, we found the two datasets are quite consistent and covered all the cardiac cell types (Fig. S3).

Cell cycle is a major determinant of expression variation in all cardiac cell types

To identify the major sources of variation between cells within each major subtype, we performed dimensional reduction of the expression data for cells from each subtype using principal component analysis. We found that, for each cell type, the first 10 principal components (PCs) captured >90% of variance in gene expression between the cells in each group (Fig. 3A). We extracted the top 50 genes from each PC based on absolute magnitude of their weightings in each PC, and performed gene ontology enrichment on these gene sets to identify possible functional themes represented by these genes. Interestingly, this analysis revealed that, within each cell type, a PC heavily weighting genes involved in the cell cycle was among the top 10 PCs (Fig. 3B). The cell cycle genes include *Mki67*, *Pcna*, *Aurka* and *Aurkb* (Fig. S4). Furthermore, cells with

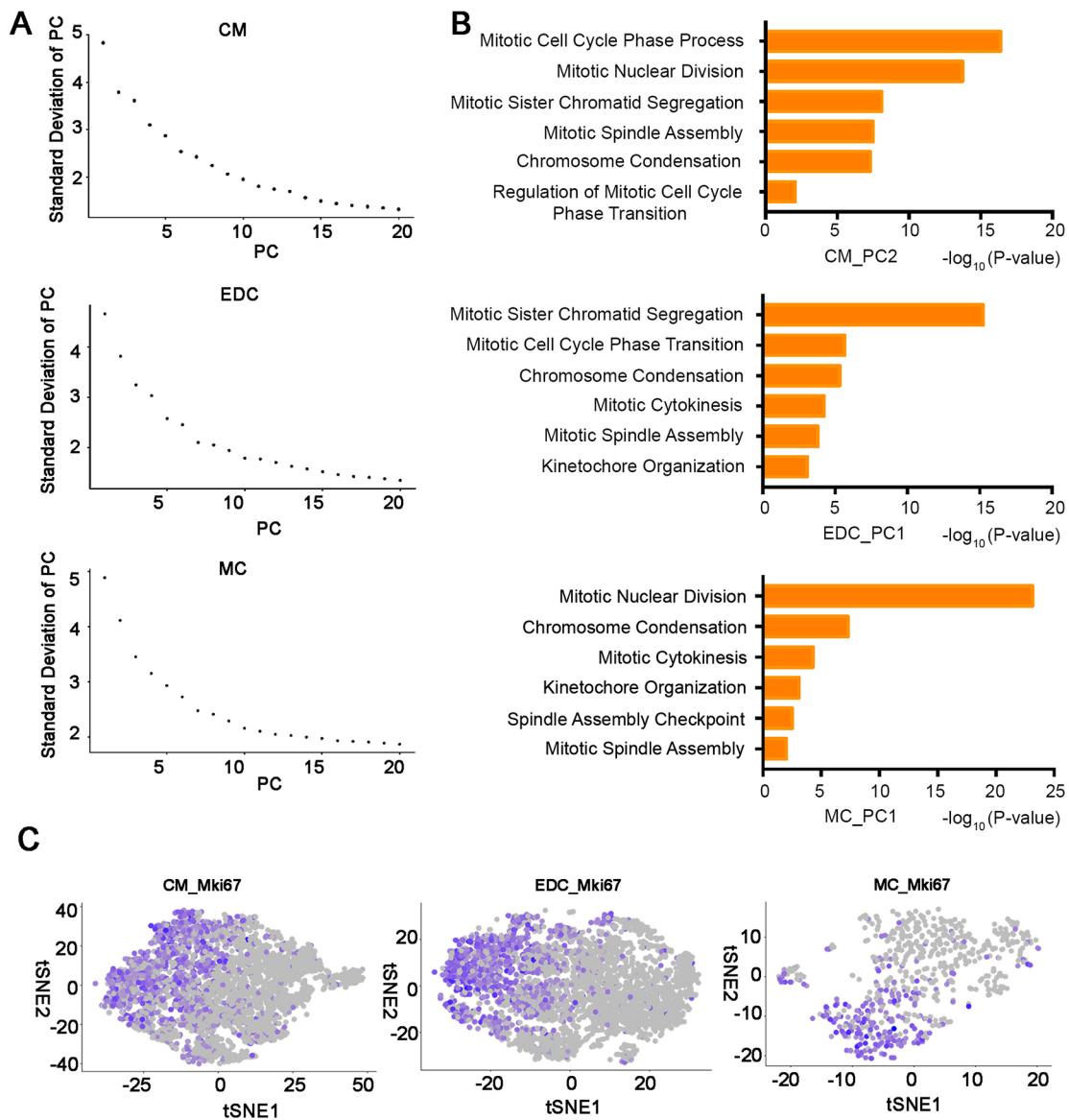


Fig. 3. Variation in the expression of cell cycle genes represents a main principal component in the cardiac cell population. (A) Each cardiac cell type exhibited multiple principle components of expression variation. The first 10 PCs captured most the variations in gene expression in all cell types. (B) The PC2 in CMs, PC1 in EDCs and PC1 in MCs were found to enrich genes in cell cycle-related pathways through gene ontology analysis of the top 50 genes in each PC. (C) Cells in each cell type that express Mki67, a key cell cycle marker, clustered with one another in tSNE plots using the whole transcriptional expression data.

positive expression of known cell cycle genes such as Mki67, clustered together in the same region of respective tSNE plots (Fig. 3C). These analyses suggest that the expression of cell cycle genes represents an important distinguishing factor in the heterogeneity found within each cardiac cell type.

Analysis of the single cell anatomical locations

To understand, ultimately, the contribution of cardiomyocyte proliferation differences within each heart region to myocardial development, we first have to identify the anatomical locations of each single cell. As the C1-profiled cells were harvested and annotated by their heart chamber of origin, we have the precise heart region information for each cell (Fig. 4A, Table S1). Using the list of zone separation genes that have been published previously (Table S1) (Li et al., 2016), we found the atrial ventricular canal (AVC) cardiomyocytes (CMs) on tSNE plot were transcriptionally distinct

from the other types of ventricular CMs. The computationally identified zone separations were further confirmed by the expression pattern of chamber-specific genes. For example, *Rspo3* serves as an AVC marker gene specifically expressed in AVC cells (Xiao et al., 2018), and *Hand1* and *Tnnt1* are expressed within left ventricular cell types, highlighting our ability to distinguish ventricular cell subtypes by gene expression patterns alone (Fig. 4B) (Li et al., 2016; McFadden et al., 2005).

From the droplet-based platform data, we annotated the GFP-positive CMs as right ventricular (RV) CMs based on our known *Isl1-Cre/Rosa26-mTmG* lineage-tracing results (Li et al., 2016). Conversely, the dTomato-positive CMs were distinguished as left ventricular (LV) CMs (Fig. 4C). Consistent with their transcriptional similarity, the RV and LV CMs grouped with one another in tSNE plots, while the *Rspo3*-positive AVC cells again cluster separately from these ventricular CMs types (Fig. 4D).

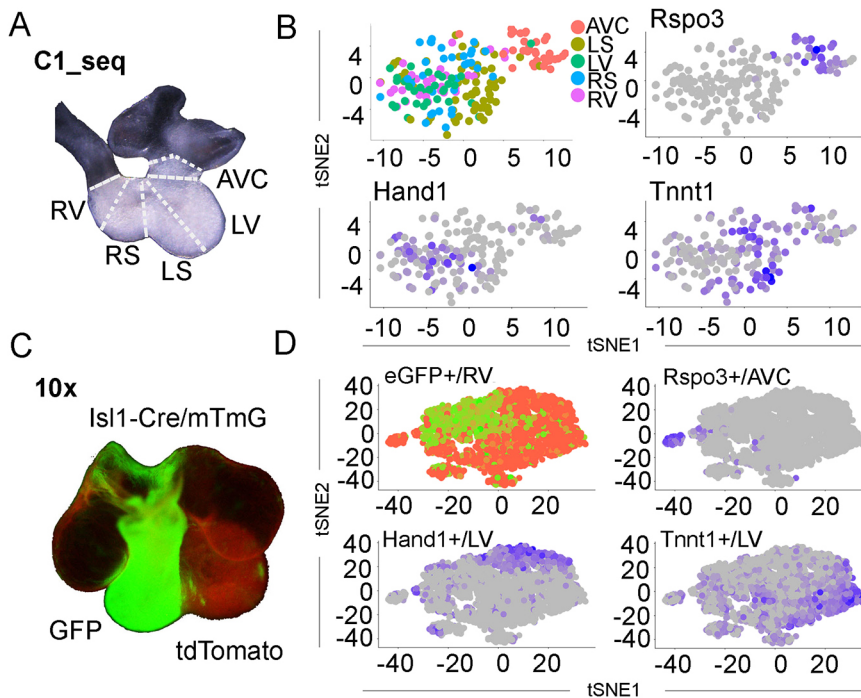


Fig. 4. Anatomical zone analysis of CMs.

(A,C) Diagram of embryonic heart dissection zones for C1 cell capture and lineage reporter boundary for droplet-based cell capture experiments. For the C1 platform, the heart ventricles and atrioventricular canal were dissected into five zones, and each zone tissue was processed separately. For the droplet-based platform, the ventricles and AVC were derived from *Isl1-cre/mTmG* transgenic embryos. (B) CMs that were profiled using the C1 approach can be separated based on their original dissection zones. The expression of *Rspo3* in AVC, *Hand1* in LV and *Tnnt1* in LS are noted. tSNE plots were made by using the panel of zone separation genes published by Li et al. (2016). (D) CMs that were profiled using the droplet platform were also separated into AVC, LV/LS and RV/RS CMs. As lineage descendants of *Isl1-Cre*, the RV/RS CMs express eGFP, whereas LV/LS CMs do not express eGFP but express LV marker *Hand1* or *Tnnt1*. The AVC CMs express *Rspo3*, consistent with data from cells captured by C1 platform. tSNE plots were made by using the panel of zone separation genes published by Li et al. (2016).

To address whether endothelial cells (ECs) and mesenchymal cells (MCs) exhibit chamber-specific expression signatures, we also performed tSNE analysis from both C1- and droplet-based platform data. Interestingly, there were no distinct EC transcriptomes based on either the dissection information from C1 data or GFP status from the droplet-based platform 10X data (Fig. S5A,C). We also found similar results for mesenchymal cells (Fig. S5B,D). Hence, we focused the remainder of our study on the cardiomyocyte populations.

Comparison of CM cell cycle activity at different heart regions

Having identified global gene expression profiles for single cells in the heart and mapped these cells to specific locations within the heart, we next determined the cell cycle state of each cell within the heart chamber. The mitotic cell cycle consists of G1, S and G2/M phases. G1 is the longest phase during which cells make RNA and protein necessary for cell growth. At S phase, cells replicate their genomic DNA, but they do not physically divide into two cells until G2/M phase.

Using panels of cell cycle genes specific to either G2/M phase or S phase (Table S2) (Macosko et al., 2015; Nestorowa et al., 2016; Tirosh et al., 2016), we assigned cardiac cells as either ‘G2/M’ phase or ‘S’ phase using the *CellCycleScoring* function of the *Seurat* package for R. Cells not expressing any of the genes within the panel were assigned as ‘G1’ phase (Fig. 5A). We were able to identify specific cell cycle phases for all CMs within each heart chamber. We then re-annotated each single cell with its cell cycle stage information using tSNE. In each heart zone, we found that cells were clearly separated by their cell cycle phases (Fig. 5B), indicating cell cycle phase dominated the cellular transcriptional profile.

We next calculated the proportion of cells in each cell cycle phase between all chambers and found a much smaller fraction of G2/M phased CMs within AVC when compared with LV and RV, suggesting that the AVC CMs were less proliferative (Fig. 5C). This finding was further confirmed by a lack of accumulation of the M phase marker phosphohistone H3 (pHH3) (Fig. 5D,E). One shortcoming of fluidic-based scRNASeq technology is the presence of technical dropouts – genes that are present but not detected

due to failure of reverse transcription primer capture within the droplet. We used the ALRA imputation algorithm (Linderman et al., 2018 preprint) to impute biological dropouts. Although imputation resulted in a marked increase in detected genes, it did not substantively affect the proportion of cells at each phase of the cell cycle at each anatomical location (Fig. S7).

We next analyzed the genes differentially expressed by cells within each cell cycle phase. In AVC-derived CMs, G2/M phased cells expressed reduced level of sarcomere genes such as *Tnni3* and *Myh9*, and lineage genes such as *Rspo3* when compared with G1 and S phased CMs (Fig. 5F). This relationship was also found in CMs from other zones (Fig. S6A,B). Interestingly, for ECs and MCs we found that their expression of lineage-specific genes was lower in G2/M phase when compared with ECs and MCs in G1 and S phases (Fig. S6C,D). Taken together, our analyses identified an intriguing inverse relationship between the expression of sarcomeric and lineage markers and cell cycle genes during cell proliferation.

We next extended our analysis of CM cell cycle activity to address variations between compact and trabecular myocardium. Specifically, we used a panel of trabecular and compact myocardium-specific genes that we have previously identified (Li et al., 2016) in order to annotate ventricular CMs isolated from single cell transcriptome data obtained using the droplet-based platform (Fig. 6A, Table S3) (Tirosh et al., 2016). While genome-wide tSNE did not show the two types of CMs to be transcriptionally distinct, a more focused PCA using the previously established panel of marker genes helps visualize the gene expression continuum between trabecular and compact myocardium (Fig. 6B, Fig. S8). Furthermore, we found the trabecular CMs showed somewhat reduced proliferative activity (i.e. trabecular CMs exhibited a higher fraction of cells in G1 phase and lower fraction of cells in G2/M and S) (Fig. 6C). To validate this finding, we stained the embryonic sections with phosphohistone H3 and found the trabecular CMs have fewer signal-positive cells than compact CMs (Fig. 6D,E). These results also support the previously reported finding that trabecular CMs are less proliferative than compact CMs (Buikema et al., 2013).

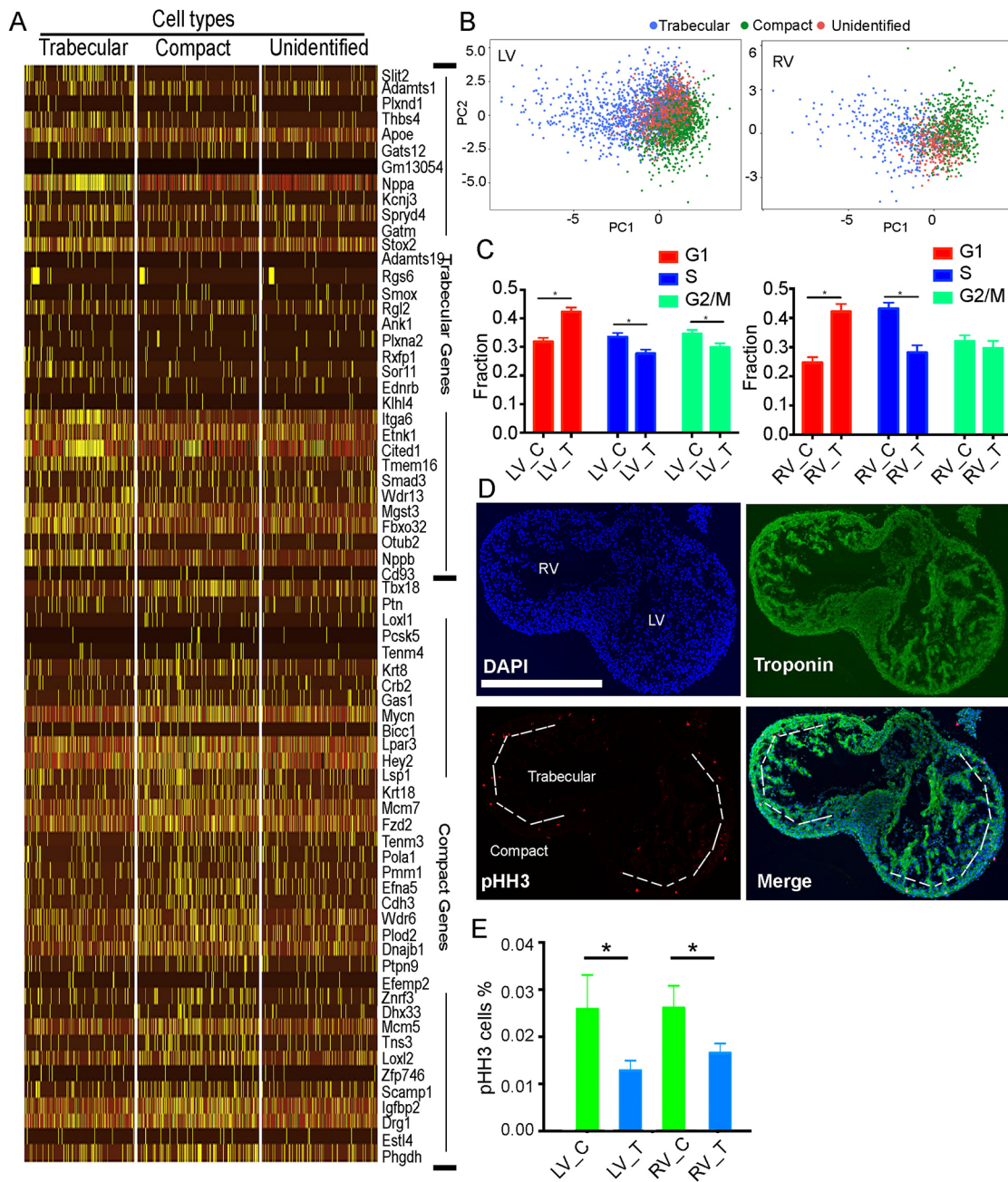


Fig. 6. Cell cycle phase analysis of CMs from compact and trabecular myocardium. (A) The list of genes used to define compact versus trabecular CMs, curated from Li et al. (2016). Cells that did not significantly express these genes were assigned as 'unidentified CMs'. (B) Compact and trabecular CMs were visualized by PCA analysis of the list of compact and trabecular genes. (C) The proportion of compact and trabecular CMs in LV and RV that are in each phase of cell cycle. Two-proportion z-test, $*P < 0.05$. LV_C and LV_T, respectively, represent compact and trabecular CMs from LV; RV_C and RV_T, respectively, represent compact and trabecular CMs from RV. (D) Phosphohistone H3 staining of trabecular and compact myocardium. The dashed lines mark the boundary of trabecular and compact myocardium. Scale bar: 500 μm . (E) Statistical analysis of pHH3-positive CMs in trabecular and compact myocardium. The number of pHH3-positive CMs in compact myocardium is significantly higher than in trabecular myocardium. Student's *t*-test, $*P < 0.05$.

in which the endocardium-secreted *Tgfb1* signal locally regulates trabecular myocardium and restricts its proliferation, and epicardium-derived *Rspo1* locally promotes the proliferation of compact myocardium (Fig. 7F).

DISCUSSION

Cardiac cells actively divide and differentiate during early development, and these processes must be tightly coordinated to ensure proper heart development. In this study, we used single cell RNA sequencing to analyze genome-wide transcriptional profiles of

individual embryonic heart cells to overcome the heterogeneity of different cell types and different cell subpopulations. After segregating each cell by type and anatomical location, we analyzed the cell cycle phase for each using a panel of cell cycle genes and found a profound transcriptional shift in individual cardiac cells based on cell cycle state. We further compared the cell cycle phases of the CMs from different heart zones and showed that the AVC CMs have reduced cell cycle gene expression when compared with LV and RV CMs, consistent with previous reports (Park et al., 2013). Finally, we analyzed the ligand-receptor expression pattern in the different

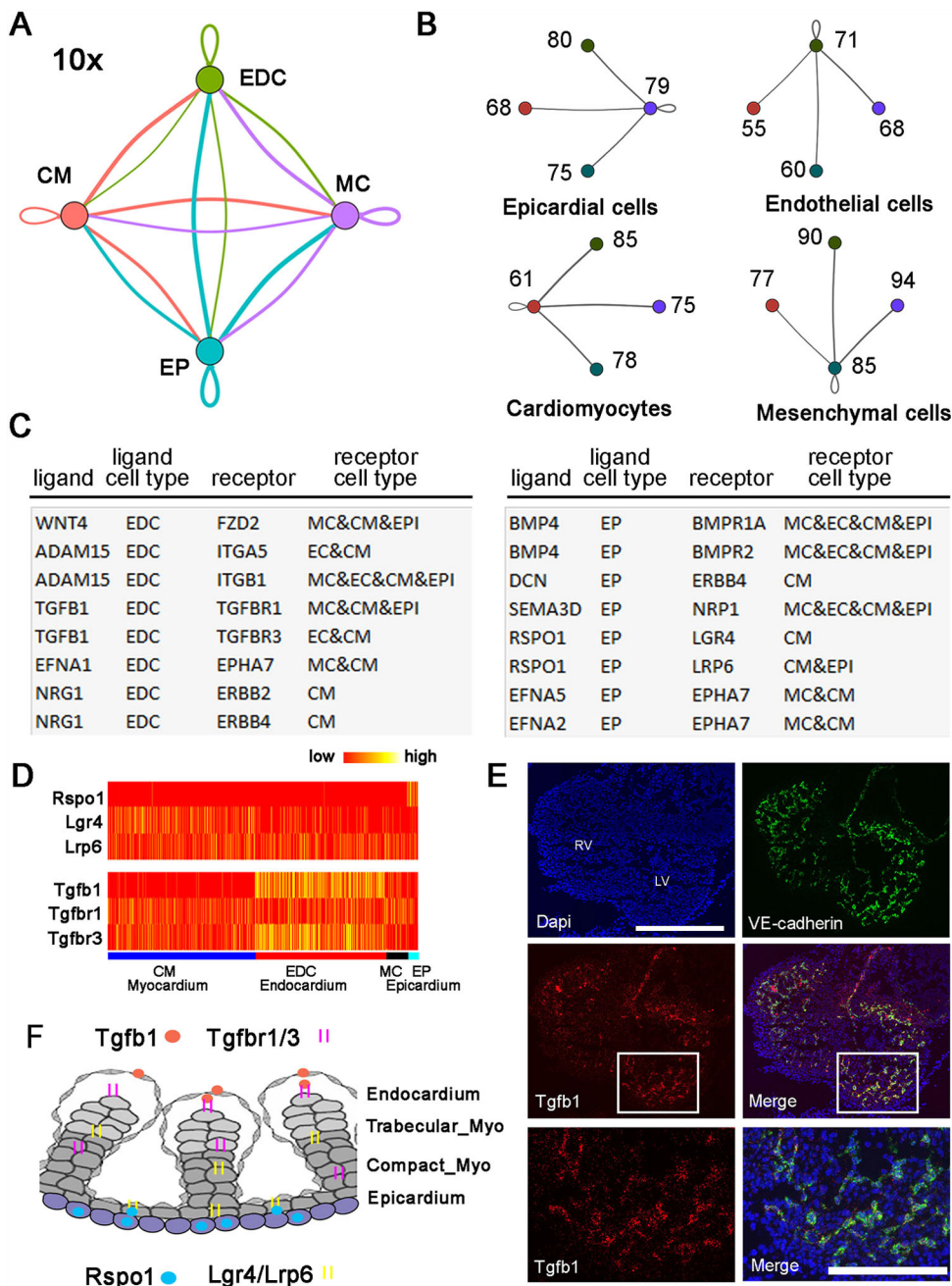


Fig. 7. Analysis of cell-cell interactions between CMs and other cell types.

(A) The overall network of ligand-receptor interactions in embryonic cardiac cells. Loops at each cell type represent autocrine signal; the lines between the cell types represent paracrine signal. Line coloring matching the color of the cell type dot indicates this cell type works as a ligand-secreting cell; line thickness correlates with the number of interactions. (B) The statistical number of ligands secreted from each cell type that interact with receptors from the same or other types of cells. (C) Ligands secreted by endocardial endothelial cells or epicardial cells to interact with the receptors expressing in CMs. (D) Detailed expression patterns of *Tgfb1*, *Rspo1* and their receptors. Each row represents a gene and each column represents a single cell. (E) Expression validation of *Tgfb1* using single molecule *in situ* hybridization. *Tgfb1* was found to be specifically expressed in endocardial endothelial cells, which were marked by the expression of the lineage gene *Cdh5*. Enlarged images of the boxed region in the second row of images are shown in the bottom row. Scale bar: 500 μ m. (F) Model of compact and trabecular myocardium development. Epicardium-derived signal promotes CM proliferation; the endocardium-derived signal repress CM proliferation.

cardiac cell types and found that the expression of *Tgfb1* from the endocardium and *Rspo1* from the epicardium may play a role in establishing a proliferation gradient between compact and trabecular myocardium; however, additional experimentation will clearly be needed to determine the necessity and/or sufficiency of these factors.

Overall, we employed both the C1 and droplet-based cell capture platforms to analyze transcriptional profiles in single cardiac cells to ensure consistency in our findings across different platforms. One advantage of the droplet-based platform was the ability to capture large cell numbers. We have profiled more than 10,000 cells in ventricles using this platform and the total numbers of cells from an E10.5 heart was reported to be \sim 180,000 cells (Meilhac et al., 2003), representing more than 5.5% coverage using this approach. Furthermore, by using a genetic reporter to mark lineage descendants of *Isl1*-expressing cells, we were able to distinguish right ventricular and outflow tract cardiomyocytes from

cardiomyocytes in the left ventricle. This provided us with chamber-specific information without having to perform laborious dissection and separately capture each cardiac chamber for analysis. However, the sequencing depth and the number of genes profiled for each cell with this platform is more limited due to the 3' or 5'-only transcript capture. The C1 platform, on the other hand, generally captures fewer cells but recovers the full transcript, allowing for detection of a greater number of genes per cell and potentially different splicing isoforms. Finally, both platforms have cell size limitations and thus cells that are too large (e.g. adult cardiomyocytes) will not be captured easily by either approach. In our study, we were able to capture embryonic cardiomyocytes for RNA sequencing due to their smaller cell size when compared with adult cardiomyocytes. Overall, our data suggest that both platforms provide sufficient gene expression output for determination of cell type, anatomical location and cell cycle status.

With regards to cardiomyocyte cell cycle, we found that AVC-derived CMs have decreased expression of proliferation markers when compared with other ventricular CMs. This was also the case when comparing trabecular from compact CMs, consistent with previous studies (Park et al., 2013). In addition, our single cell sequencing data provided the genome-wide transcriptional information at each cell cycle phase, allowing for the analysis of cell cycle phase-specific gene expression in each cell type. Using this approach, we found that expression of G2/M phase markers was significantly correlated with downregulation of their lineage-specific genes, regardless of original cell type.

We then identified cell-signaling mechanisms that underlie cell cycle differences between individual cells. Through systematic analysis of ligand-receptor interactions in all cardiac cell types, we identified hundreds of interaction pairs, and we focused on the ligands secreted from epicardium/endocardium to regulate CM proliferation. Along with other known cell signaling pathways, we found the Wnt signaling ligand (in partnership with Rspo1) is expressed specifically in the epicardium and may be involved in compact myocardium proliferation, and Tgfb1 is expressed in the endocardium and possibly influences trabecular CM development. Interestingly, these ligands have contrary effects on cell division (Wnt being pro-mitotic and Tgfb1 being anti-mitotic). We propose that, rather than differential expression of signaling receptors in cardiomyocytes, the epicardium and endocardium secrete distinct signaling molecules that may spatially regulate myocardial development into compact or trabecular myocardium. Hence, the specification into trabecular versus compact myocardium may depend on the relative distance of cardiomyocytes from the epi- or endocardium. Consistent with this, Liu et al. (2010) showed that neuregulin 1 (Nrg1), a ligand known to be expressed in the endocardium, interacts with the ErbB2 receptor that is expressed in cardiomyocytes to promote trabecular cardiomyocyte development (Fig. S12). Our single cell data showed BMP4 specifically expressed in epicardium (Fig. 7C) and, consistently, Klaus et al. (2012) demonstrated that BMP4 acted downstream of Wnt signaling in regulating heart development, and that knockout of its receptor, BMPRIa, in the epicardium and myocardium impaired compact cardiomyocyte proliferation (Stottmann et al., 2004).

In summary, we performed single cell transcriptome profiling of the ventricular chambers of mouse embryonic day 10.5 heart and revealed an intriguing role for cell cycle status to induce a major transcriptional shift in gene expression in single cells. Our data show the chamber-specific differences in CM proliferation as well as the expression of potential ligand-receptor pairs that may be driving these proliferative differences in the developing heart. We believe our findings support a crucial role for the cell cycle in the regulation of gene expression during development and provide a novel approach to understanding transcriptional events that lead to normal heart development and, by extension, may be disrupted in the pathogenesis of congenital heart defects.

MATERIALS AND METHODS

The workflow of droplet-based platform from 10X genomics

Single cells were prepared following the protocol from 10X Genomics. Briefly, E10.5 transgenic mouse *Isl1-cre/mTmG* embryos were isolated and dissected for ventricular chambers and AVC. Dissected tissue was then dissociated into single cells with 0.25% trypsin by incubating at 37°C for 10 min. After that, 10% serum was added to inactivate trypsin and cells were sequentially filtered with a 70 μ m and 40 μ m cell strainer. Cells were collected by centrifuge at 300 *g* for 5 min and washed with 0.04% FBS/HBSS^{-/-}. After suspension, cells were diluted to around 1000 cell/ μ l with 0.04% FBS/HBSS^{-/-} solution.

Prepared cells were captured with 10X Chromium by following the chromium single cell 3' reagent kits v2 user guide. Briefly, single cells were partitioned into nanoliter-scale Gel Bead-In-EMulsions (GEMs) in the chromium controller. After dissolution of the gel beads in GEMs, the primers were released and mRNA were reverse transcribed into barcoded cDNA. After further cleanup and amplification, the cDNA was enzymatically fragmented and 3' end fragments were selected for library preparation. After further end repair, A-tailing, adaptor ligation and PCR amplification, the sample index, UMI sequences, barcode sequences and sequencing primer P5 and P7 on both ends were added to cDNA. The library was sequenced with Illumina NextSeq 500 and HiSeq 4000 platforms.

The workflow of Fluidigm C1 platform

The C1-based single cell data has been described previously (Li et al., 2016). Briefly, single cells were captured with microfluidic chips (Fluidigm). After imaging and quality control of each cell, good quality cells were lysed, reverse transcribed and pre-amplified in microfluidic chambers. Amplified cDNA was transferred to a 96-well plate for quantification, and equal amounts of cDNA from each single cell was used for library preparation. The libraries were sequenced with Illumina HiSeq-2000 platform.

Bioinformatics analysis

The C1 platform data were analyzed as previously described (Li et al., 2016). Raw data were aligned to the mouse genome using STAR, and gene expression counts were calculated with HTSeq. The tSNE plots were drawn using Seurat version 2 (Macosko et al., 2015).

The Droplet platform data were de-multiplexed and mapped to mouse genome MM10 using Cell Ranger from 10x Genomics with default parameters. Cell filter, data normalization and unsupervised analysis were carried out in Seurat version 2 according to their recommended steps (Butler et al., 2018; Macosko et al., 2015). Briefly, the cells were filtered by their gene number and UMI number. The threshold we used for gene number is 500 to 25,000, and UMI number is 1000 to 5 million. Next, we used the LogNormalize function to normalize genes based on library sequencing depth followed by log transformation. Specifically, we calculated the gene expression value by following this formula: $\log[(\text{each gene expression level}/\text{total gene expression value}) \times 10,000]$. Furthermore, we scaled the data on the number of UMIs and percentage of mitochondria gene using the 'vars.to.regress' parameter in the Seurat package. These pre-processed data were then analyzed to identify variable genes, and principal component analysis was also performed. Based on plots of PC versus variance, we retained the top 10 principle components for further analysis. This dimension-reduced dataset was then processed for visualization using the tSNE algorithm. In preparing the tSNE visualization, the following parameters were used: seed.use=10, perplexity=30. Finally, we calculated the differentially expressed genes using Wilcoxon rank sum test under $P < 0.05$. For the cell type-specific analysis, we first identify the single cells of each cell type using sub-clustering analysis within Seurat and then carry out the same analysis as described above.

To score the cell cycle phases of each single cell, we used the CellCycleScoring function in Seurat. This function calculated the cell cycle score based on previously published canonical marker genes (Nestorowa et al., 2016) (Table S2). The single cells highly expressing G2/M- or S-phase markers were scored as G2/M- or S-phase cells, respectively, and the single cells not expressing any of the two categories of genes were scored as G1 phase. Compact and trabecular CMs were also defined using the CellCycleScoring function in Seurat. It defined the CMs using a list of genes we identified previously (Li et al., 2016) (Table S3). Cells not highly expressing any of these genes were defined as unidentified CMs.

To calculate the statistical differences of CM numbers at different zones, a fraction of cells that are in cell cycle x in a sample is calculated using the formula: $p = n_x/n$. The standard error (SE) of a fraction is calculated following the formula: $SE_p = \sqrt{p(1-p)/n}$. Where n_x is the number of cells predicted to be in cell cycle x and n is the total cell number in the sample. We used two-proportions z-test with two-sided hypothesis to detect the significance of the difference between two fractions.

The ligand-receptor network was analyzed as described previously (Paik et al., 2018). Briefly, a gene within each cell type was defined as expressed if its expression is higher than 0 in more than 25% of cells in that cell type. In the network, line thickness correlates with the interaction number, and the line color reflects interaction directionality. The network was drawn in igraph R package.

Immunofluorescence staining

Immunofluorescence staining was carried out by following previous protocol with minor modifications (Buikema et al., 2013). Briefly, E10.5 CD1 mouse embryos were isolated and embedded in OCT. The embryos were cut as cryosections of 10 or 20 μm and fixed with 4% PFA for 10 min. The sections were then treated with blocking buffer (5% goat serum and 0.5% saponin in PBS) for 1 h at room temperature. After that, the sections were incubated with primary antibodies [troponin (1:100, Thermofisher Scientific, MS-295-P), phosphohistone H3 (1:200, Cell Signaling, 9701S), Tgfb1 (1:200, Abcam, ab31013)] overnight at 4°C. On the second day, after washing three times with PBS, the sections were incubated with Alexa Fluor 488 goat anti-mouse secondary antibody (1:500, Invitrogen, A11001) or Alexa Fluor 647 goat anti-rabbit secondary antibody (1:500, Invitrogen, A21245) for 1 h at room temperature. After three more washes in PBS, the sections were mounted with mounting media with DAPI (Vector Laboratories, H-1200).

Single molecular *in situ* hybridization

Proximity ligation *in situ* hybridization (PLISH) was used to stain the transcriptional expression patterns of *Tgfb1* and *Cdh5* at the single-molecule level. We followed the published protocol with minor changes (Nagendran et al., 2018). Briefly, CD1 embryonic hearts were immediately embedded into OCT after isolation and cut into 20 μm cryosections. The sections were fixed with fixation buffer consisting of 3.7% formaldehyde and 0.1% DEPC, and treated with proteinase K for 10 min to retrieve antigen. After that, we stained Tgfb1 with six pairs of H probes and Cdh5 with seven pairs of H probes (Table S6). After circulation ligation, and rolling circle amplifications, we detected the hybridization signal with detection probes conjugated with Cy3 or Cy5 fluorophore.

Other data analysis

All histograms were plotted with Prism 7; all supervised clustering heatmaps were drawn in Rstudio; gene ontology analysis was completed on the gene ontology consortium website (geneontology.org/page/go-enrichment-analysis) and the plot was made in RStudio or Prism7. The immunofluorescence results and PLISH staining results were imaged using the Axioimage microscope at Neuroscience Microscope Service (NMS) facility at Stanford.

Acknowledgements

The authors thank John Collier and Dhananjay Wagh at Stanford Functional Genomics Facility (SFGF) for their help in carrying out the 10X experiment and data alignment. We also thank Megan Mayerle for helping to edit the manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: G.L., S.M.W.; Methodology: G.L., L.T., E.J.K.; Software: G.L., L.T., E.J.K., A.X.; Validation: G.L.; Formal analysis: G.L., L.T.; Investigation: G.L., S.J., S.M.W.; Resources: G.L., S.J., S.M.W.; Data curation: G.L.; Writing - original draft: G.L., S.M.W.; Writing - review & editing: G.L., W.R.G., E.J.K., J.W.B., A.X., J.C.W., S.J., S.M.W.; Visualization: G.L., S.J., S.M.W.; Supervision: G.L., J.C.W., S.J., S.M.W.; Project administration: G.L., S.J., S.M.W.; Funding acquisition: G.L., J.C.W., S.J., S.M.W.

Funding

This work was supported by the National Institutes of Health Office of Director's Pioneer Award (LM012179-03); by an American Heart Association Established Investigator Award (17EIA33410923); by the Department of Pediatrics and Division of Pediatric Cardiology at Lucille Packard Children's Hospital; by the Stanford Cardiovascular Institute; by the Stanford Division of Cardiovascular Medicine, Department of Medicine; by the Institute for Stem Cell Biology and Regenerative

Medicine; by an endowed faculty scholar award from the Stanford Child Health Research Institute/Lucille Packard Foundation for Children (S.M.W.); by the National Institutes of Health (R01 HL145676 and R01 HL141371 to J.C.W.); by the National Institutes of Health K99 award (K99HL133472 to G.L.); and by the Richard and Helen DeVos Foundation (S.J.). Deposited in PMC for release after 12 months.

Data availability

The C1 and 10X single cell RNA sequencing data have been deposited in GEO under accession numbers GSE76118 and GSE122403.

Supplementary information

Supplementary information available online at <http://dev.biologists.org/lookup/doi/10.1242/dev.173476.supplemental>

References

- Black, B. L. (2007). Transcriptional pathways in second heart field development. *Semin. Cell Dev. Biol.* **18**, 67-76. doi:10.1016/j.semcdb.2007.01.001
- Bruneau, B. G. (2002). Transcriptional regulation of vertebrate cardiac morphogenesis. *Circ. Res.* **90**, 509-519. doi:10.1161/01.RES.0000013072.5.1957.B7
- Buckingham, M., Meilhac, S. and Zaffran, S. (2005). Building the mammalian heart from two sources of myocardial cells. *Nat. Rev. Genet.* **6**, 826-835. doi:10.1038/nrg1710
- Buikema, J. W., Mady, A. S., Mittal, N. V., Atmanli, A., Caron, L., Doevendans, P. A., Sluiter, J. P. G. and Domian, I. J. (2013). Wnt/beta-catenin signaling directs the regional expansion of first and second heart field-derived ventricular cardiomyocytes. *Development* **140**, 4165-4176. doi:10.1242/dev.099325
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411-420. doi:10.1038/nbt.4096
- Cai, C.-L., Liang, X., Shi, Y., Chu, P.-H., Pfaff, S. L., Chen, J. and 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. doi:10.1016/S1534-5807(03)00363-0
- Christoffels, V. M., Habets, P. E. M. H., Franco, D., Campione, M., de Jong, F., Lamers, W. H., Bao, Z.-Z., Palmer, S., Biben, C., Harvey, R. P. et al. (2000). Chamber formation and morphogenesis in the developing mammalian heart. *Dev. Biol.* **223**, 266-278. doi:10.1006/dbio.2000.9753
- Devine, W. P., Wythe, J. D., George, M., Koshiba-Takeuchi, K. and Bruneau, B. G. (2014). Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife* **3**, e03848. doi:10.7554/eLife.03848
- Durruthy-Durruthy, R., Gottlieb, A., Hartman, B. H., Waldhaus, J., Laske, R. D., Altman, R. and Heller, S. (2014). Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. *Cell* **157**, 964-978. doi:10.1016/j.cell.2014.03.036
- Evans, S. M., Yelon, D., Conlon, F. L. and Kirby, M. L. (2010). Myocardial lineage development. *Circ. Res.* **107**, 1428-1444. doi:10.1161/CIRCRESAHA.110.227405
- Harvey, R. P. (2002). Patterning the vertebrate heart. *Nat. Rev. Genet.* **3**, 544-556. doi:10.1038/nrg843
- Klaus, A., Muller, M., Schulz, H., Saga, Y., Martin, J. F. and Birchmeier, W. (2012). Wnt/beta-catenin and Brm signals control distinct sets of transcription factors in cardiac progenitor cells. *Proc. Natl. Acad. Sci. USA* **109**, 10921-10926. doi:10.1073/pnas.1121236109
- Kodo, K., Ong, S.-G., Jahanbani, F., Termglinchan, V., Hirono, K., InanlooRahatloo, K., Ebert, A. D., Shukla, P., Abilez, O. J., Churko, J. M. et al. (2016). iPSC-derived cardiomyocytes reveal abnormal TGF-beta signalling in left ventricular non-compaction cardiomyopathy. *Nat. Cell Biol.* **18**, 1031-1042. doi:10.1038/ncb3411
- 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. doi:10.1016/j.devcel.2016.10.014
- Li, G., Dzilić, E., Flores, N., Shieh, A. and Wu, S. M. (2017). Strategies for the acquisition of transcriptional and epigenetic information in single cells. *J. Thorac. Dis.* **9**, S9-S16. doi:10.21037/jtd.2016.08.17
- Linderman, G. C., Zhao, J. and Kluger, Y. (2018). Zero-preserving imputation of scRNA-seq data using low-rank approximation. *bioRxiv*. doi:10.1101/397588
- Liu, J., Bressan, M., Hassel, D., Huiskens, J., Staudt, D., Kikuchi, K., Poss, K. D., Mikawa, T. and Stainier, D. Y. R. (2010). A dual role for ErbB2 signaling in cardiac trabeculation. *Development* **137**, 3867-3875. doi:10.1242/dev.053736
- Maaten, L. J. P. V. D. and Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579-2605.
- Macosko, E. Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M. et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202-1214. doi:10.1016/j.cell.2015.05.002
- McFadden, D. G., Barbosa, A. C., Richardson, J. A., Schneider, M. D., Srivastava, D. and Olson, E. N. (2005). The Hand1 and Hand2 transcription

- factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. *Development* **132**, 189-201. doi:10.1242/dev.01562
- Meilhac, S.M., Kelly, R. G., Rocancourt, D., Eloy-Trinquet, S., Nicolas, J. F. and Buckingham, M. E.** (2003). A retrospective clonal analysis of the myocardium reveals two phases of clonal growth in the developing mouse heart. *Development* **130**, 3877-3889. doi:10.1242/dev.00580
- Nagendran, M., Riordan, D. P., Harbury, P. B. and Desai, T. J.** (2018). Automated cell-type classification in intact tissues by single-cell molecular profiling. *eLife* **7**, e30510. doi:10.7554/eLife.30510
- Nestorowa, S., Hamey, F. K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N. K., Kent, D. G. and Gottgens, B.** (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20-e31. doi:10.1182/blood-2016-05-716480
- Paige, S. L., Plonowska, K., Xu, A. and Wu, S. M.** (2015). Molecular regulation of cardiomyocyte differentiation. *Circ. Res.* **116**, 341-353. doi:10.1161/CIRCRESAHA.116.302752
- Paik, D. T., Tian, L., Lee, J., Sayed, N., Chen, I. Y., Rhee, S., Rhee, J.-W., Kim, Y., Wirka, R. C., Buikema, J. W. et al.** (2018). Large-scale single-cell RNA-seq reveals molecular signatures of heterogeneous populations of human induced pluripotent stem cell-derived endothelial cells. *Circ. Res.* **123**, 443-450. doi:10.1161/CIRCRESAHA.118.312913
- Park, D. S., Tompkins, R. O., Liu, F., Zhang, J., Phoon, C. K. L., Zavadil, J. and Fishman, G. I.** (2013). Pocket proteins critically regulate cell cycle exit of the trabecular myocardium and the ventricular conduction system. *Biol. Open* **2**, 968-978. doi:10.1242/bio.20135785
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. and Regev, A.** (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495-502. doi:10.1038/nbt.3192
- Stottmann, R. W., Choi, M., Mishina, Y., Meyers, E. N. and Klingensmith, J.** (2004). BMP receptor IA is required in mammalian neural crest cells for development of the cardiac outflow tract and ventricular myocardium. *Development* **131**, 2205-2218. doi:10.1242/dev.01086
- Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., II, Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G. et al.** (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196. doi:10.1126/science.aad0501
- von Gise, A. and Pu, W. T.** (2012). Endocardial and epicardial epithelial to mesenchymal transitions in heart development and disease. *Circ. Res.* **110**, 1628-1645. doi:10.1161/CIRCRESAHA.111.259960
- Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., Mburu, F. M., Mantalas, G. L., Sim, S., Clarke, M. F. et al.** (2014). Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11**, 41-46. doi:10.1038/nmeth.2694
- Xiao, Y., Hill, M. C., Zhang, M., Martin, T. J., Morikawa, Y., Wang, S., Moise, A. R., Wythe, J. D. and Martin, J. F.** (2018). Hippo signaling plays an essential role in cell state transitions during cardiac fibroblast development. *Dev. Cell* **45**, 153-169.e156. doi:10.1016/j.devcel.2018.03.019
- Yang, L., Cai, C.-L., Lin, L., Qyang, Y., Chung, C., Monteiro, R. M., Mummery, C. L., Fishman, G. I., Cogen, A. and Evans, S.** (2006). Isl1Cre reveals a common Bmp pathway in heart and limb development. *Development* **133**, 1575-1585. doi:10.1242/dev.02322
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J. et al.** (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049. doi:10.1038/ncomms14049