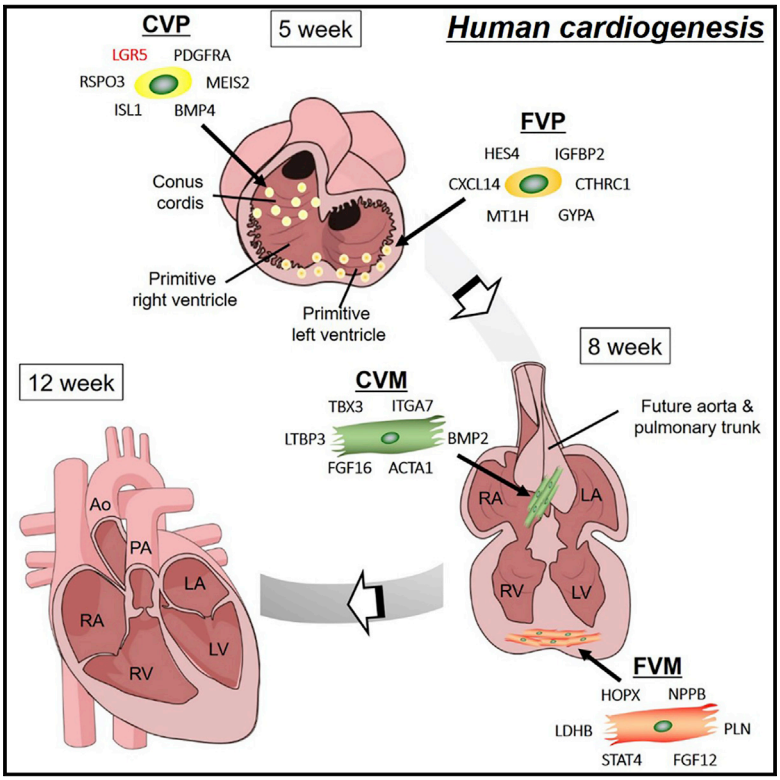
Article

Developmental Cell

Population and Single-Cell Analysis of Human Cardiogenesis Reveals Unique LGR5 Ventricular Progenitors in Embryonic Outflow Tract

Graphical Abstract



Highlights

- Comprehensive gene expression profiles on human cardiogenesis are reported
- LGR5 is identified as a key regulator on human-specific cono-ventriculogenesis
- LGR5 signaling may be associated with certain human congenital heart diseases

Authors

Makoto Sahara, Federica Santoro, Jesper Sohlmér, ..., Kristine Bylund, Peter Gruber, Kenneth R. Chien

Correspondence

makoto.sahara@ki.se (M.S.), kenneth.chien@ki.se (K.R.C.)

In Brief

Sahara et al. performed population and single-cell transcriptional analysis of *in vitro* cardiac differentiation from human embryonic stem cells and human embryonic/fetal hearts. They identified human-specific cardiogenic programs driven by LGR5, which has an important role in cono-ventricular region development in human cardiogenesis.

Sahara et al., 2019, Developmental Cell *48*, 475–490
February 25, 2019
https://doi.org/10.1016/j.devcel.2019.01.005

CellPress

Developmental Cell
Article

Population and Single-Cell Analysis of Human Cardiogenesis Reveals Unique LGR5 Ventricular Progenitors in Embryonic Outflow Tract

Makoto Sahara,[1,2,4,5,*] Federica Santoro,[1,4] Jesper Sohlmér,[1] Chikai Zhou,[1] Nevin Witman,[1,2] Chuen Yan Leung,[2] Mimmi Mononen,[2] Kristine Bylund,[2] Peter Gruber,[3] and Kenneth R. Chien[1,2,5,6,*]
[1]Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm 171 77, Sweden
[2]Department of Medicine-Cardiology, Karolinska Institutet, Stockholm 141 86, Sweden
[3]Department of Surgery, Yale University, New Haven, CT, USA
[4]These authors contributed equally
[5]Senior author
[6]Lead Contact
*Correspondence: makoto.sahara@ki.se (M.S.), kenneth.chien@ki.se (K.R.C.)
https://doi.org/10.1016/j.devcel.2019.01.005

SUMMARY

The morphogenetic process of mammalian cardiac development is complex and highly regulated spatiotemporally by multipotent cardiac stem/progenitor cells (CPCs). Mouse studies have been informative for understanding mammalian cardiogenesis; however, similar insights have been poorly established in humans. Here, we report comprehensive gene expression profiles of human cardiac derivatives from multipotent CPCs to intermediates and mature cardiac cells by population and single-cell RNA-seq using human embryonic stem cell-derived and embryonic/fetal heart-derived cardiac cells microdissected from specific heart compartments. Importantly, we discover a uniquely human subset of cono-ventricular region-specific CPCs, marked by *LGR5*. At 4 to 5 weeks of fetal age, the LGR5$^+$ population appears to emerge specifically in the proximal outflow tract of human embryonic hearts and thereafter promotes cardiac development and alignment through expansion of the ISL1$^+$TNNT2$^+$ intermediates. The current study contributes to a deeper understanding of human cardiogenesis, which may uncover the putative origins of certain human congenital cardiac malformations.

INTRODUCTION

The human heart is composed of highly diverse cell types including cardiomyocytes (CMs), pacemaker and conductive cells, vascular smooth muscle cells (SMCs), and endothelial cells (ECs). All of these cells must be assembled into discrete anatomic and functional structures at early embryonic stages (Vincent and Buckingham, 2010). This assembly is a complex and sequential morphogenetic process that is regulated in a spatiotemporal manner by multipotent cardiac stem/progenitor cells (CPCs). Mouse studies have provided many insights into forming our current understanding of mammalian cardiogenesis, as genetic fate-mapping studies using lineage tracing have revealed that MESP1, ISL1, and NKX2-5 are indispensable transcription factors for mouse cardiogenesis and thereby, represent the markers specific to early CPCs (Bu et al., 2009; Laugwitz et al., 2005; Moretti et al., 2006; Saga et al., 1999; Wu et al., 2006). However, such information and insights in humans are strikingly lacking or poorly established, in part because of the inability to access human materials at the earliest embryonic stages. Considering that there are many evolutionary divergences between mouse and human (Guigo et al., 2003), notably due to the increased complexity and size of the human heart, it is still controversial or undefined as to what the specific markers of the CPCs and their progenies at the various time stages are and what the specific molecular cues driving the CPCs into the committed intermediates and mature cardiac cells are in humans.

During mammalian embryogenesis, the formation of the ventricular heart chamber is one of the earliest and most essential steps for embryo and heart formation and survival (Sahara et al., 2015; Vincent and Buckingham, 2010). The embryonic ventricular muscle cells of the heart have the following two main critical functions: to perform the contractile work required to maintain fetal blood flow by working ventricular muscle cells in the chamber wall and to further insure the proper directional flow and alignment of the proximal outflow tract (i.e., conus) to the distal outflow tract (i.e., truncus of aorta and pulmonary artery) by cono-ventricular heart muscle cells. Although most studies have considered these two types of ventricular muscle to be largely similar (Galdos et al., 2017; Jain et al., 2015), the previous studies using mouse or chick embryos have revealed that the second heart field-derived ISL1$^+$ CPCs migrate to and form distinct cono-ventricular regions in the early embryonic stages (Takahashi et al., 2012), most likely reflecting their key roles in guiding the connection of the working chamber muscle with the outflow tract through a close coordination with a spectrum of non-cardiac muscle cells. However, the precise mechanisms that drive those progenitors into the committed cono-ventricular cells remain undetermined. In addition, the

differences of transcriptional profiles in-between the free wall- and cono-ventricular muscle cells are largely unknown.

Recent technology with next-generation sequencers enables us to conduct deep RNA sequencing (RNA-seq) of protein-coding mRNAs at the single-cell level. This serves as a promising tool to investigate the cellular and molecular basis in cell lineage subtypes that construct various tissues and organs *in vivo* (Picelli et al., 2013; Treutlein et al., 2014). Using this technology, combined with the *in vitro* cardiac differentiation system of human pluripotent stem cells (here we employed embryonic stem cells [ESCs]), together with single cardiac cells derived from human embryonic/fetal hearts in various stages, we provide a comprehensive gene expression resource, characterizing the transcriptional dynamics of human cardiac lineage specification and identifying novel markers of developing cardiac derivatives from multipotent CPCs to intermediates and mature cardiac cells. Importantly, we discover a uniquely human subset of the early CPCs marked by the Wnt signal activator *LGR5*, a leucine-rich repeat-containing G-protein-coupled receptor. At the earliest embryonic stages (4–5 weeks of fetal age), the LGR5$^+$ population appears to specifically emerge in the proximal outflow tract (cono-ventricular region) of human embryonic hearts with co-expression of a well-known CPC marker *ISL1* and thereafter promote cardiac specification and differentiation through expansion of the ISL1$^+$TNNT2$^+$ intermediates. *LGR5* has been known as a stem cell marker in various organs including intestine, colon, kidney, hair, and follicle (Barker et al., 2007; Jansson et al., 2015) but not described in the setting of cardiac development to date. Our study reveals a novel and human-specific cardiogenic program, which is driven by a newly identified sequential transcriptional network connecting a mesodermal precursor marker *MESP1* to *LGR5* and *ISL1* at the early embryonic stages, which likely works to promote cono-ventriculogenesis in humans. Collectively, we chart the developmental landscape of human cardiac formation at the cellular and molecular basis, developing our understanding of human cardiogenesis.

RESULTS

Induction of Human ESC-Derived Cardiac Derivatives *In Vitro*

Using the established *in vitro* cardiac differentiation protocol based on the Wnt signaling modulation (Figure 1A) (Burridge et al., 2014; Lian et al., 2013), we isolated the various lineages of human ESC-derived cardiac derivatives from early multipotent cardiac progenitors (MCPs) to intermediates and differentiated cardiac cells at different time points. We observed that ISL1$^+$ cells, which are well-known CPCs (Bu et al., 2009; Laugwitz et al., 2005; Moretti et al., 2006), first appear on day 3 and peak by day 6 when they occupy 80%–90% of the whole cell population. These cells continue to effectively generate beating CMs that occupy 60%–70% of the whole cell population from day 10 onward (Figures 1B–1D). By applying a multicolor cell separation approach using fluorescence-activated cell sorting (FACS) with antibodies to lineage-specific markers, we could successfully isolate and collect the stage-specific single cardiac lineages, including ISL1$^+$PDGFR-alpha (PDGFRA)$^+$ MCPs on day 3, ISL1$^+$ and lineage marker (Lin)$^+$ intermediates on day 6, and ISL1$^-$Lin$^+$ differentiated cardiac cells, including cardiac troponin T (TNNT2)$^+$ CMs, hyperpolarization-activated cyclic nucleotide-gated potassium channel 4 (HCN4)$^+$ pacemaker cells (PMs), PECAM1 (CD31)$^+$ ECs, and smooth muscle myosin heavy chain (SMMHC)$^+$ SMCs (Figures 1B–1D). We validated the sorted population-specific expression of the known marker genes for each lineage in three biological replicates by quantitative RT-PCR (qPCR) experiments (Figure S1A).

Population RNA-Seq of Human ESC-Derived Cardiac Derivatives Identified Each Lineage-Specific Genes

Transcriptomes of the 13 FACS-isolated populations, including undifferentiated ESCs (day 0); MCPs and non-MCPs (day 3); cardiac intermediates such as cardiomyocyte, pacemaker, smooth muscle, and endothelial intermediates (CI, PMI, SMI, and EI; day 6); and differentiated cells such as CMs, PMs, SMCs, and ECs (day 10) were sequenced as bulk RNA samples and libraries. Given the low amounts of extracted total RNA, because of both the small cell numbers associated with some lineages and the need to perform fixation for sorted cells for successful inner cell protein staining, we employed the "MALIS" protocol (Hrvatin et al., 2014) that had been reported to efficiently extract the RNA of fixed cells after decross-linking. We then applied the Smart-Seq2 approach for cDNA library construction and sequencing (Picelli et al., 2013).

We performed multi-dimensional scaling on all 39 population RNA-seq data (hESC-derived 13 cardiac lineages; three biological replicates) using the most variable expressed genes across all populations, selected by the mean-variance relationship (Figure S1B; see STAR Methods) (Brennecke et al., 2013). These unbiased approaches allowed us to clearly distinguish between 13 cardiac cell lineages (Figure 1E). We then performed differential expression analysis and identified 1,011 genes that were significantly differentially expressed among the lineages (false discovery rate [FDR] $\leq$ 5%). To find each lineage-specific gene, we ranked the differentially expressed genes (DEGs) with mean fragments per kilobase of transcript per million mapped reads (FPKM) $\geq$ 5 in each lineage by descending order of the calculated Z scores of $\log_2$ (FPKM) of the genes (Figure S1C). The 50 top-ranked DEGs in each lineage, including known and less investigated markers, are listed in Table S1, and the representative genes are visualized in the heatmap image (Figure 1F). To investigate the functional roles of the lineage-specific genes, we performed Gene Ontology (GO) gene set enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses on the top 100 lineage-specific genes, which showed that the MCP-specific genes were enriched for terms such as developmental protein, anterior/posterior pattern specification, gastrulation, and Wnt signaling pathway (Figure 1G; Table S2).

LGR5 Is a Novel Cardiac Progenitor Marker

To validate the findings of population RNA-seq data (Figures 1E–1G; Table S1) at a protein level, we performed confirmation assays, such as immunostaining of sectioned human embryonic/fetal hearts and flow cytometry analysis of human ESC-derived cardiac cells (Figure S2). Among the newly identified marker genes, we most notably focused on the Wnt signal activator gene *LGR5* because of its close association with the MCPs (day 3), and sought to elucidate this gene's involvement in early

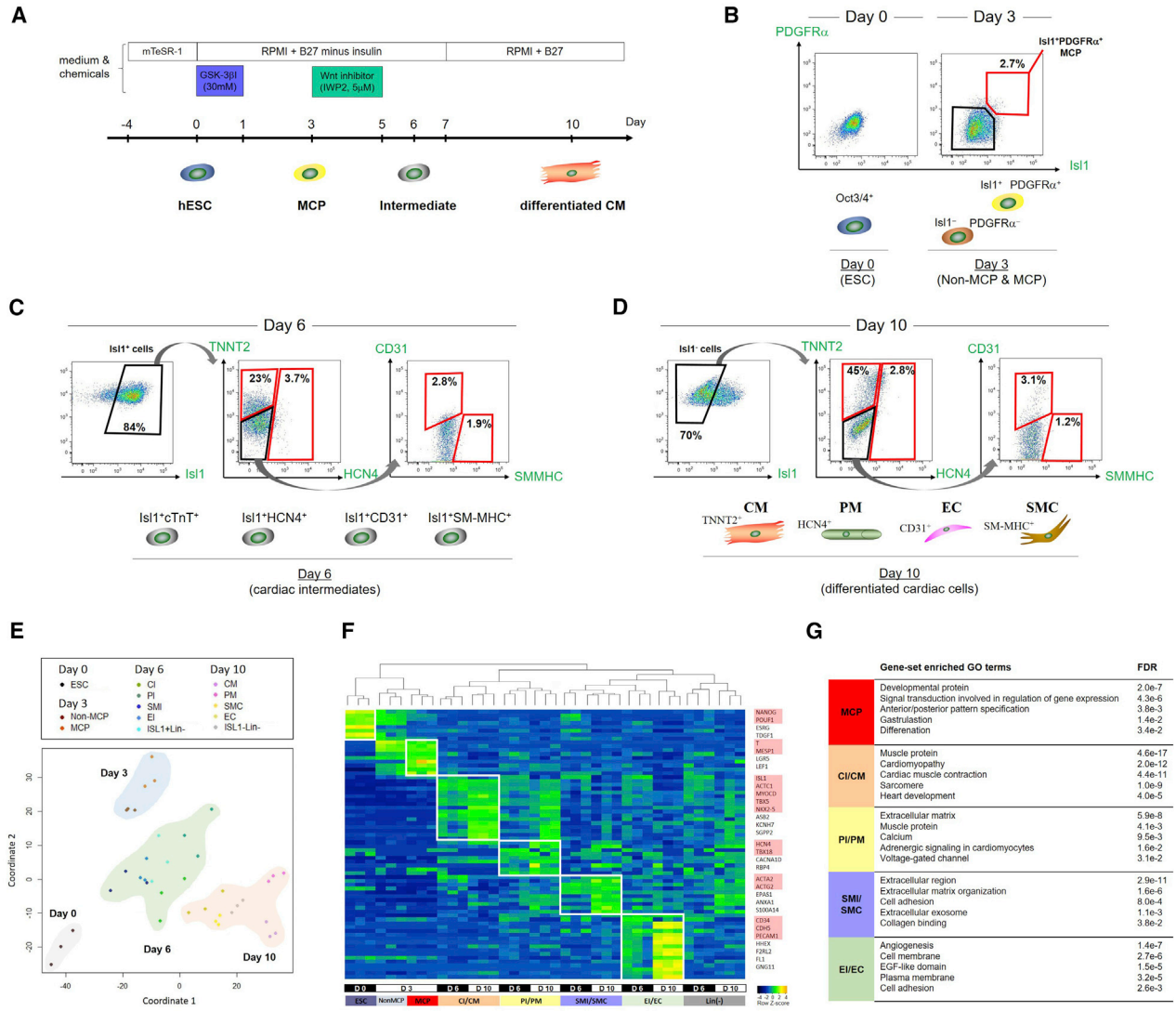


Figure 1. Population RNA-Seq Segregated the Human ESC-Derived Cardiac Lineages and Identified Each Lineage-Specific Gene

(A) Human embryonic stem cell (hESC) cardiac differentiation protocol based on the Wnt signaling modulation. CM, cardiomyocyte; GSK-3βI, glycogen synthase kinase 3 beta inhibitor; MCP, multipotent cardiac progenitor.

(B) On day 3 in the cardiac differentiation, ISL1$^+$PDGFRα^+ MCPs were showing up and sorted by FACS, separately from ISL1$^-$PDGFRα^- non-MCPs.

(C) On day 6, the expression of ISL1 in the differentiated cells reached a peak (80%–90%). Among the ISL1$^+$ population, each of the cardiac lineage marker (TNNT2, HCN4, SMMHC, or CD31)-positive cells, which are considered as intermediates destined to become specific differentiated heart cell types such as CM, pacemaker cell (PM), smooth muscle cell (SMC), and endothelial cell (EC) lineages, were separately sorted by FACS.

(D) On day 10, the ISL1$^-$ population became dominant in the cultured cells, indicating the differentiation and maturation of these cells, and each of the cardiac lineage marker-positive cells among the ISL1$^-$ population were separately sorted by FACS as differentiated CM, PM, SMC, and EC lineages.

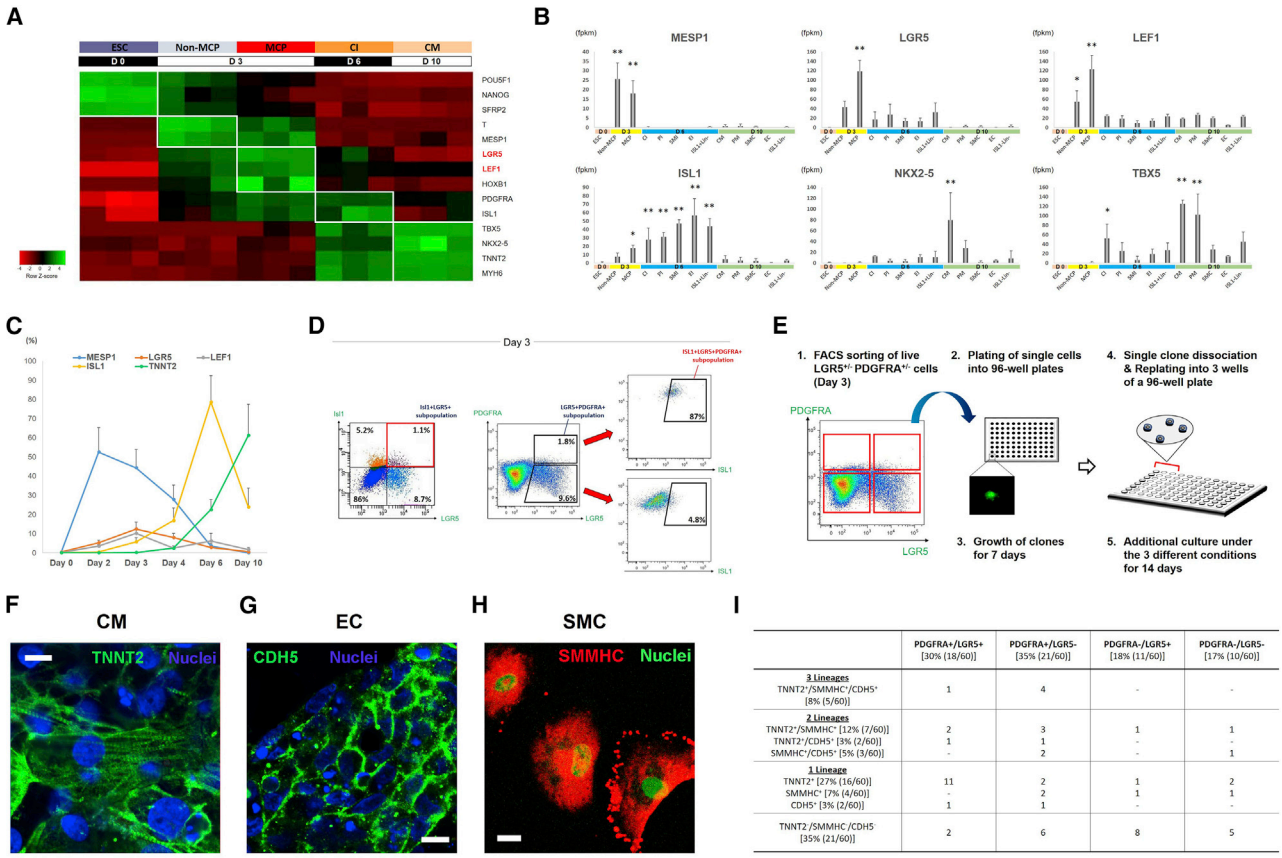
(E) Multi-dimensional scaling at analyzing the 39 bulk RNA dataset (hESC-derived 13 cardiac lineages; three biological replicates) using the most variable expressed genes across all populations. CI, cardiomyocyte intermediate; EI, endothelial intermediate; Lin, lineage; PI, pacemaker intermediate; SMI, smooth muscle intermediate.

(F) Hierarchical clustering and a heatmap image of the 13 cardiac lineages' RNA-seq data. Normalized FPKM levels of each gene (row) are shown after *Z* score normalization. The representative lineage-specific genes are highlighted in white-lined rectangles and indicated on the right column, where the known marker genes are further highlighted by red shades. "D" (e.g., D 0) indicates a "day" (e.g., day 0) in the cardiac differentiation.

(G) Results of Gene Ontology (GO) gene set enrichment analyses using the top 100 lineage-specific genes in each lineage. FDR, false discovery rate.

See also Figures S1, S2, and S4; Tables S1 and S2.

cardiac differentiation and development. First, we compared the time course of expression of *LGR5* and another Wnt signal activator (transcription factor) *LEF1* to that of other cardiogenic transcription factors, such as *MESP1*, *ISL1*, *NKX2-5*, and *TBX5* (Bu et al., 2009; Laugwitz et al., 2005; Saga et al., 1999; Wu et al., 2006) in the *in vitro* cardiac differentiation. This clearly revealed that *LGR5* and *LEF1* were strongly expressed on MCPs (day 3) and weakly on non-MCPs (day 3) or CIs (day 6) (Figures 2A



I

	PDGFRA+/LGR5+ [30% (18/60)]	PDGFRA+/LGR5- [35% (21/60)]	PDGFRA-/LGR5+ [18% (11/60)]	PDGFRA-/LGR5- [17% (10/60)]
3 Lineages				
TNNT2+/SMMHC+/CDH5+ [8% (5/60)]	1	4	-	-
2 Lineages				
TNNT2+/SMMHC+ [12% (7/60)]	2	3	1	1
TNNT2+/CDH5+ [3% (2/60)]	1	1	-	-
SMMHC+/CDH5+ [5% (3/60)]	-	2	-	1
1 Lineage				
TNNT2+ [27% (16/60)]	11	2	1	2
SMMHC+ [7% (4/60)]	-	2	1	1
CDH5+ [3% (2/60)]	1	1	-	-
TNNT2 /SMMHC /CDH5 [35% (21/60)]	2	6	8	5

Figure 2. Discovery of a Novel Cardiac Progenitor Marker, *LGR5*

(A) A heatmap image indicates the time course of FPKM expression levels of the stage-specific mesodermal/cardiogenic developing factors, including ESC, mesodermal precursor, cardiac progenitor, and mature CM markers, which are combined with *LGR5* and *LEF1* (red). All abbreviations are the same as in Figure 1.

(B) Mean FPKM levels of the selected mesodermal/cardiac genes, as well as *LGR5* and *LEF1*, in the human ESC-derived 13 cardiac lineages are shown. *$p < 0.05$ versus others with no asterisks, **$p < 0.01$ versus others with no asterisks.

(C) The time course of protein expression levels of MESP1/LGR5/LEF1/ISL1/TNNT2 in the cardiac differentiating cells, analyzed by flow cytometry (three biological replicates).

(D) The co-expression levels of LGR5, ISL1, and PDGFRA in the cardiac differentiating cells on day 3, analyzed by flow cytometry. The LGR5$^+$PDGFRA$^+$ cells were largely co-expressing ISL1 (rightmost, top), whereas the LGR5$^+$PDGFRA$^-$ cells co-expressed ISL1 to a much lesser degree (rightmost, bottom).

(E) Workflow of a clonal assay using human ESC-derived PDGFRA$^{+/-}$LGR5$^{+/-}$ single cells sorted by FACS on day 3 in the cardiac differentiation. The sorted cells were seeded onto fibronectin-coated 96-well plates at 1 cell per well. Growing clones from single cells were picked after 7 days, and single clone-derived cells were plated into 3 wells of a 96-well plate for differentiation experiments and further cultured under the 3 different culture conditions, customized for the CM, SMC, or EC differentiation, respectively, for additional 14 days (Bu et al., 2009; Moretti et al., 2006).

(F) The LGR5$^+$PDGFRA$^+$ clone-derived CMs that were TNNT2$^+$ with sarcomeric structure in the clonal assay (E). Scale bar, 10 μm.

(G) The LGR5$^+$PDGFRA$^+$ clone-derived ECs that were CDH5 (VE-cadherin)$^+$ in the clonal assay (E). Scale bar, 10 μm.

(H) The LGR5$^+$PDGFRA$^+$ clone-derived SMCs that were SMMHC$^+$ in the clonal assay (E). Scale bar, 10 μm.

(I) Summary of *in vitro* clonal differentiation of mesodermal LGR5 lineages. A total of 60 PDGFRA$^{+/-}$LGR5$^{+/-}$ clones categorized into four groups were analyzed for their differentiation into the three cardiac linages (TNNT2$^+$ CM, SMMHC$^+$ SMC, and CDH5$^+$ EC) identified by immunocytochemistry (F–H). Typical clones of each group are presented with definitive marker expression pattern as well as their total numbers. (B and C) Error bars represent SEM.

and 2B). Non-MCPs are considered as a mixed population, including intermediate populations undergoing a transition between undifferentiated ESCs and MCPs, as non-MCPs were moderately expressing both the ESC- and MCP-specific genes (Figures 1F and 2A). Both MCPs and non-MCPs (day 3) expressed *MESP1* to a higher degree, while other populations did not express *MESP1* at all (Figures 2A and 2B). This is in agreement with the notion that *MESP1* is the marker gene representing the global mesodermal lineage in the early developmental stages (Bondue and Blanpain, 2010). Expression of *ISL1* reached a peak on day 6, whereas expression of *NKX2-5* and *TBX5* reached their peaks on day 10 (Figures 2A and 2B). These findings, related to the sequential expression manners of the cardiogenic transcription factors, are in agreement with the previous reports (Birket et al., 2015). To validate the RNA-seq data, we performed FACS analysis of human ESC-derived cardiac cells under the same protocol (Figure 1A) at different time points (day 0, 2, 3, 4, 6, and 10). Similar to the RNA-seq data, the results showed that expression of MESP1, LGR5/LEF1, and ISL1 reached a peak on day 2, day 3, and day 6, respectively (Figure 2C). On day 3, there were still a number of endoderm (Sox17$^+$) and ectoderm (Nestin$^+$) lineage cells existing

in the culture; thus, the main population of LGR5$^+$ cells at this time point were non-mesodermal lineage (PDGFRA$^-$) (Chong et al., 2013) (Figure 2D; left and middle). For example, a large number of PDGFRA$^-$LGR5$^+$ cells on day 3 were endoderm (Sox17$^+$; 47.4 ± 9.6%) or ectoderm (Nestin$^+$; 33.6 ± 7.7%) lineage cells, and those cells were mainly negative for ISL1 expression (ISL1$^+$; 4.4 ± 1.3%) (Figure 2D; middle and right). In contrast, a major population of the mesodermal (PDGFRA$^+$) LGR5$^+$ cells also expressed ISL1 on day 3 (85.2 ± 11.7%) (Figure 2D; middle and right).

To explore the multipotency in the mesodermal LGR5$^+$ cells and clarify a role of LGR5 on the PDGFRA$^+$ cardiac progenitor population (Chong et al., 2013), we performed a clonal assay using FACS-sorted PDGFRA$^{+/-}$LGR5$^{+/-}$ single-cell-derived clones and quantitatively compared their capabilities for differentiation into the three cardiac lineages under the 3 different culture conditions, customized for the CM, SMC, and EC differentiation (Bu et al., 2009; Moretti et al., 2006) (Figure 2E). The PDGFRA$^+$LGR5$^+$ single-cell-derived clones showed their multipotency to differentiate into the 2 or 3 cardiac linages (4/18 clones; 22%) (Figures 2F–2H); however, those clones differentiated more favorably into CMs (15/18 clones; 83%) and less frequently into ECs or SMCs (3/18 clones; 17%, respectively) (Figure 2I). These tendencies were quite different from ones in the case of PDGFRA$^+$LGR5$^-$ single-cell-derived clones, which showed the multipotency (10/21 clones; 48%) and could also differentiate more frequently into ECs or SMCs (8/21 or 11/21 clones; 38% or 52%, respectively) ($p < 0.01$; PDGFRA$^+$LGR5$^+$ versus PDGFRA$^+$LGR5$^-$ by chi-square test) (Figure 2I). These findings support the multipotency of the *in vitro* mesodermal LGR5$^+$ cells, but at the same time suggest that LGR5 may mark preferably some type of cardiomyogenic subsets among heterogeneous cardiac progenitor populations.

In Vitro Single-Cell RNA-Seq of Human ESC-Derived Cardiac Derivatives

Next, to survey the cellular composition of human ESC-derived cells at different stages, we analyzed 366 single-cell transcriptomes from the cardiac differentiating cells, which were generated with the same protocol (Figure 1A) and harvested randomly by manually picking on days 3, 6, and 15. We performed unbiased clustering of the 366 individual cells with their highly variable genes using the Seurat program (Macosko et al., 2015) and visualized the clustering results by two-dimensional t-distributed stochastic neighbor embedding (tSNE) to reduce the complexity of the data (Figure 3A). This approach organized the cells into 5 molecularly distinct clusters (Figure 3B), and differential gene expression analysis identified each cluster-specific gene (Figure 3C). The selected mesodermal and cardiac genes' expression distribution on the 366 individual cells also clearly segregated the 5 clusters (Figure 3D). For example, *POU5F1* expression was enriched in cluster #1, whereas *EOMES* and *MESP1* were enriched in cluster #2. *PDGFRA*, *LGR5*, and *LEF1* were enriched in clusters #2 and #3, whereas *ISL1* was enriched specifically in cluster #3. *NKX2-5* and *TNNT2* were enriched in cluster #3 mildly and cluster #5 strongly (Figures 3D and 3E). Collectively, clusters #1, #2, #3, and #5 were likely to be the most immature cells, mesodermal precursors/MCPs, cardiac progenitors/intermediates, and mature CMs, respectively. Cluster #4 was considered as non-CMs because fibrotic markers (e.g., *FN1*) were enriched with little or no expression of cardiogenic markers in this cluster (Figures 3D and 3E).

Because we found that LGR5 expression was distributed into clusters #2 and #3, we compared the selected mesodermal and cardiac genes' expression between LGR5$^+$ and LGR5$^-$ cells in those clusters (Figures 3F and 3G). Interestingly, mesodermal precursor/early cardiogenic markers, such as *MESP1*, *PDGFRA* (Bax et al., 2009), *TBX6* (Christiaen et al., 2009), and *BMP2*/4 (Klaus et al., 2007) were enriched in the LGR5$^+$ cells of cluster #2, indicating the close interactions between LGR5 and these genes in this stage and setting (Figure 3F). In cluster #3, the second heart field-related genes, such as *BMP4*, *ISL1* (Bu et al., 2009), and *FGF10* (Watanabe et al., 2012) were enriched in the LGR5$^+$ cells than LGR5$^-$ cells (Figure 3G). On the other hand, expression of the pan-cardiac (*NKX2-5*, *TBX20*, *MEF2C*, and *TNNT2*) or first heart field-related (*GATA4*, *TBX5*, and *HAND1*) genes was comparative between LGR5$^+$ and LGR5$^-$ cells in both clusters #2 and #3 (Figures 3F and 3G), suggesting that LGR5 may be associated with predominantly the second heart field-related genes in human ESC-derived cardiac differentiation.

In Vivo Single-Cell RNA-Seq Identifies Molecular Distinct Subpopulations in Human Embryonic/Fetal Hearts

To develop the notion obtained from the population and single-cell RNA-seq of human ESC-derived cardiac lineages *in vitro*, we next sequenced the transcriptomes of individual cardiac cells isolated from human embryonic/fetal hearts, ranging from 4.5 to 10 weeks of fetal ages (see STAR Methods). In accordance with the embryonic/fetal stages, the obtained hearts were classified into three groups: (1) early stage at 4.5 to 5.5 weeks; (2) middle stage at 6 to 7.5 weeks; and (3) late stage at 8 to 10 weeks (Figure 4A). The hearts were carefully micro-dissected and divided into 3 compartments, such as proximal outflow tract (i.e., the cono-ventricular region), ventricle (i.e., the free wall-ventricular region), and atria, without any valve or fatty and connective tissues, which were further digested into single cells (Sahara et al., 2014) for single-cell RNA-seq. After implementing quality control (Figure S3A) (Petropoulos et al., 2016), 458 high-quality single-cell transcriptomes were retained from 7 human embryonic/fetal hearts (2 early-, 3 middle-, and 2 late-staged hearts). At the initial screening, we found 5 single cells specifically expressing endothelial markers (e.g., *PECAM1*, *CDH5*, and *VWF*), 4 single cells specifically expressing smooth muscle markers (e.g., *MYH11*, *ACTG*, and *SM22α*), and 5 single cells specifically expressing fibroblastic markers (e.g., *VIM*, *FBN1*, and *COL1A1*) without any expression of cardiac sarcomere protein genes (e.g., *MYH6*, *TNNC1*, and *ACTC1*) and other lineage markers. These cells, which were considered as ECs, SMCs, and fibroblasts, respectively, were excluded from further analyses because of low abundance.

We performed principal component analysis (PCA) and diffusion map dimensionality reduction (Haghverdi et al., 2015) on the 7-embryonic/fetal heart-derived 458 single-cell transcriptomes using highly variable expressed genes that were calculated through the average expression and dispersion for each gene. These showed the stage (early, middle, and late)- and compartment (atria, OFT, and ventricle)-dependent segregation

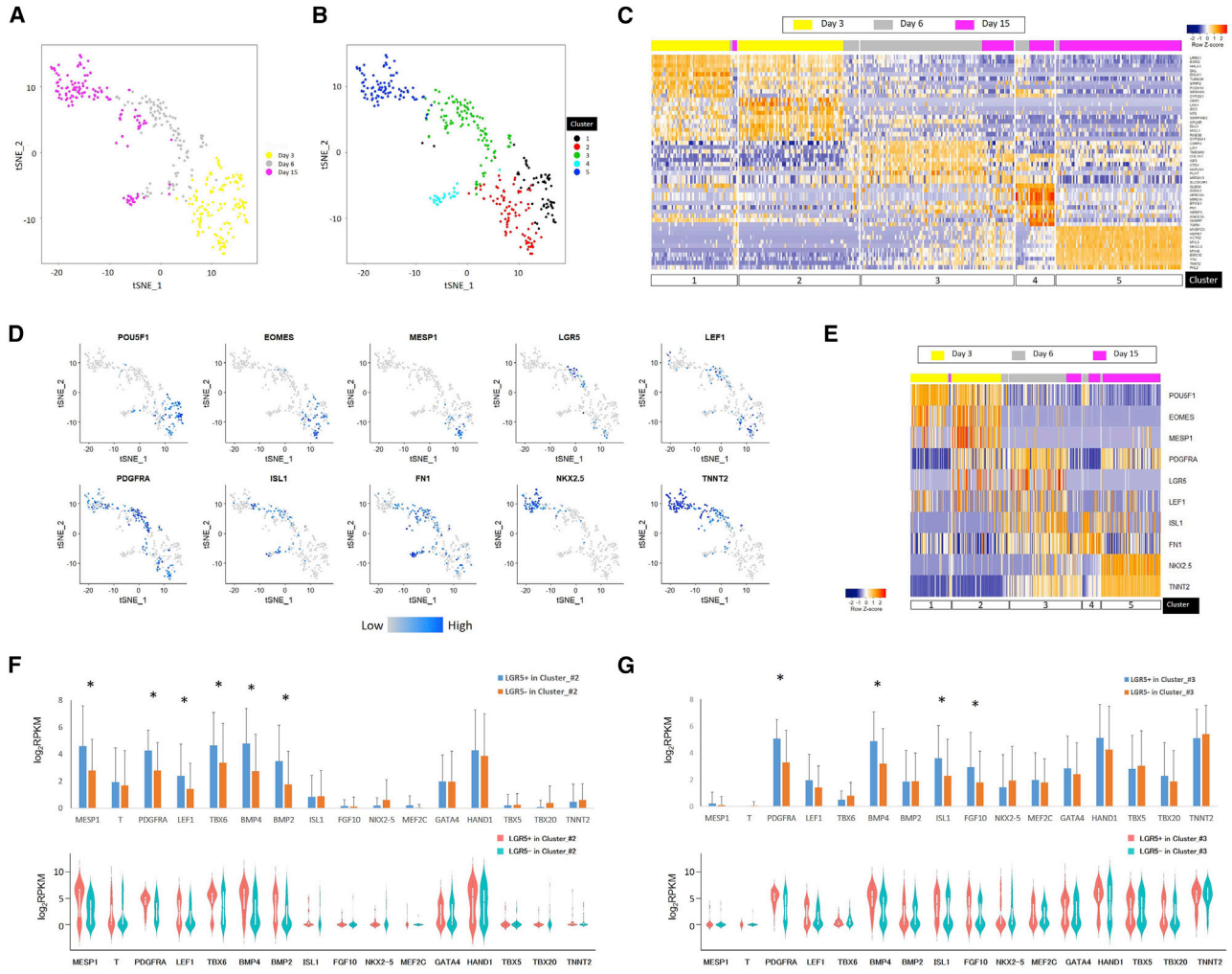


Figure 3. *In Vitro* Single-Cell RNA-Seq Stratifies hESC-Derived Cardiac Differentiating Cells

(A) Two-dimensional tSNE representation of 366 high-quality single-cell transcriptomes using highly variable expressed genes. Each dot is a single cell.

(B) Unbiased clustering of the 366 individual cells using the Seurat program (Macosko et al., 2015) revealed the 5 molecularly distinct clusters that are indicated by different colors.

(C) Differential gene expression analysis by the Seurat and edgeR programs identified differential expression genes on the 5 clusters. A heatmap image of representative differential expression genes' (shown in the right column) expression in the 366 individual cells segregated by the 5 clusters is displayed. Normalized RPKM levels of each gene (row) are shown after Z score normalization.

(D) The selected mesodermal and cardiac genes' expression distribution is displayed on tSNE plot of the 366 individual cells. Each dot is a single cell, and cells are colored based on the gene expression level.

(E) A heatmap image of the 10 selected genes' (D) expression in the 366 individual cells segregated by the 5 clusters is displayed.

(F and G) The selected mesodermal and cardiac genes' expression levels were compared between LGR5^{+} and LGR5^{-} fractions in clusters #2 (F) and #3 (G) and were showed by bar plots (top) and violin plots (bottom). *$p < 0.05$. Error bars represent SD.

of analyzed single cells (Figures 4B and S3B). To cluster the human embryonic/fetal heart-derived single cells, we implemented unbiased clustering using the Seurat program (Macosko et al., 2015) and visualized the results by tSNE, which revealed 10 molecularly distinct clusters from the 458 individual cardiac cells (Figure 4C). Differential gene expression analysis identified each cluster-specific genes (Figures 4D, 4E, and S3C–S3L). Cluster #1 that was composed of early-staged OFT-derived cells was specifically expressing early cardiogenic markers, such as *PDGFRA*, *BMP4*, *ISL1*, and *MEIS2*, as well as *LGR5*, whereas cluster #2 that was composed of early-staged ventricle-derived cells was specifically expressing cell proliferation markers (e.g., *IGFBP2*) and Wnt signaling inhibitors (e.g., *CTHRC1*) (Figures 4D, 4E, S3C, and S3D). Clusters #3 and #4 involved predominantly middle-staged OFT and ventricle cells, in which TGF signaling markers (e.g., *LTBP2*) and cardiogenic markers (e.g., *HAND1*) were enriched, respectively. Clusters #5 and #6 were composed of the atria-derived cells and expressing pacemaker markers (e.g., *SHOX2*) and atria markers (e.g., *GJA5*) (Kapoor et al., 2013; Hashem et al., 2013). Clusters #7 and #8 were expressing both sarcomere protein genes and extracellular matrix (ECM)-related proteins (e.g., *DCN*, *MMP7*, *LAMA3*, and *NEXN*), such

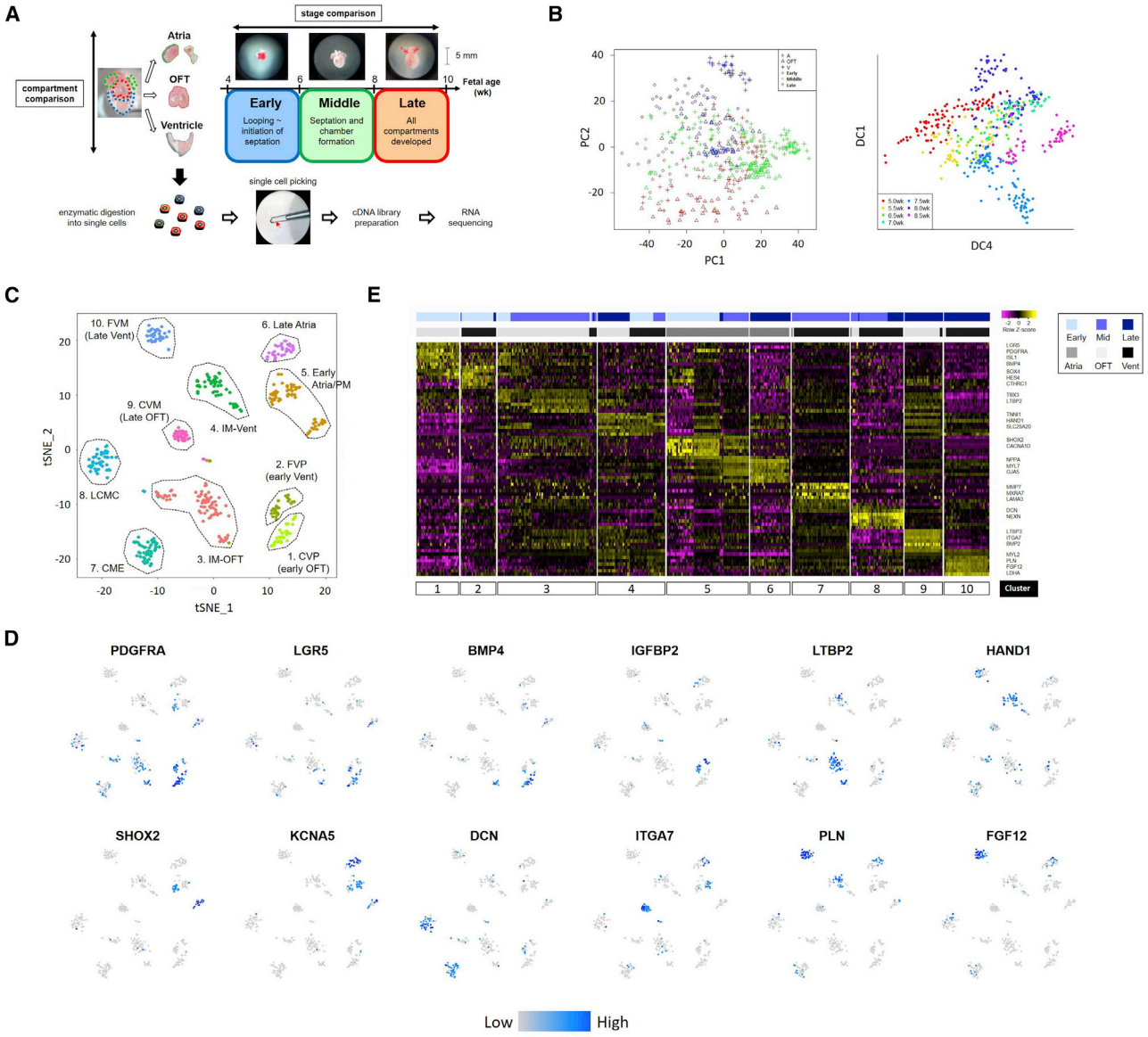


Figure 4. *In Vivo* Single-Cell RNA-Seq of Human Embryonic/Fetal Hearts Identified Molecularly Distinct and Compartment-Specific Subpopulations

(A) Workflow of human embryonic/fetal heart-derived single-cell RNA-seq experiments.

(B) After implementing quality control (Figure S3A), principal component analysis (PCA; left) and diffusion map dimensionality reduction (right) on the 7-embryonic/fetal heart-derived 4,58 high-quality single-cell transcriptomes were performed using highly variable expressed genes. Each dot is a single cell. A, atria; OFT, outflow tract; V, ventricle.

(C) To cluster the human embryonic/fetal heart-derived cells, the Seurat program (Macosko et al., 2015) was implemented, and the results were visualized using tSNE. This approach revealed the 10 molecularly distinct clusters that are indicated by different colors within the 458 individual cardiac cells.

(D) The Seurat and edgeR programs were used to perform differential gene expression analysis among the clusters. Representative marker genes enriched in each cluster are displayed. Cells are colored based on the gene expression level.

(E) A heatmap image of the differential expression genes in the 458 individual cardiac cells segregated by the 10 clusters (C) is displayed. Normalized FPKM levels of each gene (row) are shown after *Z* score normalization. The stage and heart compartment of the individual cells (A) are indicated on the top. Representative genes enriched in each cluster are shown in the right column.

See also Figures S3 and S4.

as myofibroblasts (Mayer and Leinwand, 1997; DeLaughter et al., 2016). Clusters #9 and #10 involved predominantly late-staged OFT and ventricle cells, in which cytoskeleton-related genes (e.g., *DES*, *ACTA1*, and *ITGA7*) (Cui et al., 2013) and matured CM markers (*PLN* and *MYL2*) were enriched, respectively (Figures 4D, 4E, and S3C–S3L). Judging from each cluster-enriched genes as well as the stage and heart compartment of the main population in each cluster, the 10 clusters were

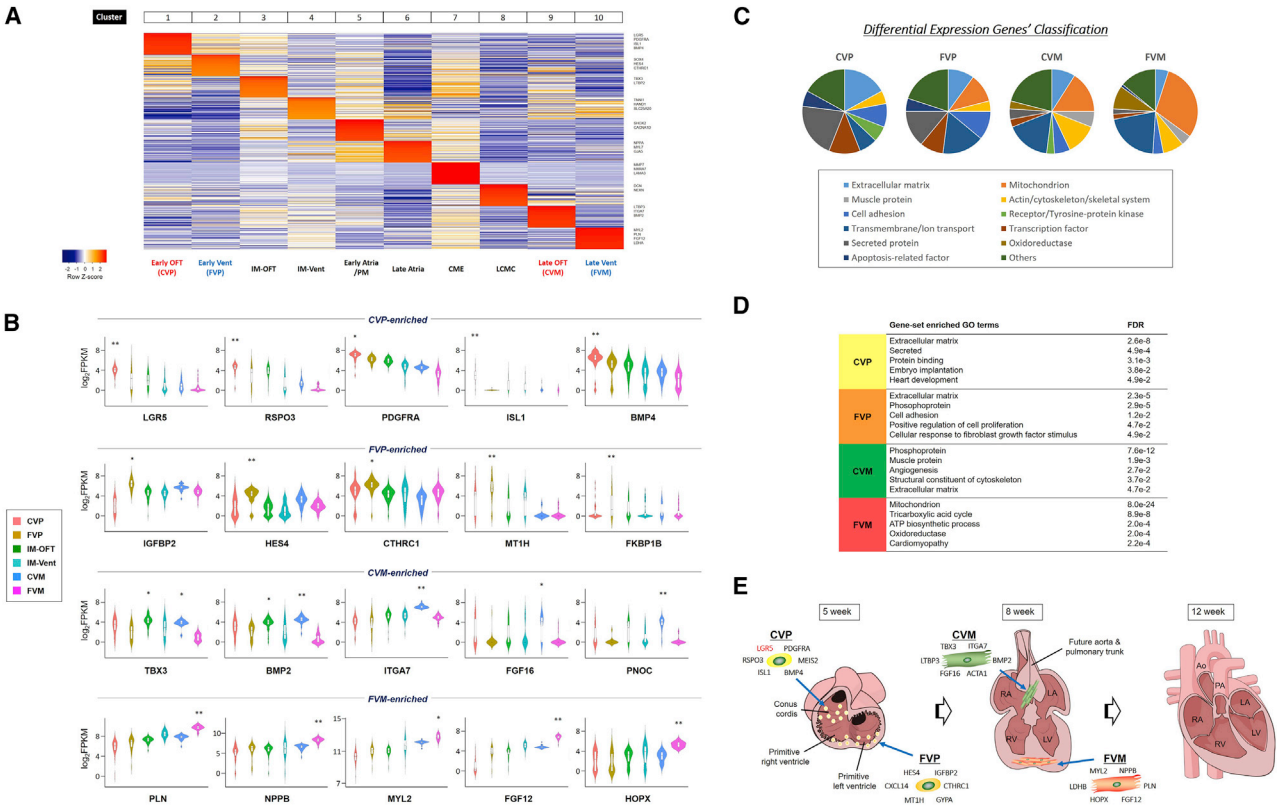


Figure 5. Transcriptional Divergence between Cono-ventricular and Free Wall-Ventricular Progenitor and Muscle Cells

(A) A heatmap image of the 10 clusters (Figure 4C) using 500 (50 × 10) top-ranked differential expression genes, highlighting the molecularly distinct phenotypes in each cluster. Representative genes enriched in each cluster are shown in the right column. All abbreviations are the same as in Figure 4.

(B) The violin plots on the first, second, third, and fourth rows indicate the genes specifically enriched in the CVP, FVP, CVM, and FVM subsets, respectively. *$p < 0.05$ versus others with no asterisks, **$p < 0.01$ versus others with no asterisks.

(C) The gene-type classification of the 100 top-ranked differential expression genes (Table S3) in each of the 4 subsets (CVP, FVP, CVM, and FVM).

(D) Results of Gene Ontology (GO) gene set enrichment analyses using the top 100 cluster-specific genes in the 4 subsets (CVP, FVP, CVM, and FVM).

(E) Schematic image of human developing hearts, highlighting the cono-ventricular and free wall-ventricular progenitor and muscle cells. Representative marker genes of the CVP and FVP (≈5 weeks) and of the CVM and FVM (≈8 weeks) are shown, respectively. Ao, aorta; LA, left atria; LV, left ventricle; PA, pulmonary artery; RA, right atria; RV, right ventricle.

See also Figures S3 and S4, and Tables S2–S4.

named as: (1) cono-ventricular progenitors (CVP; early OFT); (2) free-wall ventricular progenitors (FVP; early ventricle); (3) intermediates of OFT (IM-OFT); (4) intermediates of ventricle (IM-Vent); (5) early atria/PMs; (6) late atria; (7) CMs expressing extracellular matrix genes (CME); (8) late cardiac mesenchymal cells (LCMC); (9) cono-ventricular muscle cells (CVM; late OFT); and (10) free-wall ventricular muscle cells (FVM; late ventricle), respectively (Figure 4C).

The Transcriptional Divergence between Cono-ventricular and Free Wall-Ventricular Progenitor and Muscle Cells

A heatmap image of the 500 top-ranked differential expression genes that were specific to each of the 10 clusters in Figure 4C highlighted the molecularly distinct phenotypes in each cluster (Figure 5A). Unexpectedly, we found that Wnt signaling activators, *LGR5*, *LEF1*, and *RSPO3* (a ligand to *LGR5*) (Han et al., 2011) as well as early cardiogenic markers such as *PDGFRA*, *BMP4*, and *ISL1* were specifically enriched in cluster #1, CVPs (Figures 5A, 5B, and S3C). We then focused on the cono-ventricular and free wall ventricular lineages, including the 4 subpopulations (CVP [cluster #1 in Figure 4C], FVP [#2], CVM [#9], and FVM [#10])-specific genes to establish a molecular atlas of these 4 groups.

The 100 top-ranked DEGs in each of the 4 groups are listed in Table S3, and the gene-type classification of those DEGs are visualized in Figure 5C. In the CVPs, genes related to several developmental and proliferative signaling pathways, such as Wnt (e.g., *LGR5*, *RSPO3*), TGF-β (e.g., *BMP4* and *INHBA*), and epidermal growth factor (EGF)-related (e.g., *ERBB3*) were significantly activated, along with higher expression of the early CPC markers (e.g., *ISL1*, *MEIS2*, *GATA2*, and *GATA5*) and ECM protein genes (e.g., *HAPLN1* and *MMP16*) (Figures 5B and 5C; Table S3). In the FVPs, genes related to cell adhesion (e.g., *CCL2*, *CXCL14*, and *EMILIN1*), growth factors (e.g., *IGFBP2*), TGF-β (e.g., *TGIF1* and *INHBA*) and PI3K-Akt (e.g., *HES4*) signaling, and ECM protein genes (e.g., *HAPLN1* and *COL6A2*) were enriched. Notably, as opposed to the case in the CVPs, Wnt

signaling was inhibited rather than activated in the FVPs (e.g., *CTHRC1* and *SOX9*) (Figures 5B and 5C; Table S3). In the CVMs, genes related to muscle protein (e.g., *TNNT1*), actin/cytoskeleton-related protein (e.g., *DES*, *ACTA1*, and *ITGA7*), angiogenesis (e.g., *ROBO4* and *EPHB3*), and TGF-β signaling (e.g., *LTBP3* and *BMP2*) were enriched, whereas genes related to CM maturation implying sarcomere structure (e.g., *MYL2*, *MYL12A*, and *PLN*), mitochondrial fatty acid β-oxidation (e.g., *COX5A* and *UQCRC2*) (Vander Heiden et al., 2010), and the downstream cardiogenic transcription factors (e.g., *HOPX* and *STAT4*) were significantly enriched in the FVMs (Figures 5B and 5C; Table S3). The GO gene set enrichment and KEGG pathway analyses revealed that the CVP-specific genes were enriched for terms such as extracellular matrix, heart development, and Wnt and TGF-β signaling pathways, while the FVP-specific genes were enriched for terms such as extracellular matrix, cell adhesion, positive regulation of cell proliferation, and PI3K-Akt and TGF-β signaling pathways (Figure 5D; Table S2). In contrast, the CVM-specific genes were enriched for terms such as muscle protein, angiogenesis, and structural constituent of cytoskeleton, while the FVM-specific genes were enriched for terms such as mitochondrion, tricarboxylic acid cycle, oxidoreductase, and cardiomyopathy (Figure 5D; Table S2).

Next, we directly compared the *in vitro* population RNA-seq data of human ESC-derived cardiac lineages and the *in vivo* single-cell RNA-seq data of human embryonic/fetal heart cells (Figure S4). From comparisons of the dynamic range of transcript data, we observed better correlations between CVP and MCP/CI and between FVM and CM (Figure S4A). We then focused on gene expression patterns between the *in vitro* MCP/CI and the *in vivo* CVP. First, the genes highly up-regulated in MCPs compared to CVPs were enriched for GO terms such as "Developmental protein" and "Activator." which contained genes *EOMES*, *T*, *CEBRA*, *ZIC3*, *FOXA3*, *CDX1*, etc. In contrast, the genes highly up-regulated in CVPs compared to MCPs were enriched for GO terms such as "Cardiomyopathy" and "Extracellular matrix," which contained genes *MYL2*, *MYH7*, *MYBPO3*, *MYOZ2*, *LAMA4*, *POSTN*, *IGFBP3*, etc. (Figure S4B, left). This is likely reflecting the difference of the time stages when the two populations emerge on cardiac development/differentiation, respectively (Figures S4B and S4C). The genes highly up-regulated in CIs compared to CVPs were enriched for GO terms such as "Integral component of plasma membrane" and "Neuroactive ligand-receptor interaction," which contained genes *GRM8*, *HTR1B*, *F2RL1*, *GRIN2B*, *CYSLTR2*, *P2RY12*, etc. In contrast, the genes highly up-regulated in CVPs compared to CIs were enriched for GO terms such as "Cellular response to transforming growth factor beta stimulus," which contained genes *PENK*, *POSTN*, *SEMA3D*, *MGP*, *RNASE1*, *DCN*, etc (Figure S4B, right). Among cardiogenic genes, enhanced expression of some of the early cardiogenic/second heart field-related genes (e.g., *PDGFRA*, *BMP2/4*, *MEIS2*, *ISL1*, *FGF10*) as well as Wnt signaling (e.g., *LGR5*, *RSPO3*, and *LEF1*) was detected in both MCPs and CVPs (Figure S4C).

Collectively, these differences in the transcriptional profiles between CVPs and FVPs in the early stage, or between CVMs and FVMs in the late stage, may explain in more detail the differences in the functional roles of the two cono- and free wall-ventricular cell populations (Figure 5E). In addition, our gene sets may serve as a tool to uncover the putative origins of certain human congenital conotruncal heart malformations (Reller et al., 2008). To test the hypothesis, we examined if our gene lists were over-represented in data generated from the Pediatric Cardiac Genomics Consortium (PCGC) (Jin et al., 2017). We used a binomial test to compare top-100 gene lists from each of the 4 groups of cono-ventricular and free wall-ventricular progenitor and muscle cells subsets (Table S3) to the 2,504 PCGC genes that describe partly the human mutational landscape of congenital heart disease (CHD). We found that genes from the CVP ($p = 0.043$) and CVM ($p = 0.001$) groups were both over-represented in the PCGC genes, while those from the free wall, FVP ($p = 0.121$) and FVM ($p = 0.054$), were not. Among the genes overlapped in between the CVP/CVM-enriched genes in our dataset and the 2,504 PCGC genes, we found that one of the CVP-enriched gene *RSPO3* (a ligand to LGR5) was over-represented in the PCGC dataset (Table S4), supporting the notion that LGR5 signaling in progenitor/muscle cells at the proximal outflow tract may be associated with some type of CHD in humans.

LGR5$^+$ Progenitors in Early-Staged Cardiac Outflow Tract Are Human Specific

We then employed immunostaining of sectioned human embryonic/fetal hearts. In line with the *in vivo* single-cell RNA-seq data, this revealed LGR5$^+$ cells appeared in the proximal outflow tract of the early-staged heart (4.5 week of fetal age), co-localizing with another CPC marker ISL1 (Figures 6A and 6B) and another Wnt signaling activator LEF1 (Figures 6C and 6D). Some of those LGR5$^+$ISL1$^+$ cells in the early outflow tract also expressed the CM marker TNNT2, suggesting the developmental transition from the immature CPCs (LGR5$^+$ISL1$^+$TNNT2$^-$) to the committed cardiac intermediates (LGR5$^+$ISL1$^+$TNNT2$^+$) (Figure 6B). To investigate the proliferative status of these cardiac cells, we performed Ki67 staining and observed that the Ki67$^+$ proliferative cells were detected more dominantly within the LGR5$^+$TNNT2$^+$ cells in the OFT region, compared with the LGR5$^-$TNNT2$^+$ cells in the OFT/ventricle regions (Figures 6E and 6F). Another subset of early cardiac progenitors, such as PDGFRA$^+$ISL1$^+$ cells (Chong et al., 2013) were also detected in the early-staged OFT region where LGR5$^+$ cells appeared (Figure 6G). Unlike in the human early-staged embryonic heart, LGR5$^+$ cells were rarely or not detected in the human late-staged hearts (at later than 6 weeks of fetal age) (Figure 6H).

To evaluate whether the LGR5$^+$ CPCs would be common in the mammalian embryonic hearts or human specific, we performed immunostaining of sectioned mouse embryos at embryonic days 9.5–10.5 (E9.5–10.5) during the heart-looping phase. In contrast to the human embryonic heart, little or no LGR5$^+$ cells were found in the murine embryonic hearts, including OFT (Figure S5A), although the PDGFRA$^+$ISL1$^+$ cells were readily detected in the murine embryonic hearts including OFT, similar to the human embryonic hearts (Figure S5B). On the other hand, LGR5$^+$ cells were frequently detected in the murine primitive guts, including the primitive colon and small intestine (EPCAM$^+$) at E9.5–10.5 (Barker et al., 2007) (Figure S5C). To validate the findings of immunostaining using wild-type murine embryos, LGR5$^+$ lineage tracing experiments were performed using *Lgr5-EGFP-IRES-CreERT2* knockin mice (Barker et al., 2007)

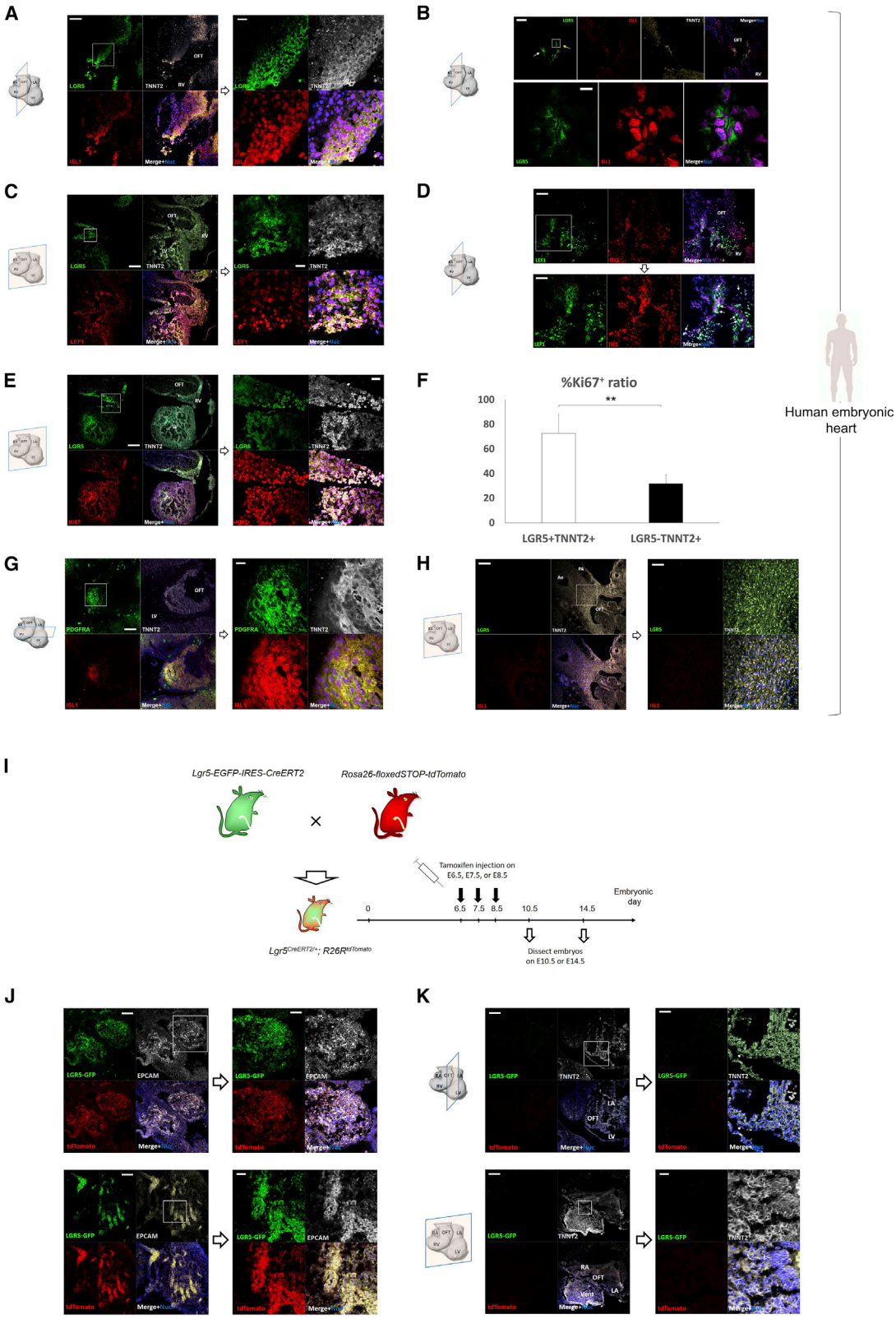


Figure 6. LGR5^{+} Progenitors in Early-Staged Cardiac Outflow Tract Are Human Specific

(A) LGR5^{+} cells were readily detected in proximal outflow tract (OFT) of the human embryonic heart (4.5 weeks) with co-expression with ISL1. Sagittal view. Scale bars, 100 μm (left) and 20 μm (right).

(legend continued on next page)

that were crossed with *Rosa26-floxed stop cassette-tdTomato* (*R26R*^{tdTomato}) reporter mice (Figure 6I). After a single intraperitoneal injection of tamoxifen (75 μg/g) into pregnant females on either embryonic day 6.5 (E6.5; gastrulation), E7.5 (cardiac crescent stage), or E8.5, the *Lgr5*$^{CreERT2/+}$; *R26R*^{tdTomato} heterozygote embryos were dissected on E10.5 or E14.5 for histological investigation. In the E10.5 embryos, tdTomato-labeled cells were readily apparent in the EPCAM$^+$ primitive guts, such as the primitive colon and small intestine (Figure 6J). In contrast, tdTomato-labeled cells did not appear either in the E10.5 or E14.5 hearts including the OFT at all time points of activation by tamoxifen (Figure 6K). Collectively, these findings indicate that unlike LGR5$^+$ cells in the primitive gut, cardiac LGR5$^+$ cells in early-staged cardiac OFT are human-specific and do not appear in mouse embryonic hearts at any stage of cardiogenesis.

LGR5$^+$ Progenitors Enhance Cardiogenesis via Generation of the Committed Cardiac Intermediates

Next, to test an essential role of *LGR5* in human cardiogenesis, we performed loss-of-function experiments, in which we first established human LGR5-knockout (KO) ESC lines (LGR5-KO) by using CRISPR-Cas9 technology (Cong et al., 2013). Using the two single guide-RNA (sgRNA) strategy (Figure 7A), we successfully obtained the LGR5-KO clones with a specific deletion (240 base pair) in the first exon of the *LGR5* gene locus, which were confirmed by PCR (Figure 7B) and sequencing. We then adopted the LGR5-KO clones into the same cardiac differentiation protocol (Figure 1A), in order to examine any cardiogenic phenotype related to the *LGR5* deletion. LGR5$^+$ cells were detected in the human wild-type (WT) ESC-differentiating cells on day 3 in cardiac differentiation but not in the LGR5-KO-differentiating cells (Figures 7C, 7D, and S6A). Interestingly, the degrees of MESP1 expression were similar in both WT and LGR5-KO ESC-derived cells on day 3 (40.6 ± 5.9% versus 37.2 ± 6.4%; p = NS); however, the degrees of LEF1 and ISL1 expression on day 3 were dramatically attenuated in LGR5-KO than in WT ESC-derived cells ([LEF1] 8.1 ± 1.9% versus 1.1 ± 0.3%; [ISL1] 4.6 ± 1.4% versus 0.4 ± 0.1%; p < 0.01, respectively) (Figures 7D and S6A). On day 6, the degrees of LEF1 and ISL1 expression were still suppressed in LGR5-KO ESC-derived cells, and in addition, the induction of TNNT2 (a CM marker) expression was significantly attenuated in LGR5-KO ESC-derived cells (20.6 ± 3.0% versus 2.3 ± 1.6%; p < 0.01) (Figures 7D and S6B). The lowered CM induction rate in LGR5-KO ESC-derived cells was not improved on day 10 or later, and the amounts of generated CMs in LGR5-KO ESC-derived cells were only approximately one-third of those in WT ESC-derived cells (Figures 7D, S6C, and S6D). Among generated ISL1$^+$Lin$^+$ intermediate populations on day 6, the percentage and cell number of only the ISL1$^+$TNNT2$^+$ population were significantly decreased in LGR5-KO than in WT ESC-derived cells (Figure 7E), although other intermediates (ISL1$^+$HCN4$^+$, ISL1$^+$SMMHC$^+$, or ISL1$^+$CD31$^+$) were not affected. In accordance with these findings, we found that the percentage and cell number of CMs (TNNT2$^+$) on day 14 were significantly decreased in LGR5-KO ESC-derived cells, while those of SMCs and ECs were not decreased or even increased a little in LGR5-KO than in WT ESC-derived cells (Figures S6E and S6F). To further demonstrate the requirement of LGR5 in human cardiac differentiation independent of the cardiac differentiation protocol types, we also tested the Activin A/BMP4-based cardiac differentiation protocol adopted from previous studies (Birket et al., 2013). Similarly, the ratio and cell number of TNNT2$^+$ CMs generated by the Activin A/BMP4-based protocol were much attenuated in LGR5-KO ESC-derived cells on day 14 (56.7 ± 11.2% [WT] versus 13.5 ± 2.7% [LGR5-KO], p < 0.01). On the other hand, we found that the ratio and cell number of CD31$^+$VE-cadherin$^+$ ECs generated by the Activin A/BMP4-based protocol were increased in LGR5-KO ESC-derived cells on day 14 (1.2 ± 0.2% [WT] versus 6.5 ± 2.3% [LGR5-KO], p < 0.01). Collectively, irrespective of the CM differentiation protocol types, LGR5 deletion significantly impaired cardiomyogenesis but instead promoted vasculogenesis *in vitro*, at least in part. This is supported by the latest report from another lab (Jha et al., 2017).

To explore the phenotypes of generated beating CMs harvested on day 14, we compared the expression of several CVM- and FVM-enriched genes (Table S3) between WT- and LGR5-KO ESC-derived CMs by qRT-PCR (qPCR). Interestingly,

(B) The LGR5$^+$ISL1$^+$ cells in proximal OFT (4.5 weeks) had already expressed a mature CM marker TNNT2 (a white arrow; top left), but in some cases, they did not (a yellow arrow; top left), suggesting the immature state of the cells. Sagittal view. Scale bars, 100 μm (top) and 20 μm (bottom).

(C) LGR5$^+$ cells in proximal OFT (4.5 weeks) frequently co-expressed another Wnt activator LEF1. Coronal view. Scale bars, 200 μm (left) and 20 μm (right).

(D) LEF1$^+$ cells in proximal OFT (4.5 weeks) also frequently co-expressed ISL1 (arrows). Sagittal view. Scale bars, 100 μm (top) and 50 μm (bottom).

(E) The Ki67$^+$ proliferative cells were detected more dominantly within the LGR5$^+$TNNT2$^+$ cells in proximal OFT (4.5 weeks), compared with the LGR5$^-$TNNT2$^+$ cells in the other regions (OFT or ventricle; 4.5 weeks). Coronal view. Scale bars, 200 μm (left) and 20 μm (right).

(F) Quantitative results in (E) are shown. Ki67$^+$ cells were counted among the LGR5$^+$TNNT2$^+$ or LGR5$^-$TNNT2$^+$ cells on multiple sections of human early-staged embryonic hearts (4.5 weeks) under a confocal microscope. **p < 0.01.

(G) The PDGFRA$^+$ISL1$^+$ cells were frequently detected in the same proximal OFT where LGR5$^+$ cells were found similarly. Transverse view. Scale bars, 100 μm (left) and 20 μm (right).

(H) LGR5$^+$ cells were not detected in the human late-staged fetal heart (at later than 6 weeks of fetal age). Coronal view. Scale bars, 200 μm (left) and 50 μm (right). Ao, aorta; LV, left ventricle; PA, pulmonary artery; RV, right ventricle. The right images in (A, C, E, G, and H) or bottom images (B and D) are the enlarged ones of white squares in the left (A, C, E, G, and H) or top (B and D) images, respectively.

(I) Schema of the mouse LGR5$^+$ lineage tracing experiments. For *Lgr5*$^{CreERT2/+}$; *R26R*^{tdTomato} heterozygote embryos, tamoxifen was given through an intraperitoneal single injection into pregnant females on either embryonic day 6.5 (E6.5; gastrulation), E7.5 (cardiac crescent stage), or E8.5.

(J) The tdTomato-labeled cells indicating the LGR5$^+$ lineage appeared in the EPCAM$^+$ primitive guts, such as the primitive colon (top) and small intestine (bottom), in the E10.5 embryos. Scale bars, 100 μm (left), 50 μm (right, top) and 20 μm (right, bottom).

(K) In contrast, tdTomato-labeled cells did not appear in the E10.5 (top, sagittal view) or E14.5 (bottom, coronal view) hearts, including OFT. LA, left atria; LV, left ventricle; RA, right atria; Vent, ventricle. Scale bars, 200 μm (left), 50 μm (right, top) and 20 μm (right, bottom). The right images in (J) and (K) are the enlarged ones of white squares in the left images, respectively.

See also Figure S5.

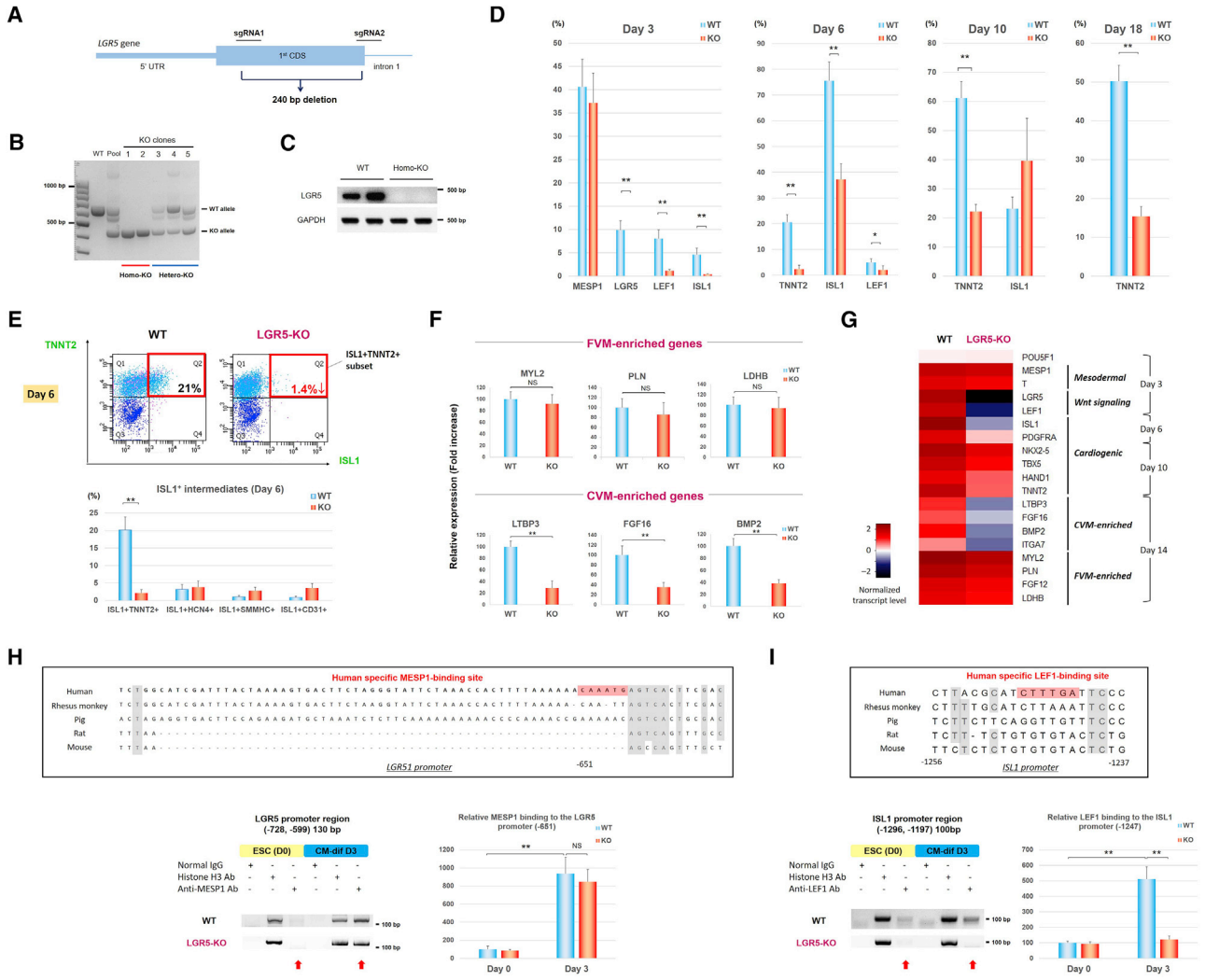


Figure 7. LGR5 Deletion in Human ESCs Abrogates ISL1 Induction and Cono-ventriculogenesis

(A) Schema of the strategy for construction of human LGR5-KO ESCs using the CRISPR/Cas9 system.

(B) Human ESCs (ESO3) were transiently co-transfected with pX330-gRNA1 and pX330-gRNA2. The panel indicates the genotyping of the transduced ESC clones, showing both homologous LGR5-KO clones (#1 and #2) and hetero-knockout clones (#3, #4, and #5).

(C) RT-PCR confirmed the absence of LGR5 mRNA expression of the LGR5-KO ESC line-derived cells on day 3 in cardiac differentiation.

(D) Human LGR5-KO ESCs exhibited their impaired cardiogenic phenotype. LGR5 deletion did not affect MESP1 expression but significantly attenuated LEF1 and ISL1 expression in the cardiac differentiating cells on day 3 and day 6, followed by lower CM ($TNNT2^+$) induction ratios than in WT ESC-derived cells on day 6, 10, and 18. *$p < 0.05$ and **$p < 0.01$ between WT and LGR5-KO ESC-derived cells.

(E) Among the induced $ISL1^+Lineage^+$ intermediate populations on day 6, the percentage and cell number of only the $ISL1^+TNNT2^+$ population were significantly decreased in the LGR5-KO than in WT ESC-derived cells, whereas other intermediate populations ($ISL1^+HCN4^+$, $ISL1^+SMMHC^+$, and $ISL1^+CD31^+$) were not affected.

(F) The comparison of the CVM- and FVM-enriched genes' (Figure 5B; Table S3) expression in between WT and LGR5-KO ESC-derived beating CMs harvested on day 14, analyzed by quantitative RT-PCR (qPCR). **$p < 0.01$ between WT and LGR5-KO ESC-derived cells.

(G) qPCR gene expression heatmap of representative genes related to mesodermal and cardiac differentiation in various time stages between WT and LGR5-KO ESC-derived cells.

(H) A top panel shows a novel and putative MESP1 binding site on the human *LGR5* promoter locus and indicates the sequence divergences of the region among several mammalian species. Chromatin immunoprecipitation (ChIP) assays demonstrated that recruitment of MESP1 onto the above binding site on the *LGR5* promoter was augmented on day 3 in cardiac differentiation (arrows, bottom). This recruitment was not affected by LGR5 deletion. **$p < 0.01$, NS, not significant.

(I) A top panel shows a novel and putative LEF1 binding site on the human *ISL1* promoter locus and indicates the sequence divergences of the region among several mammalian species. ChIP assays demonstrated that recruitment of LEF1 onto the above binding site on the *ISL1* promoter was augmented on day 3 in cardiac differentiation (arrows, bottom); however, this recruitment was significantly prevented by LGR5 deletion. **$p < 0.01$. (D, E, F, H, and I) Error bars represent SEM.

See also Figures S6 and S7.

the expression of the FVM-enriched genes (e.g., *MYL2*, *PLN*, and *LDHB*) was not changed between the two CMs; however, the expression of the CVM-enriched genes (e.g., *LTBP3*, *FGF16*, and *BMP2*) was significantly attenuated in LGR5-KO CMs (Figure 7F). In addition, qPCR gene expression heatmap of representative genes related to mesodermal and cardiac differentiation in various time stages between WT and LGR5-KO ESC-derived cardiac cells showed that expression of *LEF1* (day 3), *ISL1* and *PDGFRA* (day 6), and the CVM-enriched genes (e.g., *LTBP3*, *FGF16*, *BMP2*, and ITGA7) (day 14) was significantly attenuated in LGR5-KO cells (Figure 7G). These findings suggest that LGR5 deletion favorably induces the impaired "cono"-ventriculogenesis *in vitro*, while the "free wall"-ventriculogenesis is not affected by loss of LGR5 function.

Putative Human-Specific Mechanisms to Promote Cardiogenesis: MESP1-LGR5 and LEF1-ISL1 Transcriptional Interactions

To explore the mechanistic aspects of human LGR5-related cardiogenesis, we sought to detect novel binding sites of the earliest mesodermal transcription factor MESP1 on the *LGR5* promoter, or a Wnt signaling transcription factor LEF1 on the *ISL1* promoter through a bioinformatics approach using MatInspector software (Genomatix, Germany; http://www.genomatix.de/). This analysis found a novel MESP1 binding site, which is 651-bp upstream of the transcription start site (TSS) on the *LGR5* promoter, with a consensus motif "-CAVNTG-" (Soibam et al., 2015) and also a novel LEF1 binding site, which is 1,247-bp upstream of the TSS on the *ISL1* promoter, with a consensus motif "-YCTTTGWW-" (Hovanes et al., 2001) (Figures S7A and S7B). We identified these novel binding sites are human specific, and other mammalian species, such as rhesus monkey, pig, and rodents, do not have these sequences (Figures S7A and S7B), indicating evolutionary divergences among these species. Next, chromatin immunoprecipitation (ChIP) assays were conducted using the extracts derived from human ESCs (WT) on day 0 and day 3 in cardiac differentiation (Figure 1A). We found that recruitment of MESP1 onto the novel MESP1 binding site on the *LGR5* promoter (-651) was augmented on day 3 in the cardiac differentiation (Figure 7H). Similarly, recruitment of LEF1 onto the novel LEF1 binding site on the *ISL1* promoter (−1,247) was augmented on day 3 in cardiac differentiation (Figure 7I). These findings reveal that the cross-interactions between MESP1 and LGR5, and between LEF1 and ISL1, are vitally functioning in early cardiac differentiation of human ESCs. The ChIP assays were also conducted using the extracts derived from human LGR5-KO ESCs on day 0 and day 3 in cardiac differentiation, and we found that recruitment of MESP1 onto the novel MESP1 binding site on the *LGR5* promoter (−651) in LGR5-KO was detected similar to that found in WT ESC-derived cells on day 3 (Figure 7H), whereas recruitment of LEF1 onto the novel LEF1 binding site on the *ISL1* promoter (−1,247) was significantly prevented in LGR5-KO ESC-derived cells on day 3 (Figure 7I). Our results demonstrated that LGR5 is essential for the LEF1's stability and function (binding) to the *ISL1* promoter. Taken together, we established the novel sequential machinery mechanisms required for human-specific cardiogenesis, such as the MESP1-LGR5 and LEF1-ISL1 transcriptional interactions, which may be essential particularly for cono-ventriculogenesis in humans (Figure S7C).

DISCUSSION

Mammalian cardiogenesis is regulated spatiotemporally by multipotent CPCs (Galdos et al., 2017; Sahara et al., 2015; Vincent and Buckingham, 2010). Using a powerful technique such as single-cell RNA-seq (Petropoulos et al., 2016; Picelli et al., 2013; Treutlein et al., 2014) to deconstruct tissue heterogeneity, along with the precious materials of human embryonic/fetal hearts rarely obtained, we identified developmental cellular hierarchies and molecular signatures in human cardiogenesis. Our study design was somewhat similar to the previous report (DeLaughter et al., 2016), in which the authors employed mouse embryonic heart-derived single-cell RNA-seq studies to reveal the spatiotemporal dynamics of the transcriptomes in the developing mouse heart. Although the authors in the paper analyzed a great number of single-cell RNA-seq transcriptomes, it seems that these cells were derived from atria or the free wall-ventricular regions, and the team did not analyze outflow tract (cono-ventricular region)-derived cardiac cells. Furthermore, as there exists large evolutionary divergence between mouse and human (Guigo et al., 2003), the research to investigate the single-cell transcriptomes of human embryonic/fetal heart cells divided by various compartments, such as cono- and free wall-ventricular regions, was still desired.

LGR5 encodes a transmembrane receptor that functions as an activator of Wnt signaling by stabilizing the transcriptional coactivator β-catenin after ligand (R-spondins) binding. LGR5 is a stem cell marker for various organs and tissues, including intestine and colon (Barker et al., 2007; Jansson et al., 2015), and is overexpressed in cancer stem cells (Junttila et al., 2015). Collectively, LGR5 functions as a growth-promoting molecule and promotes cellular proliferation of several stem cells in various organs and tissues (Barker and Clevers, 2010). LGR5 appears to play a central role in Wnt signaling, as it has been previously reported that *LGR5* knockdown inhibited cellular proliferation of early differentiating cells and decreased the expression of Wnt signaling-related genes and nuclear-localized active β-catenin (Jha et al., 2017). This was further validated in our study, where LGR5 deletion decreased the expression of another Wnt activator (transcription factor) LEF1 and also blocked the LEF1's binding on the human-specific *ISL1* promoter region. Although LGR5 has not previously been considered as having any relation to or causative roles in mammalian cardiogenesis, we demonstrated that LGR5 is required for CM induction through expansion of the $ISL1^{+}TNNT2^{+}$ intermediates via human-specific transcriptional interactions such as MESP1-LGR5 and LEF1-ISL1 *in vitro*, which is partly supported by the latest report from another lab (Jha et al., 2017). One limitation in the current study is that it has not been fully addressed what the unique molecular identity of $LGR5^{+}$ CPCs is, compared to $LGR5^{-}$ CPCs, and what molecular cue drives $LGR5^{+}$ CPCs to become CMs more favorably (Figure 2I), although the *in vitro* gene expression analysis showed that early cardiogenic and second heart field-related genes were enriched in $LGR5^{+}$ cells than in $LGR5^{-}$ cells (Figures 3F and 3G). To obtain more comprehensive molecular signatures with regard to LGR5-related cardiogenesis, further investigations,

for example, lineage tracing using other species but not mice (e.g., non-human primates) might be required. On the other hand, interestingly, we discovered that the LGR5$^+$ CPCs emerge specifically in the proximal outflow tract (cono-ventricular) region in human embryonic hearts. In addition, the *LGR5* ligand *R-spondin 3* (*RSPO3*) that is required for cardiac development (Cambier et al., 2014) was also enriched in the cono-ventricular progenitor (CVP) population (Figure 5B), suggesting the molecularly distinct phenotype of the CVPs, such as activating a Wnt signaling pathway, compared to the free wall-ventricular cells, in which Wnt signaling was inhibited rather than activated.

Defects in outflow tract development are an important cause of CHD, which is the most common malformation in children affecting 1/100 live births (Reller et al., 2008). Mis-alignment of the underlying cono-ventricular region with the overlying great vessels can result in several serious forms of CHD in children, including tetralogy of Fallot, transposition of the great arteries, and other high morbidity malformations (Reller et al., 2008). While several genes that are rare causes of CHD have been identified (Jin et al., 2017), the precise mechanisms that link these genetic defects with the alterations in mal-alignment of the great vessels and the underlying cono-ventricular region are largely unknown. The lack of a definitive marker of cono-ventricular cells has hampered a detailed analysis of cono-ventricular progenitor/myocyte commitment in the setting of the comparison between cono-ventricular versus free wall-ventricular muscle cells. In this regard, our study provided such cono-ventricular cell-specific markers (e.g., *LGR5, RSPO3*) and molecular cues that might be essential for cono-ventricular muscle cell commitment and alignment with appropriate connections with the great arteries. Thus, our dataset may help to uncover novel origins of certain types of human CHD. Although further studies will be needed in the future, we have already found the close overlapping between our gene lists enriched in the cono-ventricular cells (CVPs/CVMs; Table S3) and other data sources, such as the PCGC-identified genes that describe the human mutational landscape of CHD (Jin et al., 2017). Notably, we found the CVP-enriched gene *RSPO3* was over-represented in the 2,504 PCGC genes (Table S4), suggesting that LGR5 signaling in progenitor and muscle cells at the proximal outflow tract may be closely associated with some type of CHD in humans. Interestingly, a recent study that analyzed single cell transcriptomics to investigate human pluripotent stem cell differentiation into CMs (Friedman et al., 2018) has reported that the similar Wnt modulation protocol for cardiac differentiation generated mainly two cell populations in later stages (days 15 and 30): *BMP4*$^+$/*RSPO3*$^+$ non-contractile cell types and *MYL2*$^+$/*HOPX*$^+$ CMs, which was somewhat similar to both our comparison between cono-ventricular and free wall-ventricular regions in the current study.

In summary, we provide a comprehensive gene expression resource, characterizing the transcriptional dynamics of human cardiac lineage specification and identifying novel markers of developing cardiac derivatives from multipotent CPCs to intermediates and mature cardiac cells. Importantly, we have identified the human unique subset of the early-staged proximal outflow tract (cono-ventricular region)-specific CPCs, which are marked by *LGR5* and promote cardiac differentiation, specification, development, and alignment through expansion of the ISL1$^+$TNNT2$^+$ intermediates. Our study revealed the novel and human-specific cardiogenic programs at the early embryonic stage, driven by the sequential transcriptional interactions of MESP1 to LGR5, and LEF1 to ISL1, which might be essential particularly for cono-ventriculogenesis in humans. The current study contributes to a deeper understanding of human cardiogenesis, uncovering the putative origins of certain human congenital cardiac malformations and potentially advancing cardiac regenerative medicine.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Generation and Maintenance of Human ESC Lines
  - Human Embryonic/Fetal Heart Tissues
  - Wild-Type and Genetically Modified Mouse Lines
- METHOD DETAILS
  - Cardiac-Directed Differentiation
  - Isolation of Human ESC-Derived Cardiac Derivatives with FACS
  - RNA Extraction of FACS-Sorted Cells
  - Isolation of Human Embryonic/Fetal Heart-Derived Single Cardiac Cells
  - Isolation of Human ESC-Derived Single Cardiac Cells
  - Construction of cDNA Libraries with SMART-Seq2
  - RNA Sequencing and Data Pre-processing with Bioinformatics Analyses
  - Gene Variability, Lineage Segregation of Cells, and Lineage Differential Expression Analysis
  - Immunostaining
  - Collection of Mouse Embryos
  - Mouse LGR5 Lineage Tracing
  - Tamoxifen
  - Quantitative RT-PCR
  - Clonal Assays
  - Construction of CRISPR/Cas9-Mediated Human LGR5-KO ESCs
  - Chromatin Immunoprecipitation (ChIP)
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at https://doi.org/10.1016/j.devcel.2019.01.005.

ACKNOWLEDGMENTS

We would like to thank Christer Betsholtz for his advice and careful reading of the manuscript; Mattias Karlén for help with the artwork; Shintaro Katayama and Xidan Li for help and advice for the bioinformatics analyses; Javier Avila-Cariño for help with the FACS experiments; Xiaobing He for the routine laboratory works; Katarina Drakenberg and Raffaella Giugliano for their administrative work for the laboratory; Karl-Henrik Grinnemo and Ulrika Felldin for helping us obtaining human resources; and all the members of Kenneth Chien's lab for giving us insightful discussion on a regular basis.

Work in our laboratory is supported by the Knut and Alice Wallenberg Foundation, Vetenskapsrådet (Swedish Research Council), AstraZeneca Pharmaceuticals, and Karolinska Institutet. M.S. is supported by research grants provided by Swedish Heart and Lung Foundation and Karolinska Institutet. F.S. was supported by an EMBO long-term fellowship (ALTF 620-2014). This work was supported by a grant to K.R.C. from the European Research Council Advanced Research Grant Award (AdG743225).

AUTHOR CONTRIBUTIONS

M.S. and K.R.C. conceived the study. M.S. and F.S. performed all human ESC and embryonic/fetal heart experiments. J.S. assisted in cDNA libraries construction and immunostaining. C.Z. assisted in mouse embryo experiments. N.W. assisted in human ESC culture experiments. C.Y.L. and M.M. contributed to *in vitro* single-cell RNA-seq experiments. K.B. assisted in histology works. P.G. assisted in computational experiments. M.S. and K.R.C. performed computational experiments, interpreted data, and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 7, 2018
Revised: December 18, 2018
Accepted: December 31, 2018
Published: January 31, 2019

REFERENCES

Barker, N., and Clevers, H. (2010). Leucine-rich repeat-containing G-protein-coupled receptors as markers of adult stem cells. Gastroenterology *138*, 1681–1696.

Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., et al. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003–1007.

Bax, N.A., Lie-Venema, H., Vicente-Steijn, R., Bleyl, S.B., Van Den Akker, N.M., Maas, S., Poelmann, R.E., and Gittenberger-de Groot, A.C. (2009). Platelet-derived growth factor is involved in the differentiation of second heart field-derived cardiac structures in chicken embryos. Dev. Dyn. *238*, 2658–2669.

Birket, M.J., Casini, S., Kosmidis, G., Elliott, D.A., Gerencser, A.A., Baartscheer, A., Schumacher, C., Mastroberardino, P.G., Elefanty, A.G., Stanley, E.G., et al. (2013). PGC-1alpha and reactive oxygen species regulate human embryonic stem cell-derived cardiomyocyte function. Stem Cell Rep. *1*, 560–574.

Birket, M.J., Ribeiro, M.C., Verkerk, A.O., Ward, D., Leitoguinho, A.R., den Hartogh, S.C., Orlova, V.V., Devalla, H.D., Schwach, V., Bellin, M., et al. (2015). Expansion and patterning of cardiovascular progenitors derived from human pluripotent stem cells. Nat. Biotechnol. *33*, 970–979.

Bondue, A., and Blanpain, C. (2010). Mesp1: a key regulator of cardiovascular lineage commitment. Circ. Res. *107*, 1414–1427.

Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., et al. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods *10*, 1093–1095.

Bu, L., Jiang, X., Martin-Puig, S., Caron, L., Zhu, S., Shao, Y., Roberts, D.J., Huang, P.L., Domian, I.J., and Chien, K.R. (2009). Human ISL1 heart progenitors generate diverse multipotent cardiovascular cell lineages. Nature *460*, 113–117.

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.

Cambier, L., Plate, M., Sucov, H.M., and Pashmforoush, M. (2014). Nkx2-5 regulates cardiac growth through modulation of Wnt signaling by R-spondin3. Development *141*, 2959–2971.

Chong, J.J., Reinecke, H., Iwata, M., Torok-Storb, B., Stempien-Otero, A., and Murry, C.E. (2013). Progenitor cells identified by PDGFR-alpha expression in the developing and diseased human heart. Stem Cells Dev. *22*, 1932–1943.

Christiaen, L., Stolfi, A., Davidson, B., and Levine, M. (2009). Spatio-temporal intersection of Lhx3 and Tbx6 defines the cardiac field through synergistic activation of Mesp. Dev. Biol. *328*, 552–560.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Cui, C., Chatterjee, B., Lozito, T.P., Zhang, Z., Francis, R.J., Yagi, H., Swanhart, L.M., Sanker, S., Francis, D., Yu, Q., et al. (2013). Wdpcp, a PCP protein required for ciliogenesis, regulates directional cell migration and cell polarity by direct modulation of the actin cytoskeleton. PLoS Biol. *11*, e1001720.

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.

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

Friedman, C.E., Nguyen, Q., Lukowski, S.W., Helfer, A., Chiu, H.S., Miklas, J., Levy, S., Suo, S., Han, J.J., Osteil, P., et al. (2018). Single-cell transcriptomic analysis of cardiac differentiation from human PSCs reveals HOPX-dependent cardiomyocyte maturation. Cell Stem Cell *23*, 586–598.

Galdos, F.X., Guo, Y., Paige, S.L., VanDusen, N.J., Wu, S.M., and Pu, W.T. (2017). Cardiac regeneration: lessons from development. Circ. Res. *120*, 941–959.

Guigo, R., Dermitzakis, E.T., Agarwal, P., Ponting, C.P., Parra, G., Reymond, A., Abril, J.F., Keibler, E., Lyle, R., Ucla, C., et al. (2003). Comparison of mouse and human genomes followed by experimental verification yields an estimated 1,019 additional genes. Proc. Natl. Acad. Sci. USA *100*, 1140–1145.

Haghverdi, L., Buettner, F., and Theis, F.J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics *31*, 2989–2998.

Han, X.H., Jin, Y.R., Seto, M., and Yoon, J.K. (2011). A WNT/beta-catenin signaling activator, R-spondin, plays positive regulatory roles during skeletal myogenesis. J. Biol. Chem. *286*, 10649–10659.

Hashem, S.I., Lam, M.L., Mihardja, S.S., White, S.M., Lee, R.J., and Claycomb, W.C. (2013). Shox2 regulates the pacemaker gene program in embryoid bodies. Stem Cells Dev. *22*, 2915–2926.

Hovanes, K., Li, T.W., Munguia, J.E., Truong, T., Milovanovic, T., Lawrence Marsh, J., Holcombe, R.F., and Waterman, M.L. (2001). Beta-catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. Nat. Genet. *28*, 53–57.

Hrvatin, S., Deng, F., O'Donnell, C.W., Gifford, D.K., and Melton, D.A. (2014). MARIS: method for analyzing RNA following intracellular sorting. PLoS One *9*, e89459.

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.

Jansson, L., Kim, G.S., and Cheng, A.G. (2015). Making sense of Wnt signaling-linking hair cell regeneration to development. Front. Cell. Neurosci. *9*, 66.

Jha, R., Singh, M., Wu, Q., Gentillon, C., Preininger, M.K., and Xu, C. (2017). Downregulation of LGR5 expression inhibits cardiomyocyte differentiation and potentiates endothelial differentiation from human pluripotent stem cells. Stem Cell Rep. *9*, 513–527.

Jin, S.C., Homsy, J., Zaidi, S., Lu, Q., Morton, S., DePalma, S.R., Zeng, X., Qi, H., Chang, W., Sierant, M.C., et al. (2017). Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. Nat. Genet. *49*, 1593–1601.

Junttila, M.R., Mao, W., Wang, X., Wang, B.E., Pham, T., Flygare, J., Yu, S.F., Yee, S., Goldenberg, D., Fields, C., et al. (2015). Targeting LGR5+ cells with an antibody-drug conjugate for the treatment of colon cancer. Sci. Transl. Med. *7*, 314ra186.

Kapoor, N., Liang, W., Marbán, E., and Cho, H.C. (2013). Direct conversion of quiescent cardiomyocytes to pacemaker cells by expression of Tbx18. Nat. Biotechnol. *31*, 54–62.

Klaus, A., Saga, Y., Taketo, M.M., Tzahor, E., and Birchmeier, W. (2007). Distinct roles of Wnt/beta-catenin and Bmp signaling during early cardiogenesis. Proc. Natl. Acad. Sci. USA *104*, 18531–18536.

Laugwitz, K.L., Moretti, A., Lam, J., Gruber, P., Chen, Y., Woodard, S., Lin, L.Z., Cai, C.L., Lu, M.M., Reth, M., et al. (2005). Postnatal isl1+ cardioblasts enter fully differentiated cardiomyocyte lineages. Nature *433*, 647–653.

Lian, X., Zhang, J., Azarin, S.M., Zhu, K., Hazeltine, L.B., Bao, X., Hsiao, C., Kamp, T.J., and Palecek, S.P. (2013). Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions. Nat. Protoc. *8*, 162–175.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, 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.

Mayer, D.C., and Leinwand, L.A. (1997). Sarcomeric gene expression and contractility in myofibroblasts. J. Cell Biol. *139*, 1477–1484.

Moretti, A., Caron, L., Nakano, A., Lam, J.T., Bernshausen, A., Chen, Y., Qyang, Y., Bu, L., Sasaki, M., Martin-Puig, S., et al. (2006). Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. Cell *127*, 1151–1165.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A.P., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-cell RNA-Seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell *165*, 1012–1026.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Reller, M.D., Strickland, M.J., Riehle-Colarusso, T., Mahle, W.T., and Correa, A. (2008). Prevalence of congenital heart defects in metropolitan Atlanta, 1998-2005. J. Pediatr. *153*, 807–813.

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

Sahara, M., Hansson, E.M., Wernet, O., Lui, K.O., Später, D., and Chien, K.R. (2014). Manipulation of a VEGF-Notch signaling circuit drives formation of functional vascular endothelial progenitors from human pluripotent stem cells. Cell Res. *24*, 820–841.

Sahara, M., Santoro, F., and Chien, K.R. (2015). Programming and reprogramming a human heart cell. EMBO J. *34*, 710–738.

Soibam, B., Benham, A., Kim, J., Weng, K.C., Yang, L., Xu, X., Robertson, M., Azares, A., Cooney, A.J., Schwartz, R.J., et al. (2015). Genome-wide identification of MESP1 targets demonstrates primary regulation over mesendoderm gene activity. Stem Cells *33*, 3254–3265.

Takahashi, M., Terasako, Y., Yanagawa, N., Kai, M., Yamagishi, T., and Nakajima, Y. (2012). Myocardial progenitors in the pharyngeal regions migrate to distinct conotruncal regions. Dev. Dyn. *241*, 284–293.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

Vander Heiden, M.G., Locasale, J.W., Swanson, K.D., Sharfi, H., Heffron, G.J., Amador-Noguez, D., Christofk, H.R., Wagner, G., Rabinowitz, J.D., Asara, J.M., et al. (2010). Evidence for an alternative glycolytic pathway in rapidly proliferating cells. Science *329*, 1492–1499.

Vincent, S.D., and Buckingham, M.E. (2010). How to make a heart: the origin and regulation of cardiac progenitor cells. Curr. Top. Dev. Biol. *90*, 1–41.

Watanabe, Y., Zaffran, S., Kuroiwa, A., Higuchi, H., Ogura, T., Harvey, R.P., Kelly, R.G., and Buckingham, M. (2012). Fibroblast growth factor 10 gene regulation in the second heart field by Tbx1, Nkx2-5, and Islet1 reveals a genetic switch for down-regulation in the myocardium. Proc. Natl. Acad. Sci. USA *109*, 18273–18280.

Wu, S.M., Fujiwara, Y., Cibulsky, S.M., Clapham, D.E., Lien, C.L., Schultheiss, T.M., and Orkin, S.H. (2006). Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. Cell *127*, 1137–1150.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
See Table S5 for antibodies for flow cytometry	This paper	Table S5
See Table S6 for antibodies for immunostaining	This paper	Table S6
Biological Samples		
Human embryonic/fetal heart tissues	Authorized sources in Karolinska University Hospital (Sweden)	N/A
Chemicals, Peptides, and Recombinant Proteins		
mTeSR1	STEMCELL Technologies	Cat.#05850
RPMI	ThermoFisher	Cat.#11875119
Matrigel	BD Biosciences	Cat.#354234
Dispase	STEMCELL Technologies	Cat.#07913
ROCK inhibitor Y-27632	Tocris	Cat.#1254
Accutase	STEMCELL Technologies	Cat.#07920
Collagenase type 2	Worthington	Cat.#4176
CHIR99021	Sigma	Cat.#SML1046
B-27 Supplement (50X)	ThermoFisher	Cat.#17504044
B-27 Supplement (50X), minus insulin	ThermoFisher	Cat.#A1895601
IWP-2 (Wnt inhibitor)	Tocris	Cat.#3533
Recombinant human BMP4	R&D	Cat.#314-BP
Recombinant human ActivinA	R&D	Cat.#338-AC
Horse serum	ThermoFisher	Cat.#1605130
Bovine serum albumin	Sigma	Cat.#A2153
Paraformaldehyde	Sigma	Cat.#158127
RNaseOUT Ribonuclease inhibitor	ThermoFisher	Cat.#10777019
Protease type XXIV	Sigma	Cat.#P8038
Fetal bovine serum	Sigma	Cat.#12103C
TrypLE Express enzyme	ThermoFisher	Cat.#12604013
Tryton X-100	Sigma	Cat.#X100
dNTP mix	New England Biolabs	Cat.#N0447S
SuperScript II reverse transcriptase	ThermoFisher	Cat.#18064014
KAPA HiFi HotStart ReadyMix	KAPA Biosystems	Cat.#KK2601
AMPure XP beads	Beckman Coulter	Cat.#A63881
DAPI	Sigma	Cat.#D9542
OCT compound	Tissue Tek	Cat.#25608-930
Tamoxifen	Sigma	Cat.#T5648
SuperScript III reverse transcriptase	ThermoFisher	Cat.#18080044
Fast SYBR Green Master Mix	ThermoFisher	Cat.#4385610
DMEM/F12	ThermoFisher	Cat.#11320033
KnockOut Serum Replacement	ThermoFisher	Cat.#10828028
SmGM-2	Lonza	Cat.#CC-3182
EGM2	Lonza	Cat.#CC-3162
Puromycin	ThermoFisher	Cat.#A1113802
Dream Taq DNA polymerase	ThermoFisher	Cat.#EP1701

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
RecoverAll Total Nucleic Acid Isolation Kit	Ambion	Cat.#AM1975
Smart-Seq2	Picelli et al. (2013)	N/A
Agilent RNA 6000 Nano Kit	Agilent	Cat.#5067-1511
Agilent RNA 6000 Nano Chips	Agilent	Cat.#5067-1529
Agilent High Sensitivity DNA Kit	Agilent	Cat.#5067-4626
Agilent High Sensitivity DNA Chips	Agilent	Cat.#5067-4627
Nextera XT DNA Sample Preparation Kit	Illumina	Cat.#FC-131-1024
Qubit High Sensitivity DNA Kit	ThermoFisher	Cat.#Q32854
Amaxa Human Stem Cell Nucleofector Kit #2	Lonza	Cat.#VVPH-5022
SimpleChIP Plus Enzymatic Chromatin IP Kit	Cell Signaling Technology	Cat.#9005
GeneJET Genomic DNA Purification Kit	ThermoFisher	Cat.#K0721
Deposited Data		
Bulk RNA-seq data	This paper	E-MTAB-7537
Single RNA-seq data	This paper	PRJNA510181
Experimental Models: Cell Lines		
Human ESC: ES03 line	WiCell	Cat.#ES03
Human ESC: ES03 LGR5-KO	This paper	N/A
Experimental Models: Organisms/Strains		
Mouse: wild-type: C57BL/6J	Charles River	Cat.#632
Mouse: Lgr5-EGFP-IRES-CreERT2 (*Lgr5*$^{CreERT2/+}$): B6.129P2-*Lgr5*$^{tm1(cre/ERT2)Cle}$/J	The Jackson Laboratory	Cat.#008875
Mouse: *R26R*^{tdTomato}: B6.Cg-*Gt*(ROSA) *26Sor*$^{tm14(CAG-tdTomato)Hze}$/J	The Jackson Laboratory	Cat.#007914
Oligonucleotides		
See Table S7 for quantitative PCR primer sequences	This paper	Table S7
See "Construction of cDNA Libraries with SMART-Seq2" for oligos for SMART-Seq2	This paper	N/A
See ""Construction of CRISPR/Cas9-Mediated Human LGR5-KO ESCs" for gRNA oligos	This paper	N/A
Recombinant DNA		
Plasmid: pX330	Addgene	Cat.#42230
Software and Algorithms		
R	Bioconductor	https://www.bioconductor.org/
STAR	Dobin et al. (2013)	N/A
Htseq	HTSeq by Simon Anders	https://htseq.readthedocs.io/en/release_0.10.0/
FACS Diva	Beckton Dickinson	http://www.bdbiosciences.com/
FlowJo	Tree Star	http://www.flowjo.com
CASAVA	Illumina	https://www.illumina.com/
Samtools	Sourceforge	http://samtools.sourceforge.net/
Cutadapt	Cutadapt by Marcel Martin	https://doi.org/10.14806/ej.17.1.200
Seurat	Macosko et al. (2015)	https://satijalab.org/seurat/
edgeR	Bioconductor	http://bioconductor.org/packages/release/bioc/html/edgeR.html
DAVID Bioinformatics Resources 6.8	Laboratory of Human Retrovirology and Immunoinformatics	https://david.ncifcrf.gov/
CHOPCHOP	CHOPCHOP web tool	http://chopchop.cbu.uib.no/
MatInspector	Genomatix	http://www.genomatix.de/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
FACSAria III (a flow cytometer and cell sorter)	Beckton Dickinson	http://www.bdbiosciences.com/
Bioanalyzer	Agilent	https://www.agilent.com/
Hiseq 2500	Illumina	https://www.illumina.com/
Zeiss confocal 710 microscope	Zeiss	https://www.zeiss.com/
7500 Fast Real-Time PCR System	Applied Biosystems	https://www.thermofisher.com/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kenneth R. Chien (kenneth.chien@ki.se).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Generation and Maintenance of Human ESC Lines

The human ESC line ES03 (karyotype: 46, XX) was purchased from WiCell Research Institute, and maintained on feeder-free and 0.3 mg/ml Matrigel (BD Biosciences)-coated plates in mTeSR1 medium (STEMCELL Technologies) according to manufacturers' instructions. Cells were fed daily and passaged every 5-7 days with Dispase (STEMCELL Technologies). Media was supplemented with 5 μM ROCK inhibitor Y-27632 (Tocris) for 24 hours after splitting. The ES03 cell line was used to generate the human LGR5-KO ESC line mediated by CRISPR/Cas9 technology, as described below.

Human Embryonic/Fetal Heart Tissues

Human embryonic/fetal hearts at 4.5-10 weeks of the gestation stages were obtained from authorized sources in Karolinska University Hospital with an approved ethical permission (Dnr 2015/1369-31/2) and appropriate informed consents.

Wild-Type and Genetically Modified Mouse Lines

Wild-type mice (C57BL/6J) were purchased from Charles River. Genetically modified mice, which include *Lgr5-EGFP-IRES-CreERT2* (*Lgr5*$^{CreERT2/+}$) knock-in mice (Barker et al., 2007) and *Rosa26-floxed stop cassette-tdTomato* (*R26R*^{tdTomato}) reporter mice were purchased from the Jackson Laboratory. The embryos of these mice were further investigated, as described below. All experimental procedures with animals were performed in accordance with the protocols approved by Institutional Animal Care and Use Committee at Karolinska Institutet (N103/14).

METHOD DETAILS

Cardiac-Directed Differentiation

A previously published cardiac differentiation protocol based on the Wnt signaling modulation was followed (Figure 1A) (Lian et al., 2013). Briefly, human ESCs maintained on Matrigel were dissociated into single cells with Accutase (STEMCELL Technologies) at 37°C for 5-10 min and then seeded onto Matrigel-coated 12-well plates at 250,000–300,000 cells/well in mTeSR1 supplemented with 5 μM Y-27632 for 24 h (day −4). Cells were fed daily with mTeSR1 from day −3 to −1. At day 0, cells were treated with 12 μM GSK-3β inhibitor CHIR99021 (Sigma) in RPMI medium supplemented with B27 minus insulin (RPMI/B27-ins; Thermo Fisher Scientific) for 24 h. At day 1, the GSK-3β inhibitor was removed by replacing the media with fresh RPMI/B27-ins. At day 3, half of the medium in each well was changed to RPMI/B27-ins supplemented with 5 μM Wnt inhibitor IWP-2 (Tocris), all of which was replaced with fresh RPMI/B27-ins at day 5. At day 7, cells were switched to RPMI medium with B27 supplement (RPMI/B27, Thermo Fisher Scientific). Medium was replaced thereafter every 2 or 3 days.

Separately, another well-established cardiac differentiation protocol that is based on the use of Activin A and BMP4 was applied for human ESC differentiation into CMs (Birket et al., 2013). Briefly, seeded ESCs were treated with 100 ng/mL Activin A (R&D) in RPMI/B27-ins on day 0. After 24 h, the medium was changed to fresh RPMI/B27-ins supplemented with 5 ng/mL BMP4 (R&D) for the next 4 days. On day 5, the medium was changed to RPMI/B27 and after that, medium was changed every 2 or 3 days.

Isolation of Human ESC-Derived Cardiac Derivatives with FACS

To separately isolate human ESC-derived cardiac lineage cells, including MCPs, intermediates, and differentiated cardiac cells, FACS sorting was performed with lineage-specific primary antibodies at the different time-points (day 3, 6, and 10) during cardiac differentiation (Figure 1). Cells were dissociated into single cells with Accutase for 5-10 min, washed in PBS, and blocked for 30 min in FACS buffer (1% bovine serum albumin and 10% horse serum in PBS) at 4°C. Staining for cell surface antigens was

performed for 30 min at 4°C with the following primary antibodies: anti-platelet derived growth factor receptor-α (PDGFRA, a mesodermal marker; BD Biosciences), anti-CD31 (an endothelial cell marker; BD Biosciences), and anti-HCN4 (a pacemaker/conduction system cell marker; StressMarq). Cells were then fixed with 4% paraformaldehyde, permeabilized, blocked, and stained for intracellular antigens for 30 min at room temperature with the following primary antibodies: anti-ISL1 (a CPC marker; DSHB), anti-TNNT2 (a cardiomyocyte marker; Abcam), and anti-smooth muscle myosin heavy chain (SMMHC; a smooth muscle cell marker; Biomedical Technologies). After washing in PBS, cells were stained with Alexa-Fluor 405-, 488-, 546- or 647-conjugated secondary antibodies (BD Biosciences) specific to the appropriate species for 15 min at 4°C if the used primary antibody was not conjugated with a fluorochrome. Buffers were supplemented with 1:100 RNaseOUT Ribonuclease inhibitor (Thermo Fisher Scientific) if cells were to be recovered for RNA extraction. With a multicolor cell separation technique on a flow cytometer (FACSAria III) and FACS Diva software (Beckton Dickinson), single lineage populations, such as $ISL1^{+}PDGFRA^{+}$ MCPs (day 3), ISL1+Lin (TNNT2, HCN4, SMMHC, or CD31)$^{+}$ intermediates (day 6) and $ISL1^{-}Lin^{+}$ differentiated cells (day 10) were separately sorted (Figure 1). FACS data were analyzed with FACS Diva and FlowJo software (Tree Star). All the primary antibodies used in FACS sorting, as described above, or other flow cytometry analyses are listed in Table S5.

RNA Extraction of FACS-Sorted Cells

Immediately after FACS sorting, total RNA of the cells were extracted from the cell pellets using the RecoverAll Total Nucleic Acid Isolation Kit (Ambion), starting at the protease digestion stage of the manufacturer-recommended protocol. As previously reported (Hrvatin et al., 2014), the following modification to the isolation procedure was made: cells were incubated in digestion buffer for 1 hour at 50°C. Cell lysates were frozen at -80°C overnight before continuing the isolation according to the manufacturer's instructions. RNA was extracted according to the manufacturer's instructions. RNA integrity and concentration were measured using the Agilent RNA 6000 Nano kit with the 2100 Bioanalyzer instrument (Agilent).

Isolation of Human Embryonic/Fetal Heart-Derived Single Cardiac Cells

The collected human embryonic/fetal hearts (4.5-10 weeks of the gestation stages) were classified into one of the following three groups in accordance with the embryonic/fetal stages: 1) early stage (looping ~ initiation of septation) at 4.5-5.5 week (n = 2); 2) middle stage (septation and chamber formation process) at 6-7.5 week (n = 3); and 3) late stage (all compartments developed) at 8-10 week (n = 2) (Figure 4A). The gestation stages were judged by the measurements of the embryo/fetus size, including a crown rump length and other parameters. The hearts were carefully dissected and divided into 3 compartments, such as outflow tract (= the cono-ventricular regions), ventricle (= the free wall-ventricular regions) and atria, without any valve or fatty and connective tissues. Each divided region was additionally cut in small pieces and washed repetitively in solution A (10 mM HEPES, 35 mM NaCl, 10 mM glucose, 134 mM sucrose, 16 mM Na_2HPO_4, 25 mM $NaHCO_3$, 7.75 mM KCl, and 1.18 mM KH_2PO_4 [pH 7.4]) (Sahara et al., 2014). The first digestion step was performed in solution A supplemented with 0.5% bovine serum albumin, 200 U/ml collagenase type II (Worthington), and 6 U/ml protease type XXIV (Sigma) at 37°C for 10 min to remove red blood cells and cell debris. This was followed by the next digestion step performed in solution A supplemented with 400 U/ml collagenase type II at 37°C for 25 min. Dissociated single cells were neutralized with fetal bovine serum (FBS), centrifuged, and resuspended in cell suspension buffer, containing DPBS and TrypLE Express enzyme (Thermo Fisher Scientific) in a ratio of 1:1. The divided region (OFT, ventricle, and atria)-derived single cardiac cells were harvested by manually picking with a micro capillary pipette under the microscope. At picking, the volume of liquid including a single cell was kept below 0.5 μl. The picked cells were transferred to a 0.2 ml thin-wall PCR tube containing 2 μl of a mild hypotonic lysis buffer composed of 0.2% Triton X-100 (Sigma) and 2 U/μl of RNase inhibitor (Picelli et al., 2013). Cells were then directly snap-frozen and stored at −80°C.

Isolation of Human ESC-Derived Single Cardiac Cells

In a separate experiment, human ESC-derived cardiac-differentiating cells were dissociated into single cells with Accutase for 5-10 min on days 3, 6, and 15 in the cardiac differentiation (Figure 1A). Dissociated single cells were neutralized with fetal bovine serum (FBS), centrifuged, and resuspended in cell suspension buffer, containing DPBS and TrypLE Express enzyme (Thermo Fisher Scientific) in a ratio of 1:1. *In vitro* single cells were then harvested by manually picking with a micro capillary pipette under the microscope and processed for single-cell RNA-seq, as described above.

Construction of cDNA Libraries with SMART-Seq2

cDNA libraries of the human ESC-derived cardiac bulk populations and single cardiac cells and human embryonic/fetal heart-derived single cardiac cells were generated with the Smart-seq2 approach (Picelli et al., 2013). Briefly, 100 pg total RNA from bulk populations or single-cell lysates were mixed with 1 μl of anchored oligo-dT primer (10 μM, 5′-AAGCAGTGGTATCAACGCAGAGTACT$_{30}$VN-3′, where "N" is any base and "V" is either "A", "C" or "G"; Biomers.net) and 1 μl of dNTP mix (10 mM, New England Biolabs), denatured at 72°C for 3 min and immediately placed on ice afterward. After adding 5.7 μl of the first-strand reaction mix, containing 0.5 μl SuperScript II reverse transcriptase (200 U/μl, Invitrogen) and 0.1 μl template-switching oligo (TSO; 100 μM, Exiqon), reverse transcription reaction was carried out by incubating at 42°C for 90 min, followed by 10 cycles of (50°C for 2 min, 42°C for 2 min). Then, the PCR master mix, containing 25 μl KAPA HiFi HotStart ReadyMix (2×, KAPA Biosystems) and 1 μl ISPCR primers (10 μM, 5′-AAGCAGTGG TATCAACGCAGAGT-3′; Biomers.net), was added to each sample tube, and pre-amplification PCR reaction was carried out as follows: 98°C 3 min, then 15 (bulk RNA) or 18 (single cells) cycles of (98°C 15 s, 67°C 20 s, 72°C 6 min), with a final extension at 72°C

for 5 min. PCR products were purified using a 1:1 ratio of AMPure XP beads (Beckman Coulter). The library size distribution was checked on a High-Sensitivity DNA chip (Agilent Bioanalyzer). The expected average size should be around 1.5–2.0 kb, and we relied on the amount of cDNA comprised in the interval 300–9,000 bp plot.

Five hundred picograms to 1 ng of cDNA were then used for the tagmentation reaction that was carried out with 5 μl of tagment DNA enzyme of the Nextera DNA Sample Preparation kit (Illumina) at 55°C for 5 min. To strip off the tagment enzyme, 5 μl of NT buffer (Nextera XT DNA sample preparation kit) was added to the 20-μl solution containing the tagmented DNA, and the mixture was incubated for 5 min at RT. Next, the whole volume was used for final enrichment PCR, along with 15 μl of Nextera PCR Primer Mix (NPM), 5 μl of Index 1 primers (N7xx), 5 μl of Index 2 primers (N5xx). A second amplification round was performed as follows: 72°C 3 min, 95°C 30 s, then 12 cycles of (95°C 10 s, 55°C 30 s, 72°C 30 s), with a final extension at 72°C for 5 min. Purification of PCR products was done with a 0.6:1 ratio of AMPure XP beads and samples. The quality of the final cDNA library was verified using a High-Sensitivity DNA chip and Bioanalyzer (Agilent) instrument. cDNA quantification was performed with Qubit High-Sensitivity DNA kit (Invitrogen). Libraries were diluted to a final concentration of 4 nM and pooled equally for sequencing.

RNA Sequencing and Data Pre-processing with Bioinformatics Analyses

Pooled cDNA libraries were sequenced at 150 bp paired-end or 100 bp single-end on an Illumina HiSeq 2500 instrument to a depth of (4-7) × 10^6 reads. CASAVA software was used to separate out the data for each sample by using unique barcode combinations from the Nextera XT preparation and to generate *.fastq files. Raw reads were pre-processed with the sequence-grooming tools FASTQC and Cutadapt and followed by sequence alignment with the STAR and Samtools onto human genome reference (hg19) with default settings (Treutlein et al., 2014). Mapped gene counts were carried out with HTSeq, and transcript levels were quantified for each transcript as fragments or reads per kilo base of transcript per million mapped reads (FPKM or RPKM). We filtered low expression genes by trimming genes that had $\leq$ 5 FPKM summed in all of 39 bulk RNA data set of the hESC-derived 13 cardiac lineages (*in-vitro* population RNA-seq of hESC-derived cardiac cells), that were expressed (FPKM $\geq$ 5) in only $\leq$ 5 out of 676 sequenced cells (*in vivo* single-cell RNA-seq of human embryonic/fetal heart cells), or that were expressed (RPKM $\geq$ 5) in only $\leq$ 3 out of 384 sequenced cells (*in vitro* single-cell RNA-seq of hESC-derived cardiac cells). 20,243 (population RNA-seq), 19,117 (*in vivo* single-cell RNA-seq), and 18,776 genes (*in vitro* single-cell RNA-seq) were kept and processed for further analyses.

In the *in vivo* single-cell transcriptomes, cells were quality-filtered based on the following four criteria, leaving 458 cells post-filtered out of 676 sequenced cells. First, Spearman correlations were calculated for every possible pair of cells with the FPKM expression levels of all genes and used to identify 92 outlier cells with maximum pair-wise correlations below 0.64 (Figure S3A). Second, a histogram of the number of expressed genes per cell (FPKM $\geq$ 1) was used to identify 70 outlier cells with less than 6,000 expressed genes (Figure S3A). Third, we found 28 single cells expressing hematopoietic markers (e.g., *HBB*, *HBE1*) 5 single cells expressing endothelial markers (e.g., *PECAM1*, *CDH5*, *VWF*), 4 single cells expressing smooth muscle markers (e.g., *MYH11*, *ACTG*, *SM22α*), and 5 single cells expressing fibroblastic markers (e.g., *VIM*, *FBN1*, *COL1A1*) without any expression of cardiac sarcomere protein genes (e.g., *MYH6*, *TNNC1*, *ACTC1*). These 42 cells, which were considered as hematopoietic cells, ECs, SMCs and fibroblasts, respectively, were excluded from further analyses, due to out of the range or low abundance. Fourth, 18 outlier cells were identified using principal component analysis (PCA) and diffusion map dimensionality reduction (Haghverdi et al., 2015). In the *in vitro* single-cell transcriptomes, cells were quality-filtered based on the above three criteria (1, 2, and 4), leaving 366 cells post-filtered out of 384 sequenced cells. With depth-normalized libraries' data with filtered genes and cells, all bioinformatics analyses were performed using R/Bioconductor.

Gene Variability, Lineage Segregation of Cells, and Lineage Differential Expression Analysis

A gene-variability statistic was calculated and adjusted for the mean-variance relationship present in the RNA-seq data (Figure S1B). This was done with a minor modification by assuming that the expression distribution of a gene follow a negative binomial for which the variance and coefficient of variation (CV; standard deviation divided by the mean) depends on the mean (Brennecke et al., 2013). In the *in-vitro* population RNA-seq data, we fit a simple noise model as "$\log_2$ (CV) = $\log_2$ (*mean*^a + b) + c", where the best fit was found to be a = -0.639, b = 0.120, and c = 1.58 (Figure S1B). We ranked all genes by their distance from the line and select the top 5000 highly variable genes as informative for further analysis. For analysis of the single-cell RNA-seq data (*in vivo* and *in vitro*), the Seurat package (Macosko et al., 2015) implemented in R was used, and similarly, highly variable genes were calculated through the average expression and dispersion for each gene. Finally, 5852 (*in vivo*) or 5775 (*in vitro*) highly variable genes were selected in each as informative for further analysis. Using the highly variable genes in each data set, multi-dimensional scaling, PCA, hierarchical clustering and diffusion map dimensionality reduction (Haghverdi et al., 2015) were performed for the population or single-cell transcriptomes.

To cluster the analyzed single cells, we implemented unbiased clustering using the Seurat program that is one of the established clustering methods for single cell analysis (Macosko et al., 2015) and visualized the results by two-dimensional t-distributed stochastic neighbor embedding (tSNE) to reduce the complexity of the data (Figures 3B and 4C). Among the *in vitro* human ESC-derived populations or the Seurat-defined clusters of the single cells (*in vivo* and *in vitro*), differential expression analysis was conducted using edgeR program (http://bioconductor.org/packages/release/bioc/html/edgeR.html), and differential expression genes (DEGs) were identified as ones with false discovery rate (FDR) $\leq$ 0.05. To further identify the genes specifically expressed and enriched by each population/cluster, the DEGs with mean FPKM $\geq$ 5 in each lineage (population RNA-seq) or with mean FPKM $\geq$ 1 in each cluster (single-cell RNA-seq) were ranked by descending order of the calculated Z-scores of $\log_2$ (FPKM) of the genes and listed

in Tables S1 and S3, respectively. Gene Ontology gene set enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/home.jsp).

Immunostaining

Human embryonic/fetal hearts were snap-frozen and cryosectioned at 10-μm thickness. Human embryonic/fetal heart sections and human ESC-derived cells were immunostained as follows. Samples were fixed with 4% paraformaldehyde, permeabilized in PBS with 0.1% saponin, and blocked in PBS with 1% bovine serum albumin and 10% horse serum. Samples were then stained with primary antibodies at 4°C overnight, followed by three washes with PBS and incubation with Alexa-Fluor 488-, 594-, and/or Alexa Fluor 647-conjugated secondary antibodies (Molecular Probes) specific to the appropriate species for 60 min at room temperature. After three washes with PBS, nuclei were counterstained with DAPI (Sigma), or the slides were mounted in Vectashield Mounting Medium with DAPI (Vector Laboratories). All images were obtained using a Zeiss 710 confocal microscope and its imaging system. All the primary antibodies used in immunostaining are listed in Table S6.

Collection of Mouse Embryos

Wild-type mouse (C57BL/6) embryos at Embryonic days 9.5-10.5 (E9.5-10.5) were dissected, fixed in 4% paraformaldehyde/PBS for 30 min at room temperature, processed through a sucrose gradient, and embedded in frozen OCT compound (Tissue Tek). The cryosections in 8-μm thickness were subjected to immunostaining, as described above.

Mouse LGR5 Lineage Tracing

Lgr5-EGFP-IRES-CreERT2 (*Lgr5*$^{CreERT2/+}$) knock-in mice (Barker et al., 2007) and *Rosa26-floxed stop cassette-tdTomato* (*R26R*^{tdTomato}) reporter mice were obtained from the Jackson Laboratory. *Lgr5*$^{CreERT2/+}$ mice were crossed with *R26R*^{tdTomato} reporter mice, and the pregnant females received an intraperitoneal single injection of tamoxifen (75 μg per g of murine body weight) on either embryonic day 6.5 (E6.5), E7.5, or E8.5. The *Lgr5*$^{CreERT2/+}$; *R26R*^{tdTomato} heterozygote embryos were dissected on E10.5 or E14.5. Procedures were performed blinded to genotype.

Tamoxifen

Tamoxifen (Sigma) was dissolved in corn oil at a concentration of 20 mg ml^{-1}. For lineage tracing in the embryo, a single pulse of 75 μg per g of murine body weight was given through peritoneal injection into pregnant females.

Quantitative RT-PCR

Total RNA of the FACS-sorted cells were extracted as described above and reverse-transcribed to cDNA using SuperScript III reverse transcriptase (Thermo Fisher Scientific). Quantitative real-time RT-PCR was performed with Fast SYBR Green Master Mix (Thermo Fisher Scientific) for 35 cycles on a 7500 Fast Real-Time PCR System (Applied Biosystems) under standard manufacturer's conditions. The primer sequences are listed in Table S7. Quantification was performed at a threshold detection line (Ct value). The Ct values of each target gene were normalized to a housekeeping gene glyceraldehydes-3-phosphate dehydrogenase (GAPDH) and translated to relative values. Standard deviation (SD) of the means were obtained from three independent experiments.

Clonal Assays

Human ESC-derived PDGFRA$^{+/-}$LGR5$^{+/-}$ single cells were sorted by FACS on day 3 in the cardiac differentiation and seeded onto fibronectin-coated 96-well plates at 1 cell per well in DMEM/F12 medium supplemented with 5 μM Y-27632 and 10% KnockOut Serum Replacement (KO-SR, Thermo Fisher Scientific). Growing clones from single cells were picked after 7 days in culture and trypsinized. Single clone-derived cells were plated into 3 wells of a 96-well plate for differentiation experiments and further cultured under the 3 different culture conditions, customized for the CM (RPMI/B27 with 2% KO-SR), SMC (SmGM-2, Lonza), or EC (EGM2, Lonza) differentiation, respectively, for additional 14 days (Bu et al., 2009; Moretti et al., 2006) (Figure 2E). Medium was replaced thereafter every other day. After 14 days, cells were fixed and stained as described above.

Construction of CRISPR/Cas9-Mediated Human LGR5-KO ESCs

The following sequences were selected as two single guide-RNAs (sgRNA) which target the first exon of LGR5 gene locus and have minimal off-target activity, using the CHOPCHOP software (http://chopchop.cbu.uib.no/): sgRNA1 (genomic location: Chr12, 71,440,110-71,440,091), 5'-CAGGAGCACACCGAGCCGGGAGG-3' and sgRNA2 (genomic location: Chr12, 71,440,308-71,440,289), 5'- GTGGGGAAGTACTTACAGGTAGG-3' (Figure 7A). sgRNAs were cloned into a bicistronic expression vector pX330 expressing S. pyogenes Cas9 (Addgene, #42230), following previously published protocols (Cong et al., 2013). Human ESCs (ESO3) were transiently co-transfected with pX330-gRNA1 and pX330-gRNA2 using Human Stem Cell Nucleofector Kit (Lonza), according to manufacturer's instructions. To enrich the successfully transfected cells, we also co-transfected a plasmid encoding a puromycin resistance gene and subjected cells to drug selection with 1.0 μg/ml puromycin for 3 days. Transfected cells were then expanded for genomic DNA isolation and single clone picking. Single clones were obtained by re-plating transfected pools at low density. To this end, cells were dissociated with Accutase and seeded on Matrigel-coated 10 cm dishes at the density of 5,000 cells per dish. Cells were allowed to grow for 6-10 days, until single colonies were big enough to pick and transferred to a 96-well plate. Monoclonal cell lines were then expanded and tested for their CRISPR/Cas9-mediated deletion of 240 bp fragment at the

LGR5 gene locus by PCR and sequencing. For this, genomic DNA of each clone was isolated using the GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific). The edited locus was amplified with DreamTaq DNA Polymerase (ThermoFisher Scientific) according to manufacturer's instructions, using the following primer pair: (forward) 5'- AAGCAGAGATGCTGCTCCAC-3'; (reverse) 5'- CAGGATGCCCTTGACAAACT-3'. To confirm the absence of LGR5 mRNA expression of LGR5-KO-derived cells on day 3 in cardiac differentiation, RT-PCR was performed as described above, using the following primer pair: (forward) 5'-CCTGTCCTTG CCTGTGCT-3'; (reverse) 5'-CTGCAGAGCTTCTGTGGGTA-3'.

Chromatin Immunoprecipitation (ChIP)

For ChIP experiments, human ESC-differentiating cells (WT and LGR5-KO)-derived nuclei were prepared, and the following chromatin digestion was performed using SimpleChIP Plus Enzymatic Chromatin IP Kit (Cell Signaling Technology), according to manufacturer's instructions. The lysates were immunoprecipitated with normal rabbit IgG, anti-Histone H3 (D2B12) rabbit monoclonal antibody, anti-MESP1 antibody (JH.12; sc-130461, Santa-Cruz) and anti-LEF1 (D6J2W; #76010, Cell Signaling Technology) at 4°C overnight. Immune complexes were incubated with Protein G Magnetic Beads for 2 h at 4°C with rotation. After washing and eluting chromatin from the magnetic beads, protein-DNA cross-linking was reversed in 5M NaCl with 40 μg proteinase K by overnight incubation at 65°C. After a purification process, the precipitated DNA was amplified for fragments of the *LGR5* and *ISL1* promoters by real-time PCR for 35 cycles on a 7500 Fast Real-Time PCR System (Applied Biosystems) under standard manufacturer's conditions. The following PCR primers were used: (LGR5 [-651] forward) 5'-TAGCAAACAGCCATCTGTCACT-3'; (LGR5 [-651] reverse) 5'-ACTGCTGCCTTCCTATCTCTTG-3'; (ISL1 [-1247] forward) 5'-GGGGAAGAAAGCCTCAGCTA-3'; (ISL1 [-1247] reverse) 5'-TTCGCAAAATCTAACCCTTGA-3'.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data are presented as mean ± SD. Differences between groups were examined with one-way ANOVA followed by Tukey-Kramer post-hoc test. Statistical significance is defined as $P < 0.05$. All bioinformatics analyses were performed using R/Bioconductor, as described above.

DATA AND SOFTWARE AVAILABILITY

RNA-seq data reported in this paper have been deposited in the ArrayExpress database at EMBL-EBI (http://www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-7537 or in the Sequence Read Archive (SRA, www.ncbi.nlm.nih.gov/sra/) under accession number PRJNA510181.

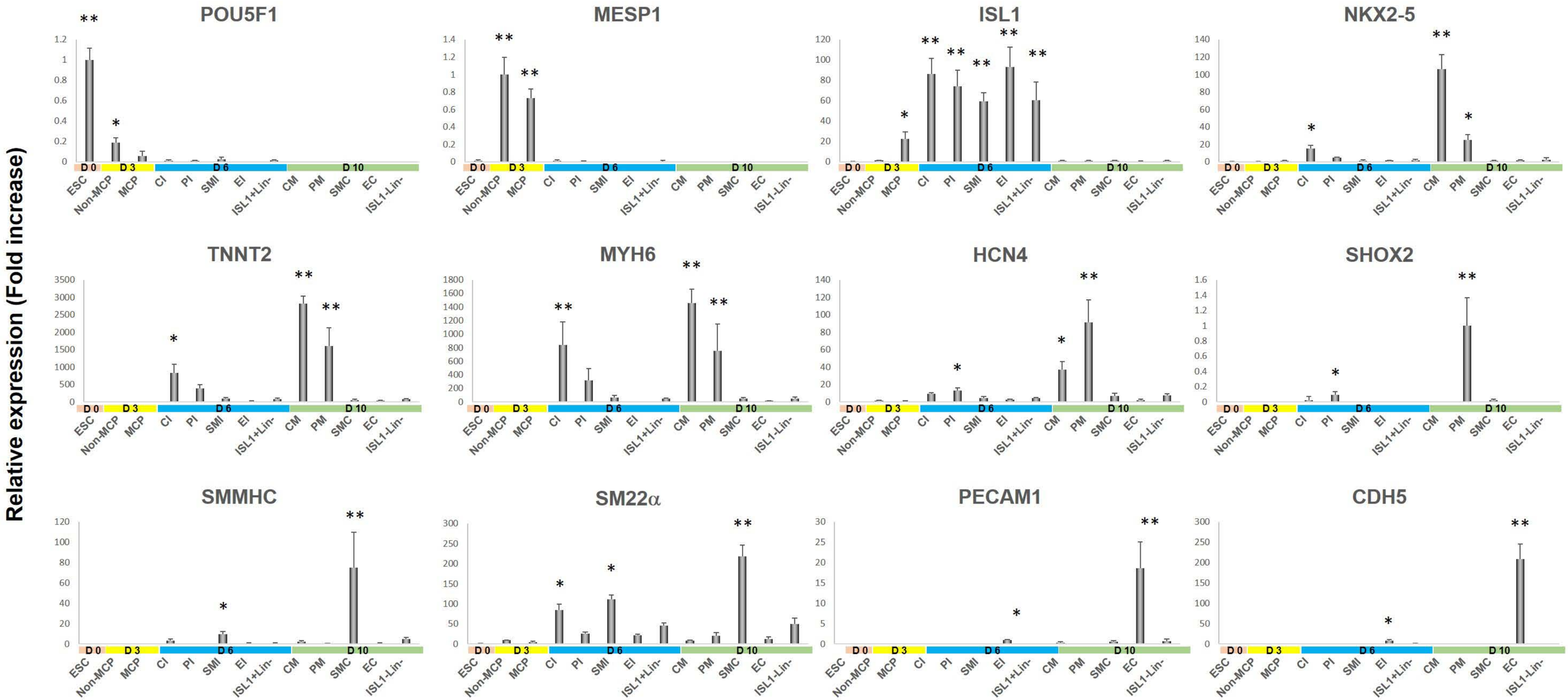
Developmental Cell, Volume *48*

Supplemental Information

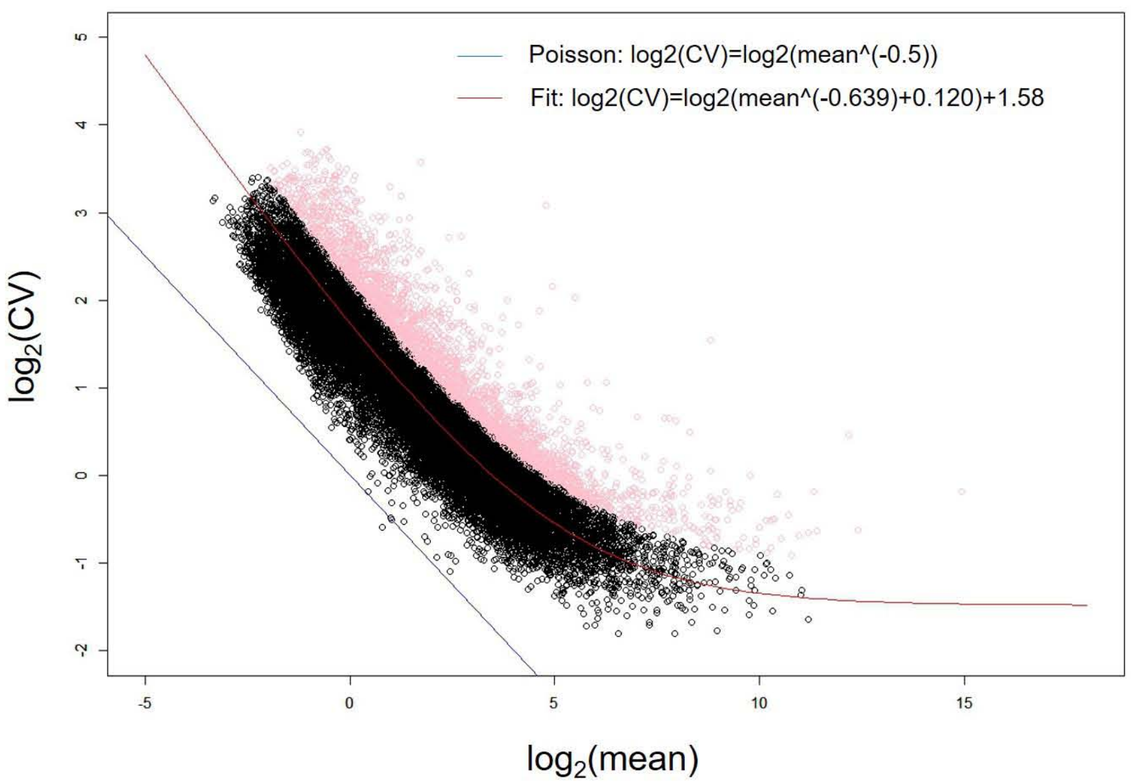
Population and Single-Cell Analysis of Human Cardiogenesis Reveals Unique LGR5 Ventricular Progenitors in Embryonic Outflow Tract

Makoto Sahara, Federica Santoro, Jesper Sohlmér, Chikai Zhou, Nevin Witman, Chuen Yan Leung, Mimmi Mononen, Kristine Bylund, Peter Gruber, and Kenneth R. Chien

A



B



C

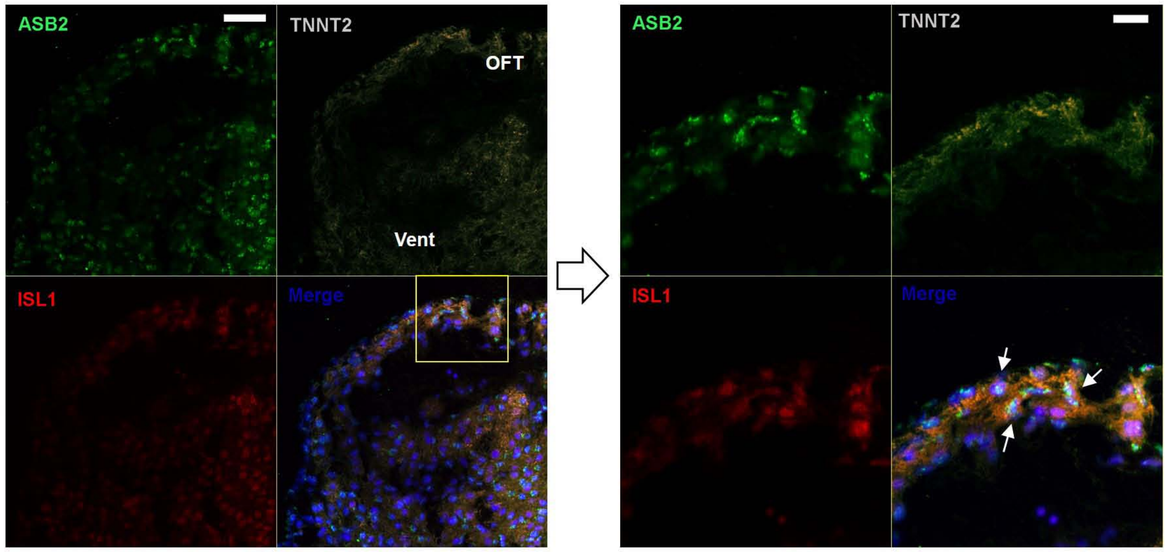
rank	ESC	MCP	CM	PM	SMC	EC
#10			TNNT2			
#13						CD34
#25						CDH5
#28			MYH6			
#41	PIM2					
#162		T				
#205			PLN			
#237					ACTG2	
#261	NANOG					
#283				TBX18		
#303		NKX2.5				
#383		TBX5				
#418					TAGLN	
#444				HCN4		
#1024		PDGFRA				
#1154	POU5F1					
#1165		ISL1				
#1263						PECAM1
#1539		MESP1				

Figure S1. Population RNA-seq Identified Highly Variable Genes among the hESC-derived Cardiac Lineages. Related to Figure 1.

(A) The identities of the FACS-sorted populations were confirmed through examinations of the lineage-specific marker genes' expression by quantitative RT-PCR (three biological replicates). The lineage-specific markers involved POU5F1 (for ESC); MESP1 (for cells on day 3); TNNT2 and MYH6 (for CM); HCN4 and SHOX2 (for PM); SMMHC and SM22α (for SMC); and PECAM1 and CDH5 (for EC). "D" (ex., D 0) indicates a "day" (ex. day 0) in the cardiac differentiation. All abbreviations are the same in Figure 1. $*P < 0.05$ vs others with no asterisks, $**P < 0.01$ vs others with no asterisks. (B) A gene-variability statistic was calculated and adjusted for the mean-variance relationship present in the population RNA-seq data set of human ESC-derived cardiac derivatives. Scatter plot of coefficient of variation (CV; standard deviation divided by the mean) versus the mean for all detected genes over the 13 lineages. Each dot represents one gene. Blue line shows expected noise from Poisson distribution and red line shows fit to a model with additive constant component. Genes were selected for further clustering analysis based on their distance from the fit line as a measure of variability (indicated by pink dots). (C) The highly variable top-ranked 5,000 genes, selected in (A), include the known lineage-specific marker genes. The variability-related ranking of each gene was shown in the leftmost column. CM, cardiomyocyte; EC, endothelial cell; ESC, embryonic stem cell; MCP, multipotent cardiac progenitor; PM, pacemaker cell; SMC, smooth muscle cell.

A

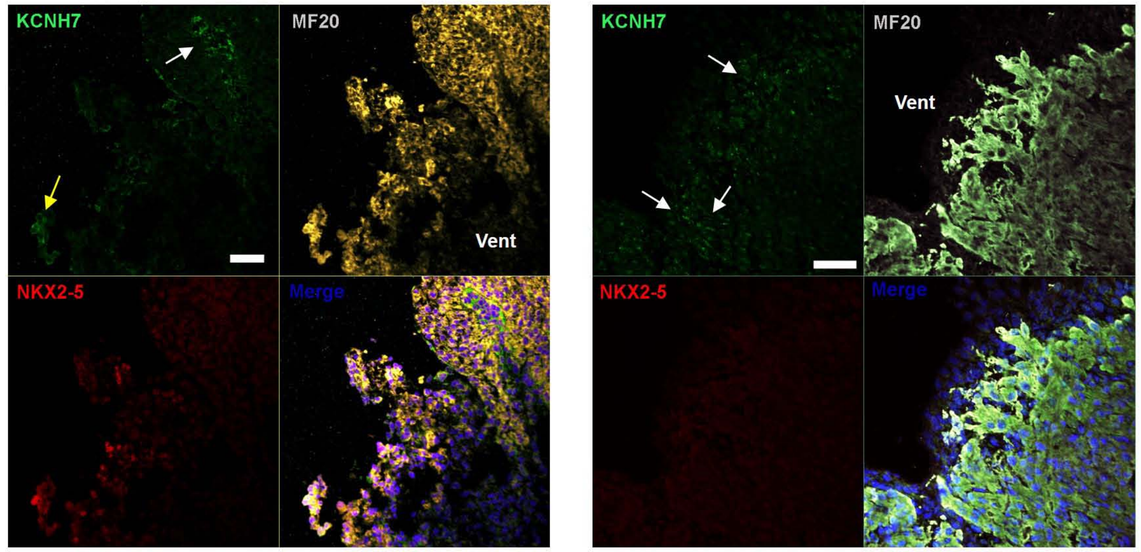
[Human embryonic heart at 6.5 week]



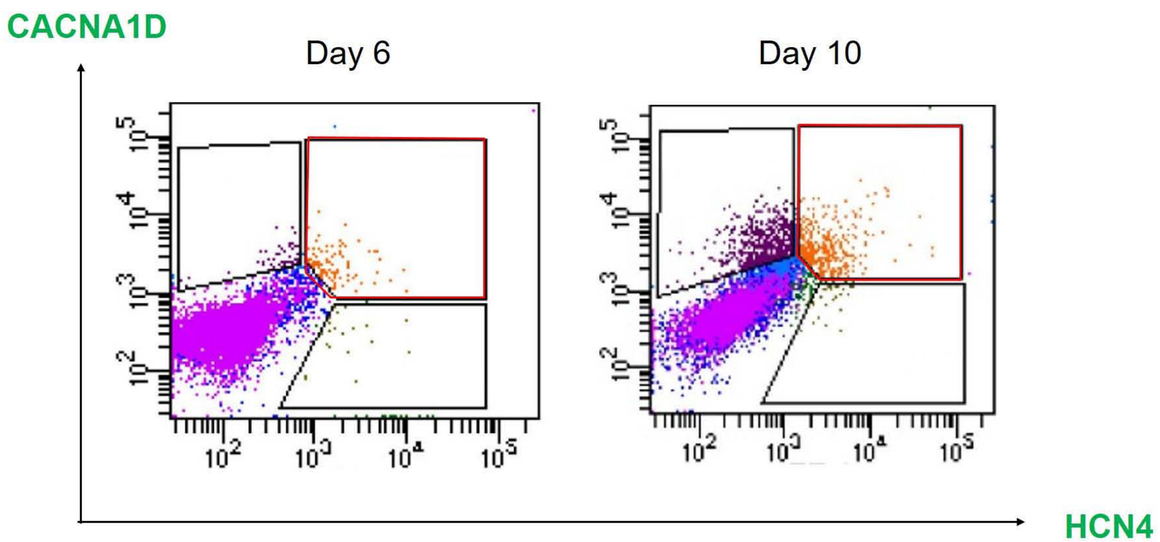
B

[Human embryonic heart at 4.5 week]

[Human fetal heart at 10 week]

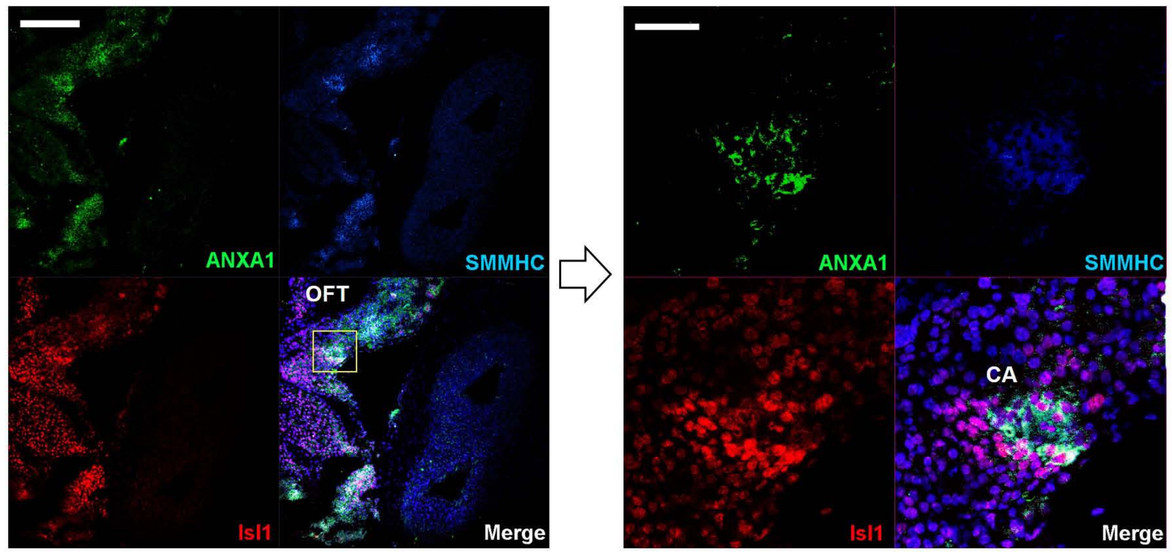


C



D

[Human embryonic heart at 6.5 week]



E

[Human embryonic heart at 6.5 week]

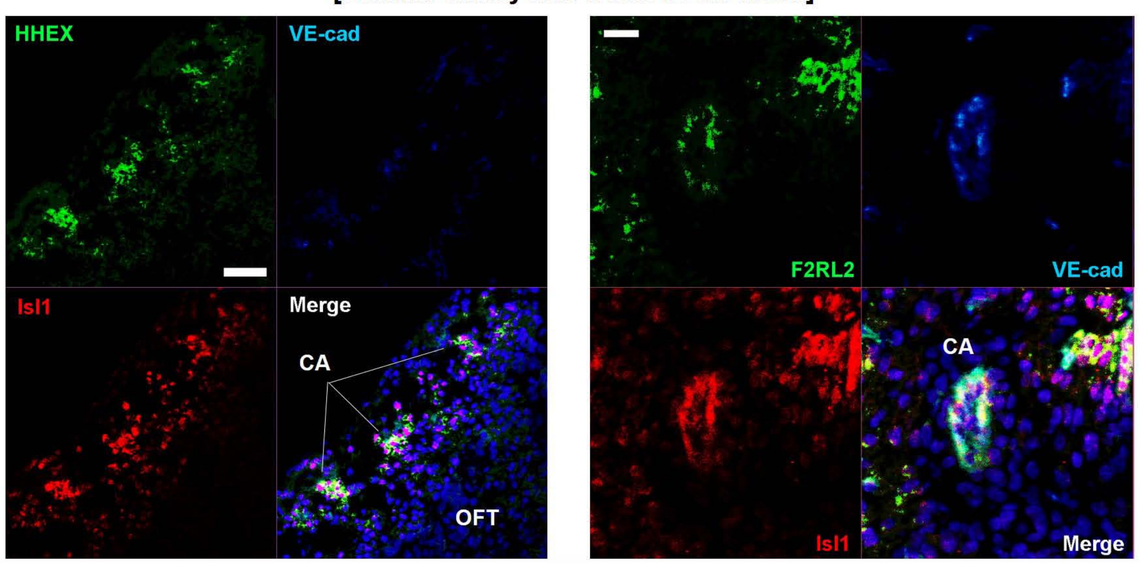
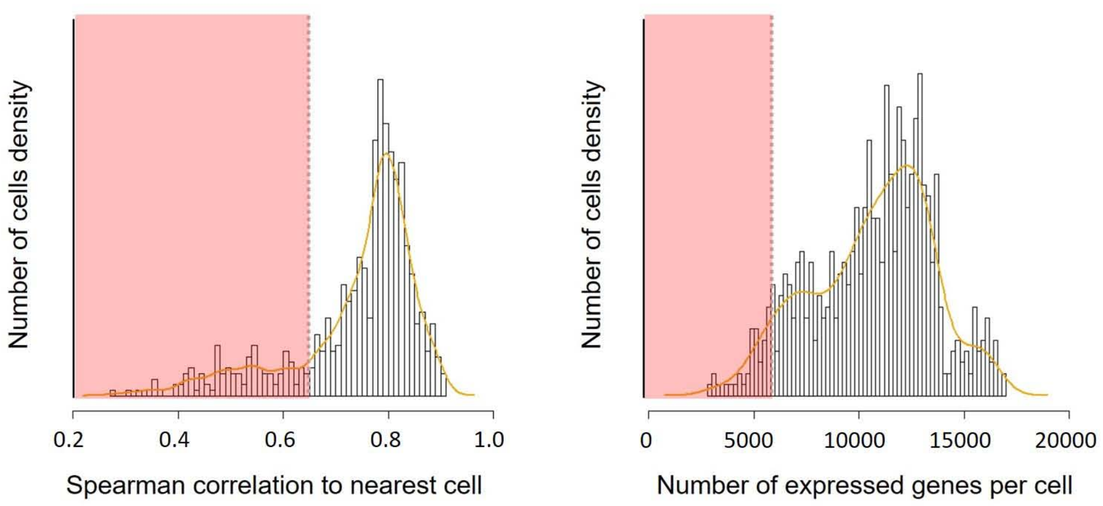


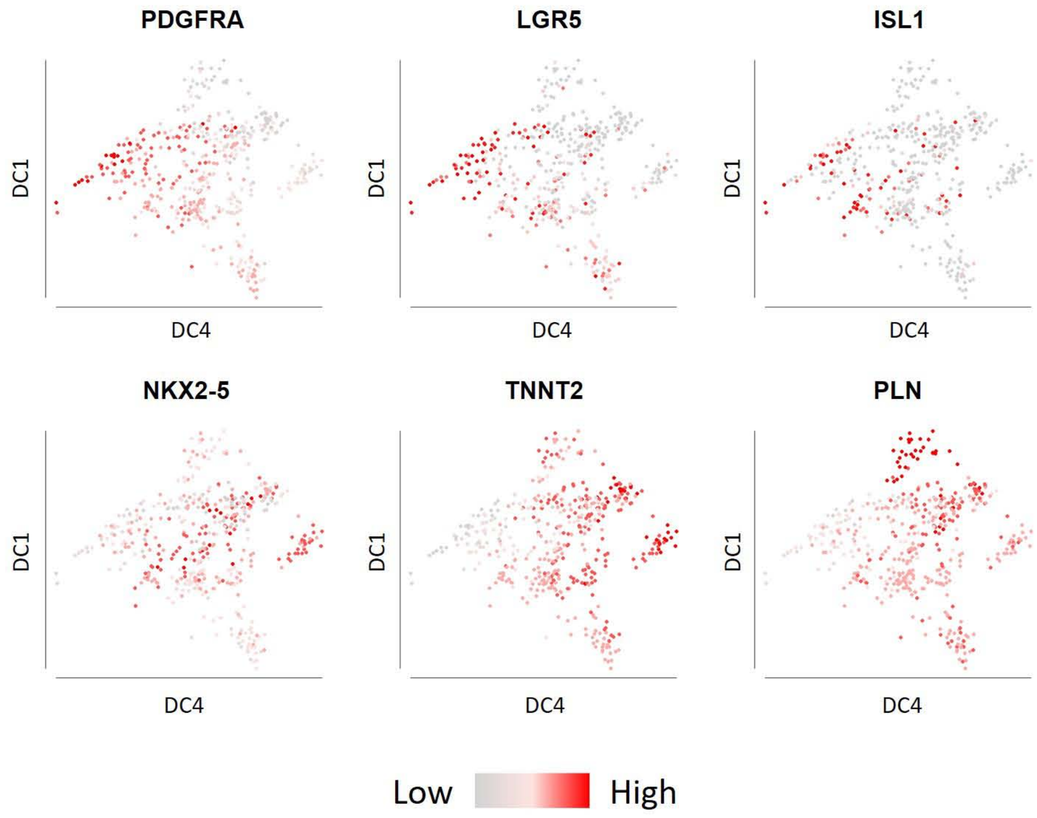
Figure S2. Unique Markers Specific to Each Cardiac Intermediate Population. Related to Figure 1.

To validate the results of the population RNA-seq data, we tested the authentic protein expression of several identified lineage-specific genes that were previously less investigated in cardiac development and differentiation, using sectioned specimens of human embryonic/fetal hearts and human ESC-derived cardiac cells. (A) Immunostaining of the human middle-staged embryonic heart (at 6.5 week) revealed that $ISL1^{+}TNNT2^{+}$ cardiomyocyte intermediates (CIs) in outflow tract (OFT) co-expressed ASB2 (Ankyrin Repeat and SOCS Box Containing 2) (*arrows, right*), enriched in the CI lineage of the population RNA-seq data. The right image is the enlarged one of a white square in the left image. Scale bars, 50 μm (*left*) and 20 μm (*right*). Vent, ventricle. (B) Immunostaining of the human early- and late-staged embryonic hearts (at 4.5 [*left*] and 10 [*right*] week) revealed that KCNH7 (Potassium Voltage-Gated Channel Subfamily H Member 7), enriched in the CM lineage of the population RNA-seq data, was weakly expressed in ventricular cardiac progenitors ($NKX2\text{-}5^{+}MF20^{+}$; *yellow arrow, left*) and myocytes ($NKX2\text{-}5^{-}MF20^{+}$; *white arrow, left*) in the early-staged heart (4.5 week), whereas it was strongly expressed in late-staged cardiac myocytes ($NKX2\text{-}5^{-}MF20^{+}$; *white arrows, right*). Scale bars, 50 μm. (C) Flow cytometry analysis showed that expression of CACNA1D (Calcium Voltage-Gated Channel Subunit Alpha1 D), enriched in the pacemaker intermediate (PI)/pacemaker cell (PM) lineages of the population RNA-seq data, was confirmed on the $HCN4^{+}$ PIs on day 6 in cardiac differentiation (*left*). On day 10, $HCN4^{+}$ cells retain the expression of CACNA1D (*right*), suggesting that it is strongly associated with the PM lineage. (D) Immunostaining of the human middle-staged embryonic heart (at 6.5 week) revealed that $ISL1^{+}SMMHC^{+}$ smooth muscle intermediates (SMIs) in the coronary vessels surrounding OFT co-expressed ANXA1 (Annexin A1), enriched in the SMI lineage of the population RNA-seq data. The right image is the enlarged one of a white square in the left image. Scale bars, 200 μm (*left*) and 50 μm (*right*). CA, coronary artery. (E) Immunostaining of the human middle-staged embryonic heart (at 6.5 week) revealed that $ISL1^{+}$VE-cadherin $(VE\text{-}cad)^{+}$ endothelial intermediates (EIs) in the coronary vessels surrounding OFT co-expressed HHEX (Hematopoietically-expressed homeobox protein, *left*) and F2RL2 (Coagulation Factor II Thrombin Receptor Like 2, *right*), enriched in the EI/EC lineages of the population RNA-seq data. Scale bars, 50 μm (*left*) and 20 μm (*right*).

A



B

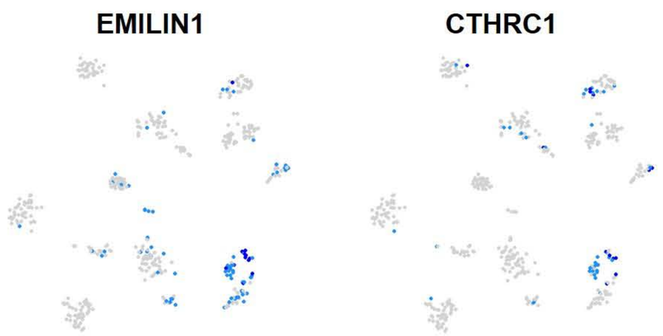


C

CVP-enriched

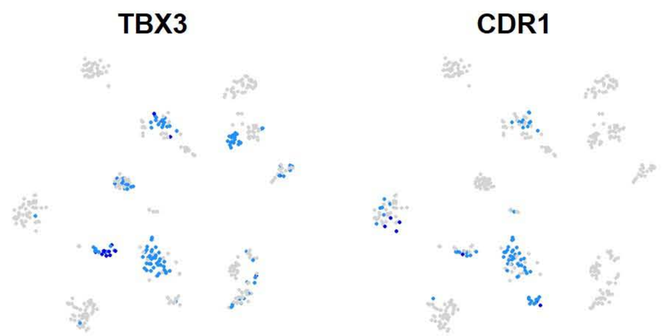
D

FVP-enriched



E

IM-OFT-enriched

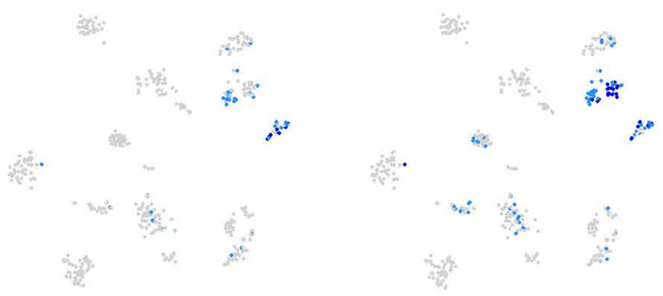


F

IM-Vent-enriched

G

Early atria/PM-enriched

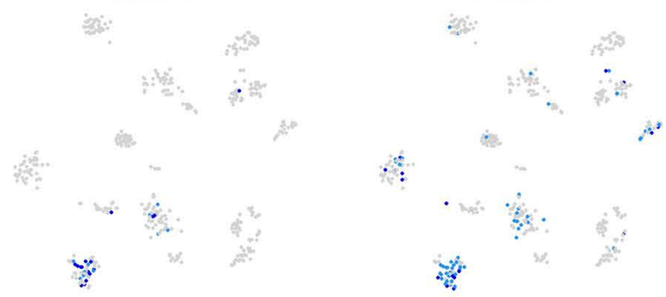


H

Late Atria-enriched

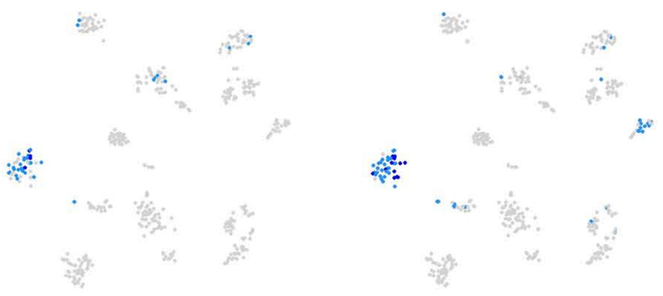
I

CME-enriched



J

LCMC-enriched

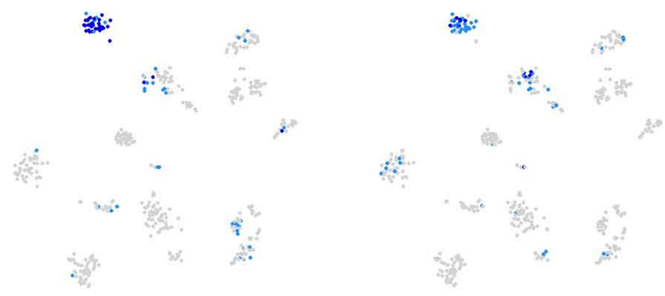


K

CVM (Late OFT)-enriched

L

FVM (Late Vent)-enriched

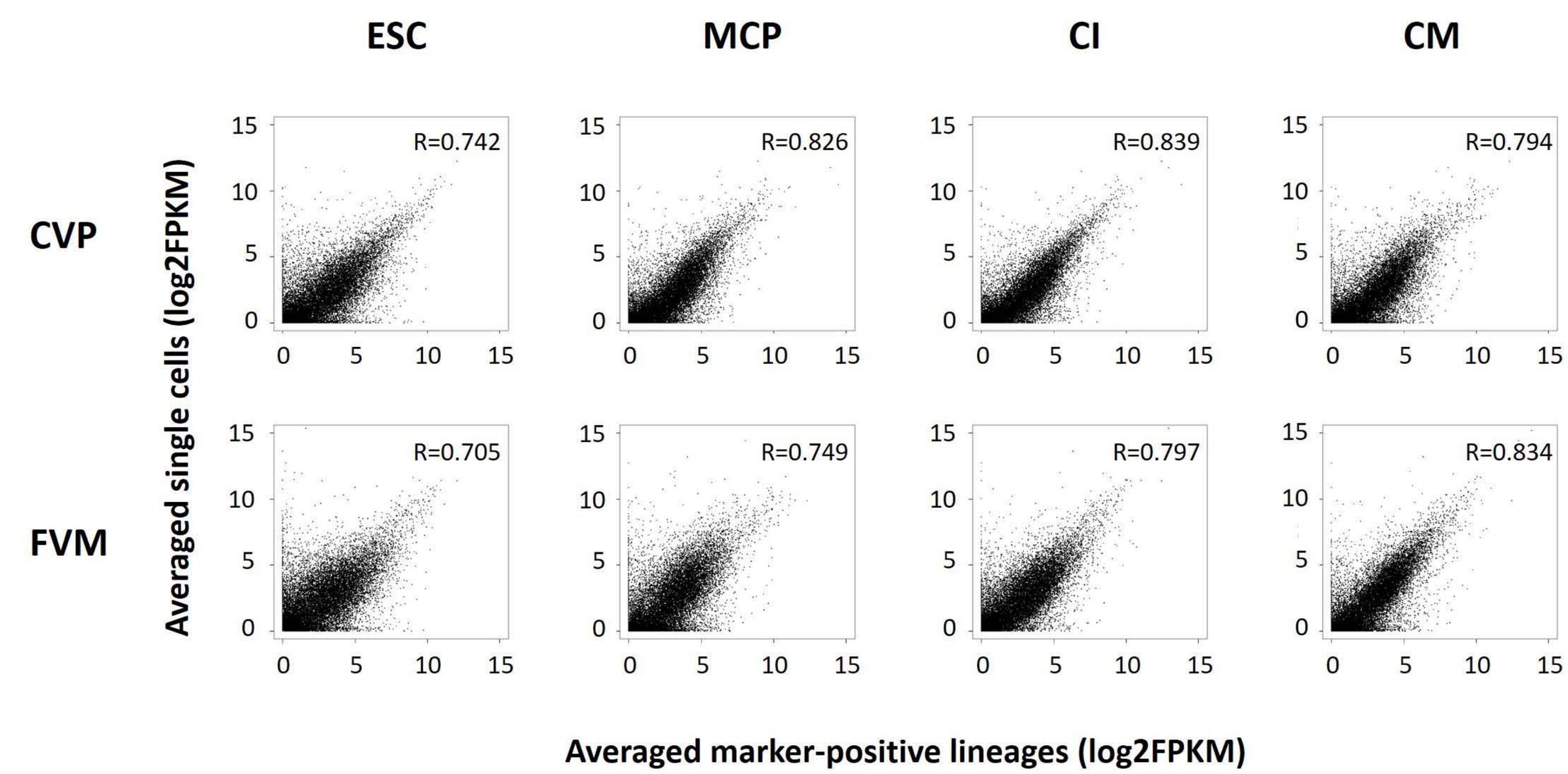


Low High

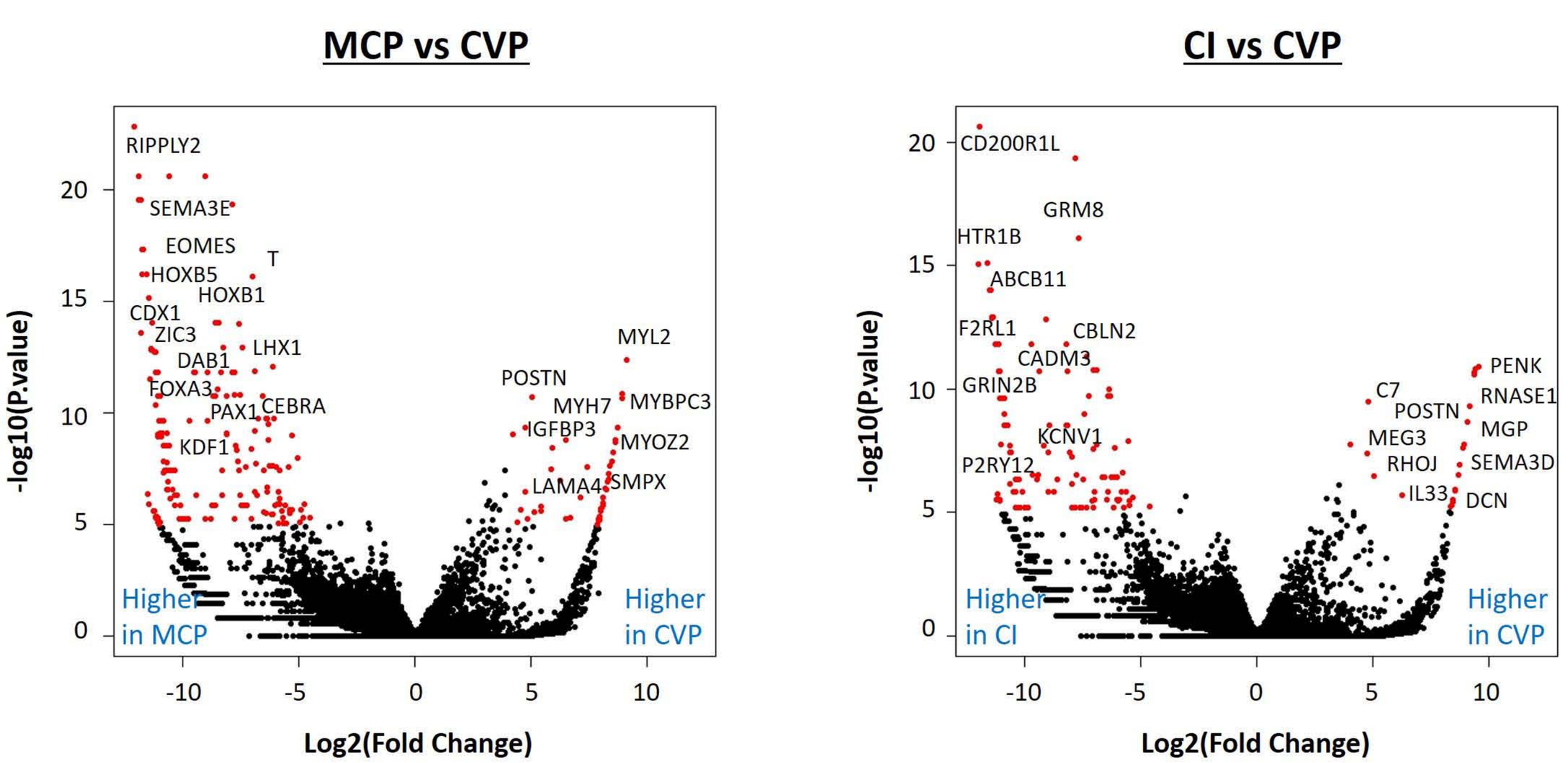
Figure S3. *In Vivo* Single-cell RNA-seq Identified the Stage/Compartment-specific Subpopulation-enriched Genes. Related to Figures 4 and 5.

(A) Quality Control of the Human Embryonic/Fetal Heart Single-cell RNA-Seq. Left: quality of *in vivo* single-cell RNA-seq experiments assessed as nearest-neighbor similarities between cells (maximum Spearman correlation per cell, using all cell-pairs and all genes). Right: histogram of the number of expressed genes per cell. Genes with FPKM ≥ 1were considered expressed (Petropoulos et al., 2016). The red color ranges in both panels involve trimmed/filtered individual cells, due to their low-quality transcriptomes. (B) The selected cardiac genes' expression distribution is displayed on diffusion map dimensionality reduction plot of the 7-embryonic/fetal heart-derived 458 high-quality single-cell transcriptomes (Figure 4B, right). Each dot is a single cell, and cells are colored based on the gene expression level. *PDGFRA*, *LGR5*, and *ISL1* were dominantly expressed in early-staged heart cells, whereas *TNNT2* and *PLN* were strongly expressed in late-staged heart cells. (C-L) Unbiased clustering of the 458 individual heart cells using the Seurat program (Macosko et al., 2015) revealed the 10 molecularly distinct clusters (Figure 4C). Differential gene expression analysis by the Seurat and edgeR programs identified differential expression genes on the 10 clusters. Representative marker genes enriched in each of the 10 clusters including CVP (C), FVP (D), IM-OFT (E), IM-Vent (F), Early atria/PM (G), Late atria (H), CME (I), LCMC (J), CVM (= Late OFT) (K), and FVM (= Late Vent) (L) are displayed. Each dot is a single cell, and cells are colored based on the gene expression level. All abbreviations are the same in Figure 4C (*also see* Figure 4D).

A


ESC
MCP
CI
CM
CVP
FVM
R=0.742
R=0.826
R=0.839
R=0.794
R=0.705
R=0.749
R=0.797
R=0.834
Averaged single cells (log2FPKM)
Averaged marker-positive lineages (log2FPKM)

B


MCP vs CVP
CI vs CVP
RIPPLY2
SEMA3E
EOMES
T
HOXB5
HOXB1
CDX1
ZIC3
LHX1
DAB1
FOXA3
PAX1
CEBRA
KDF1
MYL2
POSTN
MYH7
MYBPC3
IGFBP3
MYOZ2
LAMA4
SMPX
CD200R1L
GRM8
HTR1B
ABCB11
F2RL1
CBLN2
CADM3
GRIN2B
KCNV1
P2RY12
C7
POSTN
PENK
RNASE1
MGP
MEG3
RHOJ
SEMA3D
IL33
DCN
-log10(P.value)
Log2(Fold Change)
Higher in MCP
Higher in CVP
Higher in CI
Higher in CVP

C

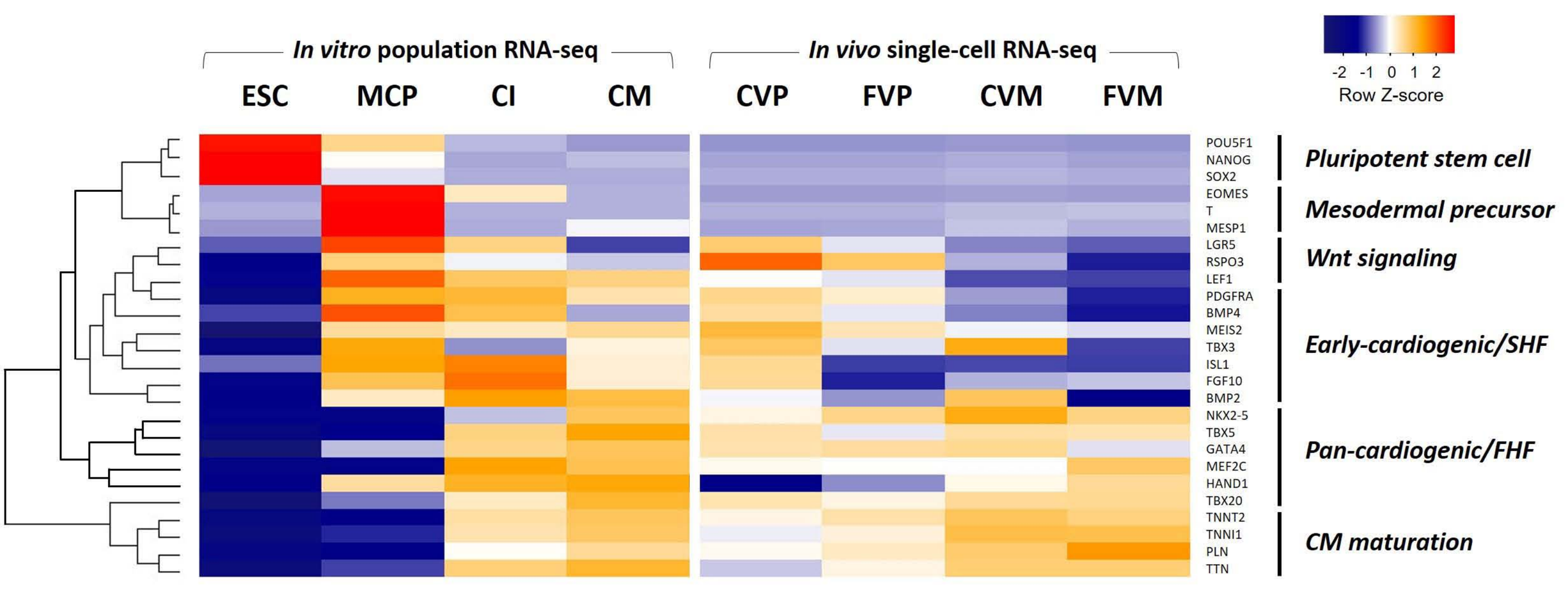

In vitro population RNA-seq
In vivo single-cell RNA-seq
ESC
MCP
CI
CM
CVP
FVP
CVM
FVM
Row Z-score
POU5F1
NANOG
SOX2
EOMES
T
MESP1
LGR5
RSPO3
LEF1
PDGFRA
BMP4
MEIS2
TBX3
ISL1
FGF10
BMP2
NKX2-5
TBX5
GATA4
MEF2C
HAND1
TBX20
TNNT2
TNNI1
PLN
TTN
Pluripotent stem cell
Mesodermal precursor
Wnt signaling
Early-cardiogenic/SHF
Pan-cardiogenic/FHF
CM maturation

Figure S4. The Similarities and Differences in Gene Expression Patterns between the *In Vitro* MCP/CI and the *In Vivo* CVP. Related to Figures 1, 4, and 5.

(A) Correlation between the *in vitro* population RNA-seq data of human ESC-derived cardiac lineages (ESC/MCP/CI/CM) and the *in vivo* single-cell RNA-seq data of human embryonic/fetal heart cells (CVP, *top*; FVM, *bottom*). For each gene in each comparison, the log of the averaged FPKM in each population/cluster was plotted. Each Pearson's correlation coefficient was then calculated and indicated on the top of the graph. (B) Volcano plots visualizing differential gene expression analysis between the *in vitro* MCP and the *in vivo* CVP (*left*), and between the *in vitro* CI and the *in vivo* CVP (*right*). For each gene, the average difference (log2[Fold Change]) between cells from each comparison was plotted against the power to discriminate between groups (-log10[P.value]). Top-scoring genes for both metrics that are indicated as red dots were further processed by Gene Ontology (GO) gene set enrichment and KEGG pathway analyses. Representative differential expression genes' names are labeled. (C) A heatmap image of cardiogenic genes' relative expression levels between the *in vitro* ESC/MCP/CI/CM and the *in vivo* CVP/FVP/CVM/FVM. Normalized FPKM levels of each gene (*row*) are shown after Z-score normalization. Among cardiogenic genes, enhanced expression of some of the early cardiogenic/second heart field-related genes (e.g., *PDGFRA, BMP2/4, MEIS2, ISL1, FGF10*) as well as Wnt signaling (e.g., *LGR5, RSPO3,* and *LEF1*) was detected in both MCPs and CVPs. On the other hand, expression of the pan-cardiac and first heart field-related genes such as *NKX2-5, TBX5, GATA4, MEF2C, TBX20, and TNNT2*, all of which the *in vitro* CI/CM and the *in vivo* CVM/FVM expressed highly, were also detected mildly in CVPs, but little or no expression of those genes detected in MCPs.

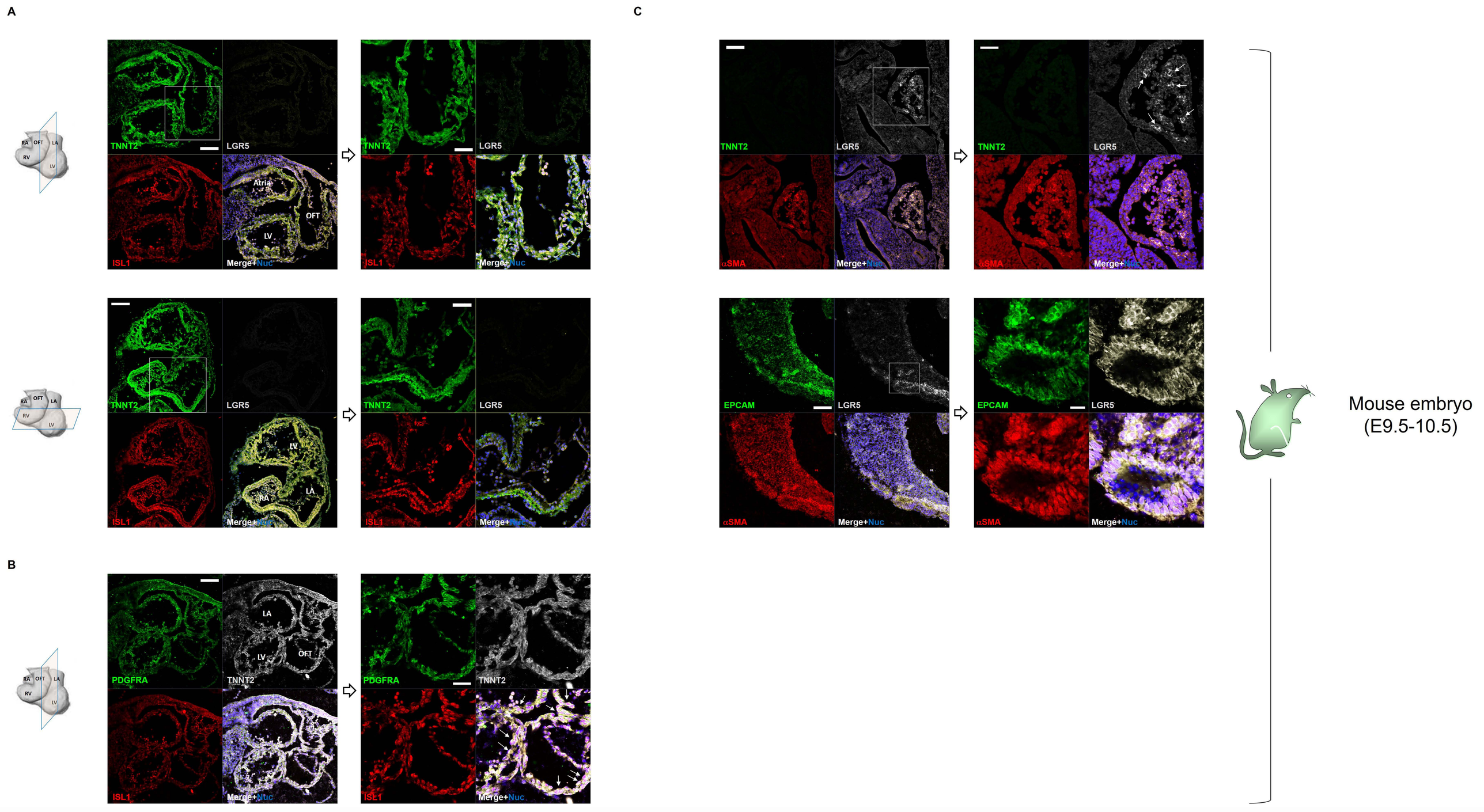

A
B
C
RA
OFT
LA
RV
LV
TNNT2
LGR5
ISL1
Merge+Nuc
Atria
OFT
LV
RA
LA
PDGFRA
TNNT2
αSMA
EPCAM
Mouse embryo
(E9.5-10.5)

Figure S5. LGR5$^+$ Cells Emerge in the Embryonic Guts but Do not Appear in the Developing Heart in Murine Embryos. Related to Figure 6.

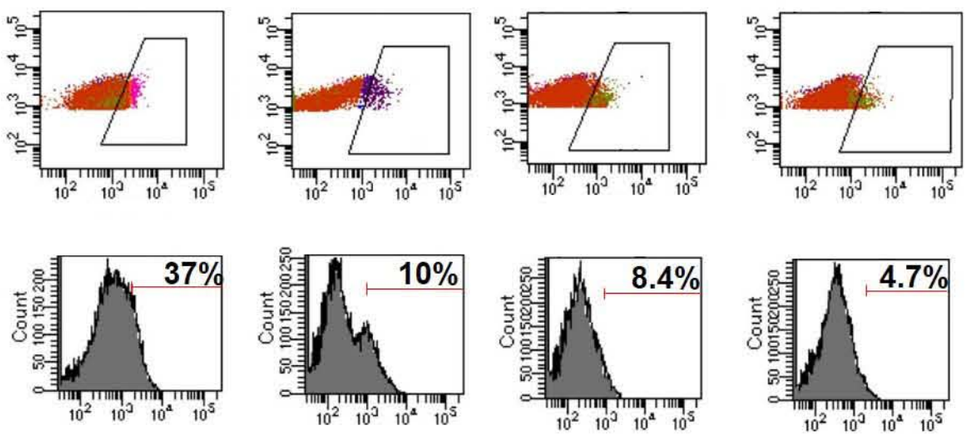
Immunostaining of the mouse embryos at embryonic day 9.5-10.5 (E9.5-E10.5) was performed with various combinations of the primary antibodies. All images were obtained using a Zeiss 710 confocal microscope and its imaging system. (A) In contrast to the human embryonic heart (Figure 6), little or no LGR5$^+$ cells were found in the murine embryonic hearts including OFT on the similar developing stage (E9.5-E10.5). Top, sagittal view; bottom, transverse view. LA, left atria; RA, right atria. (B) The PDGFRA$^+$ISL1$^+$ cells were, however, readily detected in the murine embryonic hearts including OFT, similarly to the human embryonic heart (Figure 6G). Scale bars, 100 μm (*left*) and 50 μm (*right*). (C) Unlike in the case of the murine developing hearts, LGR5$^+$ cells emerged in the murine primitive guts including the primitive colon (*top*) and small intestine (EPCAM$^+$; *bottom*). Scale bars, 100 μm (*left*), 50 μm (*right, top*) and 20 μm (*right, bottom*). The right images in (A-C) are the enlarged ones of white squares in the left images, respectively.

A

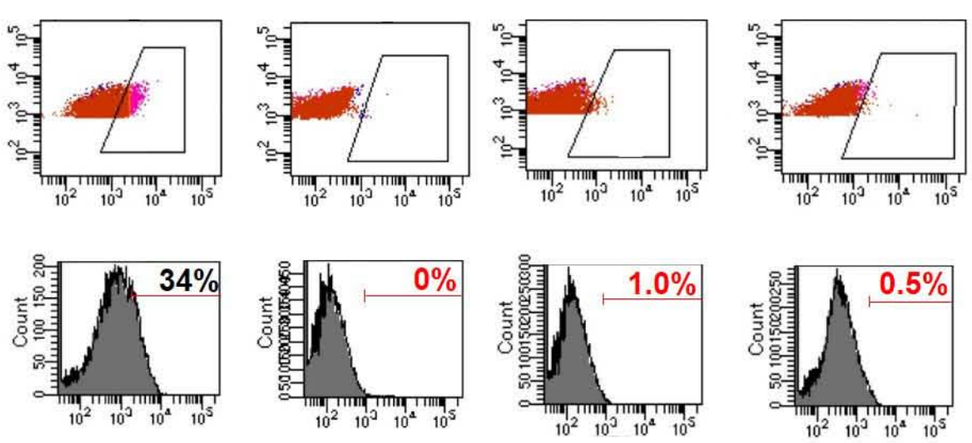
Day 3

MESP1 LGR5 LEF1 ISL1

WT



LGR5 -KO

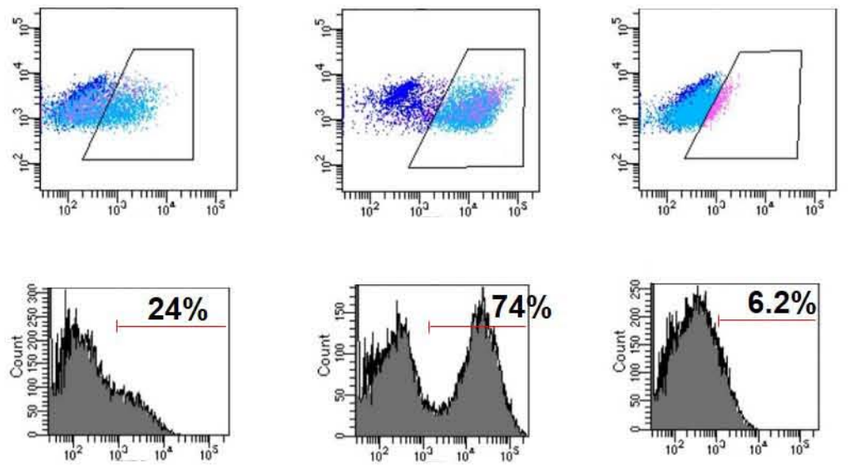


B

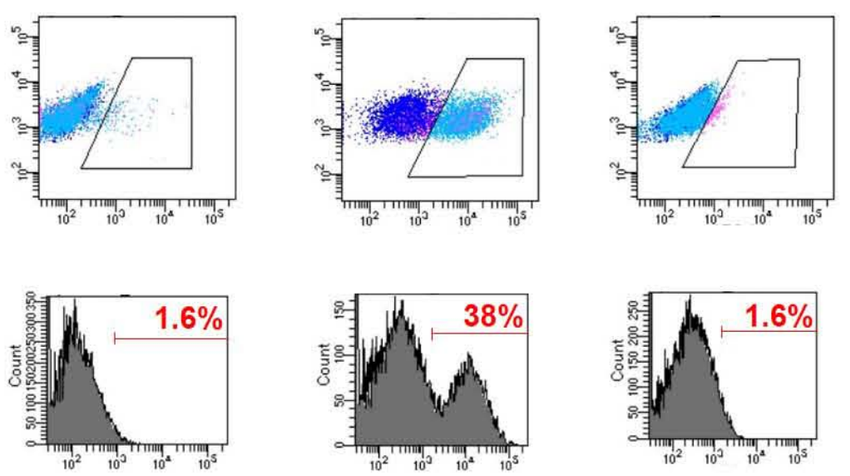
Day 6

TNNT2 ISL1 LEF1

WT



LGR5 -KO

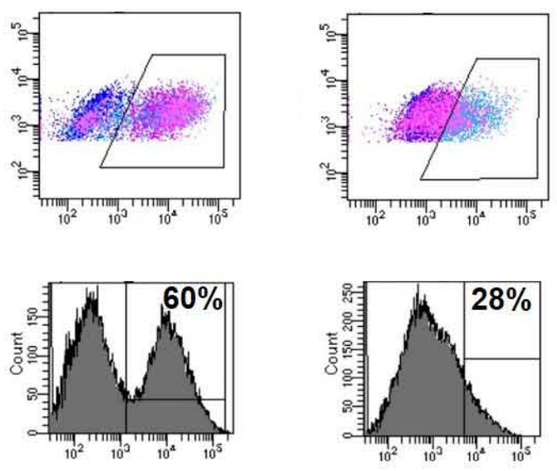


C

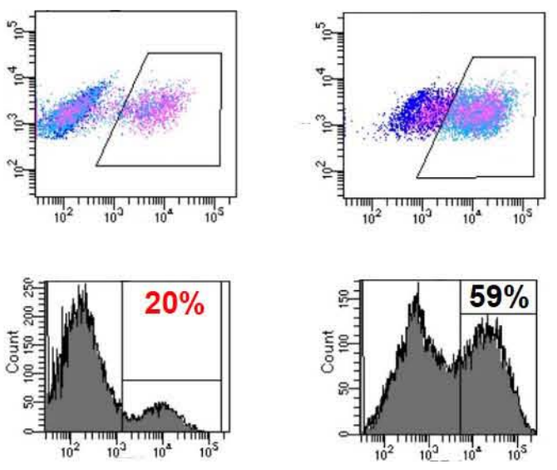
Day 10

TNNT2 ISL1

WT



LGR5 -KO

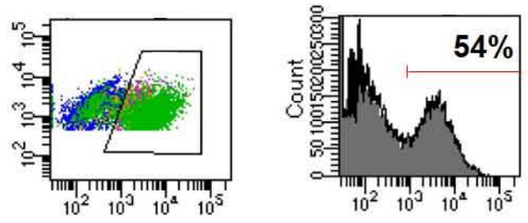


D

Day 18

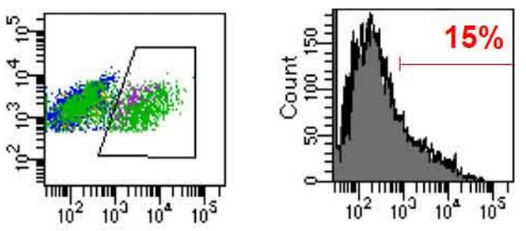
TNNT2

WT



LGR5 -KO

TNNT2

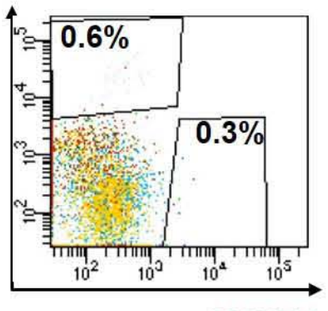


E

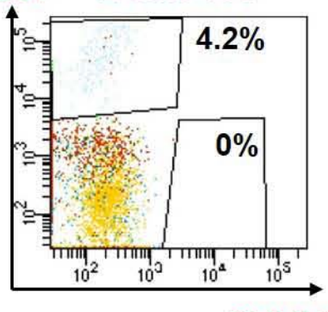
Day 14

WT

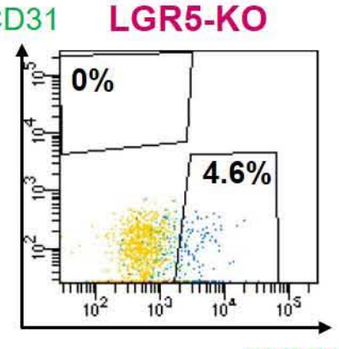
CD31



CD31 LGR5-KO



CD31 LGR5-KO



F

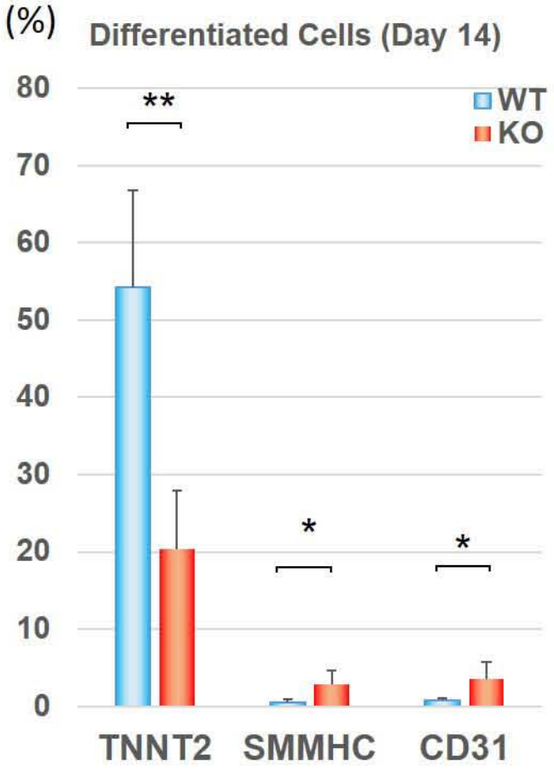
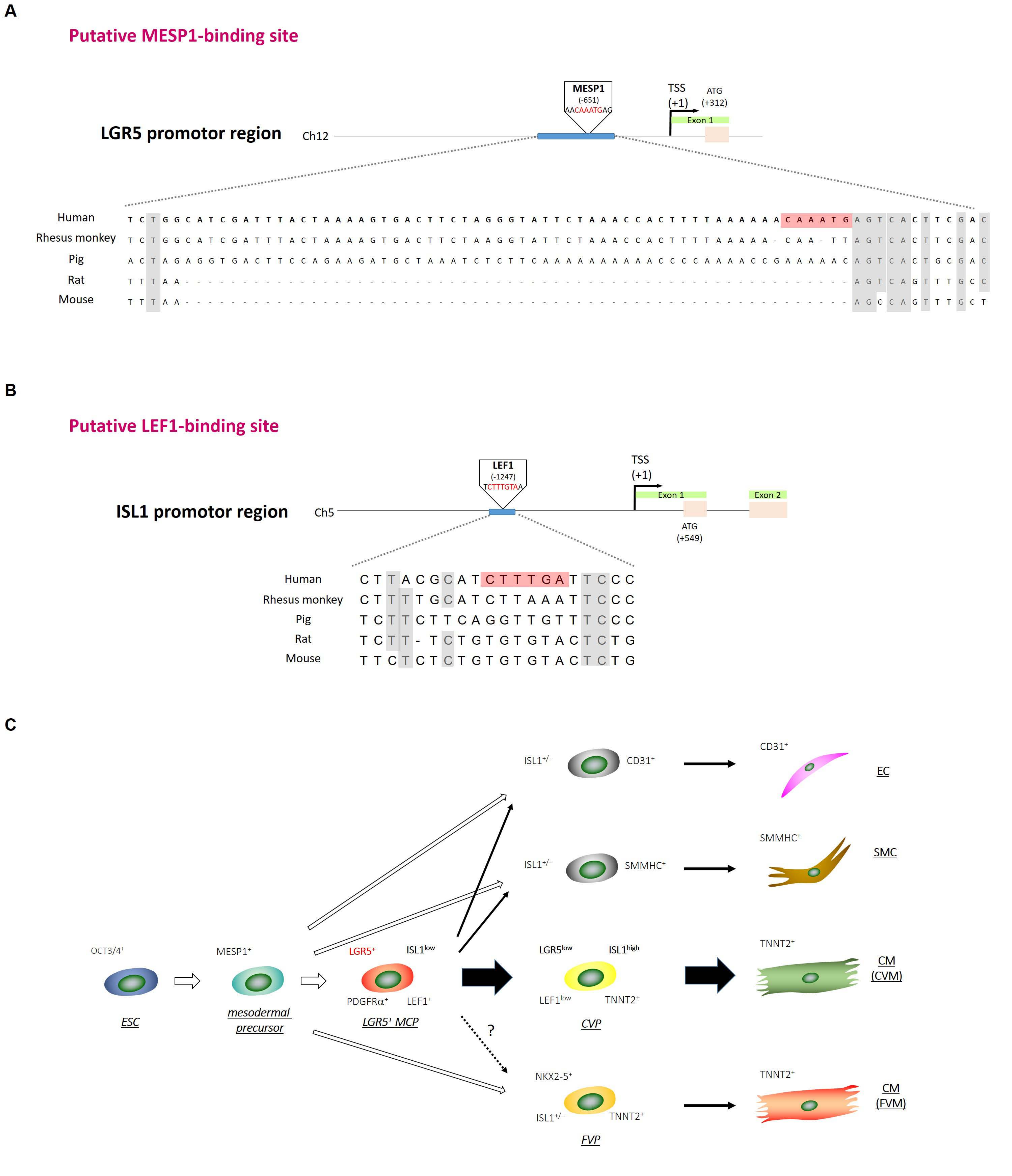


Figure S6. Human LGR5-KO ESCs Showed the Impaired Cardiogenic Capabilities. Related to Figure 7.

Human wild-type (WT) and LGR5-KO ESCs underwent cardiac differentiation simultaneously using the same protocol in Figure 1A. (A) On day 3, MESP1 expression was not affected by LGR5 deletion, whereas expression of not only LGR5 but also LEF1 and ISL1 was significantly attenuated in the LGR5-KO than in WT ESC-derived cells (*also see* Figure 7D). (B) On day 6, the degrees of LEF1 and ISL1 expression were still lowered in the LGR5-KO ESC-derived cells. Similarly, TNNT2 expression was significantly attenuated in the LGR5-KO ESC-derived cells (*also see* Figure 7D). (C and D) The lowered degree of CM induction in the LGR5-KO ESC-derived cells was not ameliorated on day 10 (C) or day 18 (D), although expression of ISL1 was rather increased in the LGR5-KO ESC-derived cells on day 10 (C), suggesting the delayed induction of ISL1, likely due to LGR5 deletion (*also see* Figure 7D). (E and F) In accordance with low induction of $ISL1^{+}TNNT2^{+}$ intermediates on day 6 (Figure 7E), the percentage and cell number of CMs ($TNNT2^{+}$) on day 14 were significantly decreased in the LGR5-KO than WT ESC-derived cells, while those of SMCs ($SMMHC^{+}$) and ECs ($CD31^{+}$) were rather increased in the LGR5-KO than WT ESC-derived cells, suggesting that LGR5 deletion may shift the mesodermal precursors' differentiation towards driving the vascular lineages. Interestingly, the increased production of SMCs and ECs on day 14 in the LGR5-KO ESC-derived cells did not occur simultaneously, but either one was induced in a somewhat stochastic manner (E, *right*).



A
Putative MESP1-binding site
MESP1
(-651)
AACAAATGAG
TSS
(+1)
ATG
(+312)
Exon 1
LGR5 promotor region
Ch12
Human TCTGGCATCGATTTACTAAAAGTGACTTCTAGGGTATTCTAAACCACTTTTAAAAAACAAATGAGTCACTTCGAC
Rhesus monkey TCTGGCATCGATTTACTAAAAGTGACTTCTAAGGTATTCTAAACCACTTTTAAAAA-CAA-TTAGTCACTTCGAC
Pig ACTAGAGGTGACTTCCAGAAGATGCTAAATCTCTTCAAAAAAAAAAACCCCAAAACCGAAAAACAGTCACTGCGAC
Rat TTTAA---AGTCAGTTGCC
Mouse TTTAA---AGCCAGTTGCT
B
Putative LEF1-binding site
LEF1
(-1247)
TCTTTGTAA
TSS
(+1)
Exon 1
Exon 2
ATG
(+549)
ISL1 promotor region
Ch5
Human CTTACGCATCTTTGATTCCC
Rhesus monkey CTTTTGCATCTTAAATTCCC
Pig TCTTCTTCAGGTTGTTTCCC
Rat TCTT-TCTGTGTGTACTCTG
Mouse TTCTCTCTGTGTGTACTCTG
C
ISL1+/–
CD31+
CD31+
EC
ISL1+/–
SMMHC+
SMMHC+
SMC
OCT3/4+
MESP1+
LGR5+
ISL1low
LGR5low
ISL1high
TNNT2+
CM
(CVM)
PDGFRα+
LEF1+
LEF1low
TNNT2+
ESC
mesodermal precursor
LGR5+ MCP
CVP
?
NKX2-5+
TNNT2+
CM
(FVM)
ISL1+/–
TNNT2+
FVP

Figure S7. Human-specific Binding Sites of MESP1 to LGR5, and LEF1 to ISL1 Promoter Regions and Developmental Cellular Path in Human LGR5-related Cardiogenesis. Related to Figure 7.

(A) Through a bioinformatics survey using MatInspector software (Genomatix, Germany; http://www.genomatix.de/), a novel MESP1 binding site with a consensus motif "-CAVNTG-" (Soibam et al., 2015) was identified on the *LGR5* promoter region. This is 651 bp upstream of the transcription start site (TSS) of LGR5 gene locus. We found that this novel binding-site is human-specific (indicated by a red shade), and other mammalian species such as rhesus monkey, pig, and rodents do not have the same sequence on the locus. (B) With a similar approach in (A), we found a novel LEF1 binding site with a consensus motif "-YCTTTGWW-" (Hovanes et al., 2001) on the *ISL1* promoter. This is 1247 bp upstream of the TSS on *ISL1* gene locus. This novel binding-site is also human-specific (indicated by a red shade), and other mammalian species do not have the same sequence on the locus. The common sequences among these mammalian species are indicated by grey shades. (C) Schema of the human LGR5-related cardiogenic cellular pathway. LGR5 would be essential for the induction of cardiomyocytes (CMs), especially cono-ventricular muscle cells (CVMs), by driving the $MESP1^{+}$ mesodermal precursor-derived $ISL1^{+}$ cono-ventricular progenitors (CVPs) to differentiate into the CVM lineage. EC, endothelial cell; ESC, embryonic stem cell; FVM, free wall-ventricular muscle cell; FVP, free wall-ventricular progenitor; MCP, multipotent cardiac progenitor; SMC, smooth muscle cell.

Table S1. List of human ESC-derived cardiac lineage-specific genes. Related to Figure 1.

MCP		CI		CM		PI		PM	
Enriched Gene	**Averaged Z-score of log2FPKM**	**Enriched Gene**	**Averaged Z-score of log2FPKM**	**Enriched Gene**	**Averaged Z-score of log2FPKM**	**Enriched Gene**	**Averaged Z-score of log2FPKM**	**Enriched Gene**	**Averaged Z-score of log2FPKM**
CCR7	3.05799	SYTL2	2.21472	TRIM63	3.00847	SPINK4	1.78319	HNF4A-AS1	2.53842
SLC8A3	2.89138	LOC101929210	2.12823	CCDC60	2.92683	ADAMTS18	1.64012	ARHGAP26-IT1	2.37042
TMEM190	2.71795	ACTA1	2.07734	C1orf186	2.89631	SNAR-C4	1.63300	VWA5B1	2.29223
TNFSF9	2.65151	GRIK1	1.84704	MB	2.87838	C8orf31	1.57248	FTCD	2.28365
HOXB1	2.59352	CMTM5	1.63756	TCAP	2.82420	C16orf54	1.49209	ACSM1	2.10213
ZAP70	2.56053	DGKI	1.60111	KLHL41	2.78493	SCARNA5	1.46619	AWAT1	2.07851
ELN	2.49926	LPPR4	1.57147	HSPB2	2.76136	VIL1	1.44931	COL14A1	2.07174
SAMD3	2.27586	CBLN2	1.55688	CRYAB	2.73000	COL19A1	1.43186	KCNMA1-AS3	2.02318
KCNK17	2.21849	REEP1	1.54519	ODAM	2.69923	KIF12	1.40070	LACAT8	2.01694
TMEM132C	2.18255	RGS4	1.53620	MYH7	2.65897	ADAM28	1.32118	OR52H1	2.00352
PTCHD2	2.17575	NPPB	1.51866	RRAD	2.63445	PTCHD1	1.29818	FUT3	1.95843
MSX1	2.14768	NKX2-6	1.50743	AHNAK2	2.61172	SCN7A	1.27911	HCN4	1.94969
SYN3	2.13385	STAC	1.50450	TMEM176A	2.59807	EEF1DP3	1.27840	FXYD7	1.90093
MIR4269	2.11024	SCARNA5	1.49411	LRRC10	2.59068	STAC	1.25936	ABCA8	1.88140
GLRA4	2.10312	MPPED2	1.48296	C10orf71	2.57547	HPX	1.24781	HIST3H3	1.85851
TBX6	2.10296	SH3BGR	1.48264	EEF1A2	2.56969	CCKBR	1.22375	CPNE5	1.84643
MESP1	2.07297	SLC16A12	1.47762	IRX4	2.56644	PALM2	1.16354	COL6A6	1.84523
WDR93	2.07277	SYNPO2L	1.47411	LAMA2	2.55715	FBLN7	1.15673	AHSG	1.84362
KCP	2.01958	NXPH2	1.46024	CKM	2.55030	CGN	1.15268	SPEG	1.84294
LINC00922	2.01815	FAM19A4	1.44697	COX6A2	2.54157	MPPED2	1.14942	TRH	1.84151
PMP22	1.94833	ACTC1	1.44389	RCAN2	2.53874	CACNA1D	1.14266	FAM13C	1.80220
IRX3	1.88114	SMPX	1.43433	ABRA	2.52813	ST8SIA2	1.13491	NCCRP1	1.79615
ANPEP	1.87838	ANXA3	1.41802	SRL	2.51773	SLCO2A1	1.13237	ADAM11	1.78298
DLL3	1.87590	ACTA2	1.40955	TNNC2	2.51666	LPPR4	1.12663	MYH7B	1.77727
NOS2	1.87445	ENOX1	1.40579	VCAM1	2.51559	GPR87	1.12633	C22orf15	1.74872
LOC100506178	1.86863	TNNC1	1.40207	TMEM176B	2.51226	SYTL5	1.12387	TSPAN32	1.74856
FAM89A	1.86049	UNC45B	1.39381	KLHL14	2.51134	METTL24	1.11650	TBX18	1.74256
ADRB2	1.84762	MYOCD	1.38745	PPP1R14C	2.50131	PPFIBP2	1.11100	CPNE7	1.73153
SNAI1	1.80674	CALB2	1.38454	PLCXD3	2.45431	C1orf168	1.10757	DPEP1	1.71683
LGR5	1.80272	BMP2	1.36435	POPDC3	2.44974	CBLN2	1.10341	C8orf49	1.71209
LOC101060019	1.77819	C3orf52	1.35277	PRSS42	2.42965	DNAJC15	1.08025	SMIM1	1.70757
ZIC3	1.76290	ASB2	1.32605	HRASLS	2.42646	HNF1B	1.03455	ADGRD1	1.70730
HAS2	1.72815	PALM2	1.31070	PPP1R3C	2.41547	BMPER	1.03335	KCNJ3	1.70583
LEF1	1.70021	SMYD1	1.29587	ADPRHL1	2.41508	MIR4292	1.02648	OR51I1	1.70473
CRIP3	1.69761	ANKRD1	1.28473	TMEM71	2.41303	SNORD42B	1.02422	C1orf95	1.69131

MLXIPL	1.69197	JPH2	1.26115	FITM1	2.40347	SVEP1	1.02320	MTUS2	1.69027
LHX1	1.68724	LOC101927497	1.24562	CSRP3	2.38542	RARB	1.01873	FGF16	1.68646
T	1.68615	RCSD1	1.24192	LAMA4	2.38444	SYNDIG1	1.00256	FER1L4	1.68453
LOC401312	1.66867	RBM24	1.23397	FSD2	2.34489	GATA3	0.99728	MYOM2	1.67528
BAALC	1.65976	PPARGC1A	1.23378	NPPA	2.33052	AMELX	0.99431	IFIT5	1.67028
GREB1L	1.64522	MEF2C	1.23307	BVES	2.32710	SH3RF2	0.98921	GIPR	1.66829
LOC100507073	1.64115	GDF3	1.22880	KCNH7	2.31535	PLXNA4	0.98758	MYO18B	1.66173
PCDH19	1.62207	DNAJC15	1.22754	LOC101927179	2.31420	FRY	0.97812	PRSS42	1.65026
HOTAIRM1	1.61496	IGFBP7	1.22374	HTR1B	2.29799	COL9A2	0.97410	RPL19P12	1.64543
RASGRP3	1.61389	LOC101928782	1.22241	RXRG	2.28188	TNC	0.95110	FGB	1.64051
CNTFR	1.60585	LYPD6B	1.21195	LRRC39	2.25748	CLSTN2	0.95006	HIST1H3E	1.63471
VWC2L-IT1	1.59490	LOC101928540	1.21162	TMEM38A	2.25508	LRP2	0.94478	RBP4	1.63147
MIXL1	1.58616	MYLK3	1.20390	C1orf95	2.25410	GRHL2	0.94477	SFTPD	1.62706
EOMES	1.58536	LOC101928418	1.19644	LINC00881	2.25360	CCDC183	0.94388	MYBPC3	1.61636
VRTN	1.57847	NKX3-1	1.18487	SYPL2	2.24524	PCDH7	0.93813	DNAJC5G	1.60833

SMI		SMC		EI		EC	
Enriched Gene	**Averaged Z-score of log2FPKM**	**Enriched Gene**	**Averaged Z-score of log2FPKM**	**Enriched Gene**	**Averaged Z-score of log2FPKM**	**Enriched Gene**	**Averaged Z-score of log2FPKM**
DLX5	2.95833	PAEP	2.66494	BPIFA1	2.21857	EMCN	3.31621
HPGD	2.77198	SERPINE3	2.55986	CST1	2.10774	PECAM1	3.31402
GRHL3	2.61447	ARSJ	2.54410	LINC00261	2.07069	KLHL6	3.31358
KRT7	2.44362	SERPINE1	2.50364	RIPPLY3	2.01306	CLEC1B	3.31342
UPK2	2.42145	CHI3L1	2.48453	IHH	1.99097	ROBO4	3.30498
GABRP	2.36238	ACTG2	2.44862	ITLN2	1.90885	RHOJ	3.29968
CLDN10	2.36173	INHBA	2.42915	CXCR4	1.79385	CD93	3.28823
TFAP2C	2.25188	S100A14	2.36601	CYP4X1	1.75520	LAPTM5	3.28494
SOX15	1.94917	BDKRB2	2.36482	COL19A1	1.73086	STAB1	3.28449
CLDN4	1.94752	NELL1	2.32572	HNF1B	1.71808	MMRN1	3.28159
LRRC3B	1.94297	PRSS35	2.26144	CGN	1.70052	LINC01594	3.28017
TRIML2	1.85739	POSTN	2.24934	SLC39A2	1.68130	SULT1C3	3.27975
DIO2	1.84824	EPHA3	2.21203	SLCO2A1	1.60871	C15orf26	3.27740
CRB3	1.73904	LINC01266	2.20017	RNU5F-1	1.59592	GPR4	3.27684
SYNDIG1	1.70072	FBLN5	2.20002	DKK1	1.58796	SHE	3.26887
GRHL2	1.68069	WNT2B	2.18688	APELA	1.56976	TFPI2	3.26807
RAB25	1.68045	ITGA11	2.16943	HHEX	1.56755	LCP2	3.26675
SHANK2-AS3	1.59428	SLC34A2	2.15047	TBX3	1.56563	PTPRB	3.26037
GPR87	1.58738	CEBPD	2.14067	VIL1	1.56476	ADCY4	3.25762
SLC7A3	1.51032	LCN1	2.11229	CAPN13	1.56163	C8orf4	3.25709
ARHGAP18	1.47786	BNC1	2.10104	PTCHD1	1.53680	EGFL7	3.25155

GATA3	1.47645	LMCD1	2.07869	SCN7A	1.53338	LOC105375734	3.24944
ANXA3	1.46933	LINC01300	2.06673	GATA3	1.51978	ITGA10	3.24599
ANXA1	1.45858	NPNT	2.06012	HABP2	1.50969	CXorf36	3.24473
SCNN1A	1.43012	FRZB	2.04618	HPX	1.50668	NT5E	3.23909
EPCAM	1.37934	AKR1B10	2.04198	KIF12	1.50044	SLC2A9	3.23901
HNF1B	1.36598	RUNX2	2.02942	CRYM	1.48719	ICAM2	3.23527
DIAPH3-AS2	1.34406	GALNT5	2.01470	SYTL5	1.43320	APLN	3.22824
EPAS1	1.34233	RBP4	2.00995	DLK1	1.40819	TDRD10	3.22611
CGN	1.31760	KLK6	1.98398	ACRV1	1.40284	TIE1	3.21905
PTCHD1	1.31644	PLAC1	1.98041	UGT3A2	1.39189	LOC101929517	3.21651
PLA2G16	1.31091	ARHGEF3	1.97343	LOC101929550	1.37174	GIMAP4	3.21018
APELA	1.30713	FGB	1.94277	CRB3	1.36116	GPR182	3.20903
MAL2	1.30285	GABRA2	1.93994	PTCHD1-AS	1.34722	LONRF3	3.20574
RARRES2	1.27740	SOX6	1.93827	MAL2	1.34242	SOX18	3.20368
ACTA1	1.27218	ARAP2	1.92916	CCKBR	1.34190	ESAM	3.19962
OR5P2	1.27115	GGT5	1.91756	ARHGAP18	1.30652	MMRN2	3.19874
PTCHD1-AS	1.26423	CNTNAP4	1.91309	EPCAM	1.30644	ACVRL1	3.18159
HABP2	1.24439	HBEGF	1.91275	NKX3-1	1.30217	THBD	3.18106
GLIPR2	1.24298	DCN	1.88981	EEF1DP3	1.26798	STARD8	3.17942
SPP1	1.23434	LOC100130264	1.88350	C1orf210	1.26466	GRAP	3.17909
AP1M2	1.23347	LRRN2	1.87789	FGFR4	1.26370	FLT4	3.17684
TUNAR	1.21634	DAAM2	1.87729	FRY	1.26175	SOST	3.17610
CLDN7	1.20713	DDR2	1.87432	TTR	1.25369	ARHGEF15	3.17506
VIL1	1.18250	RSPO1	1.87170	LYPD6B	1.23053	ERG	3.17353
EPHA1	1.17629	COBLL1	1.86553	RAB25	1.22966	SULT1C2P1	3.17149
MTUS1	1.16927	AFP	1.84118	GRHL2	1.21167	TNR	3.16999
LRRN1	1.15652	CAPN6	1.83108	PCDH17	1.19200	THSD1	3.16796
NXPH2	1.14018	DSC3	1.83077	CD34	1.18807	PLVAP	3.15005
MSX2	1.13527	AQP1	1.82451	F2RL2	1.18640	CDH5	3.14601

CI, cardiomyocyte intermediate; CM, cardiomyocyte; EC, endothelial cell; EI, endothelial intermediate; MCP, multipotent cardiac progenitor; PI, pacemaker progenitor; PM, pacemaker cell; SMC, smooth muscle cell; SMI, smooth muscle intermediate.

Table S2. Gene-set enriched KEGG pathway terms. Related to Figures 1 and 5.

Population	Gene-set enriched KEGG pathway terms	FDR
MCP	Protein digestion and absorption	3.5e-4
	Wnt signaling pathway	1.6e-2
CI/CM	Cardiac muscle contraction	9.0e-3
	cGMP-PKG signaling pathway	2.5e-2
	Adrenergic signaling in cardiomyocytes	4.7e-2
PI/PM	cAMP signaling pathway	9.9e-3
	Circadian entrainment	2.5e-2
	Cholinergic synapse	3.8e-2
	Serotonergic synapse	3.8e-2
SMI/SMC	Nicotine addiction	4.7e-3
	Tight junction	6.4e-3
	ECM-receptor interaction	6.4e-3
	Complement and Coagulation cascades	2.1e-2
	GABAergic synapse	3.6e-2
EI/EC	Cell adhesion molecules	1.1e-2
	Rap1 signaling pathway	1.5e-2
	Leukocyte transendothelial migration	2.4e-2
	PI3K-Akt signaling pathway	4.6e-2
CVP	Axon guidance	3.0e-4
	Protein digestion and absorption	4.1e-3
	Wnt signaling pathway	3.3e-2
	TGFβ signaling pathway	4.7e-2
FVP	PI3K-Akt signaling pathway	7.1e-3
	TGFβ signaling pathway	1.6e-2
	Focal adhesion	2.8e-2
	Jak-STAT signaling pathway	4.1e-2
CVM	ECM-receptor interaction	1.9e-2
	Protein digestion and absorption	2.0e-2
	Mineral absorption	4.8e-2
FVM	Glycolysis/Gluconeogenesis	5.7e-6
	Cardiac muscle contraction	2.4e-2
	Arrhythmogenic right ventricular cardiomyopathy	3.0e-2
	cGMP-PKG signaling pathway	3.8e-2

CI, cardiomyocyte intermediate; CM, cardiomyocyte; CVM, cono-ventricular myocyte; CVP, cono-

ventricular progenitor; EC, endothelial cell; EI, endothelial intermediate; FVM, free wall-ventricular myocyte; FVP, free wall-ventricular progenitor; MCP, multipotent cardiac progenitor; PI, pacemaker progenitor; PM, pacemaker cell; SMC, smooth muscle cell; SMI, smooth muscle intermediate.

Table S3. List of human embryonic/fetal heart-derived CVP/FVP/CVM/FVM-enriched genes. Related to Figure 5.

CVP		FVP		CVM		FVM	
Enriched Gene	**Averaged Z-score of log2FPKM**	**Enriched Gene**	**Averaged Z-score of log2FPKM**	**Enriched Gene**	**Averaged Z-score of log2FPKM**	**Enriched Gene**	**Averaged Z-score of log2FPKM**
PLK2	2.54239	SLC4A1	2.36598	ATP1B4	2.49313	PPP1R3A	2.12268
PDLIM3	2.47017	SLC39A8	2.16316	MTRNR2L1	2.47249	NREP	2.11067
SLC35F2	2.44254	BLOC1S4	2.16100	WBSCR17	2.40638	LDHA	2.10871
SPHK1	2.36734	FAM210B	2.15315	MARC_1	2.34126	TMEM155	2.00513
PGBD1	2.36582	MT1H	2.09816	OIT3	2.29649	SLC25A3	1.95359
HAS2	2.35086	FKBP1B	2.05120	NOB1	2.18535	FGF7	1.87069
NEDD9	2.33525	ARMC10	2.04145	FGF16	2.18030	LDHB	1.86348
TULP2	2.33312	TNIP2	1.99439	ANKRD9	2.16205	MDH1	1.86173
CYTL1	2.33201	RGS16	1.98498	LRRC20	2.15233	TKTL1	1.86076
FLYWCH1	2.32101	GOLM1	1.97420	EIF3B	2.13299	TPI1	1.82333
SEMA3D	2.32084	SNRNP35	1.93246	NATD1	2.13013	COX5A	1.80356
LGR5	2.31683	HES4	1.93174	RUNDC1	2.06693	ATP5C1	1.79460
ISL1	2.30366	MYC	1.80045	LCMT2	2.04464	CYCS	1.78186
EPHA7	2.29967	CDK5RAP1	1.77240	TTC38	2.03429	FLJ42969	1.76951
NDEL1	2.29881	THY1	1.76956	PPBP	2.00513	HIGD1A	1.70359
NLGN3	2.29437	IMPDH1	1.76834	LIMS2	1.98549	MYL12A	1.67632
FUT8	2.28769	CXCL14	1.75767	ITGA7	1.98361	ATP5B	1.67101
ITPRIP	2.27873	POP4	1.75721	FAM163B	1.96817	PLN	1.66537
ZNF468	2.27872	EMILIN1	1.74759	FSTL3	1.96672	ATP5A1	1.63079
PLAT	2.26658	SERTAD4	1.72931	ATAD3B	1.96473	LOC285043	1.57962
RFX3	2.25316	PMPCA	1.72851	VPS18	1.96278	ATP5F1	1.57746
P2RY1	2.25270	PES1	1.71872	ZNF671	1.93022	DSTN	1.57690
PLCE1-AS1	2.24405	SLC25A37	1.69613	GABRE	1.92263	NDUFA4	1.57678
DNM3	2.24032	EBF1	1.68352	RNF25	1.91956	GBAS	1.56980
SYNPO2	2.23623	GYPA	1.67404	COL18A1	1.91562	SLC25A4	1.56947
TAGLN	2.23240	TPST1	1.66581	ACAP3	1.91480	STAT4	1.53294
ERBB3	2.23000	PDP1	1.65926	TPCN2	1.91239	MYOZ2	1.53107
EPHA3	2.22652	TLE1	1.64854	ZBTB17	1.91165	TNFAIP8	1.50570
ZNF300	2.21387	RGCC	1.63087	TEX264	1.90291	CAPZA2	1.49765
S100B	2.20904	TMEM41A	1.63025	LTBP3	1.89202	HOPX	1.49103
DOCK4	2.20391	ZNF568	1.62891	TOMM40L	1.89198	MGST3	1.49030

PCDHGC3	2.20308	IGFBP2	1.61180	MOB3C	1.88701	SGCB	1.48955
IFT140	2.19969	TSPAN5	1.60822	WNK2	1.88386	NPY6R	1.48902
PTGS2	2.19899	FLT1	1.58091	MIB2	1.88288	SRP14	1.48576
TUBA1A	2.19812	RFWD2	1.57432	GRAMD1B	1.87920	CCNE2	1.48067
CTNNAL1	2.19071	TXNDC11	1.56096	SEC14L5	1.86948	NPPB	1.47916
TNFRSF19	2.18580	GPM6B	1.55051	P2RX1	1.86810	SLC25A5	1.47855
EPHB6	2.18361	VOPP1	1.53539	PLEC	1.85805	CPVL	1.47781
LOC100506990	2.18216	CTHRC1	1.52736	CASQ1	1.85218	VDAC3	1.47589
CSRP2	2.18165	ANPEP	1.50516	PLXNA2	1.83982	PERP	1.47577
SFRP4	2.17970	TNC	1.50434	ADORA2B	1.83670	CALM2	1.47052
EPB41	2.17608	NQO2	1.49980	OBSCN	1.83131	NDUFA5	1.46977
PSAT1	2.17541	BAMBI	1.49768	PCDH12	1.82643	CLIC4	1.46965
RSPO3	2.17524	LTV1	1.49292	TGM2	1.80908	CASQ2	1.46878
VMP1	2.16975	SETD4	1.48164	TDP2	1.80564	ATP5G3	1.46535
BMP4	2.16712	TOMM40	1.47959	TPRG1L	1.78962	IDH3A	1.46296
RFX7	2.16265	MARCKSL1	1.47469	TBX3	1.78707	LIN9	1.46288
SRGAP2B	2.15232	TRIB2	1.47191	EPHB3	1.77050	UQCRC2	1.45146
POLDIP3	2.15033	CCL2	1.47128	MAD1L1	1.74682	SDHD	1.45022
LINC00673	2.14913	SLC27A3	1.46900	NRAP	1.74618	SORBS2	1.44920
MMP16	2.14524	PGM2	1.46793	SLC4A3	1.73977	HSPE1	1.44373
MECOM	2.14107	SPRED2	1.46348	TNS1	1.73073	VDAC1	1.44077
CDK5RAP3	2.13773	RNF212	1.44399	BRPF1	1.72989	FHL2	1.43998
TENM3	2.13559	ATHL1	1.44063	TNRC18	1.72880	PLEK2	1.43329
GATA2	2.13410	UTP6	1.43468	FANCE	1.71225	PBK	1.42433
SLC10A1	2.12755	TPGS1	1.43199	ERCC2	1.70737	FTH1	1.41192
EGR3	2.12613	SUZ12P1	1.43076	FAM21C	1.70506	ACOT1	1.40556
FAT4	2.12260	KIT	1.42887	ACTA1	1.70447	CFL2	1.40405
ARHGEF26	2.11369	EIF3D	1.42827	RHBDF1	1.69900	ACAT1	1.39892
TRO	2.10577	TGIF1	1.42098	ADAMTSL2	1.68885	MAD2L1	1.39743
PDGFRA	2.10488	MPP1	1.41560	HSPB1	1.68711	MURC	1.39462
CYP27A1	2.10115	LXN	1.40345	BMP2	1.68543	AGL	1.38952
LOC284454	2.09942	MEAF6	1.39835	EXOC3	1.68030	SH3BGR	1.38912
TFPI2	2.09811	PVRL2	1.39760	COL18A1-AS1	1.67963	FGF12	1.38796
CPXM1	2.09294	COL9A3	1.39688	TRAIP	1.67779	VDAC2	1.38635
PPFIBP2	2.08743	KRT18	1.37940	TINAGL1	1.67730	BTF3	1.38534
TMED11P	2.07887	TASP1	1.37920	ZNF212	1.67489	CENPH	1.38254
C7	2.07818	TAL1	1.37882	TRH	1.67336	EIF1B	1.38186
SEC24D	2.07348	MORC2-AS1	1.37843	PARP12	1.67222	PGK1	1.38136

RNF212	2.06908	RAMP2-AS1	1.37384	KCNC4	1.67215	COMMD8	1.38018
ROBO1	2.06651	GOLGA5	1.35725	HSPG2	1.66875	TPM1	1.37664
LIMA1	2.06637	SOX9	1.35572	C21orf2	1.65544	HADHB	1.37506
ADAMTS9	2.06037	CES1	1.35432	FAAH	1.65144	SLC38A1	1.37248
ZNF286A	2.05869	NOV	1.33469	SBSPON	1.65024	HSPA8	1.35530
BCAT1	2.05861	INHBA	1.33143	NMRK2	1.65011	SCP2	1.35470
OLFML3	2.05531	TXNDC16	1.33115	POLR1C	1.64835	GPI	1.34893
INHBA	2.05399	RNF168	1.31538	LOC400794	1.64739	LRRC39	1.34846
HDAC1	2.05040	ZNF761	1.29518	ZNF740	1.64369	SLC16A1	1.34827
CRABP2	2.04818	COL9A2	1.28795	MB	1.64087	OTUD1	1.34825
PCIF1	2.04755	TKT	1.28787	PES1	1.63711	FH	1.34697
LRRC4C	2.04738	HAPLN1	1.28273	GMEB2	1.63704	LYRM5	1.34535
GATA5	2.04416	RBM6	1.27908	PNOC	1.63423	EIF5A2	1.34277
NFKBIZ	2.04324	IFIT1	1.27707	KLC4	1.63186	HNRNPK	1.34216
VWA5A	2.04179	CDC42EP4	1.27669	DYSF	1.63037	MPC2	1.33995
SFT2D2	2.03686	HMBS	1.26396	ROBO4	1.62275	LYRM1	1.33519
FAP	2.03652	COL6A2	1.26382	SLC26A6	1.59910	LOC100506725	1.33518
CASC15	2.03392	NFKBIA	1.26347	FGFRL1	1.59890	HSPD1	1.33433
SPATS2	2.03345	ARPC1A	1.25742	KLHL38	1.59622	PDHA1	1.32854
SLC22A3	2.02950	LITAF	1.24404	ALAS2	1.59501	RBP7	1.32788
SEMA6D	2.02882	OLFM2	1.24361	GCAT	1.59475	MPC1	1.32755
ZFP36L1	2.02799	DOLPP1	1.23863	FAM21A	1.59333	ATP1A3	1.32620
PELI2	2.02197	BICC1	1.23838	C15orf52	1.58905	MYL2	1.32603
MEIS2	2.02106	ZNF692	1.23719	NME6	1.58435	CSRP3	1.32513
ARFIP2	2.02008	SRP68	1.23206	C20orf166-AS1	1.58394	HMGB2	1.32368
TTC21A	2.02007	ARID5A	1.22933	PDGFB	1.57691	KCTD9	1.32037
FAM110B	2.01903	USP42	1.22702	TATDN2	1.57569	CISD2	1.32032
CRISPLD1	2.01657	TPBG	1.22682	TNNT1	1.56504	C5orf46	1.31803
KIAA1522	2.01509	SLC25A39	1.21278	IGFBP5	1.56137	SMYD1	1.31767
CES1	2.01480	TAF6L	1.21216	DES	1.55379	CKAP2	1.31309
HAPLN1	2.01310	SOX4	1.21023	KIF1A	1.55106	SMPX	1.30893

CVM, cono-ventricular myocyte; CVP, cono-ventricular progenitor; FVM, free wall-ventricular myocyte; FVP, free wall-ventricular progenitor.

Table S4. Over-represented genes between the current *in vivo* single-cell RNA-seq data and the PCGC dataset (Jin et al., 2017). Related to Figure 5.

CVP-enriched genes	CVM-enriched genes
SLC35F2	LIMS2
FLYWCH1	ITGA7
EPHA7	ATAD3B
DNM3	TPCN2
TAGLN	ZBTB17
ERBB3	TOMM40L
DOCK4	WNK2
IFT140	MIB2
RSPO3	PLEC
RFX7	PLXNA2
MMP16	OBSCN
MECOM	TGM2
FAT4	TNS1
PPFIBP2	BRPF1
LRRC4C	TNRC18
MEIS2	FAM21C
FAM110B	RHBDF1
	ADAMTSL2
	TRAIP
	KCNC4
	HSPG2
	SBSPON
	PES1
	SLC26A6
	C15orf52
	KIF1A

CVM, cono-ventricular muscle cell; CVP, cono-ventricular progenitor.

Table S5. Primary antibodies for flow cytometry. Related to STAR Methods.

Primary Antibody	Company	Catalog Number	Dilution
CACNA1D	Abcam	ab85491	1:100
Cardiac Troponin T/TNNT2 FITC	Abcam	ab105439	1:200
CD31 APCCY7	BD Biosciences	563653	1:150
HCN4 PE	StreeMarq Biosciences	SMC-320D-R-PE	1:200
ISL1	DSHB	39.4D5	1:100
ISL1 PE	BD Biosciences	562547	1:100
LEF1 AF488	Cell Signaling Technology	8490	1:50
LGR5 APC	R&D	FAB8078A	1:50
LGR5 PE	R&D	FAB8078P	1:100
MESP1 AF405	Novus Biologicals	MAB9219AF405	1:250
PDGFRA AF647	BD Biosciences	562798	1:100
SMMHC	Biomedical Technologies	BT-562	1:20

Table S6. Primary antibodies for immunostaining. Related to STAR Methods.

Primary Antibody	Company	Catalog Number	Dilution
ANXA1	Thermo Fisher Scientific	PA5-13530	1:100
ASB2	Abcam	ab200370	1:50
α-SMA	Sigma-Aldrich	A2547	1:400
Cardiac Troponin T/TNNT2	Thermo Fisher Scientific	MS-295-P1	1:100
CD31	Cell Signaling Technology	3528	1:50
EPCAM	HPA	067463	1:50
F2RL2	Sanrta Cruz	sc-53819	1:100
HCN4	Millipore	AB5808	1:50
HHEX	HPA	051894	1:200
ISL1	DSHB	39.4D5	1:20
ISL1 (for mouse samples)	HPA	057416	1:300
KCNH7	Thermo Fisher Scientific	PA5-26177	1:50
Ki67	BD Biosciences	550609	1:25
LEF1	Cell Signaling Technology	2230	1:50
LGR5	R&D	MAB8078	1:50
LGR5 (for mouse samples)	BD Biosciences	562733	1:50
MF20	DSHB	P13538	1:20
NKX2-5	R&D	MAB2444	1:100
PDGFRA	Thermo Fisher Scientific	710169	1:100
PDGFRA (for mouse samples)	R&D	AF1062	1:25
SMMHC	Thermo Fisher Scientific	MS-1177-P1	1:25
VE-cadherin	R&D	AF938	1:100
VIM	Millipore	AB5733	1:2000

Table S7. Primers for quantitative PCR. Related to STAR Methods.

Gene Name	Forward	Reverse
BMP2	CAGACCACCGGTTGGAGA	CCACTCGTTTCTGGTAGTTCTTC
CDH5	AAGCCTCTGATTGGCACAGT	CTGGCCCTTGTCACTGGT
FGF12	GGGACCGAAATGGAGAGC	TGGTACCATCTGGGTGCAT
FGF16	TCGGAATCCTGGAGTTTATCA	CATTCACGTGTGAGTTTCTTCG
GAPDH	AGCCACATCGCTCAGACAC	GCCCAATACGACCAAATCC
HAND1	AACTCAAGAAGGCGGATGG	GGAGGAAAACCTTCGTGCT
HCN4	GGTGTCCATCAACAACATGG	GCCTTGAAGAGCGCGTAG
HSPB7	GACTTTGGCAGCTTCATGC	GGCGTCTCCTAGGGTCTTG
ISL1	AAGGACAAGAAGCGAAGCAT	TTCCTGTCATCCCCTGGATA
ITGA7	GACGACGGTCCCTACGAG	GACCTTTCCCCGAGTCAATAG
LDHB	GATGGATTTTGGGGGAACAT	AACACCTGCCACATTCACAC
LEF1	CAGTCGACACTTCCATGTCC	ATGAGGGATGCCAGTTGTGT
LGR5	ACCAGACTATGCCTTTGGAAAC	TCCCAGGGAGTGGATTCTATT
LTBP3	CCAATGGCTCCTACAGATGTC	GGCTGCACTCATCTATGTCTTG
MESP1	CTGTTGGAGACCTGGATGC	CGTCAGTTGTCCCTTGTCAC
MYH6	CTCAAGCTCATGGCCACTCT	GCCTCCTTTGCTTTTACCACT
MYL2	GCAGGCGGAGAGGTTTTC	AGTTGCCAGTCACGTCAGG
NKX2-5	CCGGTTGGAACTGGGACT	GACGGCGAGATAGCAAAGG
PDGFRA	CCACCTGAGTGAGATTGTGG	TCTTCAGGAAGTCCAGGTGAA
PECAM1	GCAACACAGTCCAGATAGTCGT	GACCTCAAACTGGGCATCAT
PLN	ATGATCACAGCTGCCAAGG	TGAGCGAGTGAGGTATTGGA
POU5F1	GAAACCCACACTGCAGATCA	CGGTTACAGAACCACACTCG
SHOX2	AACGTAGGTGCTTTAAGGATGC	GAAAGGACAAGGGCGTCAC
SM22α	CAGTGTGGCCCTGATGTG	CACCAGCTTGCTCAGAATCA
SMMHC	CCTCAAGCTGCGGAACTG	CTGCAGCAGTGGCTTCAC
T	TTCAAGGAGCTCACCAATGA	GAAGGAGTACATGGCGTTGG
TBX5	CCAGGAGCATAGCCAAATTTAC	AGGGCTTCTTATAGGGATGGTC
TNNT2	GTCGGCAGCTGCTGTTCT	TCCTCTCTCCAGTCCTCCTCT