UNIVERSITY OF CALIFORNIA SAN DIEGO

Illuminating the diversity and organization of cardiovascular lineages forming cardiac structures during human development

A dissertation submitted in satisfaction of the requirements for the degree
Doctor of Philosophy

in

Biomedical Sciences

by

Elie Farah

Committee in charge:

Professor Neil Chi, Chair
Professor Sylvia Evans
Professor Trey Ideker
Professor Bing Ren
Professor Kun Zhang

2022

The dissertation of Elie Farah is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

For my parents and my wife, whose unconditional love and support have helped me overcome all of the trials and tribulations of my graduate studies.

EPIGRAPH

Ever tried. Ever failed. No matter. Try again. Fail again. Fail better.

– from the novella ‘Worstward Ho’ (1983) by Samuel Beckett

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge and thank Professor Neil C. Chi, my Principal Investigator and committee chair, for his support and mentorship in developing my science and writing, and for his patience and trust in my abilities.

Chapters 2 and 3, in full, are currently being prepared for submission for publication of the material. Farah, Elie N.; Hu, Robert K.; Kern, Colin; Zhang, Qingquan; Lu, Ting-Yu; Zhang, Bo; Carlin, Daniel; Li, Bin; Blair, Andrew P.; Ren, Bing; Evans, Sylvia M.; Chen, Shaochen; Zhu, Quan; Chi, Neil C. “Cardiac Single Cell Ecology Reveals Interactive Cellular Construction and Remodeling of the Human Heart”. The dissertation author was the primary investigator and author of this paper.

I would like to acknowledge the UC San Diego (UCSD) Center for Epigenomics (CFE) for technical assistance with the MERFISH experiments, including Yuanuan Han for imaging expertise, Alex Monell for bioinformatic support, and Dr. Sebastian Preissl for FACS assistance. I am grateful to the UCSD Institute of Genomic Medicine (IGM), including Dr. Kristen Jepsen and Eugenia Ricciardelli for technical assistance with library construction and sequencing support for the scRNA-seq experiments. I am also grateful for Dr. Mana Parast and Morgan Meads for helping to procure the samples used in this study. I would like to acknowledge the California Institute for Regenerative Medicine (CIRM) for funding, and sequencing and data repository support including Cesar Barragan (Salk), Dr. Joe Ecker (Salk), Dr. Joshua Stuart (UC Santa Cruz), Clay Fischer (UC Santa Cruz), Brian Aevermann (J. Craig Venter Institute), and Dr. Richard Scheuermann (J. Craig Venter Institute).

VITA

2011 Bachelor of Science, University of Rochester

2012 Master of Science, University of Rochester

2015 Master of Business Administration, University of Rochester

2015 - 2021 Graduate Student Researcher, University of California San Diego

2022 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Farah, EN; Hu, RK; Kern, C; Zhang, Q; Lu, T; Zhang, B; Carlin, D; Li, B; Blair, AP; Ren, B; Evans, SM; Chen, S; Zhu, Q; Chi, NC. "Cardiac Single Cell Ecology Reveals Interactive Cellular Construction and Remodeling of the Human Heart". *In preparation*

Hocker JD, Poirion O, Zhu F, Buchanan J, Zhang K, Chiou J, Hou X, Li Y, Zhang Y, **Farah EN**, Wang A, Gaulton KJ, Ren B*, Chi NC*, Preissl S*. Cardiac cell type-specific gene regulatory programs and disease-risk association. *Science Advances*. 2021 May 14;7(20):eabf1444.

Blair AP, Hu RK, **Farah EN**, Chi NC, Pollard KS, Przytycki PF, Kathiriya IS, Bruneau BG. Cell Layers: Uncovering clustering structure and knowledge in unsupervised single-cell transcriptomic analysis. *bioRxiv*. 2020 Nov 30. doi: https://doi.org/10.1101/2020.11.29.400614

Ren J, Han P, Ma X, **Farah EN**, Bloomekatz J, Zeng XI, Zhang R, Swim MM, Witty AD, Knight HG, Deshpande R, Xu W, Yelon D, Chen S, Chi NC. Canonical Wnt5b Signaling Directs Outlying Nkx2.5+ Mesoderm into Pacemaker Cardiomyocytes. *Dev Cell*. 2019 Sep 23; 50(6):729-743.e5.

Zhang Y, Li T, Preissl S, Amaral ML, Grinstein JD, **Farah EN**, Destici E, Qiu Y, Hu R, Lee AY, Chee S, Ma K, Ye Z, Zhu Q, Huang H, Fang R, Yu L, Belmonte JCI, Wu J, Evans SM, Chi NC & Ren B. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet*. 2019 Sep; 51(9):1380-1388.

Deacon DC*, Happe CL*, Chen C*, Tedeschi N, Manso AM, Li T, Dalton ND, Peng Q, **Farah EN**, Gu Y, Tenerelli KP, Tran VD, Chen J, Peterson KL, Schork NJ, Adler ED, Engler AJ, Ross RS, Chi NC. Genome-edited cardiac models reveal combinatorial genetic interactions in human cardiomyopathy. *Nature Biomedical Engineering*. 2019 Feb; 3(2): 147-157. doi: 10.1038/s41551-019-0348-9

Lai WJC. Kayedkhordeh M. Cornell EV, **Farah E**, Bellaousov S, Rietmeijer R, Salsi E, Mathews DH, Ermolenko DN. mRNAs and lncRNAs intrinsically form secondary structures with short end-to-end distances. *Nat Commun*. 2018 Oct 18;9(1):4328.

Veevers J*, **Farah EN***, Corselli M, Witty AD, Palomares K, Vidal JG, Emre N, Carson CT, Ouyang K, Liu C, van Vliet P, Zhu M, Hegarty JM, Deacon DC, Grinstein JD, Dirschinger RJ, Frazer KA, Adler ED, Knowlton KU, Chi NC, Martin JC, Chen J, Evans SM. Cell-Surface Marker Signature for Enrichment of Ventricular Cardiomyocytes Derived from Human Embryonic Stem Cells. *Stem Cell Reports*. 2018 Sep 11;11(3):828-841.

Salsi E, **Farah E**, Ermolenko DN. EF-G Activation by Phosphate Analogs. *J Mol Biol*. 2016 May 22;428(10 Pt B):2248-58.

Salsi E, **Farah E**, Netter Z, Dann J, Ermolenko DN. Movement of elongation factor G between compact and extended conformations. *J Mol Bio*. 2015 Jan 30;427(2):454-67.

Salsi E, **Farah E**, Dann J, Ermolenko DN. Following movement of domain IV of elongation factor G during ribosomal translocation. *Proc Natl Acad Sci U S A.* 2014 Oct 21;111(42):15060-5.

Gibbs PE, Lerner-Marmarosh N, Poulin A, **Farah E**, Maines MD. Human biliverdin reductase-based peptides activate and inhibit glucose uptake through direct interaction with the kinase domain of insulin receptor. *FASEB J*. 2014 Jun;28(6):2478-91.

FIELDS OF STUDY

Major Field: Biomedical Sciences

ABSTRACT OF THE DISSERTATION

Illuminating the diversity and organization of cardiovascular lineages forming cardiac structures during human development

by

Elie Farah

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2022

Professor Neil C. Chi, Chair

The heart, which is the first organ to develop, is highly dependent on its form to function. Disruption of cardiac structures contributing to this form can lead to a wide range of adult/congenital heart diseases and possible fetal demise. However, how diverse cardiovascular cell types spatially interact and organize into complex morphological communities/structures that are critical for heart function remains to be fully illuminated. Here, we interrogate the interactive cellular mechanisms that direct the morphogenesis and remodeling of heterogeneous cellular communities during the construction of the human heart by providing a high-resolution spatial and single-cell

transcriptomic human cardiovascular cell atlas. We integrated high-throughput multiplexed fluorescent *in situ* hybridization (MERFISH) spatial transcriptomics with corresponding single-cell RNA-seq and spatially mapped 27 major cardiac cell types that could be further refined into specialized cellular subtypes and states including previously uncharacterized cell populations. These major cell types were spatially organized into unique cellular communities that compose of combinations of specific cardiac cell types and correspond to distinct anatomic cardiac structures. Detailed examination of the cardiac ventricles revealed an unexpected cellular heterogeneity and organization of the ventricular wall, which was composed of at least three regional cellular communities corresponding to outer-compact, mid, and inner-trabecular layers. Analysis of these communities uncovered thousands of cellular interactions among 11 cell types within the left ventricular wall, including a Semaphorin-Plexin mediated cardiomyocyte-cardiac fibroblast-endothelial multi-cellular interaction that coordinates the precise allocation of migrating *ERBB2*/*4*+*PLEXINA2*/*4*+ cardiomyocyte subtypes during the critical cardiac developmental process of ventricular compaction. These identified morphogenetic events could be recapitulated in a human pluripotent stem cell (hPSC) cardiac organoid system thus confirming the role of Semaphorin-Plexin during human ventricular wall morphogenesis. Our findings provide the foundations for analyzing the morphogenesis of the critical structures comprising the heart and serves as a guide to improve the construction of complex structures during cardiovascular development and generate hPSC-derived cardiac tissues composed of multiple cardiac cell-types for the study of human cardiovascular development, tissue replacement therapy and disease modeling.

CHAPTER 1:

Introduction

1.1 Cardiac cells that construct the human heart

The heart is the first organ to form in the developing embryo[1]. Throughout development, it continues to grow and function to support the maturing fetus by circulating nutrients to all of the developing organs. Understanding this process in full requires knowledge of the individual cell types and states that play a role in cardiac organogenesis and how those cell types and states are spatially assembled to form the various structures of the heart, which are important for function. Defects in this process can lead to congenital heart defects that affect up to 1-3% of all live births[2]. Recent advances in quantitative molecular and spatial measurements monitoring global cellular identity and positioning in thousands to tens of thousands of single cells from an experiment are allowing for the discovery of cardiac cell types and states as well as the interactive cellular communities in which they reside[3–6]. These recent advances have contributed to our knowledge of many of the cell types within the human heart, including the developing heart[3,4,6,7]. Understanding the building blocks of the human heart and how they are assembled into complex morphological structures will give insights for therapies to treatments when these structures become disorganized in disease.

Recent advances in high throughput multidimensional tools like single cell sequencing technologies have characterized many of the cell types within the human heart, including the developing heart[3,4,6,7]. These cell types include many endogenous to the heart including cardiomyocytes, cardiac fibroblasts, endocardial, and epicardial cells, as well as others including vascular (smooth muscle and endothelial),

glia/schwann cells, and immune populations. Cardiomyocytes are the engine of the heart where energy is converted into mechanical work, thereby providing the contractile force of the heart muscle. Cardiomyocytes are marked by their expression of proteins involved in the contractile machinery such as troponins (*TNNT2, TNNI1, TNNI3*), myosin light (*MYL7, MYL9*), myosin heavy (*MYH6, MYH7*), and titin (*TTN*). Atrial and ventricular myocardium differs in developmental, structural, hemodynamic, and physiological characteristics, and this is reflected by distinct gene expression profiles, with atrial and ventricular cardiomyocytes expressing nuclear receptor subfamily 2 group F member 1 and 2 (*NR2F1/NR2F2*) and *IRX4/MYL2*, respectively[3,4,8]. Beyond the specialized contracting cardiomyocytes of the chambers, there are 'non-chambered' cardiomyocytes, including those which participate in the cardiac conduction system, which transduces electrical stimuli to drive cardiac contraction. Conduction system related cardiomyocytes have traditionally been identified using hyperpolarization-activated cyclic nucleotide-gated potassium channel 4 (*HCN4*), ISL LIM homeobox 1 (*ISL1*), or T-box transcription factor 3 (*TBX3*), although the expression of these markers shows significant heterogeneity depending on their anatomical location within the conduction system (nodes, atrioventricular bundle, Purkinje fiber network)[9–12].

Cardiac fibroblasts provide important biomechanical support to the heart tissue by synthesizing and remodeling the extracellular matrix, as well as provide signaling factors to the various cells in the heart[13,14]. In adult human heart tissue, canonical markers for cardiac fibroblasts include many extracellular matrix-related proteins including decorin (*DCN*), collagen type I alpha 1 chain (*COL1A1*), and vimentin (*VIM*), as well as receptors (platelet derived growth factor receptor alpha-*PDGFRA* and Thy-1

cell surface antigen-*THY1*)[3]. Fetal-derived cardiac fibroblasts express transcription factor 21 (*TCF21*) as well as extracellular matrix protein lumican (*LUM*)[6]. Similar to cardiomyocytes, fibroblasts also exhibit transcriptional differences between the atria and ventricles, with atrial fibroblasts expressing gelsolin (*GSN*), and ventricular fibroblasts expressing complement 7 (*C7*)[3]. Consistent with the expression of the complement system, fibroblasts have been proposed to serve immune-modulatory roles disease[15].

Endothelial cells within the heart line the inner lumen of the cardiac chambers, blood, and lymphatic vessels, and fulfill many functional roles including regulating the flow of solutes, fluid, and macromolecules as well as modulating vascular constriction and relaxation[16]. Despite the heterogeneity of endothelial cells, these cells can be identified by platelet and endothelial cell adhesion molecule 1 (*PECAM1*), and further subdivided into lymphatics by prospero homeobox 1 (*PROX1*) and endocardial cells by natriuretic peptide receptor 3 (NPR3)[4]. Further classification of blood endothelial cells between arterial and venous is possible in the adult heart with the examination of EPH receptor B4 (*EPHB4*), *NR2F2*, *HEY1*/*2*[3,16]. This granularity of endothelial cell definition requires further investigation within the fetal human heart to confirm if these markers can be used throughout human cardiac development.

Smooth muscle cells support the extensive vasculature of the heart and are mainly divided into pericytes (which reside within micro vessels) and vascular smooth muscle cells (which reside in the vascular wall of larger vessels). In adult human hearts, pericytes are classically marked by the expression of platelet-derived growth factor receptor beta (*PDFRB*), ATP binding cassette subfamily C member 9 (*ABCC9*), and potassium voltage-gated channel subfamily J member 8 (*KCNJ8*). Proteins related with

contractile function such as *MYH11* and Smooth muscle aortic alpha-actin 2 (*ACTA2*) are used to identify vascular smooth muscle cells.

Peripheral glial cells, such as Schwann cells, have been particularly difficult to analyze in the heart until the development of scRNA-seq methodology due to their low abundance and density. In the adult human heart, these Schwann cells express proteins related to the myelin sheath including proteolipid protein 1 (*PLP1*), while in the fetal heart these cells can be identified with the expression of aldehyde dehydrogenase 1 family member A1 (*ALDH1A1*)[3,6].

Immune populations within the heart have been investigated more frequently recently, due in part to the ease of isolating blood cells that do not require complex dissociation strategies. Practically all known types of blood cells have been described within the heart including monocytes, macrophages, T-cells, B-Cells, and mast cells, although their exact roles are yet to be fully defined[17]. In both adult and fetal human hearts, myeloid related populations represent the highest proportion of blood cells within the heart[3,4,6]. Monocytes are identified by the expression of basic leucine zipper ATF-like transcription factor 3 (*BATF3*) and lysozyme (*LYZ*), while macrophages can be identified by the expression of interleukin 1β (*IL1B*), lymphatic vessel endothelial hyaluronan receptor 1 (*LYVE1*), and proteins involved in the complement system (including complement C1q B Chain – *C1QB*). The knowledge about the significance of these blood populations during development and disease is still limited and is under active investigation[17].

Recent advances in spatial transcriptomic technologies have expanded our knowledge of multiple organ systems, including the developing human heart. Cardiac

development is a spatially complex process and comprehensive understanding of regional and structural changes in gene expression during heart maturation is of great interest. Recently a study by Asp et al. combined RNA-seq data of spatial transcriptomics (ST-seq) of human fetal tissues, scRNA-seq data of human fetal cardiac cells, and *in situ* sequencing data to map cell type distribution and spatial organization in the developing human heart and generate a 3D gene profile atlas[6]. Utilizing their spatial atlas, they were able to map the cell types identified within their scRNA-seq data to single cell resolution within the developing human heart. This investigation led to the discovery of distinct spatial locations of certain cell types including epicardial-derived cells (EPDCs), Schwann progenitor cells, and cardiac neural crest cells, with the former residing primarily within the atrioventricular sub-epicardial mesenchyme, and the latter two mainly located near the outflow tract (OFT) and atria. Several limitations of the study restrict further insights into the cellular composition and spatial organization of the heart, including lack of spatial resolution of the transcriptome-wide sequencing (ST-seq), which limits the unbiased spatial identification of cells on a single cell level. The *in situ* sequencing study addresses some of these limitations (e.g. resolution), but does not allow for identification of cell types beyond those identified in the scRNA-seq (15 cell types), partially because of the limited gene panel (only probed for 69 genes). Further work is necessary to spatially identify cell types at a high resolution (single cell) and in a more unbiased approach to allow for cell type discovery and analysis of spatial organization into cardiac structures.

1.2 Ventricular wall development and semaphorin signaling

The heart is composed of four chambers that have different structural, physiological, as well as developmental origins. The ventricle in particular undergoes trabeculation and compaction during development. Trabeculation in particular is one of the first signs of ventricular chamber development and starts around E8.5 in mice. During this developmental process, cardiomyocytes form protrusions toward the ventricular lumen that eventually form a complex meshwork lining the inner ventricular chamber[18,19]. Trabeculae are present in all vertebrate hearts and are necessary to increase the surface area of the early myocardium to aid oxygen and nutrient exchange before coronary vascularization. The formation of trabeculae is dependent on the interaction between the endocardium and myocardium, and includes signaling pathways such as a neuregulin 1 (*NRG1*)-*ERBB2* interaction. Functional studies in mice, chick, and zebrafish have provided a great deal of knowledge of the signaling pathways and transcription factors involved in trabeculation, however how all of the precise cellular processes such as oriented cell division, delamination, EMT-like processes, migration and proliferation coordinate and give rise to the trabecular network is still under intense investigation[20,21].

As the ventricle continues to develop, the trabeculae compress and become integrated within the ventricular wall through the process of compaction, which functions to thicken the myocardial wall. This process coincides with the vascularization of the myocardial wall by the coronary vessels, and in mice, this process starts around E13.5 and lasts well after birth, while in humans it starts around week 12 of gestation[22]. Defects in this chamber maturation process may lead to various cardiac disorders,

including left ventricular noncompaction (LVNC), which is characterized prominent trabeculae and thin myocardium[23]. Lineage-tracing studies in mice utilizing Cre recombinase under the control of genes specific to either trabeculae (*Nppa*) or compact (*Hey2*) have shown contribution to the endocardial zone and epicardial zone of the postnatal ventricular wall, respectively[24]. These lineage-tracing studies further identified a heterogeneous region between the endocardial and epicardial zone termed the hybrid myocardial zone, which was composed of cardiomyocytes derived from both *Nppa*+ and *Hey2*+ cardiomyocytes. Therefore, the compacting myocardial wall contains at least three separate zones: 1) a subendocardial inner myocardial wall, or trabecular myocardium, 2) a mid myocardial zone, or hybrid myocardium, and 3) a outer myocardial wall, or compact myocardium. Defects in either the contribution of *Hey2*+ cells or *Cx40*+ cells (marker of trabecular cardiomyocytes) results in the persistence of trabeculae in the postnatal heart, suggesting that both the expansion of the compact myocardium as well as the coalescence of trabecular myocardium contribute to the process of compaction[24,25].

Signaling pathways that regulate the process of compaction are poorly understood, with recent studies suggesting that Notch and semaphorin signaling contribute to this process[26]. Semaphorins are a large family of secreted or membrane-bound proteins that bind to either plexin receptors or plexin-neuropilin receptor complexes to propagate their signal. Semaphorin signaling has been extensively studied in nervous system where it has a well established role in axon guidance. Within the cardiovascular system, defects in semaphorin signaling lead to various cardiovascular defects, mainly related to the migration of cardiac neural crest cells, and

include persistent truncus arteriosus, ventricular septal defects, and coronary and aortic arch defects[27]. However, a recent study identified that an endothelial specific plexin receptor *PLXND1* plays a role in ventricular compaction, and loss of function of this gene within endothelial cells causes a noncompaction phenotype in mice[28]. The authors suggest that the possible mechanism is that the semaphorin signaling modulates extracellular matrix dynamics and Notch signaling. Whether semaphorin signaling plays additional roles in ventricular compaction, and whether these pathways are conserved in human heart development is incompletely understood.

Human pluripotent stem cell (hPSC)-derived cardiomyocytes are a powerful system to study the mechanisms of human cardiomyocyte development and disease in a controlled setting in a reproducible way. Protocols have been established for more than a decade and used by many labs to test various aspects of cardiomyocyte development and function. To test how certain semaphorin signaling pathways may be regulating the migration and organization of cardiomyocytes during human ventricular wall development, my co-authors and I created a multilayered hPSC cardiac organoid system to model aspects of the developing ventricular wall. We demonstrate that *SEMA3C* acts as an attractive signaling cue to direct the migration of *PLXNA2/A4* expressing trabecular-like cardiomyocytes toward the source of *SEMA3C*. Additionally, we find that *SEMA6A/B* act as a repulsive cue that blocks the migration of *PLXNA2/A4* expressing trabecular-like cardiomyocytes. Taken together, our results support a model of cardiomyocyte organization where *SEMA3C/D*-expressing compact ventricular fibroblasts attract *PLXNA2/A4* expressing trabecular (and hybrid) cardiomyocytes to the Mid/Outer myocardial community layers, whereas *SEMA6A/B*-expressing blood

endothelial cells may prevent these trabecular (and hybrid) cardiomyocytes from further migrating by repelling them when contacting blood endothelial cells, and that this may be what is happening *in vivo* during compaction of the ventricular wall in humans.

CHAPTER 2:

Single-cell transcriptomic analysis reveals diverse cell lineages constructing the developing human heart

2.1 INTRODUCTION

In the developing embryo, the heart is the first organ to form. By week 3 it is a linear heart tube, and between weeks 3 and 8 it undergoes a complex looping process to form the four chambers of the heart, the (Left Atrium (LA), Right Atrium (RA), Left Ventricle (LV), Right Ventricle (RV)). These chambers are separated by the interatrial and interventricular septa, which prevent the mixing of oxygenated and deoxygenated blood. A conduction system rapidly propagates electrical impulses from the sinoatrial node (SAN) to the atrioventricular node (AVN), and along His-Purkinje fibers to the ventricular apex where contraction starts. Between weeks 8 up until birth, the heart undergoes a period of rapid growth and expansion as the various cardiovascular (CV) lineages mature and organize to support the developing fetus by continuously circulating the blood supply. Understanding how the different CV cell-types develop during this critical period has been difficult to determine in bulk analysis of heterogeneous tissue samples.

Here, we characterize the developing heart at an unprecedented resolution, combining scRNA-seq data and an imaging based spatial transcriptomic technique called multiplexed error-robust fluorescence *in situ* hybridization (MERFISH) that allows for the spatial identification of cells at single cell resolution. By sequencing more cells within developing human hearts than previous studies[6,7,29], we identified 75 subpopulations of cardiac cells, including many known in addition to previously

uncharacterized cell types within the developing heart, such as non-chambered cardiomyocytes. Spatial mapping validates and resolves many of the cell types at high resolution, and we find that chamber distinct cardiomyocyte populations co-localize with chamber distinct non-cardiomyocyte counterparts. To understand how they are spatially organized, we identified cellular communities that correspond to cardiac structures, including the inner and outer layers of the ventricle, and the left and right atria. These cellular communities are composed of distinct cell types, suggesting that these specialized communities influence not only the specialized functionalization of the cardiac cells, but also direct the formation of anatomic structures critical for regulating overall cardiac function.

2.2 RESULTS

Single-cell transcriptomic analysis reveals diverse cell lineages constructing the developing human heart

To examine how diverse cardiovascular cell types coordinate to form complex structures critical for regulating human heart function, we initially investigated and identified the specific cell lineages comprising the developing human heart. To this end, single-cell RNA sequencing (scRNA-seq) was analyzed from human hearts between 9 to 15 weeks of gestation in replicate, using droplet-based methods (**Figure 1a**). Because these developing hearts were substantially smaller than adult human hearts, each collected heart was dissected into intact cardiac chambers and the interventricular septum (IVS), which were then dissociated into single cells for examination of all potential cell lineages within these hearts by RNA-sequencing (**Figure 1a**). As a result, we collected and analyzed the transcriptomes of 142,946 individual cardiac cells, which grouped by cell identity rather than donor sample after graph-based clustering[30] (**Figure 1b-d**). Detected cell clusters primarily consisted of cells from each donor and were classified into five major developmental cell populations including cardiomyocyte, mesenchymal, endothelial-like, blood and neuronal-like cell lineages (**Figure 1a**). Further clustering of these major cell populations uncovered 12 subclasses comprising 75 subpopulations (**Figures 2-7**), which were typically defined by combinations of genes rather than by single markers (**Figures 2-7**). Consistent with their overall contributions to the heart (**Figure 1b**), cardiomyocyte and mesenchymal-related lineages displayed the highest number of subpopulations (22 of 75 each), whereas neuronal-like lineages contained the fewest (7 of 75).

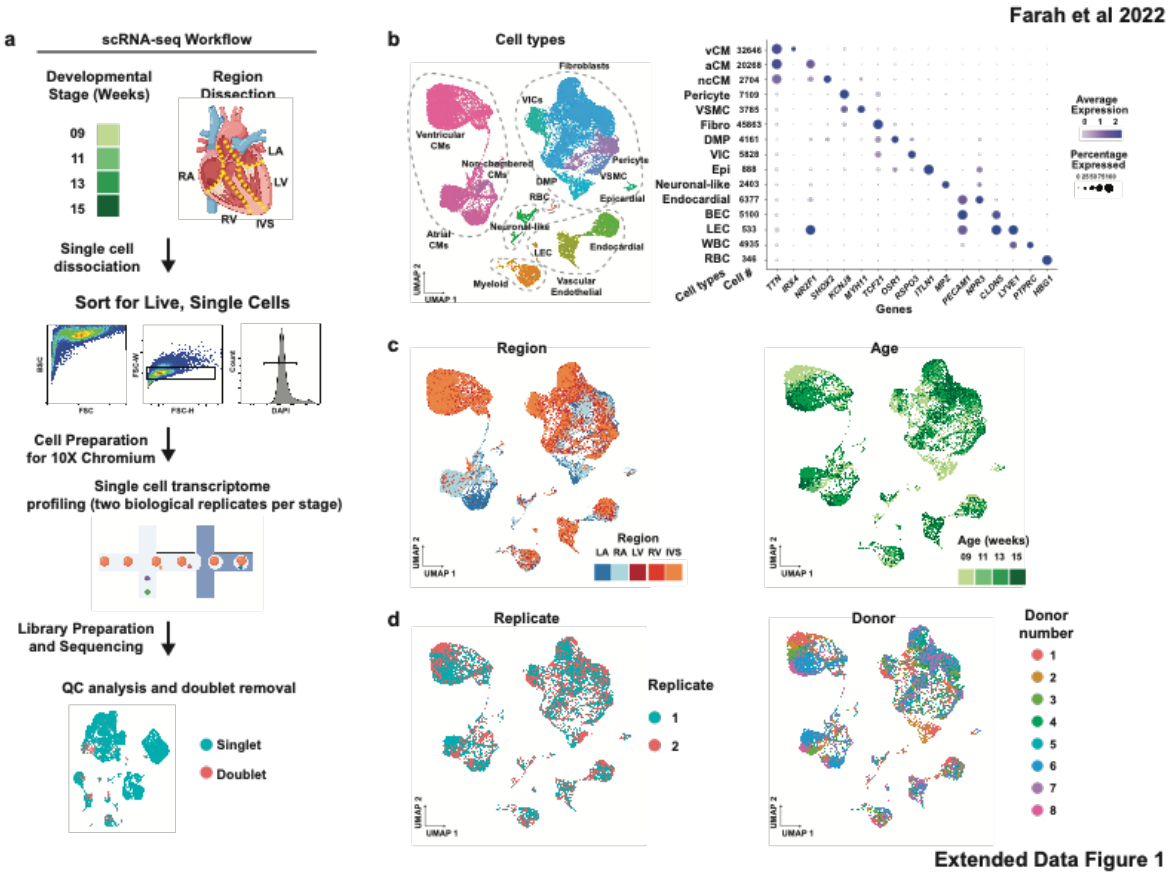
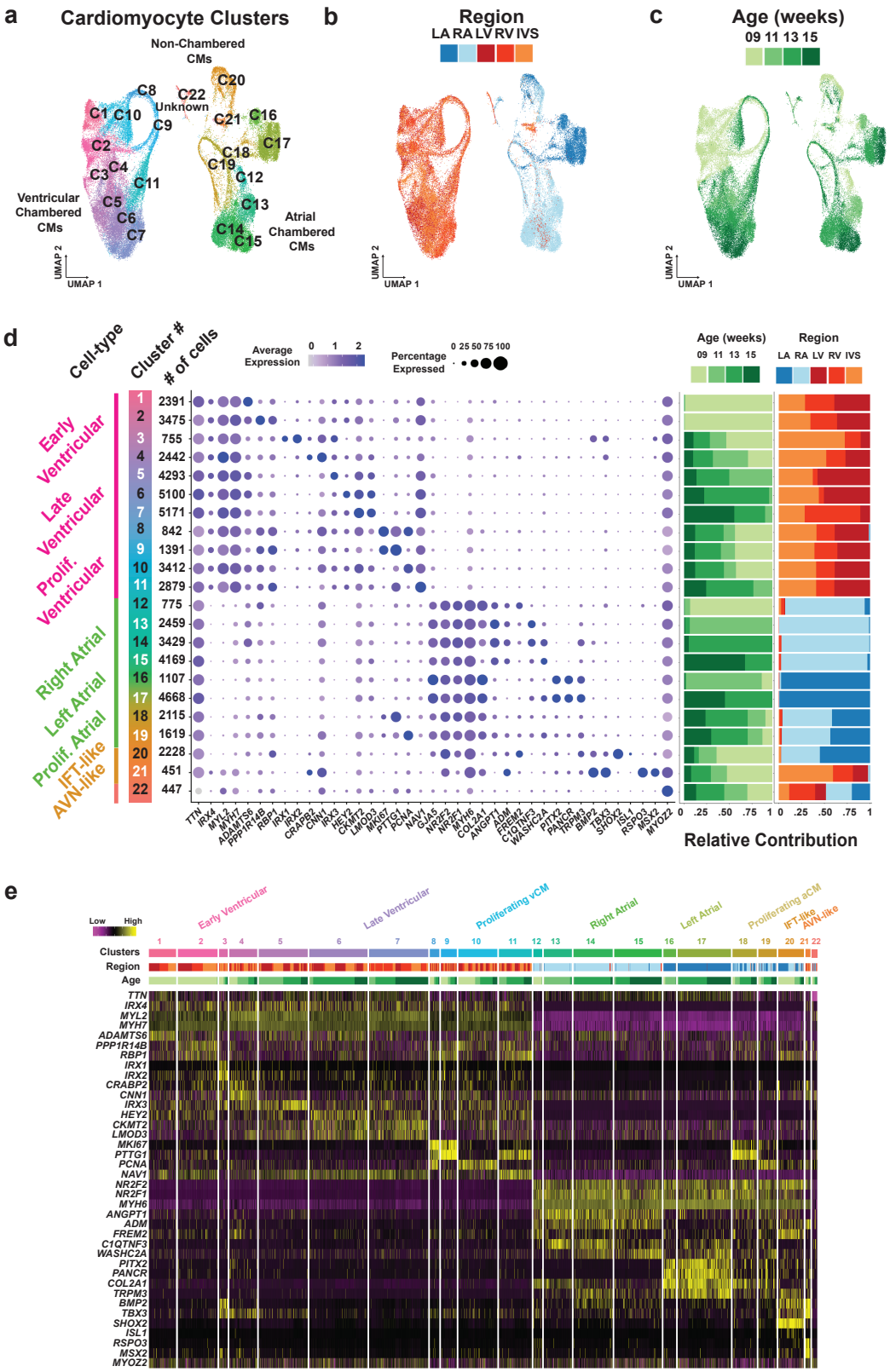


Figure 1. Developing human heart scRNA-seq experimental details show the sequencing and quality control for scRNA-seq datasets.
a, Single-cell RNA sequencing (scRNA-seq) experimental design illustrating sample collection over time (9, 11, 13, and 15 weeks of gestation) and region (LA, RA, LV, RV, IVS) during human heart development. **b**, Scatterplot of ~143k cells after principal component analysis and Uniform Manifold Approximation and Projection (UMAP), colored by major cell type (left), region (middle), and developmental age (right). **c**, Heatmap of marker gene expression and relative contribution of Age and Region to each cluster. aCM, atrial cardiomyocyte; BEC, blood endothelial cell; CMs, Cardiomyocytes; DMP, dorsal mesenchymal protrusion; Endo, endocardial; Epi, epicardial; LEC, lymphatic endothelial cell; Neuro, neuronal-like; ncCM, non-chambered cardiomyocyte; Peri, Pericyte; RBC, Red Blood Cell; vCM, ventricular cardiomyocyte; VICs, valve interstitial cells; VSMC, vascular smooth muscle cells; WBC, white blood cell.

Although cardiomyocytes displayed diverse cellular heterogeneity, they grouped into three major subclasses which included not only distinct cardiomyocytes from the atrial and ventricular cardiac chambers as previously described[3–7], but also non-chambered cardiomyocytes (**Figure 2a**). A large proportion of the cardiomyocyte diversity was observed unexpectedly in atrial and ventricular cardiomyocyte subpopulations, which consisted of more specific cell types and cell states that were associated with developmental stages or regions of the heart (**Figure 2d,e**). In particular, *NRF2F2* expressing atrial cardiomyocytes segregated into non-proliferating left and right atrial cardiomyocytes, as detected by *PITX2/PANCR* and *ANGPT1*/*ADM* markers, respectively, as well as *MKI67*+ proliferative cardiomyocytes that were present in both left and right atria but lowly expressed either left or right atrial markers (**Figure 2d,e**). In contrast to the proliferative atrial cardiomyocytes, these left and right atrial cardiomyocytes further subdivided into several cell states that corresponded to developmental ages, suggesting their continuing differentiation during heart development (**Figure 2d,e**). On the other hand, *MYL2*/*IRX4*-expressing ventricular cardiomyocytes initially clustered into subpopulations correlating to developmental age and the differential expression of maturity-related sarcomeric and metabolic genes such as *LMOD3* and *CKMT2,* which are involved in cardiomyocyte actin filament organization and oxidative phosphorylation[31,32] (**Figure 2d,e**). Finally, our examination of all regions of the heart combined with our relatively high number of cells sequenced compared to prior cardiac developmental studies[6,7], enabled the identification of more rare cardiomyocyte populations including *BMP2*+ non-chamber cardiomyocytes and an unknown non-descript *TTN*+ cardiomyocyte subpopulation that may correspond to the

recently reported MYOZ2-enriched cardiomyocyte[6] (**Figure 2d,e**). These non-chambered cardiomyocytes primarily segregated into *RSPO3/MSX2*+ atrioventricular canal/node and *SHOX2/TBX18*+ inflow tract/pacemaker cardiomyocytes, which differentially expressed *ISL1, PITX2* and *TBX3,* known transcription factors involved in regulating pacemaker sinoatrial node versus inflow tract development[33,34]. Thus, these findings provide new developmental insight into not only previously reported chamber-related cardiomyocytes[3,4,6,7] but also more specialized cardiomyocyte subtypes critical for regulating electrical cardiac conduction.

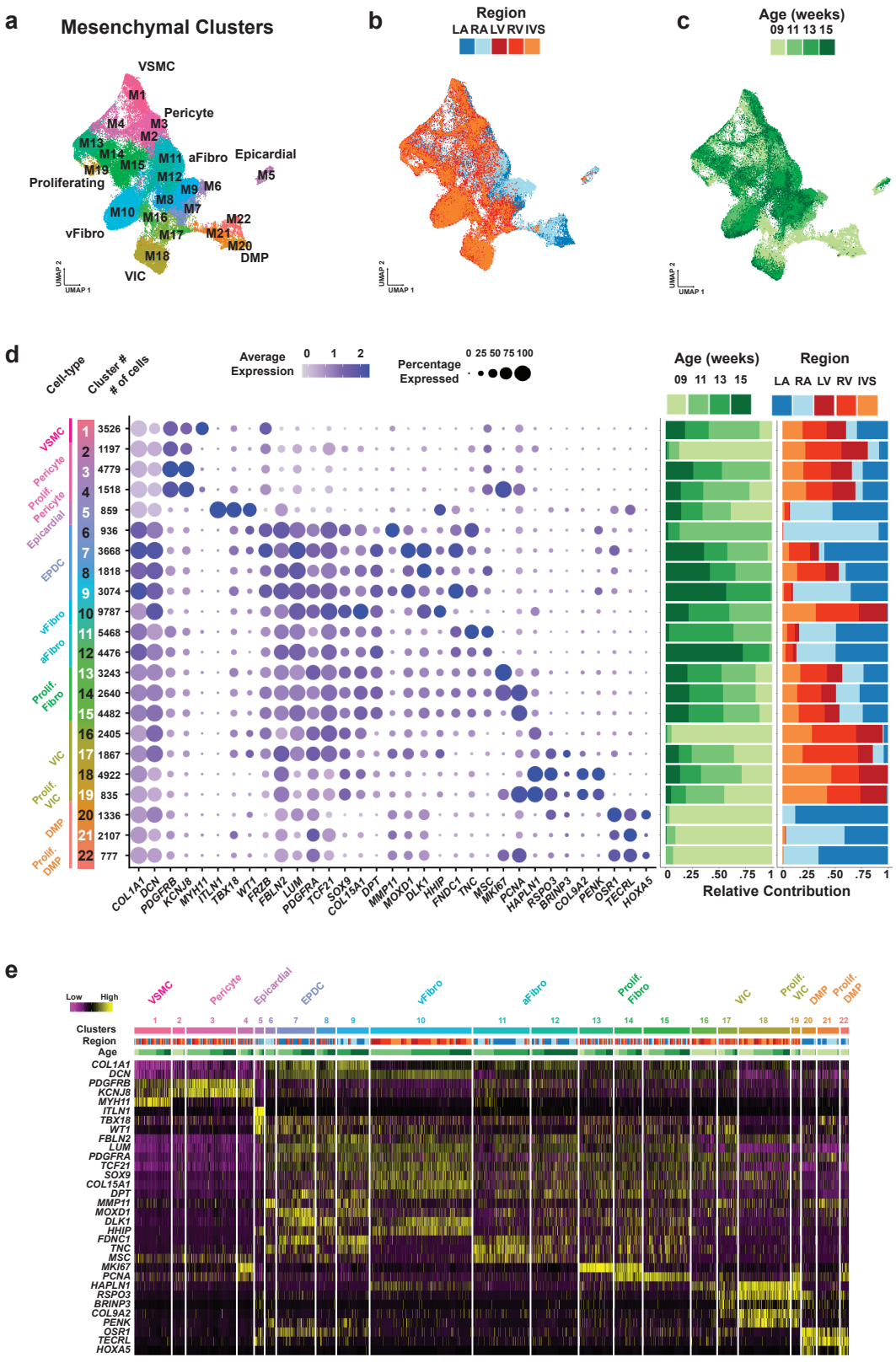
Figure 2. Single-cell transcriptomic profiling reveals heterogeneity within the cardiomyocyte partition of the human heart.
a-c, UMAP of subclustered cardiomyocyte cells colored by cluster b, Region, and c, Age. **d**, Dotplot of marker gene expression and relative contribution of Age and Region to each cluster, labeled by putative cell types and states. **e**, Heatmap of marker gene expression and relative contribution of Age and Region to each cluster, showing the information on a single cell level. Prolif., proliferating.


a
Cardiomyocyte Clusters
Non-Chambered CMs
C8
C22
C20
C1
C10
Unknown
C21
C16
C9
C17
C2
C18
C19
C4
C3
C12
C11
Ventricular Chambered CMs
C5
C13
C6
Atrial Chambered CMs
C7
C14
C15
UMAP 2
UMAP 1
b
Region
LA RA LV RV IVS
c
Age (weeks)
09 11 13 15
d
Cell-type
Cluster #
# of cells
Average Expression
Percentage Expressed
Early Ventricular
Late Ventricular
Prolif. Ventricular
Right Atrial
Left Atrial
Prolif. Atrial
IFT-like
AVN-like
Relative Contribution
e
Low
High
Early Ventricular
Late Ventricular
Proliferating vCM
Right Atrial
Left Atrial
Proliferating aCM
IFT-like
AVN-like
Clusters
Region
Age

The mesenchymal population exhibited cellular heterogeneity comparable to the cardiomyocyte population and clustered into three major subclasses: epicardial, fibroblast-like and vascular support cells. In particular, these cells displayed transcriptional similarity with each other as visualized by Uniform Manifold Approximation and Projection (UMAP) (**Figure 3a**), suggesting that many of them are developmentally related and may derive from epicardium-derived progenitor cells (EPDCs) as previously reported[1]. Consistent with this notion, EPDCs, vascular support cells, and many fibroblast-like cells expressed *TBX18* and *FRZB*, and could be partitioned into *TCF21+/PDGRFA*+ EPDCs and fibroblast-like cells versus *TCF21-/PDGFRB*+ vascular support cells (**Figure 3d,e**). While vascular support cells (4 subclusters) were composed of *MYH11*+ vascular smooth cells and *MYH11-/KCNJ8*+ pericyte subgroups, fibroblast-like cells exhibited greater cellular complexity (8 subclusters) and segregated into *DPT*+ fibroblasts and *HAPLN1*+ valvular interstitial cells (VICs). These fibroblasts clustered primarily based on their anatomic location (atrium, *TNC+/MSC+/FDNC1*+; ventricle, *HHIP+/TNC-/MSC-/FDNC1-*) and expression of proliferative markers (*MKI67+/PCNA+),* whereas VICs clustered into subpopulations correlating to developmental age (**Figure 3d,e**). In contrast, a subset of *HAPLN1+/PENK*+ VICs as well as *OSR1+/TECRL*+ dorsal mesenchymal protrusion (DMP) cells were also discovered that transcriptionally differed from the aforementioned EPDC-related fibroblasts and VICs, particularly by their lower expression of *TBX18* and *TCF21* (**Figure 3d,e**)*,* suggesting that these VICs and DMP cells may derive from alternative sources such as endocardial cells and the second heart field, respectively, as described[1,35,36]. Supporting their transient developmental role in creating the atrial

septum and atrioventricular cushions[35,36], these DMP cells were observed in the atria and only at 9 and 11 weeks of gestation, (**Figure 3c,d,e**). On the other hand, all VICs were mainly observed in the ventricle but a small proportion of *TBX18*+ EPDC-related VICs could be detected in the atria, suggesting differing roles and contributions between *TBX18*+ versus *TBX18*- VIC subpopulations. Despite these mesenchymal cells displaying transcriptional overlap and potential functional similarity, our detailed single cell analysis provides new insight into the developmental complexity of these similar cell types.

Figure 3. Single-cell transcriptomic profiling reveals heterogeneity within the mesenchymal partition of the human heart.
a-c, UMAP of subclustered mesenchymal cells colored by cluster b, Region, and c, Age. **d**, Dotplot of marker gene expression and relative contribution of Age and Region to each cluster, labeled by putative cell types and states. **e**, Heatmap of marker gene expression and relative contribution of Age and Region to each cluster, showing the information on a single cell level. aFibro, atrial fibroblast; DMP, dorsal mesenchymal protrusion; EPDC, epicardial-derived cell; Prolif., proliferating; vFibro, ventricular fibroblast; VICs, valve interstitial cells; VSMC, vascular smooth muscle cells.


a
Mesenchymal Clusters
VSMC
M1
Pericyte
M4
M3
M2
M13
M14
M11
aFibro
Epicardial
M19
M15
M12
M5
Proliferating
M9
M6
M8
M7
M10
M16
M22
vFibro
M17
M21
M20
DMP
M18
VIC
UMAP 2
UMAP 1
b
Region
LA RA LV RV IVS
c
Age (weeks)
09 11 13 15
d
Cell-type
Cluster #
# of cells
Average Expression
0 1 2
Percentage Expressed
0 25 50 75 100
Age (weeks)
Region
VSMC
Pericyte
Prolif. Pericyte
Epicardial
EPDC
vFibro
aFibro
Prolif. Fibro
VIC
Prolif. VIC
DMP
Prolif. DMP
1 3526
2 1197
3 4779
4 1518
5 859
6 936
7 3668
8 1818
9 3074
10 9787
11 5468
12 4476
13 3243
14 2640
15 4482
16 2405
17 1867
18 4922
19 835
20 1336
21 2107
22 777
COL1A1 DCN PDGFRB KCNJ8 MYH11 ITLN1 TBX18 WT1 FRZB FBLN2 LUM PDGFRA TCF21 SOX9 COL15A1 DPT MMP11 MOXD1 DLK1 HHIP FNDC1 TNC MSC MKI67 PCNA HAPLN1 RSPO3 BRINP3 COL9A2 PENK OSR1 TECRL HOXA5
0 .25 .50 .75 1
Relative Contribution
e
Low High
VSMC
Pericyte
Epicardial
EPDC
vFibro
aFibro
Prolif. Fibro
VIC
Prolif. VIC
DMP
Prolif. DMP
Clusters
Region
Age

Finally, endothelial-like, blood and neuronal-like cell populations together contributed to ~40% of the observed cardiac cellular heterogeneity but each displayed distinctive cellular diversity (**Figures 4-6**). *PECAM1*+ endothelial/endocardial cells grouped into three major subclasses including *CLDN5+/LEPR-/LYVE*- blood endothelial cells (BECs), *CLDN5+/LEPR-/LYVE1*+ lymphatic endothelial cells and *CLDN5*-/*LEPR*+/*LYVE1*-endocardial cells. While BECs subdivided into *HEY1*+ arterial, *NR2F2*+ venous, *PGF*+ capillary and *MKI67*+ proliferative endothelial cells, we also observed a spectrum of endocardial cell subpopulations that in many cases were enriched in either the atrium or ventricle. In line with these regional findings, neuronal-like lineages, which clustered into Schwann-like and neural crest-like cells, were primarily observed in the atria. On the other hand, many known blood cell lineages were discovered throughout all regions of the heart[3–6], but macrophages represented the highest population of blood cells, supporting their tissue resident-like role in heart homeostasis as recently reported[3].

Figure 4. Single-cell transcriptomic profiling reveals heterogeneity within the endothelial partition of the human heart.
a-c, UMAP of subclustered endothelial cells colored by cluster b, Region, and c, Age. **d**, Dotplot of marker gene expression and relative contribution of Age and Region to each cluster, labeled by putative cell types and states. **e**, Heatmap of marker gene expression and relative contribution of Age and Region to each cluster, showing the information on a single cell level. aEndocardial, atrial endocardial; BEC, blood endothelial cell; LEC, lymphatic endothelial cell; Prolif, proliferating; vEndocardial, ventricular endocardial.

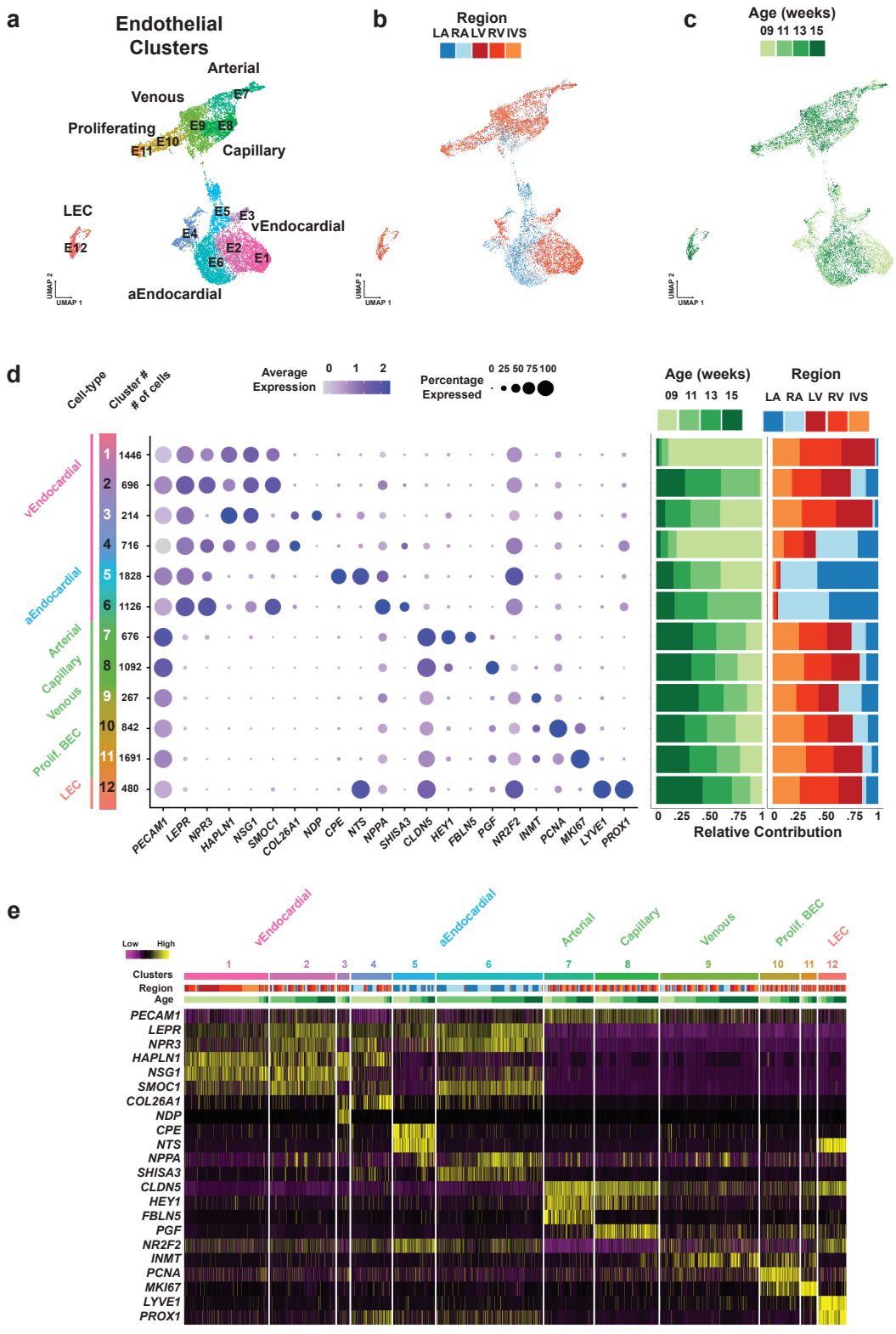

a
Endothelial Clusters
Arterial
Venous
Proliferating
Capillary
LEC
vEndocardial
aEndocardial
E1
E2
E3
E4
E5
E6
E7
E8
E9
E10
E11
E12
UMAP 1
UMAP 2
b
Region
LA RA LV RV IVS
c
Age (weeks)
09 11 13 15
d
Cell-type
Cluster #
# of cells
Average Expression
0 1 2
Percentage Expressed
0 25 50 75 100
Age (weeks)
09 11 13 15
Region
LA RA LV RV IVS
vEndocardial
aEndocardial
Arterial
Capillary
Venous
Prolif. BEC
LEC
1 1446
2 696
3 214
4 716
5 1828
6 1126
7 676
8 1092
9 267
10 842
11 1691
12 480
PECAM1 LEPR NPR3 HAPLN1 NSG1 SMOC1 COL26A1 NDP CPE NTS NPPA SHISA3 CLDN5 HEY1 FBLN5 PGF NR2F2 INMT PCNA MKI67 LYVE1 PROX1
0 .25 .50 .75 1
Relative Contribution
e
Low High
vEndocardial
aEndocardial
Arterial
Capillary
Venous
Prolif. BEC
LEC
Clusters
Region
Age
PECAM1 LEPR NPR3 HAPLN1 NSG1 SMOC1 COL26A1 NDP CPE NTS NPPA SHISA3 CLDN5 HEY1 FBLN5 PGF NR2F2 INMT PCNA MKI67 LYVE1 PROX1

Figure 5. Single-cell transcriptomic profiling reveals heterogeneity within the neuronal-like partition of the human heart.
a-c, UMAP of subclustered neuronal-like cells colored by cluster b, Region, and c, Age. **d**, Dotplot of marker gene expression and relative contribution of Age and Region to each cluster, labeled by putative cell types and states. **e**, Heatmap of marker gene expression and relative contribution of Age and Region to each cluster, showing the information on a single cell level. Prolif., proliferating.

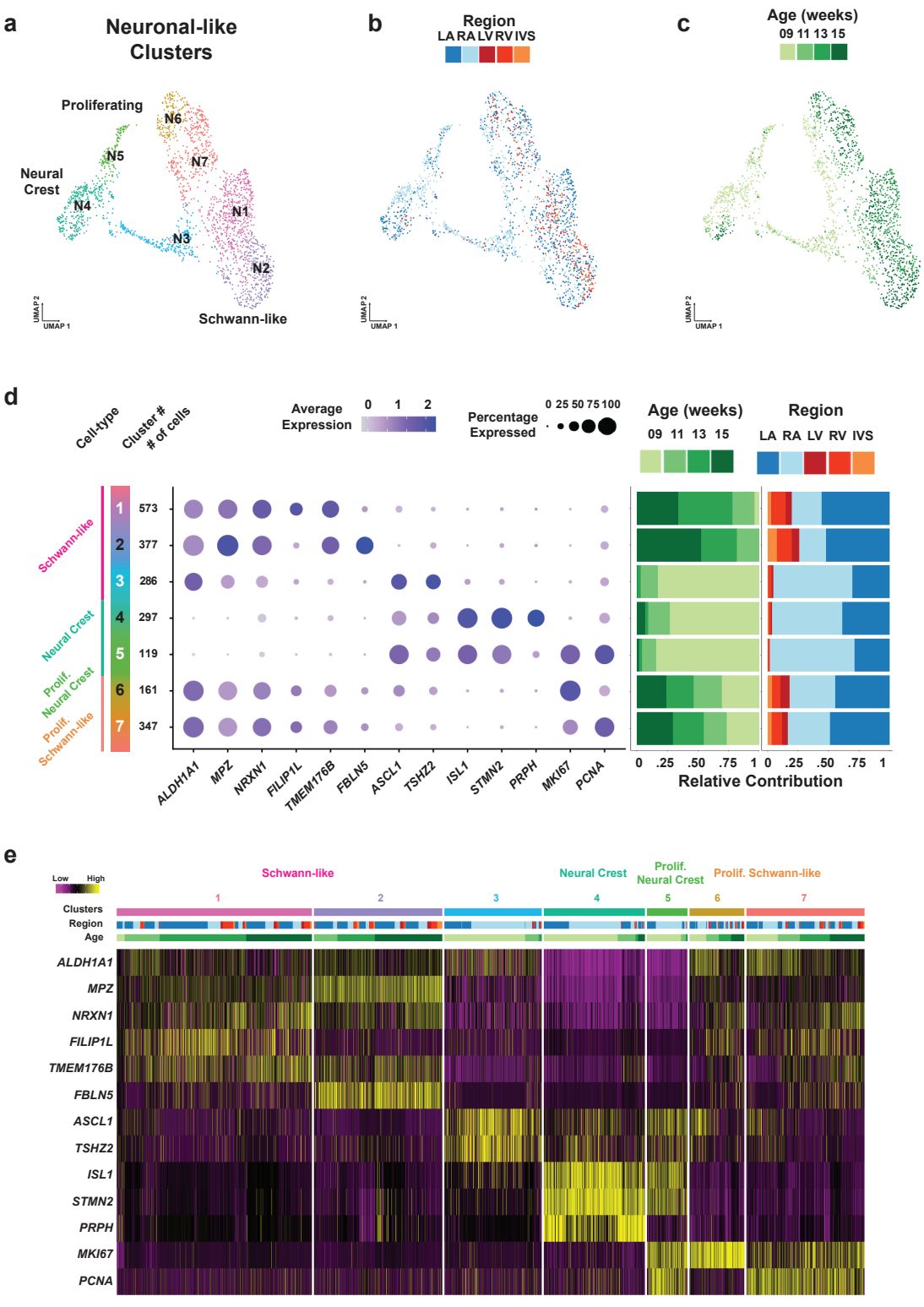
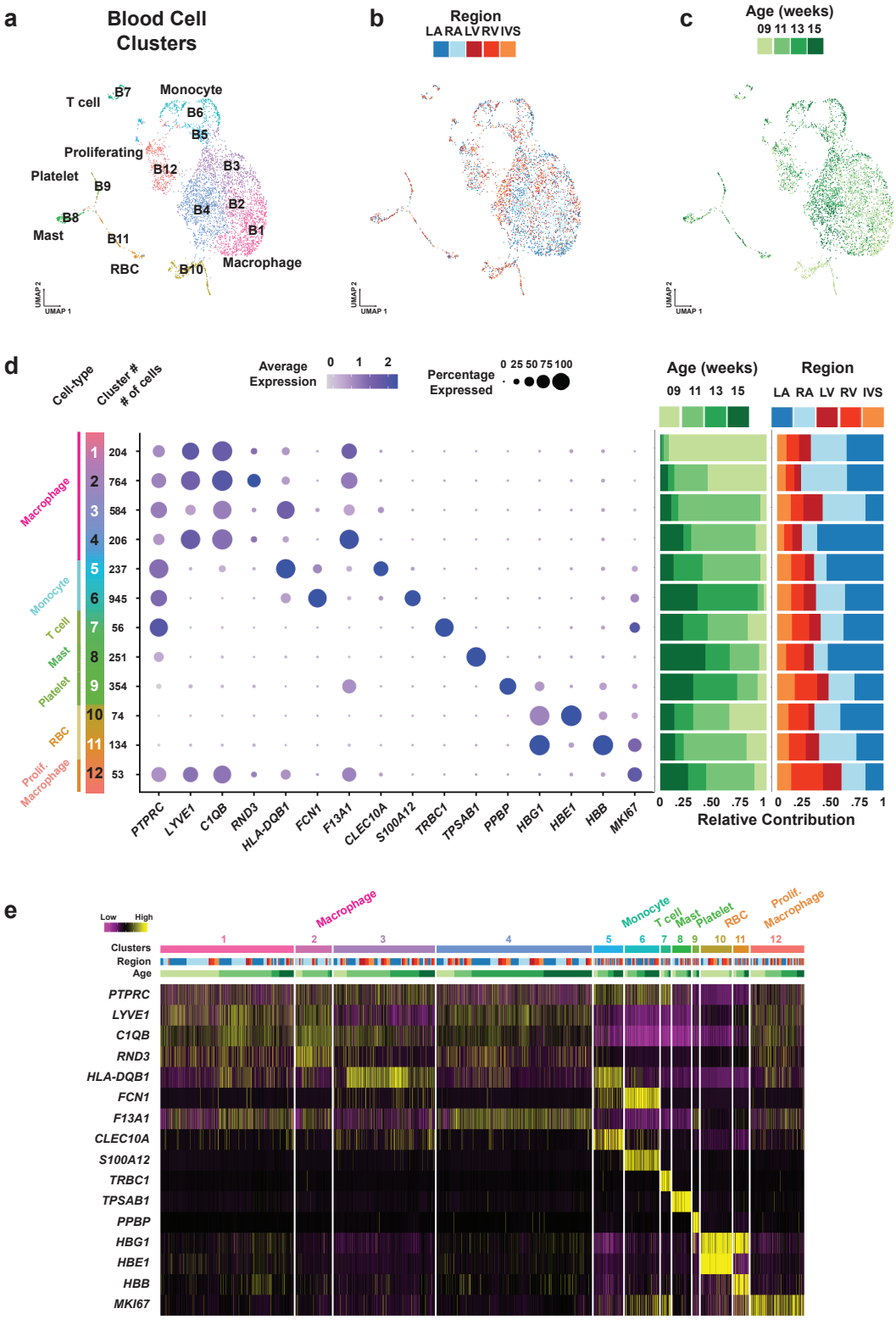

a
Neuronal-like Clusters
Proliferating
N6
N5
N7
Neural Crest
N4
N1
N3
N2
Schwann-like
UMAP 2
UMAP 1
b
Region
LA RA LV RV IVS
c
Age (weeks)
09 11 13 15
d
Cell-type
Cluster # # of cells
Average Expression
0 1 2
Percentage Expressed
0 25 50 75 100
Age (weeks)
09 11 13 15
Region
LA RA LV RV IVS
Schwann-like
Neural Crest
Prolif. Neural Crest
Prolif. Schwann-like
1 573
2 377
3 286
4 297
5 119
6 161
7 347
ALDH1A1
MPZ
NRXN1
FILIP1L
TMEM176B
FBLN5
ASCL1
TSHZ2
ISL1
STMN2
PRPH
MKI67
PCNA
0 .25 .50 .75 1
Relative Contribution
e
Low High
Schwann-like
Neural Crest
Prolif. Neural Crest
Prolif. Schwann-like
1 2 3 4 5 6 7
Clusters
Region
Age
ALDH1A1
MPZ
NRXN1
FILIP1L
TMEM176B
FBLN5
ASCL1
TSHZ2
ISL1
STMN2
PRPH
MKI67
PCNA

Figure 6. Single-cell transcriptomic profiling reveals heterogeneity within the blood partition of the human heart.
a-c, UMAP of subclustered blood cells colored by cluster b, Region, and c, Age. **d**, Dotplot of marker gene expression and relative contribution of Age and Region to each cluster, labeled by putative cell types and states. **e**, Heatmap of marker gene expression and relative contribution of Age and Region to each cluster, showing the information on a single cell level. Prolif., proliferating; RBC, red blood cell.


a
Blood Cell Clusters
T cell
B7
Monocyte
B6
B5
Proliferating
B12
B3
Platelet
B9
B4
B2
B8
B1
Mast
B11
RBC
B10
Macrophage
UMAP 2
UMAP 1
b
Region
LA RA LV RV IVS
c
Age (weeks)
09 11 13 15
d
Cell-type
Cluster #
# of cells
Average Expression
0 1 2
Percentage Expressed
0 25 50 75 100
Age (weeks)
09 11 13 15
Region
LA RA LV RV IVS
Macrophage
Monocyte
T cell
Mast
Platelet
RBC
Prolif. Macrophage
1 204
2 764
3 584
4 206
5 237
6 945
7 56
8 251
9 354
10 74
11 134
12 53
PTPRC
LYVE1
C1QB
RND3
HLA-DQB1
FCN1
F13A1
CLEC10A
S100A12
TRBC1
TPSAB1
PPBP
HBG1
HBE1
HBB
MKI67
0 .25 .50 .75 1
Relative Contribution
e
Low High
Macrophage
Monocyte
T cell
Mast
Platelet
RBC
Prolif. Macrophage
1 2 3 4 5 6 7 8 9 10 11 12
Clusters
Region
Age
PTPRC
LYVE1
C1QB
RND3
HLA-DQB1
FCN1
F13A1
CLEC10A
S100A12
TRBC1
TPSAB1
PPBP
HBG1
HBE1
HBB
MKI67

Overall, our comprehensive human heart cell atlas uncovered a multitude of abundant and rare (<1%) cell subpopulations creating the human heart (**Figures 2-7**). Because our tissue dissection and single-cell isolation protocol provided the opportunity to examine the entire heart, our single cell analyses identified additional cardiac cell subpopulations beyond those recently reported[3–6,37]. As a result, we not only annotated new cell populations but also discovered other populations that remain to be defined. Interrogating each dissected cardiac chamber/IVS revealed that many identified cardiac lineages, particularly those inherent to the heart (cardiomyocytes, mesenchymal and endocardial), were region/chamber specific, suggesting that these distinct cardiac lineages are cell types that become specialized during development to accommodate the function of each cardiac structure (**Figures 2-4**). Consistent with this notion, these specialized cardiac subpopulations expressed combinations of genes specific to each cardiac structure but typically no proliferative markers, and further segregated in many instances according to developmental age, thus supporting that these cells may have differentiated and acquired specific cellular roles based on their regional environment (**Figures 2-4**). In contrast, cardiac subpopulations expressing proliferative markers were frequently observed across regions/structures, but expressed lowly all cardiac structure-specific genes, suggesting that these cells may be progenitor-like with the potential to expand and then further differentiate into more refined cell subpopulations within specific cardiac structures. While these scRNA-seq findings provide new developmental insights into the diverse cell populations creating the heart, how these cells interact and organize into complex morphologic communities/structures crucial for heart function and cell specialization remains to be illuminated.

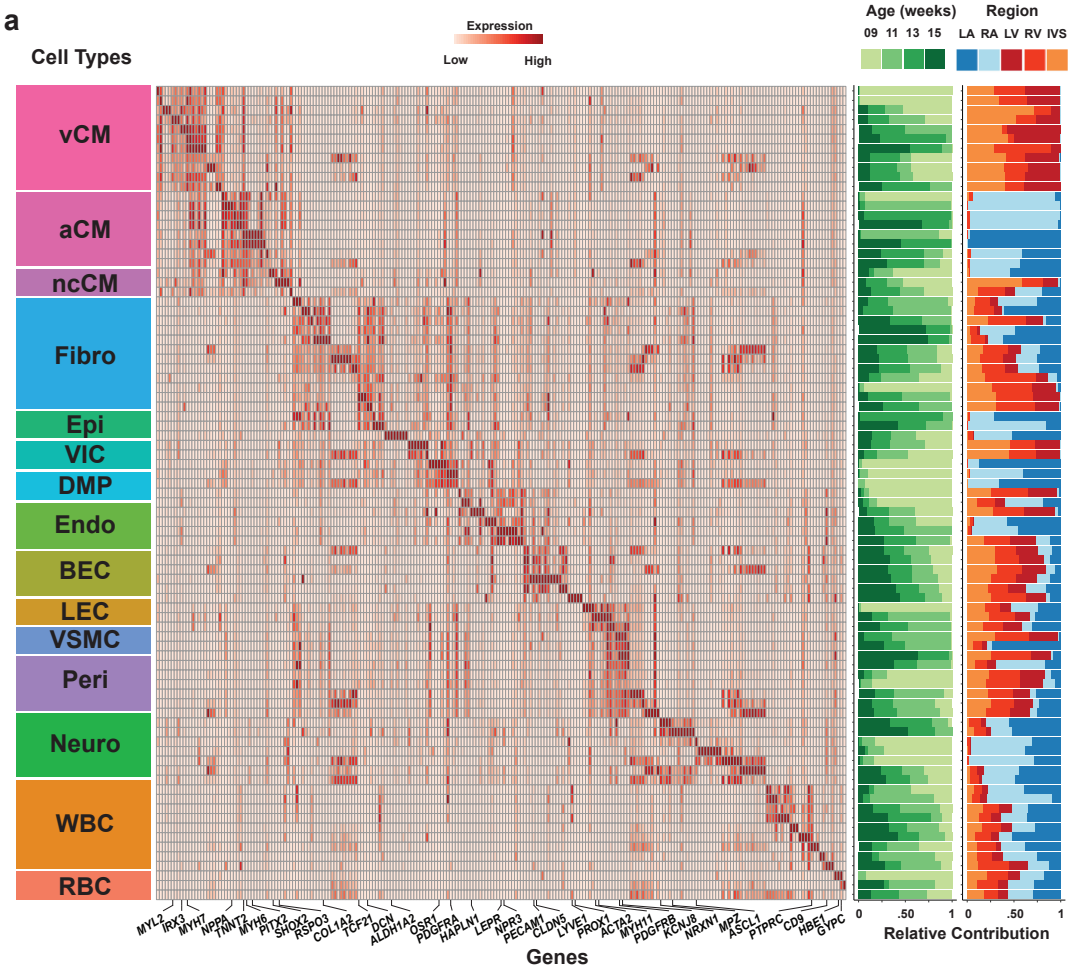


Figure 7. Single-cell transcriptomic profiling reveals a multitude of diverse cellular populations in the developing human heart.

a, Heatmap of the 75 subpopulations across all of the subclustered partitions, with marker gene expression defined by NS-Forest3[38], and relative contribution of Age and Region to each cluster. aCM, atrial cardiomyocyte; BEC, blood endothelial cell; CMs, Cardiomyocytes; DMP, dorsal mesenchymal protrusion; Endo, endocardial; Epi, epicardial; LEC, lymphatic endothelial cell; Neuro, neuronal-like; ncCM, non-chambered cardiomyocyte; Peri, Pericyte; RBC, Red Blood Cell; vCM, ventricular cardiomyocyte; VICs, valve interstitial cells; VSMC, vascular smooth muscle cells; WBC, white blood cell.

MERFISH imaging reveals spatial organization of cardiac cell types and their unanticipated specialization within specific cardiac structures/regions.

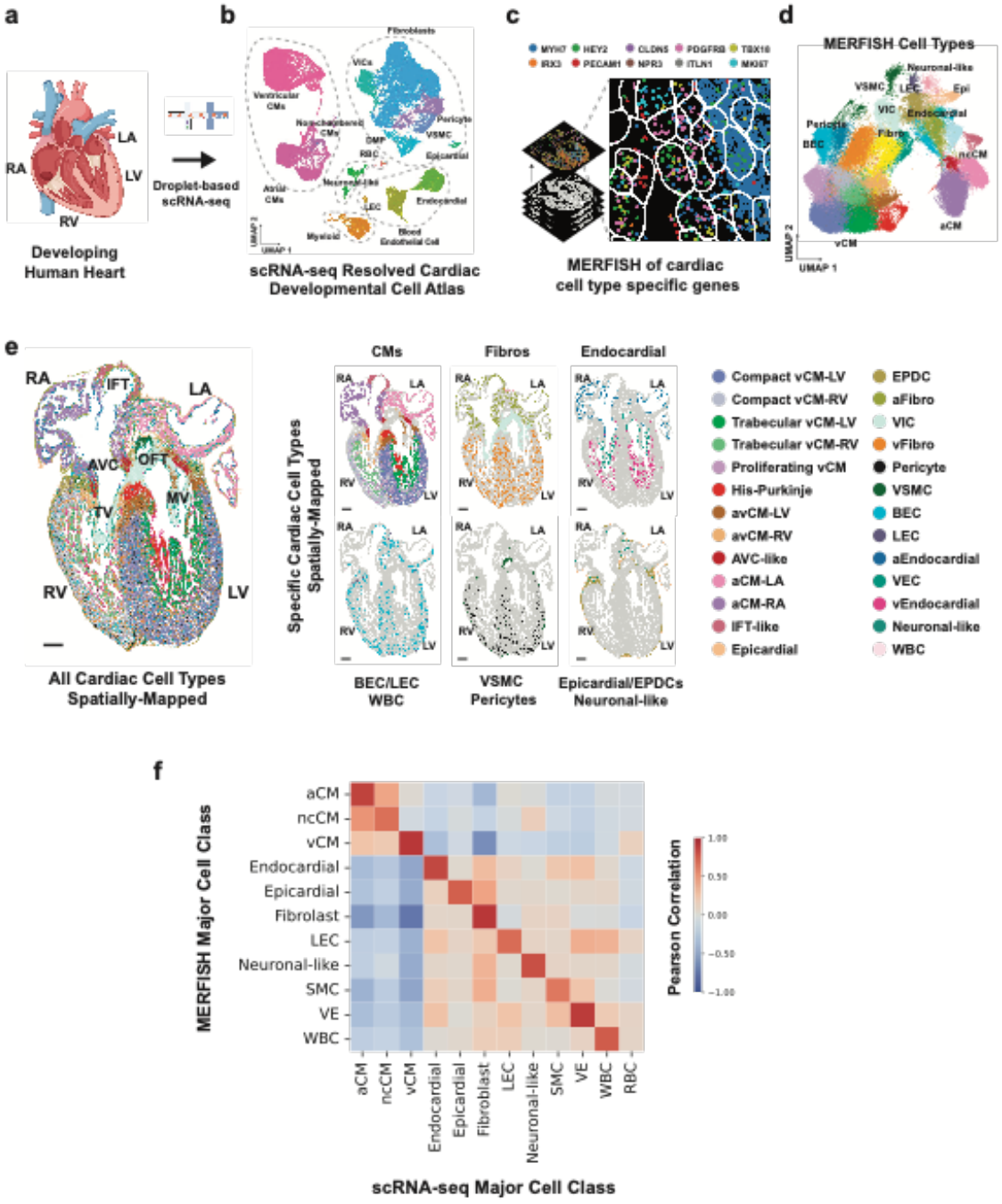
To explore the interactive cellular mechanisms directing cardiac morphogenesis and remodeling, we interrogated the spatial organization of cardiovascular cell types that we characterized by scRNA-seq using multiplexed error-robust fluorescence in situ hybridization (MERFISH)[39]. To this end, we identified a list of 238 cell-type/cell-state specific genes after applying a NS-Forest3 classifier[38] on our scRNA-seq analysis (**Figure 7**). In particular, these genes were selected based on, but not limited to the following strategies: 1) cell-type marker genes that are based on prior knowledge[3–7], 2) differential/specific gene expression of more refined subpopulations identified by scRNA-seq and 3) transcription factors that play a role in cardiac development. Using smFISH probes designed for these selected genes, we performed MERFISH studies on a series of coronal slices at approximately 600 mm along the anterior-posterior axis of 12-13 weeks of gestation human hearts which captured all major cardiac structures (**Figure 8c-e**). Distinct RNA molecules were clearly identified and assigned to individual cells as segmented by total polyadenylated RNA staining and 4′,6-diamidino-2-phenylindole (DAPI) staining, which allowed for the detection of cellular boundaries[40] (**Figure 9a**). After segmentation and adaptive filtering, we obtained 108.2 million transcripts from 258,237 cells across three experiments. On average, 365 transcripts from 85 genes per cell were detected from this analysis, whereas only 208 transcripts from 51 genes per cell were discovered by scRNA-seq using the same target gene list, highlighting the high RNA capture efficiency of MERFISH[39]. Additionally, the levels of RNA transcripts identified by each MERFISH experiment showed high correlation

(Pearson correlation coefficient > 0.95) between experimental replicates (**Figure 9b**) and our scRNA-seq datasets (**Figure 9c**).

To discover cell populations from these MERFISH studies, a semi-unsupervised, community-detection-based clustering algorithm[30] was applied to MERFISH single cell expression data. This analysis identified 27 cell clusters that grouped into subclasses closely correlating to the major developmental subclasses discovered by scRNA-seq, except for red blood cells, potentially because of the exclusion of their marker genes from the MERFISH gene library (**Figure 8f**). These 27 cell clusters were defined by marker gene expression as well as spatial location (**Figure 10, 11**).

Figure 8. Molecular and spatial atlas reveal a multitude of diverse cellular populations cooperating in the human heart during development

a, Single-cell RNA sequencing (scRNA-seq) experimental design during human heart development, with **b,** Uniform Manifold Approximation and Projection (UMAP) of ~143k cells colored by major cell type. **c**, Schematic of MERFISH measurements, where combinatorial smFISH imaging was used to identify 238 genes, with pseudo-colored dots marking localizations of individual molecules of ten example RNA species marking distinct major cell populations. **d,** UMAP of ~250k imaged cells colored by major cell type. **e**, (Left) Spatial distribution of all cell types across the heart. (Right) Spatial distribution of major cell classes the heart. Cells are marked with cell segmentation boundaries and colored by cell types as indicated. **f,** Heatmap of Pearson correlation between the expression profiles of major cell classes within the scRNA-seq and the MERFISH datasets. aCM, atrial cardiomyocyte; AVC, atrioventricular canal; avCM, muscular leaflet of the valve cardiomyocyte; BEC, Blood Endothelial cell; CMs, Cardiomyocytes; Endo, endocardial; EPDC, epicardial-derived cell; Epi, epicardial; IFT-like, inflow tract-like; LA, left atrium; LEC, lymphatic endothelial cell; LV, left ventricle; MV, mitral valve; Neuro, neuronal-like; ncCM, non-chambered cardiomyocyte; OFT, outflow tract; Peri, Pericyte; RA, right atrium; RBC, Red Blood Cell; RV, right ventricle; TV, tricuspid valve; vCM, ventricular cardiomyocyte; vEndocardial, ventricular endocardial; VICs, valve interstitial cells; VSMC, vascular smooth muscle cells; WBC, white blood cell. Scale bar, 250 μm.


a
RA
LA
RV
LV
Developing
Human Heart
Droplet-based
scRNA-seq
b
scRNA-seq Resolved Cardiac
Developmental Cell Atlas
c
MERFISH of cardiac
cell type specific genes
d
MERFISH Cell Types
e
All Cardiac Cell Types
Spatially-Mapped
Specific Cardiac Cell Types
Spatially-Mapped
CMs
Fibros
Endocardial
BEC/LEC
WBC
VSMC
Pericytes
Epicardial/EPDCs
Neuronal-like
Compact vCM-LV
Compact vCM-RV
Trabecular vCM-LV
Trabecular vCM-RV
Proliferating vCM
His-Purkinje
avCM-LV
avCM-RV
AVC-like
aCM-LA
aCM-RA
IFT-like
Epicardial
EPDC
aFibro
VIC
vFibro
Pericyte
VSMC
BEC
LEC
aEndocardial
VEC
vEndocardial
Neuronal-like
WBC
f
MERFISH Major Cell Class
aCM
ncCM
vCM
Endocardial
Epicardial
Fibrolast
LEC
Neuronal-like
SMC
VE
WBC
RBC
scRNA-seq Major Cell Class
Pearson Correlation

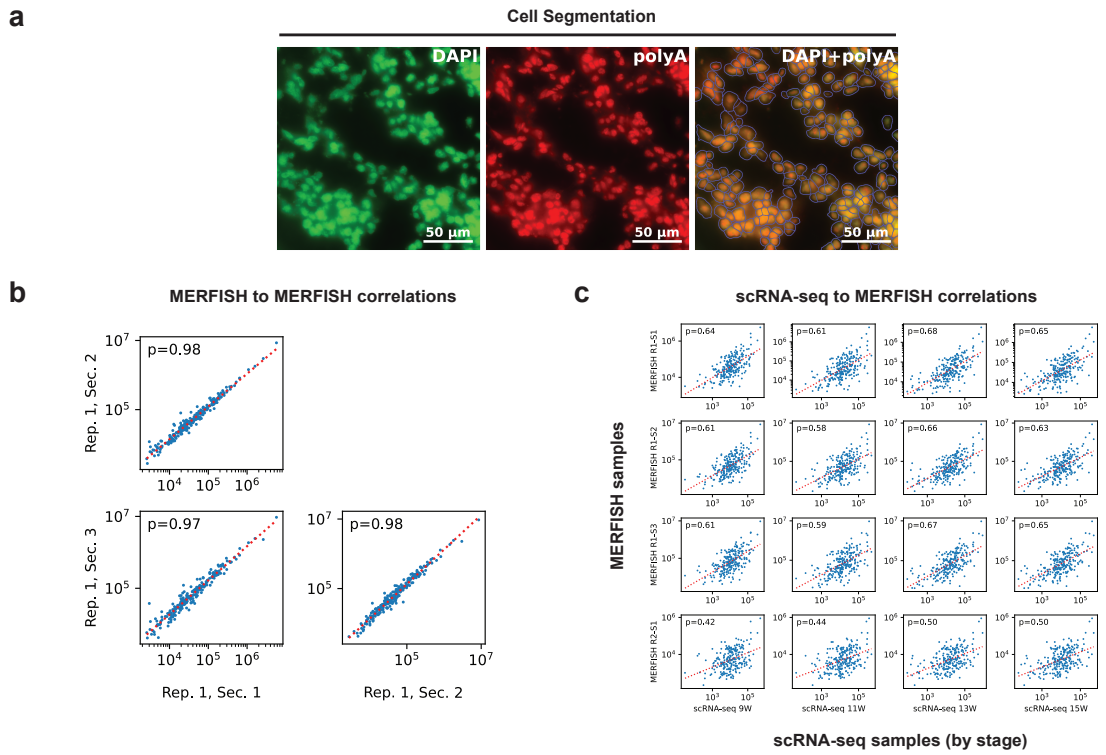


Figure 9. Cell segmentation of MERFISH images and replicate reproducibility of MERFISH data

a, Example images of cell segmentation after MERFISH imaging. Cell boundaries were defined using CellPose[40] with DAPI and polyA staining as input images. **b**, Pearson correlation of the counts of each of the 238 target genes for each of the three replicate sections used for MERFISH. **c**, Pearson correlation between each of the MERFISH replicate sections and the scRNA-seq timepoints of the counts of each of the 238 target genes. Scale bar, 50 µm.

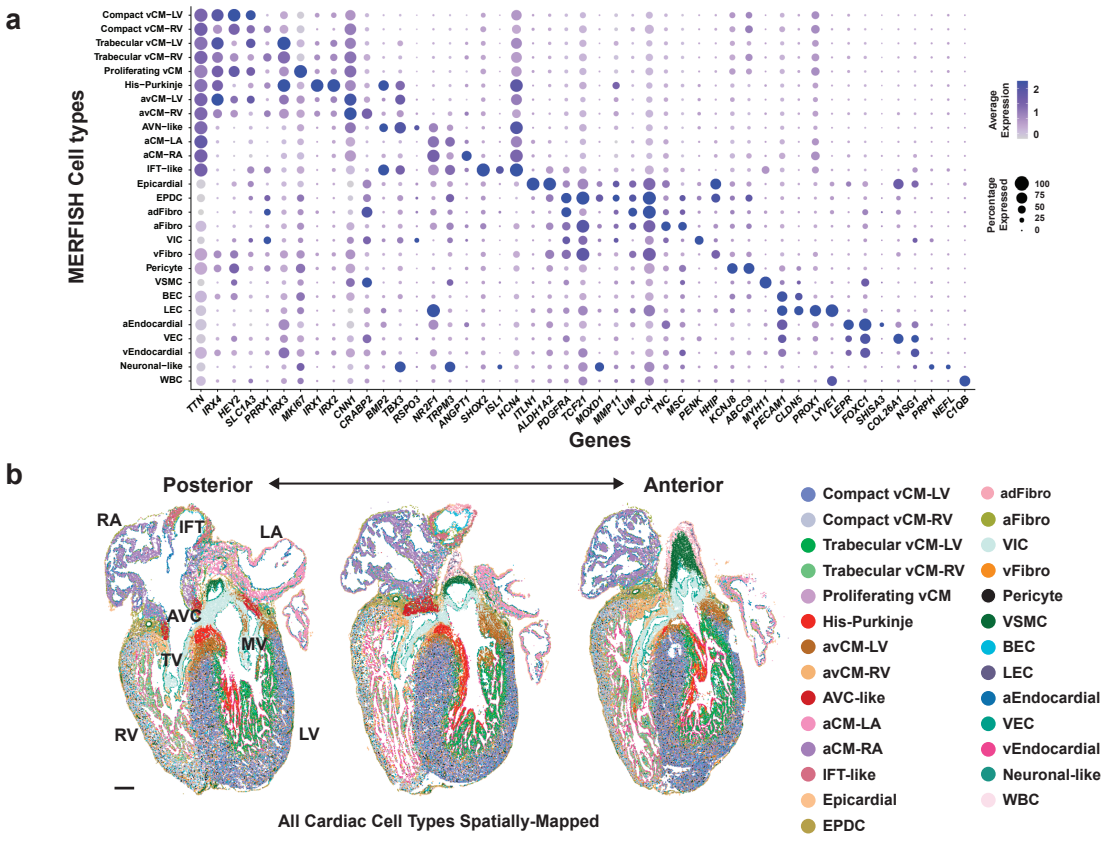


Figure 10. Identification of the 27 populations identified in the MERFISH dataset
a, Dotplot of marker gene expression for each cell type identified in the MERFISH dataset. **b**, Spatial plots of the three replicate MERFISH sections, ordered left to right from posterior to anterior, colored by identified MERFISH cell types. Scale bar, 250 μm.

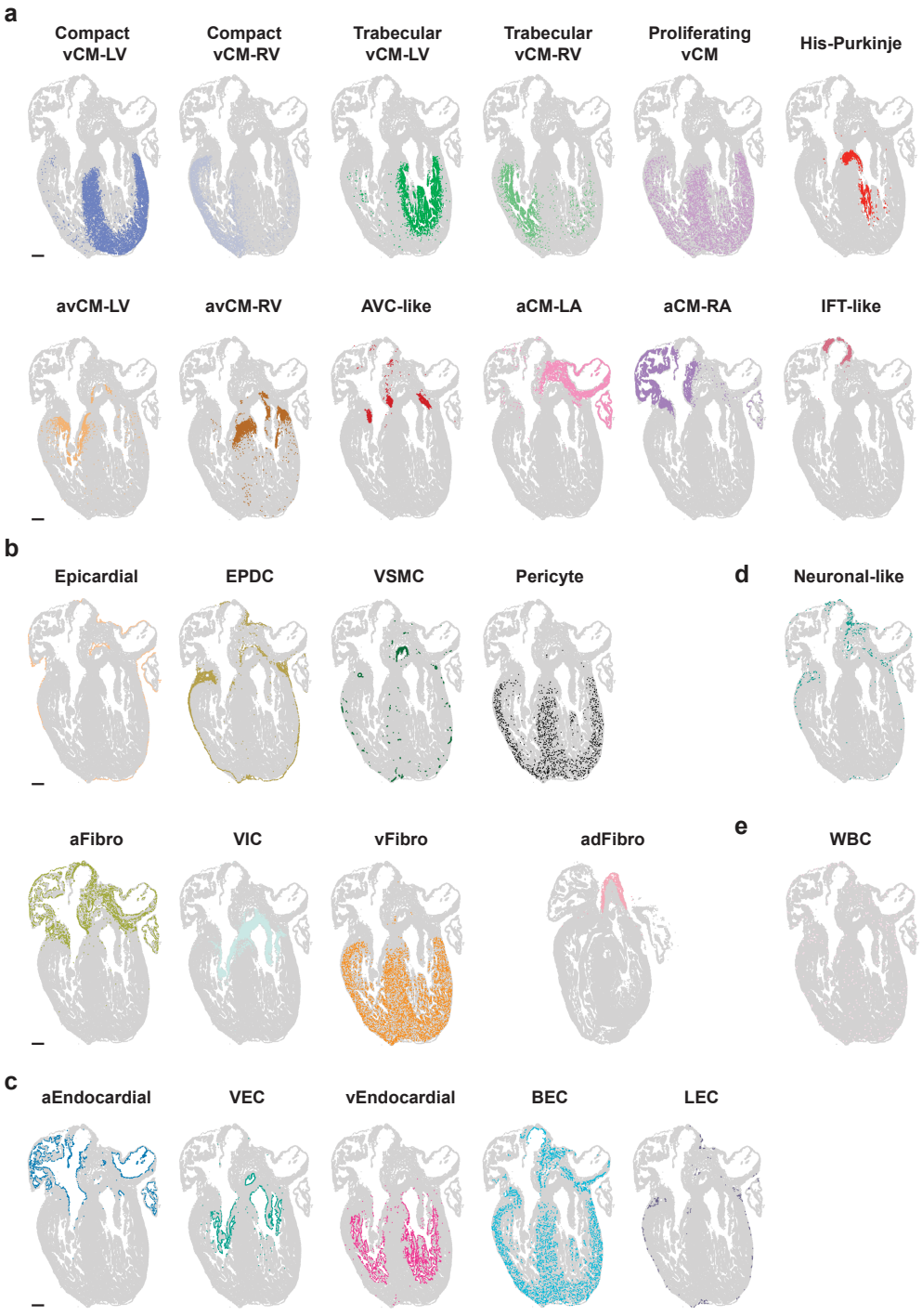


Figure 11. Spatial mapping of the 27 individual populations identified in the MERFISH dataset

MERFISH clusters individually mapped back onto the heart for **a**, cardiomyocyte related clusters, **b**, epicardial, EPDC, and vascular support related clusters, **c**, endothelial related clusters, **d**, a neuronal-like cluster, and **e**, a blood related cluster. Scale bar, 250 μm.

Specific cardiomyocyte lineages are spatially localized to specific anatomic cardiac structures.

In line with our scRNA-seq data, cardiomyocyte lineages represented the largest portion of major cell clusters identified from MERFISH analyses (12 of 27 clusters). Spatial mapping of these transcripts revealed that identified cardiomyocyte clusters displayed distinct regional and structural distributions across the heart, corroborating our scRNA-seq regional findings (**Figure 10b, 11a**). Unlike recent *in situ* RNA-seq studies of the heart[6,41], these MERFISH results additionally provided high-resolution spatial imaging that enabled tracking of individual cells to detailed structures of the heart. Combined with gene marker analyses, this spatial anatomic and histologic information not only confirmed cardiomyocytes identified from scRNA-seq analysis but also resolved the classification of others not clearly defined (**Figure 10b, 11a**). As a result, we discovered that these cardiomyocyte clusters populated distinct anatomic domains of chambered and non-chambered regions of the heart and frequently did not spatially overlap with each other (**Figure 11a**).

Chambered cardiomyocytes were broadly divided into *NR2F1+* and *IRX4+* cardiomyocytes that contributed mutually exclusively to the atrial and ventricular chambers, respectively (**Figure 10b, 11a**). Atrial cardiomyocytes spatially segregated into those residing in the left and right atria which were transcriptionally distinguished by *ANGPT1* as observed in our scRNA-seq analyses (**Figure 3d,e**). On the other hand, ventricular cardiomyocytes displayed more cellular complexity than observed in the scRNA-seq analysis, and clustered into those specifically occupying not only the left and right ventricles but also more distinct anatomic subdomains within the outer and

inner layers of the ventricles (**Figure 11a**). Both known and new markers were discovered to be enriched in these ventricular cardiomyocyte clusters including *SLC1A3* and *PRRX1/IGFBP4*, which were expressed in the left and right ventricular cardiomyocytes, respectively, as well as *HEY2* and *IRX3*, which marked outer and inner layer ventricular cardiomyocytes, thus resolving them as compact and trabecular cardiomyocytes, respectively[42–44]. Within the inner ventricle layer, we discovered an additional cardiomyocyte class not defined by scRNA-seq analyses, that extended along the luminal portion of the ventricle to the atrioventricular canal (AVC), and co-expressed *IRX3*, *TBX3* and *HCN4* known markers of the His-Purkinje fast cardiac conduction system of the ventricle[33], as well as *IRX1* and *IRX2*, which have been observed along the subendocardial layer of the interventricular septum of mouse hearts[43] (**Figure 10a**). Although most ventricular cardiomyocyte populations were observed in specific regions of the ventricle, we discovered a cardiomyocyte population that was present throughout the ventricle and displayed moderate expression of proliferative markers but diffuse expression of cardiac-structure specific genes (**Figure 10a**), suggesting that these cardiomyocytes may be progenitor-like with the capacity to further specialize within specific cardiac structures.

While our scRNA-seq uncovered cardiomyocyte populations (i.e., *BMP2*+ non-chambered) beyond those of the atrial and ventricle chambers as previously reported[3,4], our MERFISH analyses further resolved and confirmed the identification of these relatively rare but diverse specialized cardiomyocytes. In particular, we discovered that these distinct cardiomyocyte populations were located above the atria within the presumptive inflow tract as well as between the atria and ventricle where the

atrioventricular canal (AVC) typically resides[1,33,45] (**Figure 11b**). Inflow tract-like cardiomyocytes above the right atria co-expressed *ISL1* and *SHOX2,* known transcriptional regulators of inflow tract development and more specifically the sinoatrial node (SAN) pacemaker[1,33], whereas AVC cardiomyocytes co-expressed *TBX3* and *RSPO3*, which regulate AVC and atrioventricular node development[33,34,45] (**Figure 10a**). Additionally, a class of *CNN1+/CRABP2+* cardiomyocytes that was also identified in our scRNA-seq analysis (**Figure 2d,e**) but not well defined[46], was spatially resolved, and observed specifically within the atrioventricular valve leaflets. These cardiomyocytes further subdivided to those specifically populating the tricuspid and mitral valves of the right and left ventricles, respectively, suggesting that these cardiomyocyte subclasses may exhibit functional differences between valves (**Figure 10, 11a**).

Non-cardiomyocyte populations display distinct spatial regionalization and co-localization with specialized cardiomyocyte lineages.

Although displaying less diversity than cardiomyocytes, MERFISH imaging revealed that non-cardiomyocyte cell populations, particularly those endogenous to the heart, also further segregated and contributed to specific regions/structures of the heart, supporting similar observations from our scRNA-seq analysis. However, MERFISH analyses provided additional detailed spatial information at single-cell resolution that further resolved the identity of other less well-defined cell populations by scRNA-seq (**Figure 8e, 10a-b, 11a-e**). For the fibroblast class, we observed distinct *PDGFRA+/TCF21+* fibroblast clusters populating specifically either the atria or the ventricle which expressed *TNC* and *HHIP,* respectively, as well as VICs contributing to

the cardiac valves (**Figure 10a, 11b**). Similarly, we unexpectedly discovered three distinct *LEPR*+ endocardial populations that particularly lined the luminal surfaces of the atria, ventricle or cardiac valves and could be molecularly distinguished by *SHISA3*+, *NSG1+/COL26A1-* and *NSG1+/COL26A1*+ genes, respectively (**Figure 10a, 11c**). Vascular-related cell populations including *CLDN5+/LYVE1-* blood endothelial cells, *MYH11*+ vascular smooth muscle cells and *KCNJ8*+ pericytes were widely distributed throughout the ventricle revealing vessels in some cases but less so within the atria, supporting that the atria may be less vascularized possibly due to its thin myocardial walls (**Figure 10a, 11b**). On the other hand, *PRPH*+ neuronal-\like populations were primarily observed in the outflow tract and atria, particularly near the inflow tract, consistent with their role in outflow tract development and innervation of the venous pole of the heart[1] (**Figure 11d**). *CLDN5+/LYVE1*+ lymphatic, *MOXD1+/MMP11*+ EPDC and *ITLN1*+ epicardial cell populations were localized on the surface of the heart, and EPDCs were particularly enriched within the AVC regions as previously observed[6]. Finally, many of these non-cardiomyocyte cell populations exclusively co-localized with each other as well as corresponding cardiomyocyte counterparts within distinct cardiac regions, suggesting that they may assemble into cellular communities, which not only influence their specialized cellular functionalization but also form anatomic structures critical for regulating overall cardiac function.

Gene imputation of MERFISH data provides spatial map of inferred gene expression of the heart.

While our MERFISH studies have enabled the detailed examination of the spatial expression of specific genes that mark identified cardiac cell types, extending such spatial information beyond these genes would provide the opportunity to interrogate how all expressed genes may influence specific cell types and their spatial allocation during cardiac development and particularly heart morphogenesis. However, measuring the entire transcriptional profile of each cell would be challenging to accomplish with current multiplexed RNA FISH strategies. Thus, to create a spatial map of expressed genes for the human heart, we inferred the full transcriptome for each spatially mapped MERFISH cell by matching similar gene expression profiles between MERFISH cells and cells from corresponding age-matched scRNA-seq data (**Figure 12a, 13a**). Toward this end, we initially constructed a shared embedding space between cells of these datasets based on the 238 MERFISH target genes using Harmony, an algorithm that integrates multi-modal datasets to facilitate their joint analyses[47]. Nearest neighbor mapping of cells within the joint embedding matched MERFISH cells with similar scRNA-seq cells, thus allowing for the imputation of the expression of missing genes in MERFISH cells using transcriptomic data from corresponding scRNA-seq cells, as well as transfer of cell type annotations between datasets (**Figure 12a, 13a**). Comparing the cell type annotations assigned to the MERFISH cells *de novo* to those transferred from the scRNA-seq in the joint embedding revealed a strong correspondence of related cell populations between MERFISH and scRNA-seq datasets (**Figure 12b**), supporting that these datasets are well integrated.

To determine the proficiency of our algorithm to infer the expression of each gene, we analyzed how accurately the transcriptome for any particular cell identified by

scRNA-seq could be predicted by extrapolating the scRNA-seq gene expression profiles of its nearest cell neighbor in the shared embedding space (**Figure 13b**). To this end, we used this imputation strategy to predict the transcriptome for each cell identified by scRNA-seq and then tested the accuracy of this prediction by calculating the Pearson correlation between the imputed and the measured scRNA-seq gene expression levels for each analyzed cell ('scRNA-seq gene predictability score'). We observed that when predicting the entire transcriptome for these cells, scRNA-seq gene predictability scores were generally low (median = 0.024). This outcome may be expected, as many genes are either not appreciably expressed in the heart or are housekeeping genes with little variation in expression, and therefore most of the variability in their observed expression is likely due to technical variables (variation in sequencing depth per cell, drop-out rate, etc.) rather than biological variation. However, when considering only the 3,000 most highly variable genes in the scRNA-seq dataset, thus enriching for genes more relevant to the heart, the scRNA-seq gene predictability scores were markedly higher (median = 0.276) (**Figure 13b**). This result highlights the importance of developing a confidence metric to assess how well the expression of each gene can be imputed based on the 238 MERFISH target genes, as only a subset of the full transcriptome can be imputed with accuracy.

To determine how well this imputation method can accurately infer gene expression in MERFISH cells, we imputed the expression of the 238 genes imaged by MERFISH using this algorithm and compared these imputed values to corresponding MERFISH imaged expression data. Pearson correlation was then calculated between the imputed expression and measured image expression for all MERFISH cells to

determine the MERFISH gene predictability score. Because the shared embedding is constructed using these 238 genes, it is expected that these genes would be imputed more accurately than genes not used in the construction. To avoid this bias, 10-fold cross-validation was used to calculate independently the MERFISH and scRNA-seq gene predictability scores. To this end, a new shared embedding utilizing only 90% of the 238 MERFISH target genes was used to calculate the MERFISH and scRNA-seq gene predictability scores for the remaining 10% of genes that were not included for constructing the embedding. This process was repeated 10 times with a different 10% of genes being imputed by a different shared embedding each time to cover the full set of 238 genes. The median MERFISH and scRNA-seq gene predictability scores for the 238 MERFISH target genes was 0.247 and 0.416, respectively (**Figure 13c**). To further determine how well the scRNA-seq gene predictability score estimates the MERFISH gene predictability score, we calculated the Pearson correlation between these predictability scores, which was determined to be 0.71 (**Figure 13d**), thus, while we observed a disparity in the magnitude of predictability scores between MERFISH and scRNA-seq, genes with higher scRNA-seq predictability were strongly correlated with higher MERFISH predictability scores. Therefore, scRNA-seq gene predictability scores can be a useful metric for determining which genes are the most predictable in MERFISH cells. Supporting this analysis, the spatial patterns of imputed and measured MERFISH genes that were specifically examined displayed both high correspondence visually (**Figure 12c**) and high MERFISH gene predictability scores (**Figure 13e**). To further validate the approach, we re-integrated the scRNA-seq and MERFISH datasets using several quantities of genes (25, 50, 75, 100, 150, and 200 genes) and compared

the MERFISH gene predictability score for each quantity. We discovered that increasing the number of genes used for the integration improved the MERFISH gene predictability score up to 150 genes, after which the Pearson correlation plateaus (**Figure 13f**), suggesting that the number of genes used for MERFISH is sufficient to maximize the accuracy achievable by this method to impute the spatial expression of genes not imaged in the MERFISH.

Thus, utilizing MERFISH we were able to construct a spatial cell atlas of the developing human heart and spatially validate many of the cell types identified in scRNA-seq as well as further refine other cell type definitions using single-cell-resolved spatial information. By integrating the MERFISH and scRNA-seq datasets, we were able to impute and spatially map additional genes in the transcriptome, and developed a metric, scRNA-seq predictability score, which can be used as a confidence measure for the imputation of these genes, providing a valuable resource for the field. While interrogating the spatial patterns of cell types and gene expression, we noticed substantial spatial intermixing of different cell populations across the various regions of the heart. This spatial intermixing appeared highly organized, suggesting that there was a network of complex structures across the entire heart.

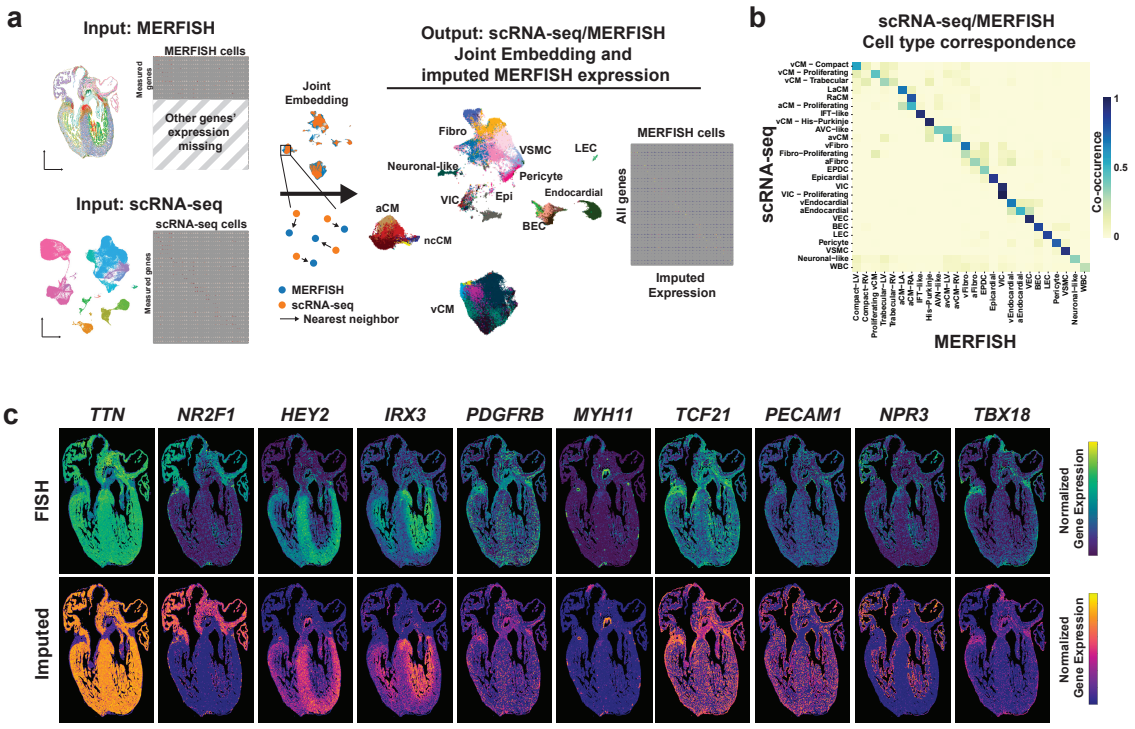
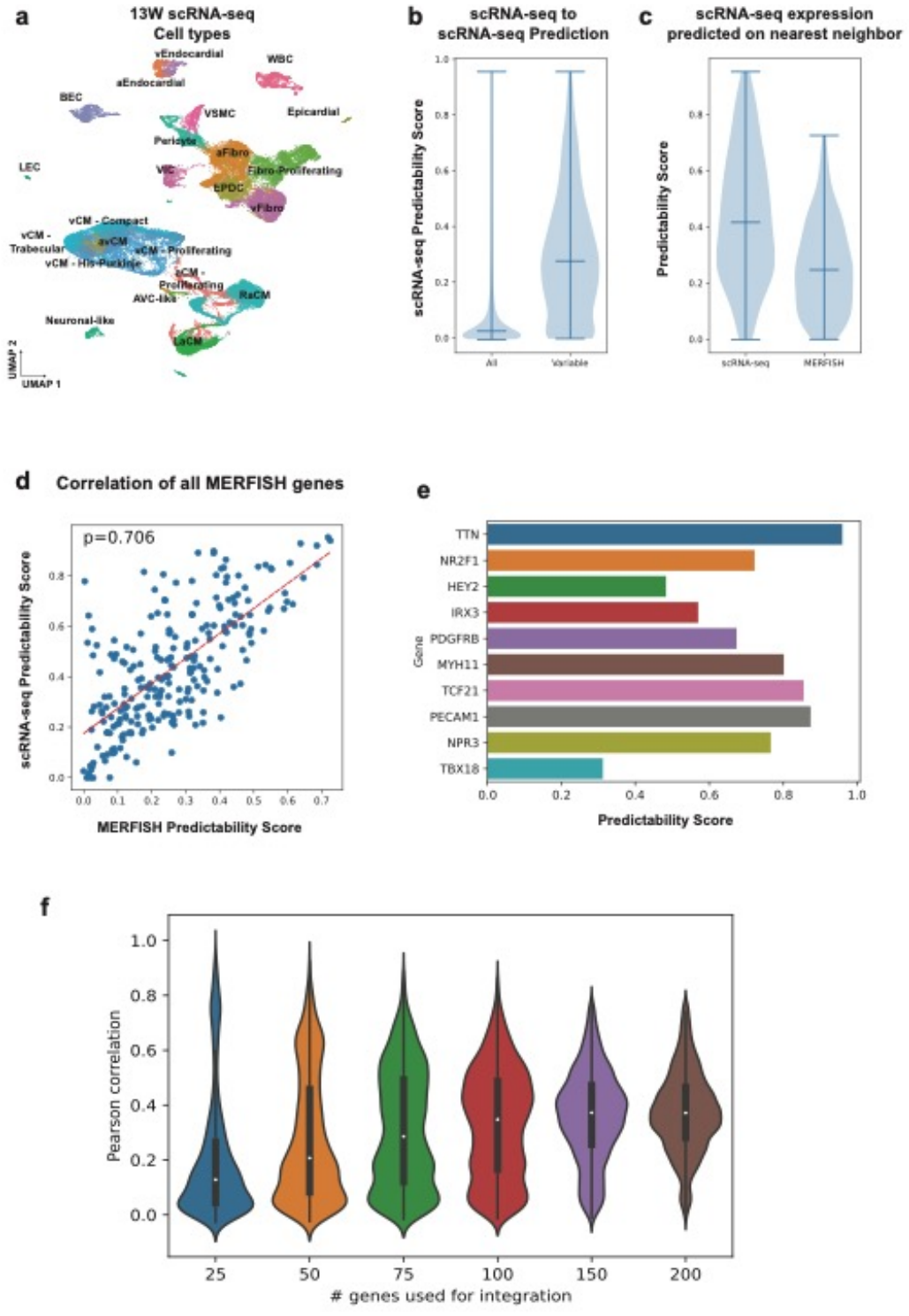


Figure 12. Integration of scRNA-seq and MERFISH datasets and gene imputation validation

a, Schematic representation of the cell type label transfer and gene imputation strategy. Both datasets were put into a joint embedding space based on the expression of the 238 MERFISH genes, and then the transcriptome of each MERFISH cell was inferred from the nearest scRNA-seq cell **b**, Heatmap of a co-occurrence matrix showing the count of cells in MERFISH with corresponding scRNA-seq labels. **c**, Validation of gene imputation performance by comparing normalized gene expression profiles of marker genes measured by MERFISH with the corresponding imputed gene expression profiles.

Figure 13. Validation and QC of gene imputation by predictability scores
a, scRNA-seq data from 13W heart re-clustered and relabeled with refined cell type labels. **b**, Violin plots showing the scRNA-seq predictability score of expression in each scRNA-seq cell with its nearest neighbor scRNA-seq cell in the Harmony integrated PCA space for all 26,581 and the 3,000 highest variable genes. **c**, Violin plots showing the predictability score measured in (b) and the predictability score of expression in each MERFISH cell with its nearest neighbor scRNA-seq cell in the Harmony integrated PCA space for the 238 genes from the MERFISH gene panel. **d**, Scatterplot where each dot is a gene from the MERFISH gene panel. The x-axis is the MERFISH predictability score of the imputed expression of the gene in the MERFISH cells with the expression measured by imaging, while the y-axis is the scRNA-seq predictability score of the imputed expression in scRNA-seq cells with the expression measured by scRNA-seq. The Pearson correlation shown by the trendline was 0.706. **e**, The correlation values from (b) for the ten marker genes whose imputed values were used for validation with spatial expression plots. **f**, Violin plots of the calculated predictability scores for the 238 MEFISH genes after performing the integration with different number of genes.


a
13W scRNA-seq
Cell types
vEndocardial
aEndocardial
BEC
WBC
VSMC
Epicardial
Pericyte
aFibro
LEC
vEC
Fibro-Proliferating
EPDC
vFibro
vCM - Compact
vCM - Trabecular
aCM
vCM - Proliferating
vCM - His-Purkinje
aCM - Proliferating
AVC-like
RaCM
Neuronal-like
LaCM
UMAP 2
UMAP 1
b
scRNA-seq to scRNA-seq Prediction
scRNA-seq Predictability Score
All
Variable
c
scRNA-seq expression predicted on nearest neighbor
Predictability Score
scRNA-seq
MERFISH
d
Correlation of all MERFISH genes
p=0.706
scRNA-seq Predictability Score
MERFISH Predictability Score
e
TTN
NR2F1
HEY2
IRX3
PDGFRB
MYH11
TCF21
PECAM1
NPR3
TBX18
Gene
Predictability Score
f
Pearson correlation
genes used for integration

Diverse cardiac cell types organize into cellular communities that form the morphologic structures of the human heart.

Cardiac development is a spatially complex process and comprehensive understanding of regional and structural changes in cellular organization and gene expression during human development is important for understanding cardiac function. To understand how the cardiovascular cell populations are spatially organized across the heart into specific cardiac structures, we performed a previously published analytical framework[48] to detect cellular communities across the entire heart. Cellular communities (CCs) are identified as regions of the heart with a specific local composition of neighboring cardiac cell types. For every cell, its nearest spatial cellular neighbors within a 150 μm radius is identified, which we label as a 'zone'. Zones were defined as a 150 μm radius around a cell, representing a typical diffusion distance for extracellular signaling molecules for which the cell would be interacting with[49]. The zones were clustered based on their cell type composition to identify CCs (**Figure 14a**), using a statistically determined number of clusters (**Figure 14c**).

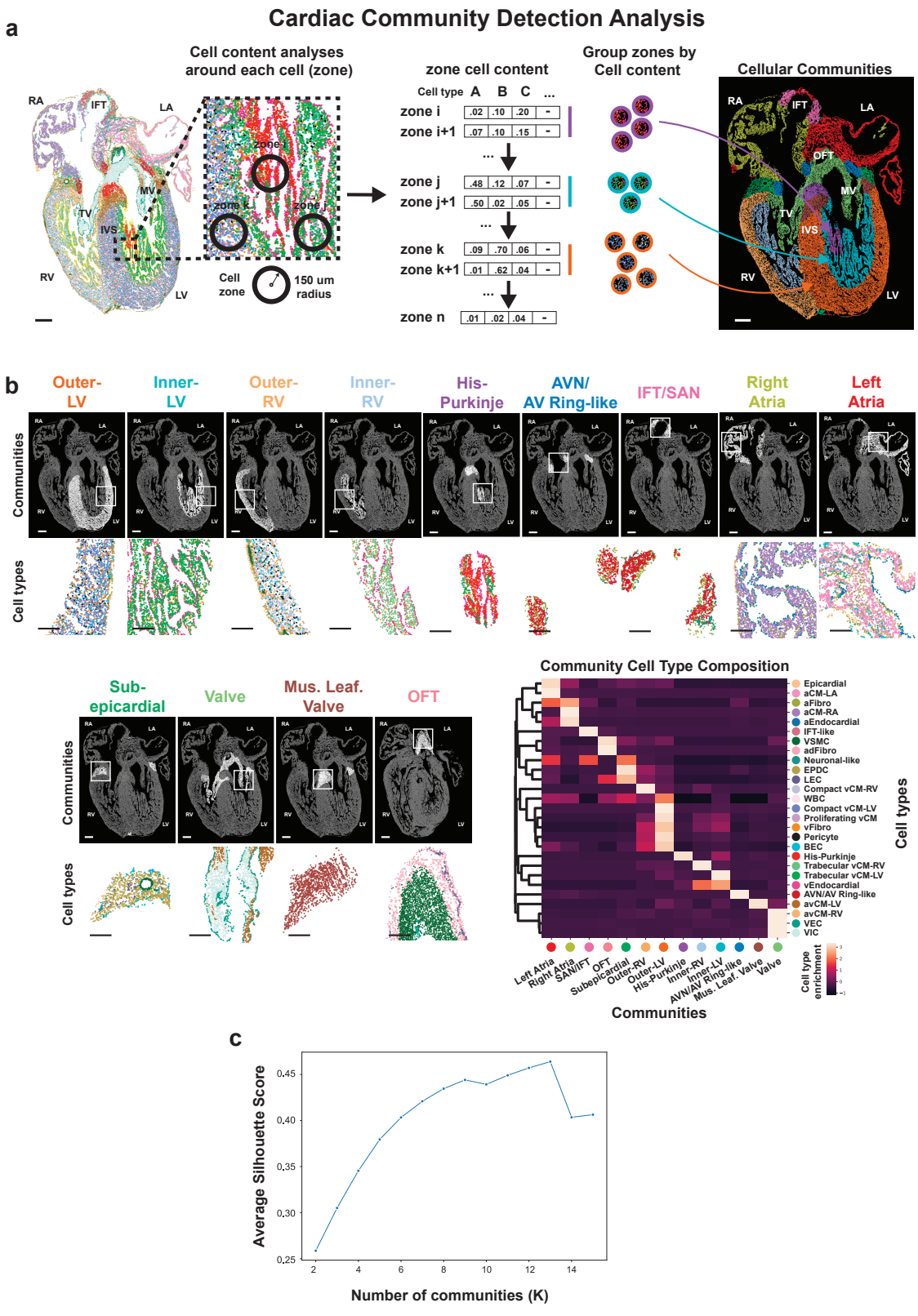
We identified thirteen distinct CCs that corresponded to cardiac structures and thus were labeled as: Left and Right Atria, Outer-LV and Inner-LV, Outer-RV and Inner-RV, the Subepicardium, the Outflow Tract (OFT), the Atrioventricular Node/Ring (AVN/AV Ring-like), the Inflow Tract (IFT), His-Purkinje, the Valves, and the Muscular Leaflet of the Valve. We visualized the relative cell type composition of each community (**Figure 14b**). The atria consisted of two main communities (Right Atria CC and Left Atria CC) corresponding to the left and right atria, while the ventricles each contained an Inner and Outer CC, with a Subepicardial CC and a Valve CC spanning both ventricles

(**Figure 14b**). The left ventricle contained a subendocardial community corresponding to the developing His-Purkinje fast conduction system of the ventricle (**Figure 14b**). The absence of this community from the right ventricle suggests that the fast conduction system of the ventricle develops earlier on the left side. The Valve CC spatially encompassed the mitral and tricuspid valves within the ventricles, with the mitral valve region containing a CC that corresponded to the muscular valve leaflet (Muscular Leaflet of the Valve CC), which may also reflect that cell types within the left ventricle are specializing earlier in development than the right ventricle (**Figure 14b**). Near the right atrium, there is a CC consisting mainly of IFT-like cells and due to its location above the right atria, corresponds to the superior vena cava may contain the SAN (**Figure 14b**). Furthermore, a CC flanking the AVC may correspond to the AV ring, which is developmental structure that gives rise to the atrioventricular node (AVN) (**Figure 14b**). The Subepicardial CC correlates well with the AV mesenchyme previously reported to be present within the heart around this time in development[6]. These analyses revealed diverse cellular communities that correspond to morphological structures in the heart and uncover complex substructure within the ventricles, particularly around the valves.

We noticed that certain CCs were composed of only one or two cell types, such as the His-Purkinje which is composed of just one main cell type (the His-Purkinje cells), whereas other communities like the Outer-LV are composed of many different cell types (**Figure 14b**). To further assess the differences in cell type composition between the CCs, we quantified cellular complexity by defining a complexity metric as the number of unique cell types in the zone around each cell (**Figure 15a**). Zones containing many

unique cell types are considered to be more complex, and vice versa. The distribution of zone complexities in each of the CCs was calculated and visualized (**Figure 15b**). On average, the communities within the ventricles including the Subepicardial and Outer communities are more cellularly complex, whereas the Atrial and OFT CCs are less cellularly complex (**Figure 15b**). Zone complexity was also spatially visualized on the MERFISH data, which revealed that the ventricle wall exhibited high cellular complexity compared to the atria. Comparison of the ventricles and atria uncovered that the ventricular wall is more cellularly complex where it is composed of many different cell types, and appears to increase in complexity going from the Inner to the Outer ventricular wall (**Figure 15b**), suggesting that the cell type composition changes across the ventricular wall, which may lead to differences in cell interactions and function.

Figure 14. Cellular communities illuminate the spatial organization of cell types coordinating in the formation of cardiac structures.
a, Schematic of the cellular community identification illustrating the definition and clustering of cell zones. **b**, Spatial distribution of 13 distinct CCs corresponding to cardiac structures. (Bottom right) The spatial neighborhood frequencies (enrichment score) of each of the MERFISH identified cell types within each CC. **c**, Average silhouette score calculated for a range of K, which is the parameter used in K-means clustering to determine the number of communities detected, with the highest value being 13. aCM, atrial cardiomyocyte; AVC, atrioventricular canal; avCM, muscular leaflet of the valve cardiomyocyte; BEC, Blood Endothelial cell; CMs, Cardiomyocytes; Endo, endocardial; EPDC, epicardial-derived cell; Epi, epicardial; IFT-like, inflow tract-like; LA, left atrium; LEC, lymphatic endothelial cell; LV, left ventricle; MV, mitral valve; Mus. Leaf. Valve, muscular leaflet of the valve; Neuro, neuronal-like; ncCM, non-chambered cardiomyocyte; OFT, outflow tract; Peri, Pericyte; RA, right atrium; RBC, Red Blood Cell; RV, right ventricle; TV, tricuspid valve; vCM, ventricular cardiomyocyte; vEndocardial, ventricular endocardial; VICs, valve interstitial cells; VSMC, vascular smooth muscle cells; WBC, white blood cell. Scale bar, 250 μm.


a
Cardiac Community Detection Analysis
Cell content analyses around each cell (zone)
zone cell content
Group zones by Cell content
Cellular Communities
RA
IFT
LA
MV
TV
IVS
RV
LV
OFT
zone i
zone i+1
zone j
zone j+1
zone k
zone k+1
zone n
Cell type A B C ...
Cell zone
150 um radius
b
Outer-LV
Inner-LV
Outer-RV
Inner-RV
His-Purkinje
AVN/AV Ring-like
IFT/SAN
Right Atria
Left Atria
Communities
Cell types
Sub-epicardial
Valve
Mus. Leaf. Valve
OFT
Community Cell Type Composition
Epicardial
aCM-LA
aFibro
aCM-RA
aEndocardial
IFT-like
VSMC
adFibro
Neuronal-like
EPDC
LEC
Compact vCM-RV
WBC
Compact vCM-LV
Proliferating vCM
vFibro
Pericyte
BEC
His-Purkinje
Trabecular vCM-RV
Trabecular vCM-LV
vEndocardial
AVN/AV Ring-like
avCM-LV
avCM-RV
VEC
VIC
Left Atria
Right Atria
SAN/IFT
OFT
Subepicardial
Outer-RV
Outer-LV
His-Purkinje
Inner-RV
Inner-LV
AVN/AV Ring-like
Mus. Leaf. Valve
Valve
Cell type enrichment
Communities
Cell types
c
Average Silhouette Score
0.45
0.40
0.35
0.30
0.25
2
4
6
8
10
12
14
Number of communities (K)

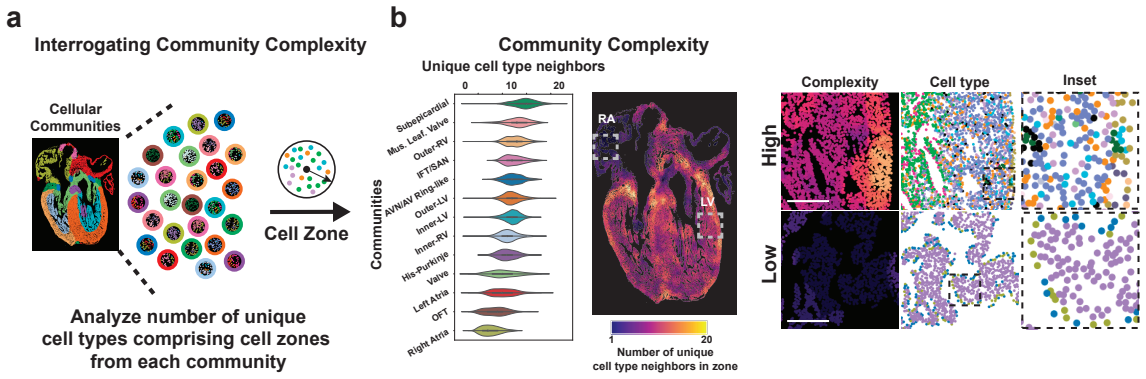


Figure 15. Cellular complexity of cardiac communities
a, Analysis of spatial mixing of distinct cell types within each zone. We define the complexity of a zone surrounding any given cell as the number of unique cell types present within that zone. **b**, (Left) Distribution of zone complexities within each community, displayed both spatially and plotted. (Right) Examples of two regions (one in the atria and one in the ventricle) with high and low complexity zones, colored by complexity and cell type. Scale bar, 100 μm.

2.3 DISCUSSION

Understanding the cell type composition, their spatial organization, and how that forms structures to regulate organ function requires both cellular information and spatial context. Recent work done on the developing human heart has made progress in our understanding of these aspects, limitations in both cellular and spatial resolution hindered further understanding of the complex network of cells and structures that the developing heart is composed of. Here, we investigate the developing heart at single cell resolution, combining a larger scRNA-seq and spatial dataset than previously reported to construct a comprehensive molecular and spatial atlas.

Interrogation of the molecular atlas uncovered 75 subpopulations across time and space of the developing heart, and included many known cell populations, in addition to previously uncharacterized cell populations within the developing heart such as the non-chambered cardiomyocyte populations (IFT-like and AVC-like). These populations are a small proportion of the heart, and so have not been captured in any human heart datasets[3,4,6]. Cardiomyocytes and mesenchymal populations were the most prevalent cardiac cells within the developing heart, with both containing subpopulations exhibiting transcriptional differences between regions of the heart (atrial and ventricular cardiomyocytes and fibroblasts), indicating different developmental origins and specialized functions in cardiac chambers. This regionalization was further displayed by cell types not previously described to be region specific before, such as the atrial and ventricular endocardium, suggesting that the endocardial layer is specialized for the function of each chamber. Much of the cellular diversity correlated with region and age of the heart, suggesting that as these distinct cardiac lineages

mature, they are becoming specialized during development to accommodate the function of each cardiac structure

Spatial mapping of cardiac cells with a high resolution spatial transcriptomic technique (MERFISH) validated and resolved many of the cell populations at high resolution, and in general we see specialization for the regions and structures in the heart in which the cells are located. We found that cardiomyocytes display a high amount of molecular diversity, particularly within the ventricles where molecularly distinct populations were seen within the ventricular walls and valves.

Finally, to understand how the cardiac cell types are spatially organized, we used a community detection to identify 13 structures across the heart, including the inner and outer layers of the ventricle, as well as the subluminal structure of the Purkinje fiber ventricular conduction system. Many of these structures have been defined morphologically in the past, however with our MERFISH dataset, we were able to identify the structures with high molecular detail, which was previously unattainable with existing technologies. Investigation of the cellular complexities within the communities showed that the ventricle has high cellular complexity, suggesting it may be undergoing developmental changes.

In conclusion, we have generated a molecular and spatial atlas of the developing human heart that is the most comprehensive to date. We identified known, as well as previously uncharacterized cell types, spatially mapped them, and identified cardiac structures with unprecedented molecular and spatial detail. Our approach makes it possible to explore the cellular heterogeneity and organization of cardiac structures,

which will provide meaningful insights for understanding the human heart during development and disease.

2.4 MATERIALS AND METHODS

Tissue samples

De-identified tissue samples were collected with previous patient consent in strict observance of the legal and institutional ethical regulations. Protocols were approved by the #101021X (Institutional Review Board) at the University of California, San Diego.

Tissue Processing

Tissue samples were dissected in buffer containing 125 mM NaCl, 2.5 mM KCl, 1mM $MgCl_2$, and 1.25 mM NaH_2PO_4 under a stereotaxic dissection microscope (Leica).

For single cell dissociation, tissue samples were further cut into small piec es and enzymatically digested by incubating with collagenase type IV (Gibco) and Accutase (ThermoFisher) at 37^{o}C for 60 min. After removing the dissociation media, cells were resuspended in PBS supplemented with 5% FBS and sorted on a Sony sorter. Samples were diluted to approximately 1,000 cells per µl before processing for scRNA-seq. Single-cell droplet libraries from this suspension were generated in the 10X Genomics Chromium controller according to the manufacturer's instructions in the Chromium Single Cell 3′ Reagent Kit v2 User Guide.

Samples for MERFISH were washed 1 time with ice-cold PBS, then fixed in 4% PFA at 4°C overnight. On the second day, the sample was washed in ice-cold PBS three times, 10 minutes each, and were incubated in 10% and 20% sucrose at 4°C for 4 hours each, and in 30% sucrose overnight, followed by immersion with OCT (Fisher,

cat# 23-730-571) and 30% sucrose (1 vol:1 vol) for 1 hour. The sample was then embedded in OCT and stored at -80°C until sectioning.

Single-cell transcriptome library preparation and sequencing.

Libraries were prepared according to the manufacturer's instructions using the Chromium Single Cell 3′ Library and Gel Bead Kit v2 (PN-120237) and Chromium i7 Multiplex Kit (PN-120262). All libraries were sequenced on the HiSeq 4000 to a mean read depth of at least 75,000 total aligned reads per cell.

Processing of raw sequencing reads.

Raw sequencing reads were processed using the Cell Ranger v3.0.1 pipeline from 10X Genomics. In brief, reads were demultiplexed, aligned to the human hg38 genome and UMI counts were quantified per gene per cell to generate a gene-barcode matrix.

Cell quality control and filtering

After generating the gene-barcode matrix file from cellranger, the individual count matrices were merged together and processed using the Seurat v4.0.1 R package[30] (https://satijalab.org/seurat/). Further filtering and clustering analyses of the scRNA-seq cells were performed with the Seurat package, as described in the tutorials (https://satijalab.org/seurat/). Cells with at least 1000 genes detected, and a mitochondrial read percentage of less than 30% were used for downstream processing. Potential doublets were removed using DoubletFinder[50] (https://github.com/chris-

mcginnis-ucsf/DoubletFinder), using an anticipated doublet rate of 5%, which is the expected rate reported by 10X Genomics for the amount of cells loaded onto the 10X Controller. For the aggregated dataset, gene expression was normalized for genes expressed per cell and total expression by the NormalizeData function. The top 3000 variable genes were detected with the FindVariableFeatures function with default parameters. All the genes were subsequently scaled using the ScaleData function, which utilizes a linear regression model to eliminate technical variability due to number of genes detected, replicate, and mitochondrial read percentage. Principal components were calculated using RunPCA and the 50 significant principal components (determined by ElbowPlot) were used for creating the nearest neighbor graph utilizing the FindNeighbors function with k.param = 50. The generated nearest neighbor graph was then used for graph-based, semi-unsupervised clustering (FindClusters, default resolution of 0.8) and uniform manifold approximation and projection (UMAP) to project the cells into two dimensions. Marker genes were identified using a Wilcoxon rank-sum test (FindAllMarkers, default parameters) for pairwise comparisons between cell clusters. Cell identities were assigned to the clusters by cross-referencing their marker genes with known cardiac cell type markers from both human and mouse studies, in addition to *in situ* hybridization data from the literature. Typically one cell cluster would emerge that expressed marker genes representing multiple populations, as well as contained cells with low UMI and gene counts that escaped the first filtering step. These cells were removed from downstream analyses. The clustering approach was then repeated for each partition of cell types (Cardiomyocyte, Mesenchymal, Endothelial, Neuronal-like, Blood) as described above.

MERFISH gene selection

To identify the transcriptionally distinct subpopulations identified in the scRNA-seq dataset, we designed a panel of 238 genes to be imaged with combinatorial barcoded imaging. We first identified differential gene markers for each of the 75 subpopulations in the scRNA-seq with NS-Forest3[38], combining all markers from the binary gene analysis. The list was then filtered for genes that were either not long enough to construct 60 target regions (each 30-nucleotide long) without overlap or whose expression levels were outside the range of 0.01 to 300 average UMI per cluster, as measured by scRNA-seq.

MERFISH imaging

MERFISH measurements of 238 genes with 10 non-targeting blank controls was done as previously described, using the encoding sequences and published readout probes[51]. Briefly, 12-μm-thick tissue sections were mounted on 40 mm #1.5 coverslips that are silanized and poly-L-lysine coated[52] and subsequently pre-cleared by immersing into 50% (vol/vol) ethanol, 70% (vol/vol) ethanol. 100% ethanol, each for 5 minutes. The tissue was then air dried for 5 minutes prior to be treated with Protease III (ACDBio), at 40 °C, for 30 minutes prior to be washed with PBS for 5 minutes. Then the tissues were preincubated with hybridization wash buffer (30% (vol/vol) formamide in 2x SSC) for ten minutes at room temperature. After preincubation, the coverslip was moved to a fresh 60 mm petri dish and residual hybridization wash buffer was removed with a Kimwipe lab tissue. In the new dish, 50 uL of encoding probe hybridization buffer

(2X SSC, 30% (vol/vol) formamide, 10% (wt/vol) dextran sulfate, 1 mg/ml yeast tRNA, and a total concentration of 5 uM encoding probes and 1 μM of anchor probe: a 15-nt sequence of alternating dT and thymidine-locked nucleic acid (dT+) with a 5′-acrydite modification (Integrated DNA Technologies). The sample was placed in a humidified 37C oven for 36 to 48 hours then washed with 30% (vol/vol) formamide in 2X SSC for 20 minutes at 37°C, 20 minutes at room temperature. Samples were post-fixed with 4% (vol/vol) paraformaldehyde in 2X SSC and washed with 2X SSC with murine RNase inhibitor for five minutes. To anchor RNAs in place, the encoding-probe–hybridized samples were embedded in thin, 4% poly-acrymide gels as described[52]. Briefly, the hybridized samples on coverslips were first washed with a de-gassed 4% poly-acrylamide solution, consisting of 4% (vol/vol) of 19:1 acrylamide/bis-acrylamide (BioRad, 1610144), 60 mM Tris·HCl pH 8 (ThermoFisher, AM9856), 0.3 M NaCl (ThermoFisher, AM9759), and a 1:1,000 dilution of 0.1-μm-diameter blue beads (Life Technologies, F-9800). The beads served as fiducial markers for the alignment of images taken across multiple rounds of imaging. The coverslips were then washed again for 2 min with the same 4% poly-acrylamide solution gel solution supplemented with the polymerizing agents ammonium persulfate (Sigma, A3678) and TEMED (Sigma, T9281) at final concentrations of 0.03% (wt/vol) and 0.15% (vol/vol), respectively. The gel was then allowed to cast for 1.5 h at room temperature. The coverslip and the glass plate were then gently separated, and the PA film was incubated with a digestion buffer consisting of 50 mM Tris·HCl pH 8, 1 mM EDTA, and 0.5% (vol/vol) Triton X-100 in nuclease-free water and 1% (vol/vol) proteinase K (New

England Biolabs, P8107S). The sample was digested in this buffer for >36h in a humidified, 37°C incubator and then washed with 2× SSC three times.
The samples were finally stained with a Alexa 488-conjugated anchor probe-readout oligo (Integrated DNA Technologies) and DAPI solution at 1 μg/ml. MERFISH measurements were conducted on a home-built system as described[51].

Processing of MERFISH images

Individual RNA molecules were decoded using MERlin v0.6.1 as previously described[53]. Images were aligned across hybridization rounds by maximizing phase cross-correlation on the fiducial bead channel to adjust for drift in the position of the stage from round to round. Background was reduced by applying a high-pass filter and decoding was then performed per-pixel. For each pixel, a vector was constructed of the 16 brightness values from each of the 16 rounds of imaging. These vectors were then L2 normalized and their euclidean distances to each of the L2 normalized barcodes from MERFISH codebook was calculated. Pixels were assigned to the gene whose barcode they were closest to, unless the closest distance was greater than 0.512, in which case the pixel was not assigned a gene. Adjacent pixels assigned to the same gene were combined into a single RNA molecule. Molecules were filtered to remove potential false positives by comparing the mean brightness, pixel size, and distance to the closest barcode of molecules assigned to blank barcodes to those assigned to genes to achieve an estimated misidentification rate of 5%. The exact position of each molecule was calculated as the median position of all pixels consisting of the molecule.

Cellpose v1.0.2[40] was used to perform image segmentation to determine the boundaries of cells and nuclei. The nuclei boundaries were determined by running Cellpose with the 'nuclei' model using default parameters on the DAPI stain channel of the pre-hybridization images. Cytoplasm boundaries were segmented with the 'cyto' model and default parameters using the polyT stain channel. RNA molecules identified by MERlin were assigned to cells and nuclei by applying these segmentation masks to the positions of the molecules. Any segmented cells that did not have any barcodes assigned were removed before constructing the cell-by-gene matrix.

Cell clustering analysis of MERFISH

With the cell-by-gene matrix, we followed a standard procedure as suggested in the Scanpy version 1.8[54] tutorial using python version 3.9 for processing MERFISH data. Count normalization, Principal Component Analysis (PCA), neighborhood graph construction, and Uniform Manifold Approximation and Projection (UMAP) were performed with SCANPY's default parameters. We performed Leiden clustering utilizing a resolution of 2. The top 20 differential genes identified by the rank_gene_groups() function were used to annotate each cluster. We further subclustered the ventricular cardiomyocytes (vCM) clusters using Leiden clustering at a resolution of 1. This allowed us to further annotate vCM cells as Compact and Trabecular vCMs for both the left and right ventricles. After manually defining the ventricular region, we subset the MERFISH dataset to the ventricular cells and performed Leiden clustering at a resolution of 5. Again, the top 20 genes identified by the rank_gene_groups() function were used to annotate each cluster.

Integration of single-cell RNA-seq and MERFISH

We first subsetted both the scRNA-seq and MERFISH to only genes interrogated by both modalities. We then utilized SCANPY's implementation of Harmony to project both the scRNA-seq and MERFISH dataset into a shared PCA space. The dimensionality of the joint embedding was further reduced using UMAP (min_dist=0.3) to visualize the space in two dimensions using a k-nearest neighbor graph constructed with a k of 30 and the Pearson correlation distance metric to match the parameters used by the Harmony authors[47].

To impute a complete expression profile and cell type for each MERFISH profile, we assigned the expression profile and cell type of the closest scRNA-seq cell in the Harmony PCA space using the euclidean distance metric.

Identifying cellular communities

We sought to define cellular communities which represented unique shared cellular signaling environments defined by the neighboring cell types in close proximity. To this end, we clustered each MERFISH cell based on the cell type composition of neighboring cells within 150 μm, representing a typical diffusion distance for extracellular signaling molecules. This radius sampled approximately 300 neighbors. Each cell's community was therefore represented by a vector of length 27, containing the relative proportions of each of the 27 previously identified cell types. We then clustered the zones using Python's scikit-learn version 0.22 implementation of KMeans

with k = 13, chosen by silhouette score. Thus, each MERFISH cell was assigned to one of the 13 cellular communities.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Farah, Elie N.; Hu, Robert K.; Kern, Colin; Zhang, Qingquan; Lu, Ting-Yu; Zhang, Bo; Carlin, Daniel; Li, Bin; Blair, Andrew P.; Ren, Bing; Evans, Sylvia M.; Chen, Shaochen; Zhu, Quan; Chi, Neil C. “Cardiac Single Cell Ecology Reveals Interactive Cellular Construction and Remodeling of the Human Heart”. The dissertation author was the primary investigator and author of this paper.

CHAPTER 3:

Uncovering mechanisms of ventricular wall organization utilizing bioprinted multilayer cardiac organoids

3.1 INTRODUCTION

The heart contains four chambers that serve different physiological functions and decades of research have uncovered that the different chambers have distinct developmental origins and processes. The ventricles specifically undergo structural changes during development known as trabeculation and compaction. Trabeculation is the process where cardiomyocytes form a meshwork of long finger-like projections into the lumen of the heart known as trabeculae, which function to provide increased oxygenation to the growing myocardium[20,21,55]. As the heart grows, the trabeculae compress and integrate into the thickening myocardial wall in the process known as compaction. The wall thickening serves to increase the contractile force of the heart to be able to circulate blood through the growing body. Defects in either process (trabeculation or compaction) is extremely detrimental to the function of the heart and leads to early fetal demise[23].

Various signaling pathways have been discovered to play a role in ventricular wall development, including Notch and semaphorin signaling[27,55]. Semaphorins are a class of both secreted and membrane-bound signaling molecules that bind to their cognate receptors (plexins or plexin-neuropilin complexes)[27,56,57]. Semaphorin signaling canonically play a role in axon guidance within the nervous system, although within the cardiovascular system it is thought to direct the migration of cardiac neural crest cells[27]. Defects in many semaphorin signaling molecules result in cardiac defects including

persistent truncus arteriosus, ventricular septal defects, and aortic arch defects[27]. Recently, loss of function of an endothelial specific plexin (*PLXND1*) has been shown to cause defects in compaction[28]. Whether semaphorin signaling serves other functions during heart development remains unclear.

Here, we uncover a potential role of semaphorin signaling in the developing heart in the organization of cardiomyocytes within the ventricular wall. Subclustering of spatial transcriptomic data within the developing human ventricle identified that ventricular cardiomyocytes (CMs) display a gradient of distribution across the ventricular wall, with correlated gene expression profiles and wall depths of individual CMs. This further identified a molecularly distinct 'hybrid' cardiomyocyte that displays gene expression that are a mix of compact and trabecular CM markers, and may correspond to the CMs present in the recently described Mid/hybrid myocardium[24]. To understand how the cell types are spatially organized into structures, we performed our community detection algorithm and discovered that the left ventricular wall contains Inner, Mid, and Outer cellular communities with distinct cellular compositions. Investigating cell-cell interactions may be influencing the organization of the cardiomyocytes within each of these communities revealed semaphorin signaling to be specific and diverse within the Mid community, and included *SEMA3C/D*-expressing compact ventricular fibroblasts, *SEMA6A/B*-expressing blood endothelial cells, and *PLXNA2/A4* expressing trabecular (and hybrid) cardiomyocytes to the Mid/Outer myocardial community layers. To test how these interactions may regulate the migration and organization of these cardiomyocytes, we created an *in vitro* hPSC cardiac trabecular system, and discovered that *SEMA3C*

acts as an attractive signaling cue to *PLXNA2/A4* expressing trabecular-like cardiomyocytes, whereas *SEMA6A/B* act as a repulsive cue. These data suggest that whereas *SEMA6A/B*-expressing blood endothelial cells may prevent these trabecular (and hybrid) cardiomyocytes from further migrating by repelling them when contacting blood endothelial cells. These results support a model cardiomyocyte organization within the ventricular wall during compaction where *SEMA3C/D*-expressing compact ventricular fibroblasts attract *PLXNA2/A4* expressing trabecular (and hybrid) cardiomyocytes to the Mid/Outer myocardial community layers, whereas *SEMA6A/B*-expressing blood endothelial cells may prevent these trabecular (and hybrid) cardiomyocytes from further migrating by repelling them when contacting blood endothelial cells.

3.2 RESULTS

Cardiomyocytes Within the Ventricular Wall Display a Spatial and Molecular Gradient

During heart development, the ventricular wall contains an inner sponge-like network of cardiomyocytes known as trabeculae. These structures are critical for contraction and conduction, ventricular septation, and wall thickening through the process of compaction[55]. In the human, this process of compaction starts around week 12-13 of gestation[22], which corresponds well with our MERFISH sample. Defects in the development of the ventricular wall, including the compact and trabecular layers, lead to serious cardiomyopathies, including non-compaction cardiomyopathy[58]. Recent studies in model organisms have elucidated initializing events in the process of trabeculation and compaction, however the how compaction happens in humans remains elusive [20,21]. Toward this end, subclustered the MERFISH cells within the ventricular and reidentified the cell populations. Subclustering uncovered new sub-populations corresponding to cell types and possibly cell states of vCM and vFibro populations (**Figure 16a**). Cardiomyocytes, which constitute approximately 60% of all cells in the ventricle, were classified into 6 populations within the LV, and appeared to display a spatial gradient across the wall depth (**Figure 16b)**. Notably, MERFISH identified a cluster residing between the compact and trabecular populations that may correspond to the transient hybrid myocardial zone[24], therefore we named it hybrid vCM (**Figure 16b**). Gene expression analyses of these new populations revealed that the populations of vFibros were both spatially and molecularly distinct, with trabecular vFIbros expressing a transcriptional repressor *MSC* (musculin), and compact vFibros

expressing *HHIP*, an inhibitor of hedgehog signaling (**Figure 17a,b**). A third population of vFibros expressed markers consistent with a proliferating population, including *MKI67* and *PCNA* (**Figure 17a,b**). Conversely, for the cardiomyocytes we observed a largely gradual transition of gene expression profiles (**Figure 17b**). All vCMs expressed the known markers of *IRX4* and *MYH7*. Compact vCM populations were marked by their expression of a known marker *HEY2*, while trabecular and purkinje fiber populations appear to express *IRX3* and *GJA5*. The hybrid vCM population expressed markers of both the compact and trabecular vCMs (**Figure 17b**). Corresponding ventricular populations were identified in the scRNA-seq dataset that displayed similar gene expression patterns (**Figure 18a,b**). The spatial and molecular heterogeneity among the CM clusters led us to further investigate whether the CMs have a gradient of spatial and molecular profiles.

To this end, we measured the connectivity between CM clusters using PAGA and found that the CM populations formed highly connected networks with populations that were spatially close, meaning that cells that were spatially close had the highest similarity in gene expression (**Figure 19a,b**). By using a spatial distance metric for wall depth, defined as distance away from the epicardium, we ordered the cardiomyocytes by their relative position in the wall and identified genes whose expression changed substantially with wall depth (**Figure 17c, 19c**). This analysis further supported a gradual change of gene expression profiles of cells along the wall depth. Using a pseudotime analysis that approximates the distance between cells within expression space by walking along the nearest neighbor graph to order the CMs on the basis of their expression profiles (using Compact vCM-LV II as the starting point), we observed

that the expression distance of cells was highly correlated with their wall depth (**Figure 19d**). We observed more substantial separation in expression distance between Trabecular I/II and Purkinje Fiber populations, when compared to other CM populations.

Because of the gradient of expression within the CMs of the wall, we can use combinations of markers to define populations that are spatially adjacent, one example being the hybrid CM population that can be visualized by the co-expression of *HEY2* and *IRX3*, where cells that exhibit high expression of both correspond well to the cells defined as the Hybrid vCM population (**Figure 19e**). Taken together, these results suggest that CMs within the ventricle form a continuous distribution across the ventricular wall, with correlated molecular profiles and wall depths. But how these cardiomyocytes organize to form the distinct layers across the ventricle and why that is important for compaction and ventricular wall development is still unclear, therefore we looked at what interactions may be influencing the organization of the cardiomyocytes.

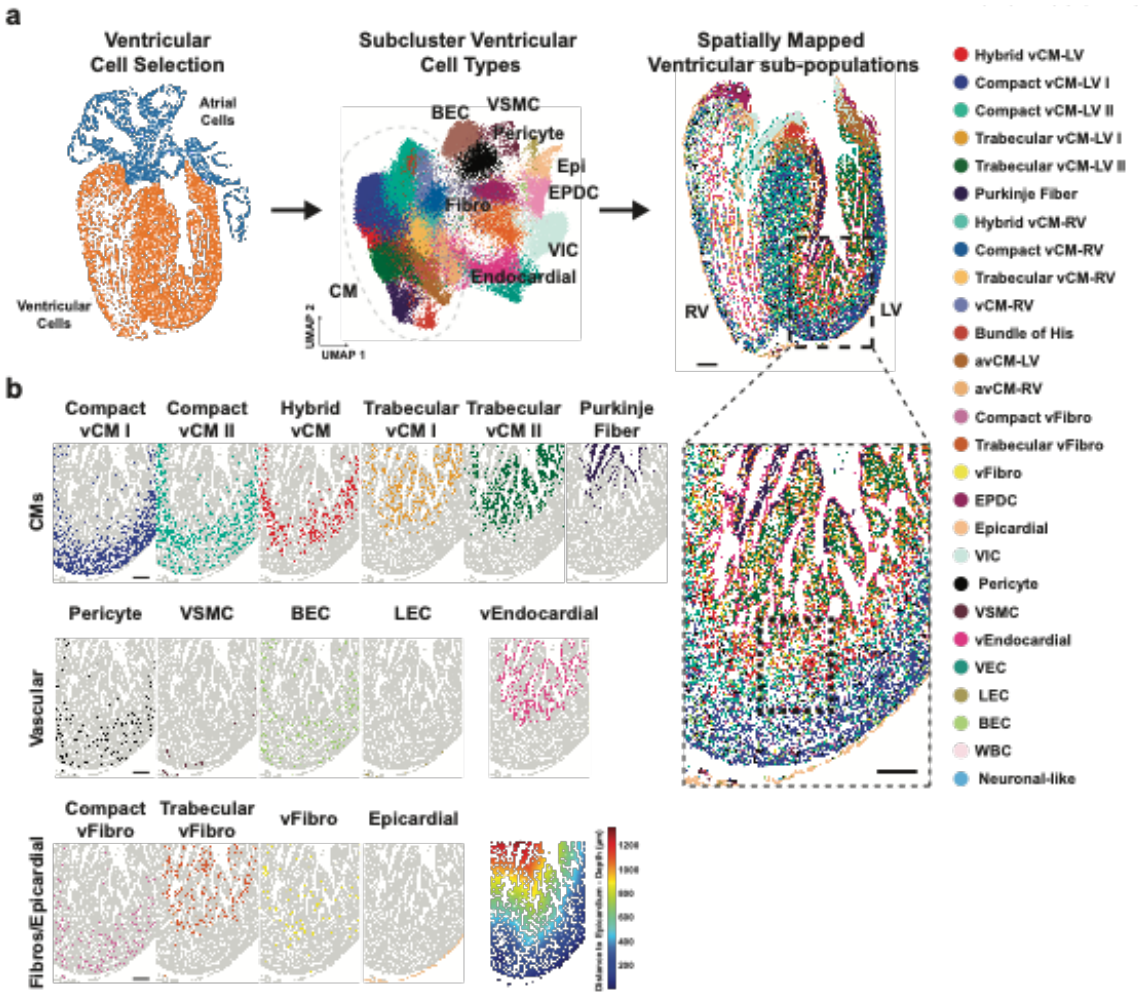


Figure 16. Subcluster analysis of the cardiac cells within the ventricles reveals new cardiomyocyte and fibroblast subpopulations

a, (Left) Cells selected as spatially located within the ventricle to perform subcluster analysis on. (Right) UMAP and corresponding spatial plot of re-identified clusters, colored by cell type. **b**, Spatial distribution of major cell types within the left ventricular free wall. (Bottom right) Plot of each cell's distance metric defined as a cell's distance to the epicardium. avCM, muscular leaflet of the valve cardiomyocyte; BEC, blood endothelial cell; EPDC, epicardial-derived cell; LEC, lymphatic endothelial cell; vCM, ventricular cardiomyocyte; VEC, valve endocardial cell; vEndocardial, ventricular endocardial; VICs, valve interstitial cells; VSMC, vascular smooth muscle cells; WBC, white blood cell. Scale bar, 250 μm.

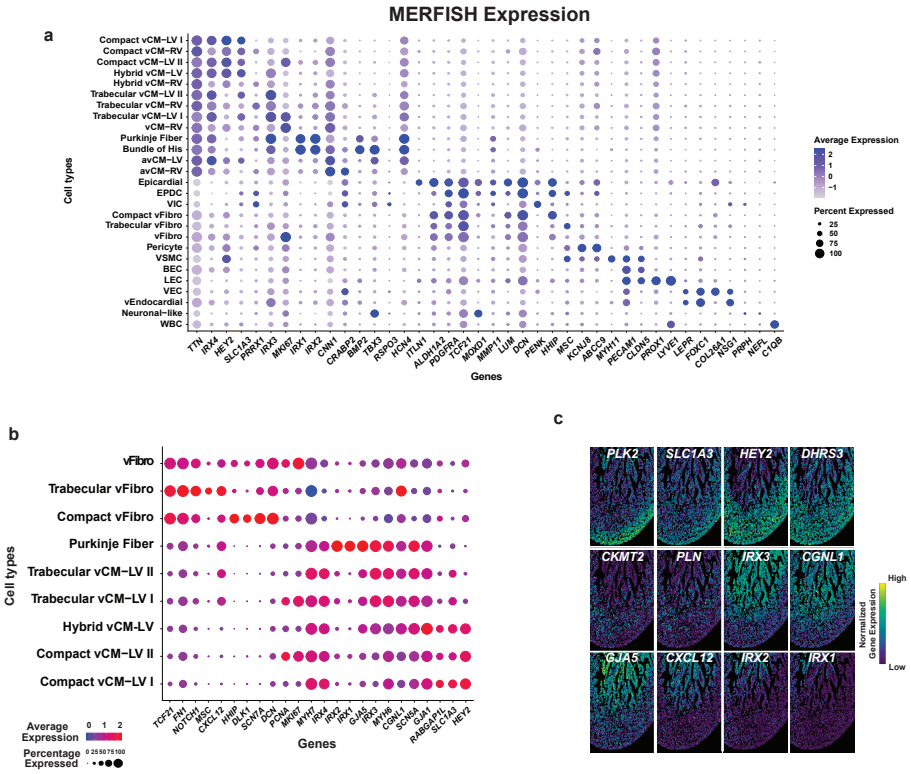


Figure 17. Identification of subclustered ventricular MERFISH cell types
a, Dotplot of marker genes identifying subclustered MERFISH populations within the developing ventricle. **b,** Dotplot of differentially expressed genes for the newly identified cell populations not originally identified in the overall heart MERFISH dataset. **c,** Spatial gene expression plots of twelve genes that correlate with ventricular wall depth.

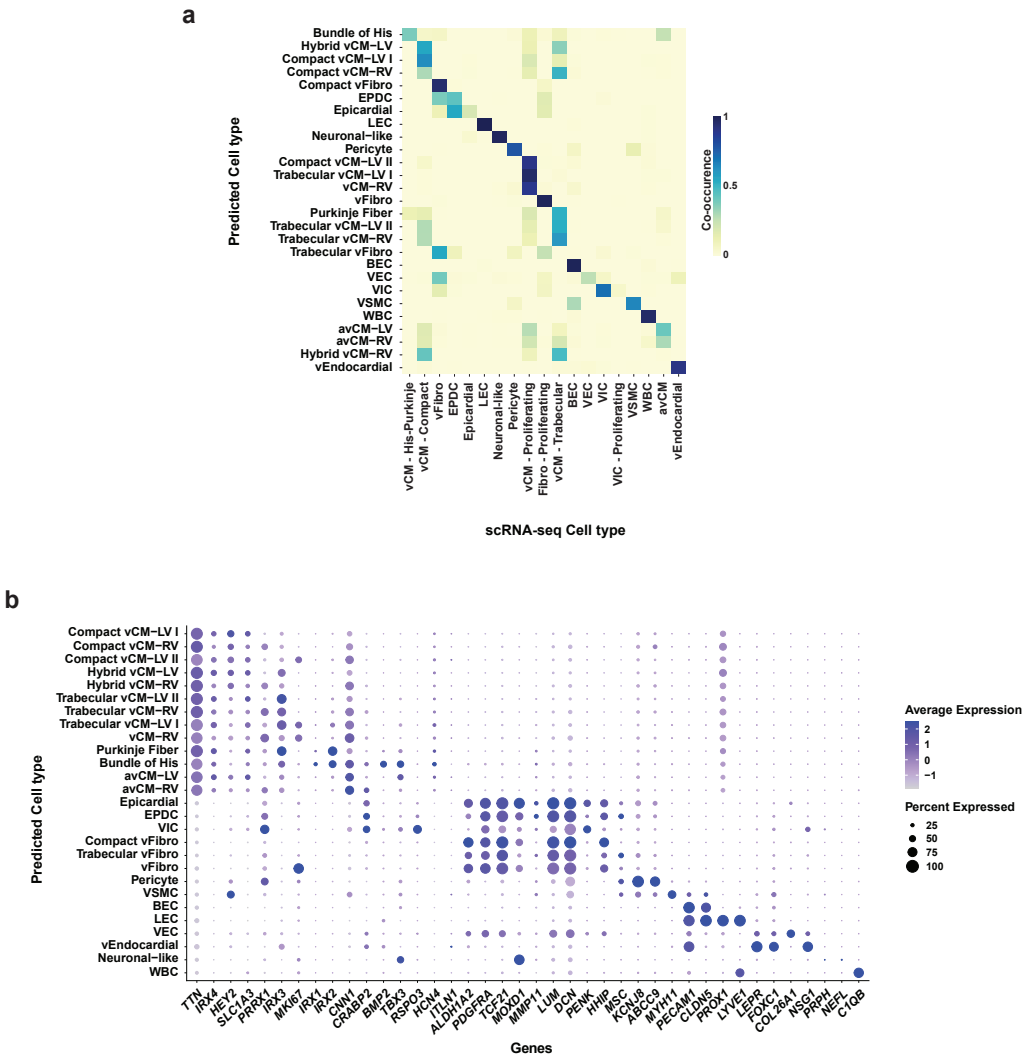


Figure 18. Integration of scRNA-seq and MERFISH can identify corresponding ventricular cell types in the transcriptomic dataset

a, Co-occurrence heatmap showing the counts of MERFISH identified ventricular subpopulations and nearest scRNA-seq cell (labeled by original label in Figure 13) in joint embedding space. **b,** Dotplot of marker gene expression for scRNA-seq cells grouped by transferred labels from MERFISH identified ventricular subpopulations, using the same gene markers as in Figure 17 to compare expression between MERFISH and scRNA-seq datasets.

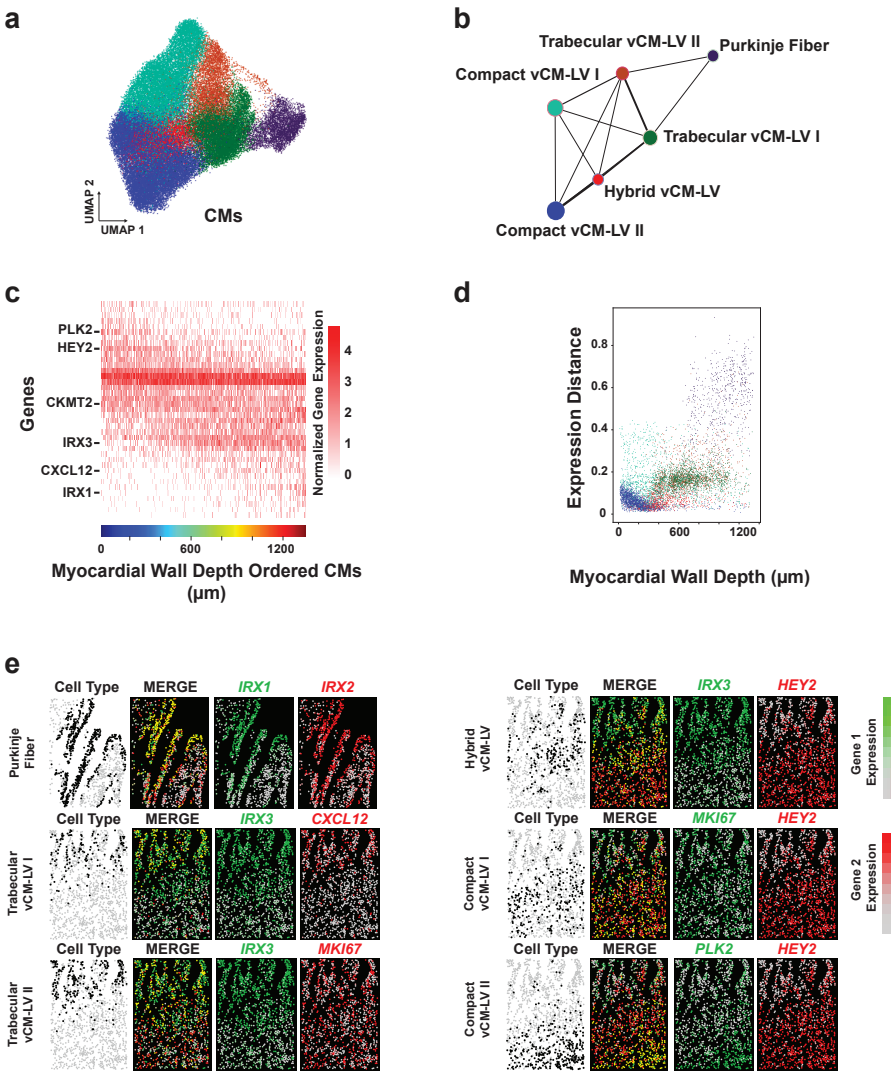


Figure 19. Cardiomyocytes within the ventricular wall are organized into a spatial gradient with correlated gene expression

a, UMAP of vCMs within the LV colored by cell type identity **b,** The degree of connectivity between cell types in a k-nearest neighbor graph for the vCM cells, with each cell type represented as a node colored as in **a**, and the weighted edges between nodes representing their connectivity. **c,** Heatmap of normalized expression of differentially expressed genes across ventricular wall depth. Differentially expressed genes refer to genes which the expression correlated substantially with wall depth (Methods). The colored bar at the bottom indicates the wall depth of the cell, calculated as the distance from the nearest epicardial cell. **d,** Scatter plot of the pseudotime versus ventricular wall depth for individual vCMs in the LV, exhibiting a Pearson correlation of 0.6. **e**, Spatial plots of zoomed in regions of the LV, with each of the six vCM populations colored separately by cell type, two gene markers (expression of gene 1 in green and expression of gene 2 in red), and the merge of the expression of the two marker genes.

Cell-Cell Communication Analysis Reveals Community Specific Interactions Mediating the Organization of Cardiomyocytes Along the Ventricular Wall

To further understand how the human ventricular wall is formed, we further investigated how distinct cardiac cell types coordinately interact to organize into the ventricular cardiac chambers. Toward this end, we defined communities of cardiac cell types which may be interacting with each other within the ventricle. By using a 150 μm distance for the zone, we allow for the detection of juxtacrine, autocrine, and paracrine signaling (**Figure 20a**). We discovered that the ventricular walls comprise at least three major cellular communities (Outer, Mid, Inner), with the left ventricle also containing a fourth subendocardial community that consists of purkinje fibers forming the fast cardiac conduction system of the ventricles (**Figure 20a**). Focusing on the left ventricle, we observed a broad range of specific cardiac cell types spatially located in these communities (**Figure 20b**). By measuring the spatial distribution of cell types within the ventricular wall in the context of the cellular communities, we discovered that the Mid community contained the highest number of unique cell types, and therefore exhibited the highest cellular complexity and potentially the most complex interactions. Because of our interest in compaction of the ventricular wall, we decided to focus on the Outer, Mid, and Inner communities because of their demonstrated roles in the process of compaction[23,41,55].

By utilizing the integrated MERFISH and scRNA-seq data (**Figure 18a,b**), we were able to run a cell-cell communication pipeline called CellChat[59] on the full transcriptome between cell types within the same community. Supporting the finding that the mid community was most cellularly complex, these cell-cell signaling analyses

revealed that the Mid community also displayed the highest number of cell-cell signaling interactions, nearly four times that of the next highest community (Outer) (**Figure 21a-d**).

As distinct cardiomyocyte cell types are spatially layered along the myocardial wall (**Figure 20b**), we further examined signals within these ventricular communities that may organize the spatial distribution of these cardiomyocytes (**Figure 22**). We discovered a wide array of cardiac cell types signaling to distinct cardiomyocytes with fibroblast subtypes displaying the strongest and highest number of interactions with these cardiomyocyte populations (**Figure 23a**). Consistent with these findings, we observed that the highest number of signaling pathways received by these cardiomyocytes were growth and ECM-related pathways that were broadly derived from fibroblast subtypes across the ventricular wall communities (**Figure 21b-d, 22**).

On the other hand, we also identified specific interactions differentially received by cardiomyocytes between these communities (**Figure 22, 23b**). Specifically, we observed Neuregulin-ErbB signaling between *NRG1*+ endocardial cells and *ERBB2/4*+ trabecular cardiomyocytes in the Inner community as previously reported[20,21] (**Figure 22**). Interestingly, we discovered Plexin-Semaphorin axon guidance signaling pathways as specific interactions within the Mid community, directing interactions between not only *PLXNA2/A4*+ trabecular cardiomyocytes and *SEMA3C/D*+ compact fibroblasts but also *PLXNA2/A4*+ trabecular cardiomyocytes and *SEMA6A/B*+ blood endothelial cells (**Figure 23c**). Supporting that these signaling pathways may mediate a complex multi-cellular interaction among cardiomyocytes, fibroblasts and endothelial cells in the Mid community, hybrid cardiomyocytes also receive these Plexin-Semaphorin signaling

pathways from the same corresponding cell types (**Figure 23c**). The strongest interactions appear to be between the Compact vFibro and Trabecular vCM-LV II and the VE and Trabecular vCM-LV II (**Figure 23c, 24**).

In line with these analyses, we observed that cell types participating in these signaling interactions are spatially co-localized (**Figure 23d**). Furthermore, by using our integrated dataset, we can spatially visualize the localization of specific semaphorins and plexins, and find that the corresponding ligands and receptors are expressed in both complementary and overlapping domains in the Mid community, consistent with a ligand-receptor interaction (**Figure 23e**). As *SEMA3C/D* and *SEMA6A/B* have been shown to attract and repel *PLXNA2/A4* expressing cells via paracrine and juxtacrine signaling pathways, respectively[60–64], we hypothesize that *SEMA3C/D*-expressing compact ventricular fibroblasts attract *PLXNA2/A4* expressing trabecular (and hybrid) cardiomyocytes to the Mid/Outer myocardial community layers, whereas *SEMA6A/B*-expressing blood endothelial cells may prevent these trabecular (and hybrid) cardiomyocytes from further migrating by repelling them when contacting blood endothelial cells. This interaction may be influencing cardiomyocyte organization within the ventricular wall. To further explore this possibility, we created an *in vitro* hPSC multilayer cardiac organoid system.

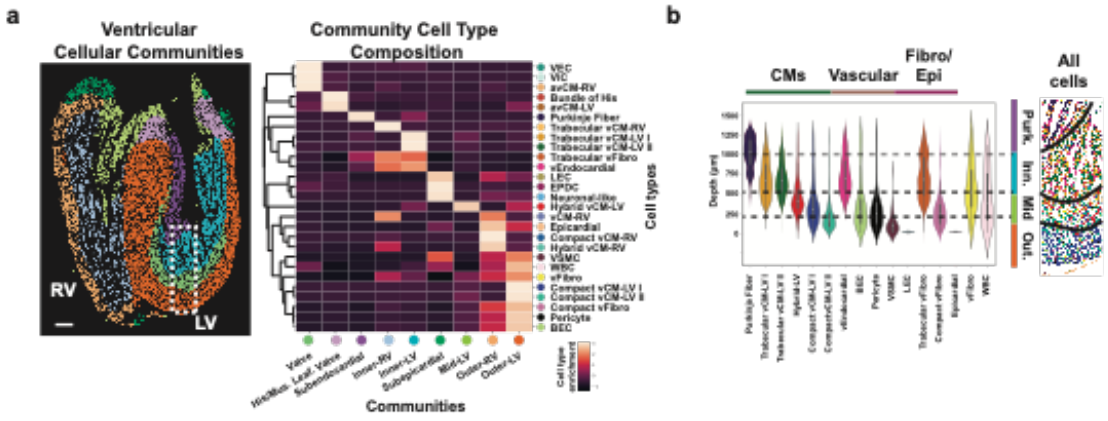
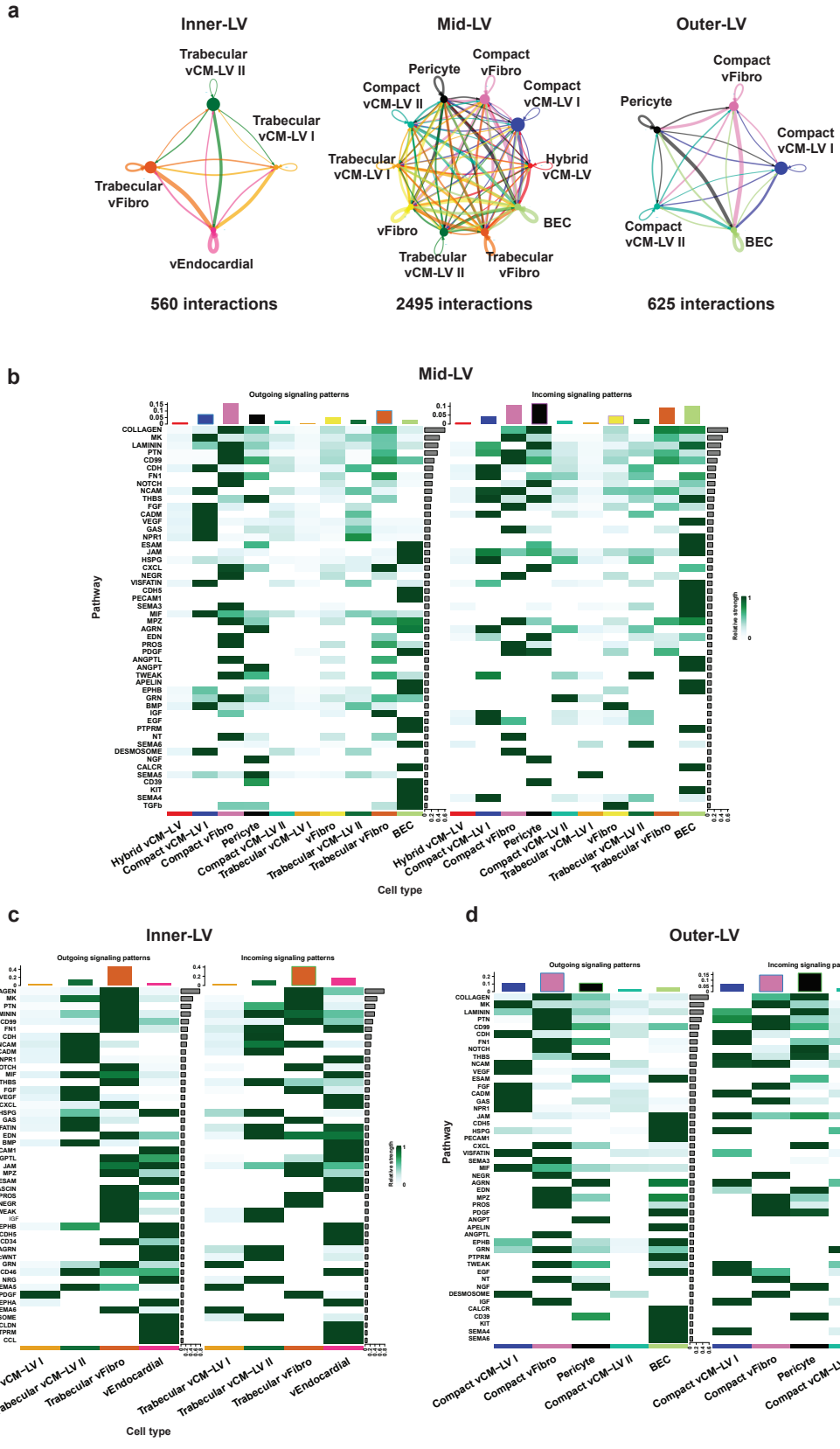


Figure 20. Identification of cellular communities within the ventricles reveals a Mid community within the left ventricular wall

a, (Left) Spatial distribution of nine distinct cellular communities within the ventricle identified based on the ventricle subset defined cell types. (Right) The spatial neighborhood frequencies (enrichment score) of each of the MERFISH identified ventricular cell types within each ventricular cellular community. **b**, Violin plots showing the distribution of wall depths for all of the cell types within the ventricular wall of the LV, with dashed lines denoting the approximate wall depth of each cellular community identified within the LV. (Right) Spatial plot of the LV, colored by cell type, with black lines marking the cellular community. Scale bar, 250 μm.

Figure 21. Cell-cell signaling analysis of Inner-LV, Mid-LV, and Outer-LV communities

a, Chord diagrams of the cell-cell signaling interactions detected in the Inner, Mid, and Outer communities of the left ventricle. The size of the node represents the size of the cluster, and the width of the edge represents the number of interactions between a pair of clusters. **b-d,** Heatmaps of the signaling strength of each signaling pathway for each pair of clusters within the b, Mid-LV community, c, Inner-LV community and d, Outer-LV community. Signaling pathways are ordered by overall strength of signaling. The left side of the heatmap represents outgoing (ligand) signaling and the right side represents incoming (receptor) signaling.


a
Inner-LV
Trabecular vCM-LV II
Trabecular vCM-LV I
Trabecular vFibro
vEndocardial
560 interactions
Mid-LV
Pericyte
Compact vFibro
Compact vCM-LV II
Compact vCM-LV I
Trabecular vCM-LV I
Hybrid vCM-LV
vFibro
BEC
Trabecular vCM-LV II
Trabecular vFibro
2495 interactions
Outer-LV
Compact vFibro
Pericyte
Compact vCM-LV I
Compact vCM-LV II
BEC
625 interactions
b
Mid-LV
Outgoing signaling patterns
Incoming signaling patterns
Pathway
Relative strength
Hybrid vCM-LV
Compact vCM-LV I
Compact vFibro
Pericyte
Compact vCM-LV II
Trabecular vCM-LV I
vFibro
Trabecular vCM-LV II
Trabecular vFibro
BEC
Cell type
c
Inner-LV
Outgoing signaling patterns
Incoming signaling patterns
Pathway
Trabecular vCM-LV I
Trabecular vCM-LV II
Trabecular vFibro
vEndocardial
Cell type
d
Outer-LV
Outgoing signaling patterns
Incoming signaling patterns
Pathway
Compact vCM-LV I
Compact vFibro
Pericyte
Compact vCM-LV II
BEC

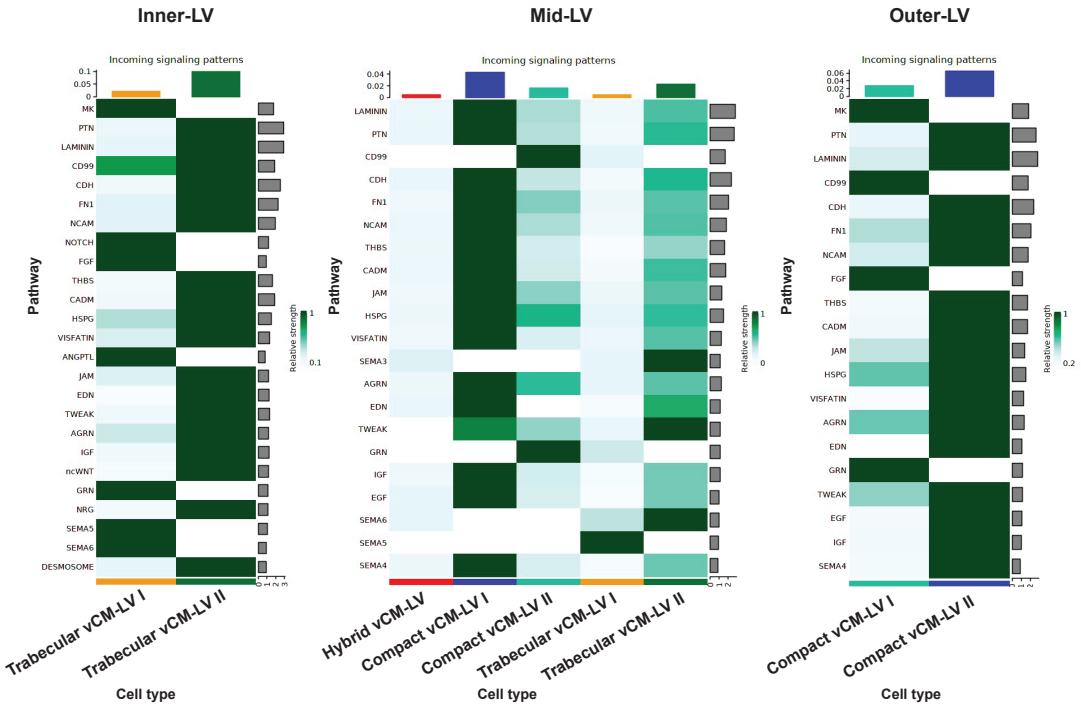
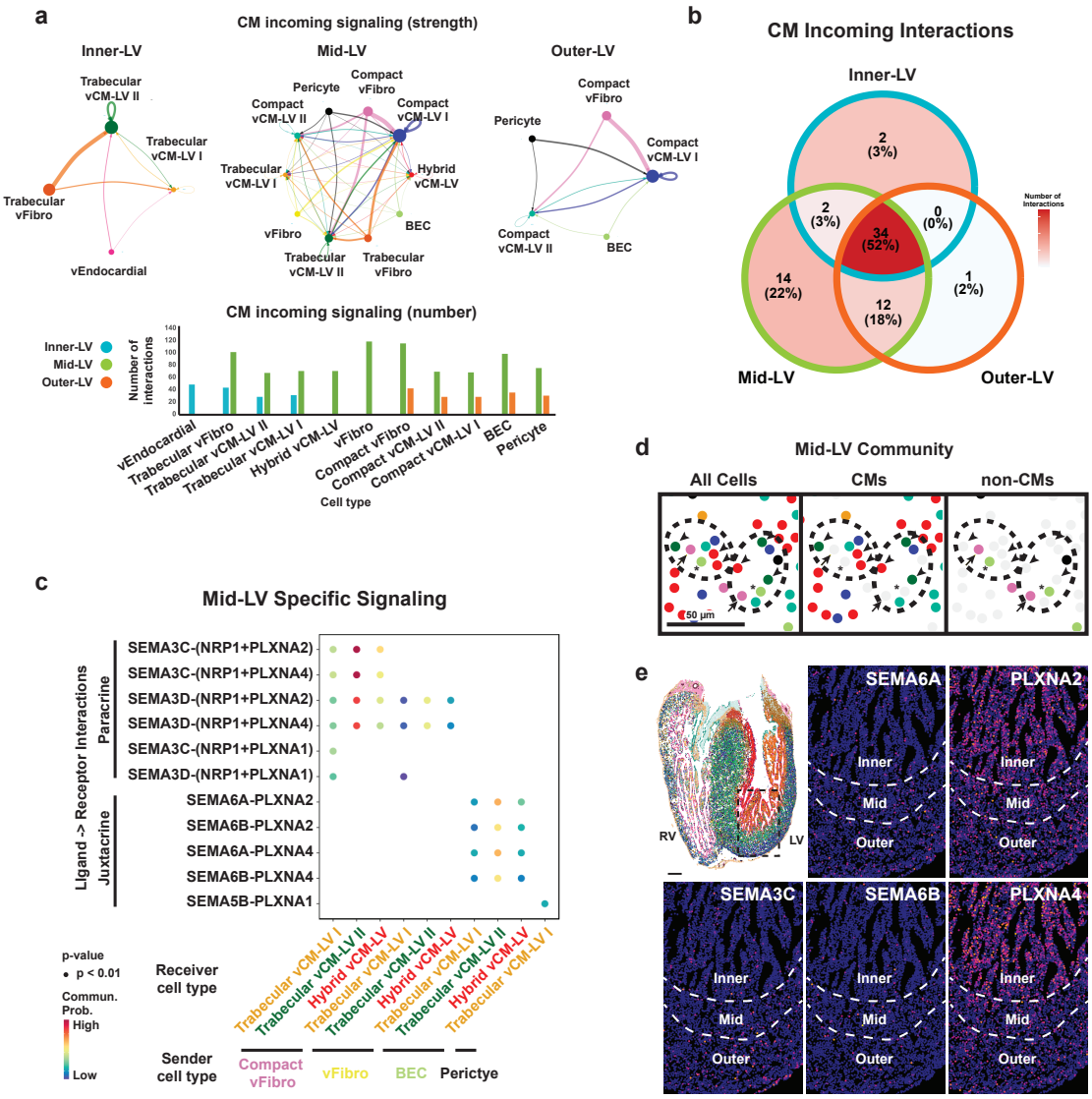


Figure 22. Interactions regulating cardiomyocytes in Inner-LV, Mid-LV, and Outer-LV communities

Heatmaps of the incoming (receptor) signaling strength of the major signaling pathways received by cardiomyocyte populations within the Mid-LV, Inner-LV, and Outer-LV communities. Each row is labeled by a separate signaling pathway received by the cardiomyocyte populations within the LV.

Figure 23. Semaphorin interactions are specific for the Mid-LV community
a, (Top) Chord diagrams of the cell-cell signaling interactions received by cardiomyocytes in the Inner, Mid, and Outer communities of the left ventricle. The size of the node represents the size of the cluster, and the width of the edge represents the strength of interaction between a pair of clusters. (Bottom) Bar chart showing the number of interactions received by cardiomyocytes from other cell populations within the LV, separated by community. **b**, Venn diagram of the cell-cell interactions received by cardiomyocytes within the Inner-LV, Mid-LV, and Outer-LV communities, showing that most are common between the communities but a few are specific. **c,** Dotplot of specific significant semaphorin interactions within the Mid-LV community. The dots are colored by communication probability, based on the expression of the ligand and receptor. The x-axis shows the specific ligand-receptor interactions, and the y-axis show the pair of cell populations involved in the cell-cell interaction. **d**, Spatial plots of a region within the left ventricle showing the close proximity of cells involved in the semaphorin interaction, colored by all cells, cardiomyocytes, and non-cardiomyocytes. Scale bar, 50 μm. **e**, Spatial gene expression plots of the ligands and receptors of the semaphorin interaction, inferred by the integrated scRNA-seq and MERFISH datasets. Scale bar, 250 μm.


a
CM incoming signaling (strength)
Inner-LV
Mid-LV
Outer-LV
Trabecular vCM-LV II
Trabecular vCM-LV I
Trabecular vFibro
vEndocardial
Pericyte
Compact vFibro
Compact vCM-LV II
Compact vCM-LV I
Hybrid vCM-LV
vFibro
BEC
CM incoming signaling (number)
Number of interactions
Cell type
b
CM Incoming Interactions
Inner-LV
Mid-LV
Outer-LV
2 (3%)
2 (3%)
0 (0%)
34 (52%)
14 (22%)
1 (2%)
12 (18%)
c
Mid-LV Specific Signaling
Ligand -> Receptor Interactions
Paracrine
Juxtacrine
SEMA3C-(NRP1+PLXNA2)
SEMA3C-(NRP1+PLXNA4)
SEMA3D-(NRP1+PLXNA2)
SEMA3D-(NRP1+PLXNA4)
SEMA3C-(NRP1+PLXNA1)
SEMA3D-(NRP1+PLXNA1)
SEMA6A-PLXNA2
SEMA6B-PLXNA2
SEMA6A-PLXNA4
SEMA6B-PLXNA4
SEMA5B-PLXNA1
p-value
p < 0.01
Commun. Prob.
High
Low
Receiver cell type
Sender cell type
Compact vFibro
vFibro
BEC
Pericyte
d
Mid-LV Community
All Cells
CMs
non-CMs
50 μm
e
RV
LV
SEMA6A
PLXNA2
SEMA3C
SEMA6B
PLXNA4
Inner
Mid
Outer

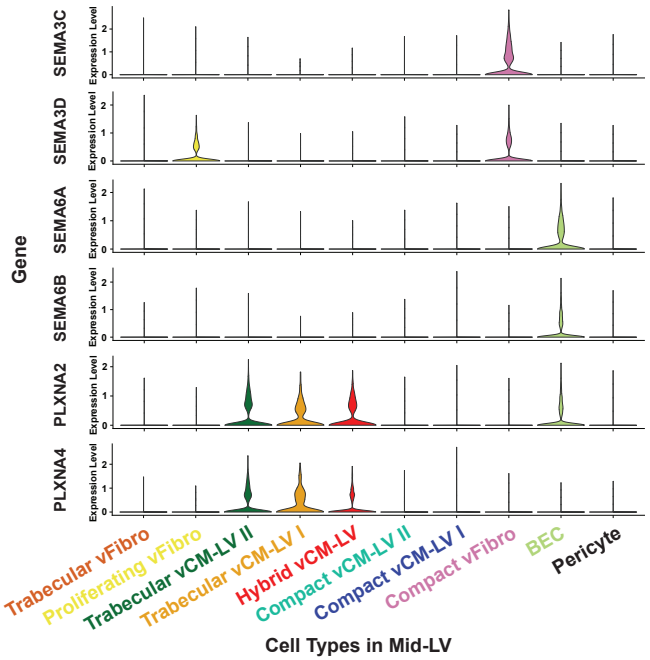


Figure 24. Expression of semaphorin ligands and receptors
Violin plots of major semaphorin ligands and receptors detected within the Mid-LV community, showing the expression of SEMA3C, SEMA3D, SEMA6A, SEMA6B, PLXNA2, and PLXNA4 across all cell types in the Mid-LV community. The cells grouped based on the predicted cell type labels from Figure 18.

PLXNA2/PLXNA4+ hPSC Trabecular-like cardiomyocytes can migrate toward *SEMA3C* in a bioprinted model of the ventricular wall

To test the how semaphorin signaling may be influencing cardiomyocyte organization, we utilized a highly efficient monolayer cardiac differentiation[65] system to create human pluripotent stem cell derived cardiomyocytes (hPSC-cardiomyocytes) to a stage where ventricular cardiomyocytes can be identified[66]. To this end, we differentiated H9 hPSCs containing the *TNNT2:EGFP* cardiomyocyte-specific transgene to Day 25 of differentiation and performed scRNA-seq (**Figure 25a**). Gene expression analyses revealed that the majority (>80%) of the cells were *TTN+/TNNT2*+ cardiomyocytes, with a smaller subpopulation of fibroblasts (**Figure 25a**). Of the cardiomyocytes, the vast majority (>90%) were IRX4+ ventricular cardiomyocytes, and these ventricular hPSC-cardiomyocytes weakly expressed markers of compact (*HEY2*) and trabecular (*IRX3*) cardiomyocytes, suggesting that these may correspond to 'hybrid' vCMs, potentially because of a lack of signaling to specify these cardiomyocytes (**Figure 25a**). To investigate how these hPSC-cardiomyocytes compare to those found *in vivo*, we utilized our Harmony integration method to project both datasets into the same embedding space and found that the majority of hPSC-derived cells are transcriptionally similar to *in vivo* vCMs but do not overlap, suggesting that the hPSC-cardiomyocytes are not fully *in vivo*-like, consistent with their low level of expression of compact (*HEY2*) and trabecular markers (*IRX3*) (**Figure 25b**). Closer examination of marker expression of *in vivo* vs hPSC-cardiomyocytes revealed that hPSC-cardiomyocytes express lower levels of lineage specific and semaphorin signaling

specific markers compared to their *in vivo* counterparts, suggesting a lack of maturity within the hPSC system which has been well documented (**Figure 25c**)[67].

To differentiate trabecular-like hPSC-cardiomyocytes, we utilized a published protocol that employs *NRG1* treatment[68], an established signaling pathway influencing trabecular development[20,21] and one that we also find as specific to the Inner-LV community between endocardial cells and Trabecular CMs (**Figure 22**). Using this method, hPSC-cardiomyocytes displayed an upregulation of markers of trabecular vCMs (*IRX3*) and downregulation of markers of compact vCMs (*HEY2*), which is consistent with previous studies[68] (**Figure 25d**). Consistent with this, trabecular-like hPSC-cardiomyocytes also upregulated the expression of plexin receptors expressed by *in vivo* trabecular vCMs, *PLXNA2* and *PLXNA4*, suggesting that there is a general trabecular vCM program that may be activated on in these cells upon treatment with *NRG1* (**Figure 25d**).

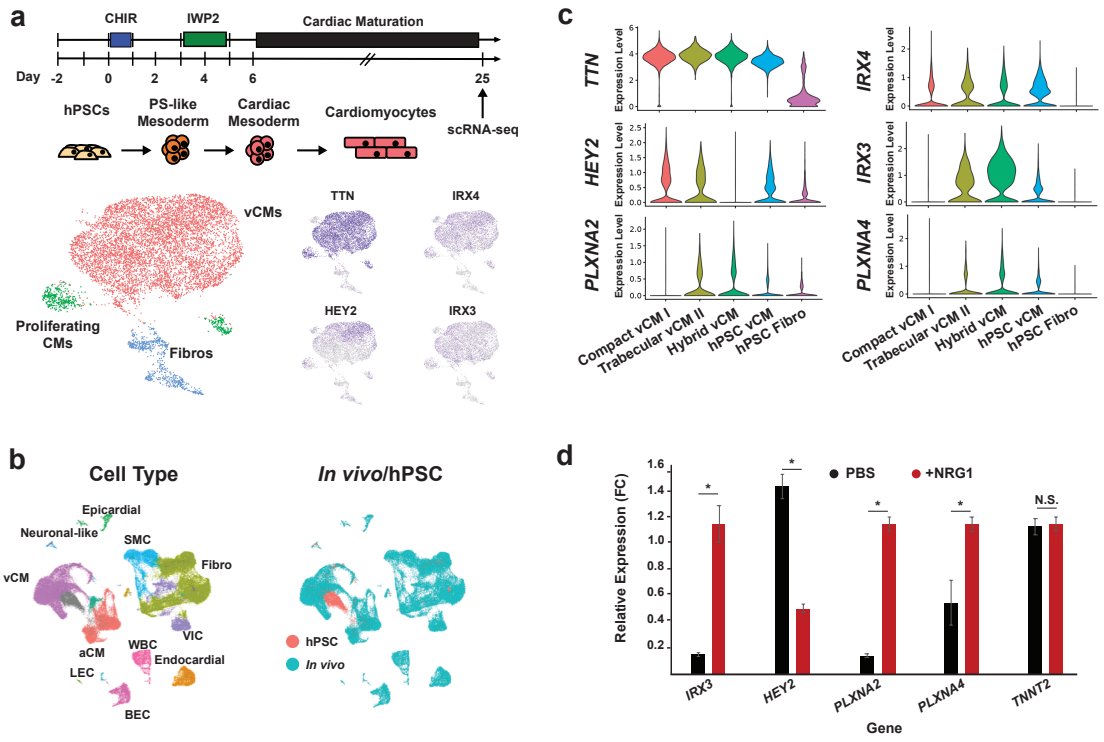


Figure 25. *NRG1* treatment can differentiate Trabecular-like cardiomyocytes that express the receptors *PLXNA2* and *PLXNA4*.

a, Schematic of differentiation of hPSC-cardiomyocytes using the Wnt-based protocol, with scRNA-seq performed at Day 25 of differentiation. (Bottom) UMAP plots showing the expression of genes related to ventricular cardiomyocytes showing that most hPSC-cardiomyocytes are ventricular (IRX4+). **b**, UMAP of integrated scRNA-seq datasets of the Day 25 hPSC-cardiomyocytes with the 13W *in vivo* dataset from the developing human heart, colored by (Left) *in vivo* cell type and (Right). **c**, Violin plots showing the expression of key marker genes of ventricular cardiomyocyte populations and plexin receptors showing that hPSC-cardiomyocytes express lower levels of *IRX3*, *PLXNA2*, and *PLXNA4* compared to *in vivo* trabecular cardiomyocytes. d, Gene expression analysis (RT-qPCR) of *IRX3*, *HEY2*, *PLXNA2*, *PLXNA4*, and *TNNT2* performed on Day 25 hPSC-cardiomyocytes treated with 50 ng/ml NRG1 for 10 days (Day 15 to 25 of differentiation), showing that the expression of *IRX3*, *PLXNA2*, and *PLXNA4* increases with NRG1 treatment. N = 4 per condition. Error bars are SEM. *p $\leq$ .05 by Student's t-test, N.S., not significant.

To investigate whether these trabecular-like hPSC-cardiomyocytes could migrate in response to semaphorin guidance cues, we utilized a rapid 3D bioprinting technique. To this end, we bioprinted hPSC-cardiomyocytes into a multi-layered cardiac organoid to model the ventricular wall, with 3 separate layers mimicking the inner (300 μm), mid (150 μm), and outer (150 μm) layers of the ventricular wall (**Figure 26a**). Within the top 'inner' layer, we bioprinted either control (PBS) or *NRG1* treated hPSC-cardiomyocytes to examine whether *NRG1*-treated hPSC cardiomyocytes specifically migrate in response to semaphorin guidance cues (**Figure 26a**). This was done by adding purified semaphorin signaling factors (either *SEMA3C* or a mixture of *SEMA6A/SEMA6B*) to either the 'mid' layer, the bottom or 'outer' layer, or both. In contrast to organoids with bioprinted control hPSC-cardiomyocytes, which do not migrate out of the 'inner' layer, organoids with *NRG1*-treated hPSC-cardiomyocytes consistently migrate out of the 'inner' layer, through the 'mid' layer, and into the 'outer' layer of the organoid (248 μm by Day 5) (**Figure 26b,c**). Neither control nor *NRG1*-treated hPSC-cardiomyocytes migrated in response to the printing matrix alone or to *SEMA6A/SEMA6B*, suggesting that *SEMA6A/SEMA6B* may be blocking migration (**Figure 26b,c**). To test whether *SEMA6A/SEMA6B* can block the migration towards *SEMA3C*, we printed control and NRG1-treated hPSC-cardiomyocytes with two different combinations of semaphorins across the 'mid' and 'outer' layer: 1) *SEMA3C* in the 'mid' layer and *SEMA6A/SEMA6B* in the 'outer' layer, 2) *SEMA6A/SEMA6B* in the 'mid' layer and *SEMA3C* in the 'outer' layer. When *NRG1* treated hPSC-cardiomyocytes are printed in condition 1), they are able to migrate to the 'mid' layer but do not migrate to the 'outer' layer containing *SEMA6A/SEMA6B* (110 μm by Day 5) (**Figure 26b,c**). On the other hand, when *NRG1*

treated hPSC-cardiomyocytes are printed in condition 2), they are not able to migrate at all, supporting the model that *SEMA3C* acts as an attractive guidance cue for migrating trabecular cardiomyocytes, whereas *SEMA6A/SEMA6B* acts as a repulsive cue to block this migration (**Figure 26b,c**). Organoids with bioprinted control hPSC-cardiomyocytes do not migrate in either condition 1) or 2), consistent with their low level of *PLXNA2/PLXNA4* expression (**Figure 26b,c**). Altogether, these results reveal that hPSC-derived trabecular-like cardiomyocytes that express *PLXNA2/PLXNA4* migrate towards an attractive SEMA3C guidance cue, and that this migration is blocked by *SEMA6A/SEMA6B*, suggesting that this interaction may be what is happening during human development to integrate the trabecular cardiomyocytes within the Inner layer with the compact cardiomyocytes within the Outer layer to contribute to ventricular wall thickening during ventricular compaction (**Figure 27**).

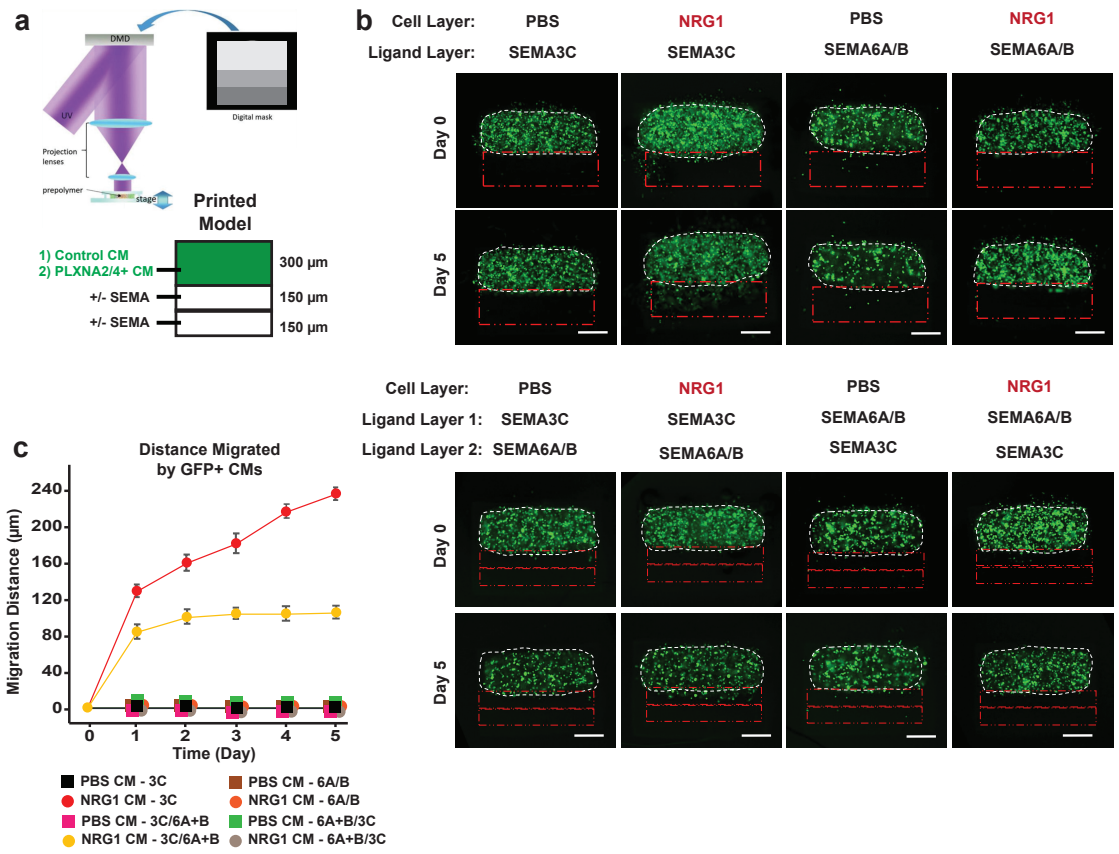


Figure 26. Trabecular-like CMs differentiated from human pluripotent stem cells (hPSCs) migrate toward an attracting semaphorin signal.

a, Schematic of the bioprinted multilayered cardiac organoid, with a layer of printed cardiomyocytes (Cell layer, 300 µm), followed by two layers of extracellular matrix (Ligand layer, 150 µm each) that can be mixed with recombinant proteins. **b**, Bioprinted multilayered cardiac organoid of GFP+ cardiomyocytes that have been either treated with *NRG1* or PBS (control), with fluorescent images showing the position of the printed cardiomyocytes on the day of printing (Day 0) and five days after printing (Day 5), with the red dashed lines showing the extracellular matrix layers with either (Top) one semaphorin protein added or (Bottom) a combination of semaphorin proteins added to each of the extracellular ligand layers. **c**, Quantification of the distance migrated by the different conditions shown in b. Cardiomyocytes treated with *NRG1* that express *PLXNA2*/*PLXNA4* migrate the furthest when the ligand layer with mixed with *SEMA3C* protein, and this migration is reduced when a layer of *SEMA6A*/*SEMA6B* is added. N = 3 per condition. Error bars are SEM. Scale bar, 250 µm.

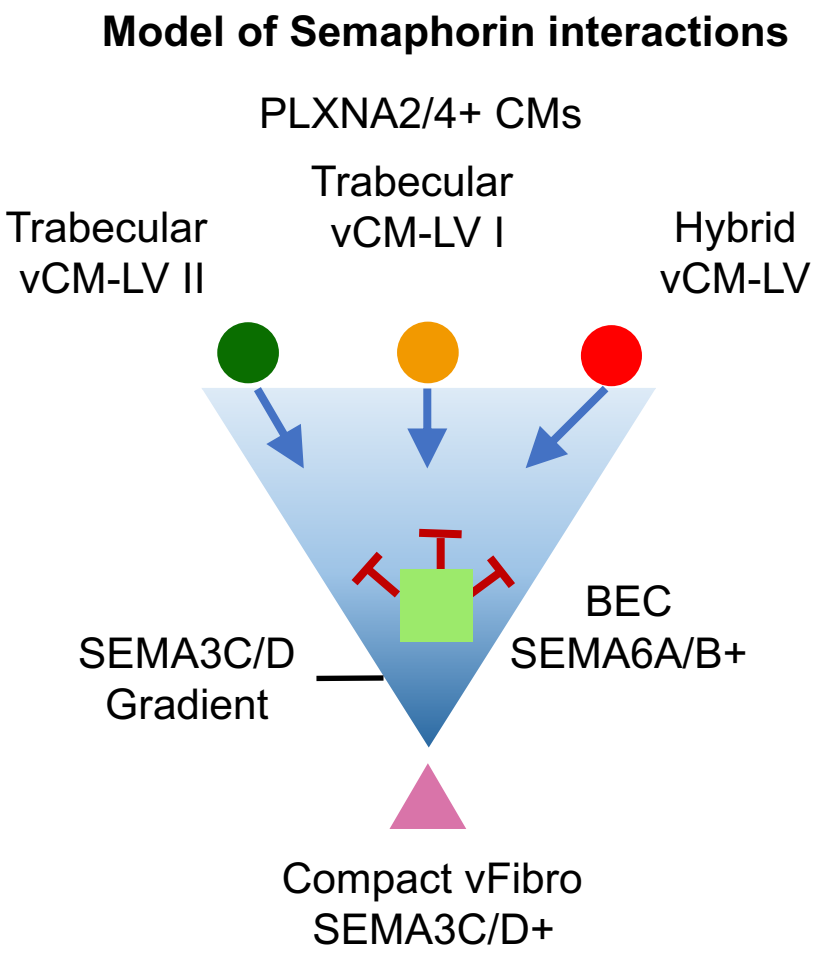


Figure 27. Model of semaphorin signaling regulating Trabecular and Hybrid vCM organization within the ventricular wall of human heart development
Model depicting how semaphorin signaling may influence cardiomyocyte organization within the ventricular wall during ventricular compaction. As compaction proceeds and trabecular vCMs within the inner layer become integrated into the outer layer, there is a gradient formed by the attractive *SEMA3C/D* guidance cue for trabecular and hybrid vCMs to migrate down into the mid community secreted by compact vFibro. As these vCMs migrate down into the mid layer, they encounter *SEMA6A/B*-expressing blood endothelial cells, and this interaction may prevent these trabecular and hybrid vCMs from migrating further into the outer layer.

3.3 DISCUSSION

Overall, through identifying multiple communities within the ventricular wall, and the cell interactions within those communities, we discover that semaphorin signaling is active in the Mid community of the left ventricular wall, which instructs the allocation of trabecular cardiomyocytes to migrate from the Inner layer and integrate with the developing Mid and Outer layers. In addition to the semaphorin signaling within the Mid community, we also find that in general across the ventricle, the highest number of signaling pathways received by these cardiomyocytes were growth and ECM-related pathways that were broadly derived from fibroblast subtypes across the ventricular wall, which is consistent with the fact that the human heart is undergoing rapid expansion during this time in development. Specific to the Inner community, we observed Neuregulin-ErbB signaling between *NRG1*+ endocardial cells and *ERBB2/4*+ trabecular cardiomyocytes, which is consistent with previous reports in other model systems[20,21].

Semaphorin signaling has been reported to regulate multiple aspects of cardiac development, mainly through the regulation of cardiac neural crest cell migration, and more recently through endocardial regulation of ECM dynamics[27,28]. *PLXNA2* and *PLXNA4* have been previously reported to be expressed in cardiac neural crest cells, regulating their migration to the outflow tract[60]. The expression of these plexin receptors have not been reported in trabecular cardiomyocytes previously, and therefore suggests a new role of semaphorin signaling in regulating trabecular cardiomyocyte migration in developing human hearts. This attractive signaling cue of *SEMA3C* supports a model where trabecular cardiomyocytes migrate to the Mid/Outer layers of the ventricular wall, causing the trabeculae to integrate with the growing compact myocardium and

supporting that the process of compaction whereby trabecular cardiomyocytes contribute to the thickening myocardium does happen in humans, which has been an area of great debate recently[69]. The thickening of the myocardial wall through compaction is composed of multiple processes including the expansion of the compact myocardium as well as the coalescence of trabecular myocardium[24,25]. More studies will be required to fully understand the contribution of each process to the thickening of the ventricular wall in humans. Within the mouse heart, the proliferation of compact cardiomyocytes is major contributing factor to the thickening ventricular wall[24], but whether this process is enough to support the size of the human ventricular wall warrants further investigation. Conversely, further investigation is needed to confirm if the process of semaphorin regulated organization of cardiomyocytes is an evolutionarily conserved pathway regulating compaction in mouse hearts.

In conclusion, we have uncovered a spatial and cellular diversity of cardiomyocytes within the ventricular wall and discovered a new role of semaphorin signaling in regulating the organization within the developing human ventricle that may play a role in ventricular compaction. Our findings support a model where *PLXNA2* and *PLXNA4* expressing trabecular (and hybrid) cardiomyocytes migrate into Mid and Outer myocardial layers, attracted by a secreted *SEMA3C* ligand that is expressed by compact ventricular fibroblasts. These migrating cardiomyocytes are then come into contact with *SEMA6A/B*-expressing blood endothelial cells may prevent these trabecular (and hybrid) cardiomyocytes from further migration by the juxtacrine mediated repulsion (**Figure 27**).

3.4 MATERIALS AND METHODS

Cell clustering analysis of ventricular cells in MERFISH

After manually defining the ventricular region, we subset the MERFISH dataset to the ventricular cells and performed Leiden clustering at a resolution of 5. The top 20 genes identified by the rank_gene_groups() function, as well as the spatial information, were used to annotate each cluster.

Integration of single-cell RNA-seq and MERFISH on ventricular dataset

We first subsetted both the scRNA-seq and MERFISH to only genes interrogated by both modalities, as well as to only cells within the ventricles of the heart. We then utilized SCANPY's implementation of Harmony to project both the scRNA-seq and MERFISH dataset into a shared PCA space. The dimensionality of the joint embedding was further reduced using UMAP (min_dist=0.3) to visualize the space in two dimensions using a k-nearest neighbor graph constructed with a k of 30 and the Pearson correlation distance metric to match the parameters used by the Harmony authors[47].

To impute a complete expression profile and cell type for each MERFISH profile, we assigned the expression profile and cell type of the closest scRNA-seq cell in the Harmony PCA space using the euclidean distance metric.

Ventricular wall depth measurement

Each MERFISH cell's ventricular wall depth was determined as the distance to the nearest epicardial or EPDC cell, which both lie on the outer layer of the heart.

Connectivity and pseudotime analyses of ventricular cardiomyocytes

To visualize the similarity in the gene expression profiles of the ventricular cardiomyocyte subclusters, we isolated the Compact-vCM I, Compact-vCM II, Hybrid vCM, Trabecular-vCM I, Trabecular-vCM II, and Purkinje Fiber clusters, reperformed PCA, nearest neighbor graph construction, and UMAP utilizing SCANPY with default parameters. We then used Scanpy's implementation of Partition Based Graph Abstraction (PAGA) to construct a graph where the nodes represent different ventricular cardiomyocyte clusters, and the edges represent the degree of connectivity between the clusters in the neighborhood graph. This captures the a measure of similarity between the cell type clusters.

To identify depth correlated genes, we computed the Pearson correlation coefficient between expression and ventricular wall depth for each gene. We considered genes with a correlation greater than 0.2 to be depth correlated.

Next, the diffusion pseudotime distance on the vCM nearest neighbor graph from the Compact-vCM I were calculated using Scanpy's scanpy.tl.dpt() function with default parameters. We chose the Compact-vCM I cell with the highest HEY2 expression as the root cell for the pseudotime calculation. We note that this metric is often reported as

pseudotime to represent developmental changes, but in this case we use it simply as a metric for expression similarity to Compact-vCM I cells.

Cell-cell interaction analysis of ventricular wall communities

We applied CellChat to our scRNA-seq dataset to identify community specific cell-cell interactions. Community annotations derived above for each MERFISH cell were transferred to the nearest scRNA-seq profile in the Harmony shared embedding space. The ventricular communities Inner-LV, Mid-LV and Outer-LV were each analyzed separately, by analyzing the scRNA-seq profiles assigned to those communities. For each ventricular cellular community we only considered cell types that represented at least 3.5% of the community in the MERFISH dataset. We followed CellChat's suggested workflow, using CellChat's database of human ligand-receptor interactions, identifying overexpressed genes, computing interaction probabilities, and identifying significant interactions, utilizing default parameters.

hPSC culture and cardiac differentiations

The H9-hTnnTZ-pGZ-D2 (WiCell) human embryonic stem cell line (hPSCs) was cultured in E8 medium and grown on Geltrex (Gibco) coated plates. Differentiation of hPSCs from this line into cardiomyocytes was performed using established protocols as previously described[70]. Briefly, hPSCs were grown to 80% confluency, and on D0, cells were cultured with RPMI/B27 without insulin (B27-insulin) containing 10 mM CHIR. 24 hours post CHIR application, the media was replaced with fresh B27-insulin and the cells were cultured for another 48 hours. Next (D3), 5 mM IWP2 was supplemented to

B27-insulin and cultured for another 48 hours. Following this (D5), the B27-insulin with IWP2 was replaced with fresh B27-insulin for another 48 hours. From D7 onwards, cells were maintained in RPMI/B27 with insulin (B27+insulin). On D15, this B27+insulin media was then either supplemented with *NRG1* (50 ng/ml) or PBS, and further cultured until D30+, refreshing the media every 3 days.

RNA expression was measured by real-time quantitative polymerase chain reaction (RT-qPCR). RNA was extracted using the TRIzol Reagent (ThermoFisher) and Direct-zol RNA MiniPrep Kit (Zymo Research). Next, the cDNA was generated using 1000 ng RNA and iScript Reverse Transcription Supermix (Bio-Rad), then diluted 1:10 in UltraPure DNase/RNase-Free Distilled Water (ThermoFisher). RT-qPCR was performed using Power SYBR Green Master Mix (ThermoFisher) according to the manufacturer's recommendations, and a two-step amplification CFX_2stepAmp protocol on a Bio-Rad CFX Connect Real-Time PCR Detection System. Data was analyzed using the $2^{-\Delta\Delta Ct}$ method. All gene expression was normalized to the expression of TATA box binding protein (*TBP*). The following RT-qPCR primers were used (Gene ID, forward primer, reverse primer):

HEY2, GGGTAAAGGCTACTTTGACG, GCAACTTCTGTTAGGCACTC

IRX3, TCTGCCTTTGTGTGTGTG, GTGGCAGCAGCTCATTTA

PLXNA2, AAGTATAGTGAGGAGCTCATCG, CTGCTCCACCTTATAAGCCA

PLXNA4, AATGACCGCATTAAGGAGC, AATGGTTAAGAGCGCACTG

TBP, TGAGTTGCTCATACCGTGCTGCTA, CCCTCAAACCAACTTGTCAACAGC

TNNT2, GGAGGAGTCCAAACCAAAGCC, TCAAAGTCCACTCTCTCTCCATC

Single cell isolation and sequencing of D25 hPSC cardiomyocytes

Single cell RNA-sequencing (scRNA-seq) analyses were performed as previously described in Chapter 2. Briefly, D25 hPSC-cardiomyocytes were enzymatically digested by incubating with collagenase type IV (Gibco) and Accutase (ThermoFisher) at 37^{o}C for 60 min. After removing the dissociation media, cells were resuspended in PBS supplemented with 5% FBS and sorted on a Sony sorter. Single-cell droplet libraries from this suspension were generated in the 10X Genomics Chromium controller according to the manufacturer's instructions in the Chromium Single Cell 3′ Reagent Kit v2 User Guide. Raw sequencing reads were processed using the Cell Ranger v3.0.1 pipeline from 10X Genomics. Cell quality control, filtering, clustering, and downstream analyses were performed utilizing the Seurat v3.1 R package. The integration of the hPSC-cardiomyocyte and in vivo datasets was performed as described in the integration of single-cell RNA-seq and MERFISH section.

3D Bioprinting

To bioprint multilayered organoids constructs, we first differentiated TNNT2:GFP hPSCs into D30+ cardiomyocytes that were either treated NRG1 or PBS, as described in the hPSC studies section. These hPSC-cardiomyocytes (90%+ efficiency by flow cytometry) were then enzymatically dissociated and resuspended at 100 million cardiomyocytes/ml. The method to bioprint multilayered organoid construct involved printing D30+ hPSC-cardiomyocytes that were treated with *NRG1* or PBS into a rectangle-shape that was 500 μm x 700 μm x 300 μm (H x W x L), followed by printing

an adjacent rectangular structure (500 μm x 700 μm x 150 μm) containing GelMA[71] mixed with either 1) 100 ng/ml *SEMA3C* (R&D Systems) 2) 100 ng/ml each *SEMA6A* (R&D Systems) and *SEMA6B* (R&D Systems), or 3) PBS (control). The cell-encapsulated layer was fabricated using 6.25% GelMA and 0.33% LAP with 15s exposure time, and the cells were added directly before printing. The adjacent layer containing the signaling factors was printed using 4% GelMA, 0.4% LAP with 15s exposure time. Using a methacrylated coverslip fixed to the controller stage, a 20 μl cell-material mixture was placed into the space between the coverslip and a polydimethylsiloxane (PDMS) film attached to a glass slide. This cell-material mixture was then exposed to UV light (365 nm, 88 mW/cm2) to fix the first pattern into a digital micromirror device (DMD) chip. After printing each layer, the construct was washed three times with warm PBS and aspirated dry. After printing, the multilayered organoid construct was washed in both PBS and media, and then stored in a cell culture incubator (37^{o}C, 5% CO_2). Media was refreshed every other day.

Migration distance measurement

The distance traveled by the printed hPSC-cardiomyocytes was measured using the Day 0 position as the starting point, and the distance migrated was calculated daily. Briefly, brightfield and green fluorescent confocal images were taken on a Leica SP5 of the GFP+ hPSC-cardiomyocytes, with the brightfield used to visualize the construct. The length of the construct divided into 10 blocks, and we used the top 10 farthest hPSC-cardiomyocyte positions migrated from the Day 0 position to each block and averaged those distances to calculate the migration distance.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Farah, Elie N.; Hu, Robert K.; Kern, Colin; Zhang, Qingquan; Lu, Ting-Yu; Zhang, Bo; Carlin, Daniel; Li, Bin; Blair, Andrew P.; Ren, Bing; Evans, Sylvia M.; Chen, Shaochen; Zhu, Quan; Chi, Neil C. “Cardiac Single Cell Ecology Reveals Interactive Cellular Construction and Remodeling of the Human Heart”. The dissertation author was the primary investigator and author of this paper.

REFERENCES

1. Meilhac, S. M. & Buckingham, M. E. The deployment of cell lineages that form the mammalian heart. *Nat. Rev. Cardiol.* **15**, 705–724 (2018).

2. Gelb, B. *et al.* The Congenital Heart Disease Genetic Network Study. *Circ. Res.* **112**, 698–706 (2013).

3. Litviňuková, M. *et al.* Cells of the adult human heart. *Nature* **588**, 466–472 (2020).

4. Tucker, N. R. *et al.* Transcriptional and Cellular Diversity of the Human Heart. *Circulation* **142**, 466–482 (2020).

5. Hocker, J. D. *et al.* Cardiac cell type-specific gene regulatory programs and disease risk association. *Sci. Adv.* **7**, (2021).

6. Asp, M. *et al.* A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**, 1647-1660.e19 (2019).

7. Cui, Y. *et al.* Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. *Cell Rep.* **26**, (2019).

8. Kääb, S. *et al.* Global gene expression in human myocardium - Oligonucleotide microarray analysis of regional diversity and transcriptional regulation in heart failure. *J. Mol. Med.* **82**, 308–316 (2004).

9. Liang, X. *et al.* HCN4 dynamically marks the first heart field and conduction system precursors. *Circ. Res.* **113**, 399–407 (2013).

10. van Eif, V. W. W. *et al.* Transcriptome analysis of mouse and human sinoatrial node cells reveals a conserved genetic program. *Dev.* **146**, (2019).

11. Hoogaars, W. M. H. *et al.* The transcriptional repressor Tbx3 delineates the developing central conduction system of the heart. *Cardiovasc. Res.* **62**, 489–499 (2004).

12. Marín-Sedeño, E., de Morentin, X. M., Pérez-Pomares, J. M., Gómez-Cabrero, D. & Ruiz-Villalba, A. Understanding the Adult Mammalian Heart at Single-Cell RNA-Seq Resolution. *Frontiers in Cell and Developmental Biology* **9**, 1117 (2021).

13. Tallquist, M. D. & Molkentin, J. D. Redefining the identity of cardiac fibroblasts. *Nature Reviews Cardiology* **14**, 484–491 (2017).

14. Tallquist, M. D. Cardiac fibroblasts: from origin to injury. *Curr. Opin. Physiol.* **1**, 75–79 (2018).

15. Davidson, S. *et al.* Fibroblasts as immune regulators in infection, inflammation and cancer. *Nature Reviews Immunology* **21**, 704–717 (2021).

16. Aird, W. C. Phenotypic heterogeneity of the endothelium: I. Structure, function, and mechanisms. *Circulation Research* **100**, 158–173 (2007).

17. Steffens, S., Nahrendorf, M. & Madonna, R. Immune cells in cardiac homeostasis and disease: emerging insights from novel technologies. *Eur. Heart J.* **43**, 1533–1541 (2022).

18. Sedmera, D., Pexieder, T., Vuillemin, M., Thompson, R. P. & Anderson, R. H. Developmental patterning of the myocardium. *Anat. Rec.* **258**, 319–337 (2000).

19. Sedmera, D., Pexieder, T., Hu, N. & Clark, E. B. Developmental changes in the myocardial architecture of the chick. *Anat. Rec.* **248**, 421–432 (1997).

20. Han, P. *et al.* Coordinating cardiomyocyte interactions to direct ventricular chamber morphogenesis. *Nature* **534**, 700–704 (2016).

21. del Monte-Nieto, G. *et al.* Control of cardiac jelly dynamics by NOTCH1 and NRG1 defines the building plan for trabeculation. *Nature* **557**, 439–471 (2018).

22. Wessels, A. & Sedmera, D. Developmental anatomy of the heart: A tale of mice and man. *Physiol. Genomics* **15**, 165–176 (2004).

23. Choquet, C., Kelly, R. G. & Miquerol, L. Defects in Trabecular Development Contribute to Left Ventricular Noncompaction. *Pediatr. Cardiol.* **40**, 1331–1338 (2019).

24. Tian, X. *et al.* Identification of a hybrid myocardial zone in the mammalian heart after birth. *Nat. Commun.* **8**, 87 (2017).

25. Choquet, C. *et al. Deletion of Nkx2-5 in trabecular myocardium reveals the developmental origins of pathological heterogeneity associated with ventricular non-compaction cardiomyopathy*. *PLoS Genetics* **14**, (2018).

26. Luxán, G. *et al.* Mutations in the NOTCH pathway regulator MIB1 cause left ventricular noncompaction cardiomyopathy. *Nat. Med.* **19**, 193–201 (2013).

27. Epstein, J. A., Aghajanian, H. & Singh, M. K. Semaphorin signaling in cardiovascular development. *Cell Metab.* **21**, 163–173 (2015).

28. Sandireddy, R. *et al.* Semaphorin 3E/PlexinD1 signaling is required for cardiac ventricular compaction. *JCI Insight* **4**, (2019).

29. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science (80-.).* **370**,

(2020).

30. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

31. Yuen, M. *et al.* Leiomodin-dysfunction results in thin filament disorganization and nemaline myopathy. *J. Clin. Invest.* **124**, 4693–4708 (2014).

32. Whittington, H. J. *et al.* Over-expression of mitochondrial creatine kinase in the murine heart improves functional recovery and protects against injury following ischaemia–reperfusion. *Cardiovasc. Res.* **114**, 858–869 (2018).

33. van Weerd, J. H. & Christoffels, V. M. The formation and function of the cardiac conduction system. *Development* **143**, 197–210 (2016).

34. Li, G. *et al.* Transcriptomic Profiling Maps Anatomically Patterned Subpopulations among Single Embryonic Cardiac Cells. *Dev. Cell* **39**, 491–507 (2016).

35. Snarr, B. S., Wirrig, E. E., Phelps, A. L., Trusk, T. C. & Wessels, A. A spatiotemporal evaluation of the contribution of the dorsal mesenchymal protrusion to cardiac development. *Dev. Dyn.* **236**, 1287–1294 (2007).

36. Xie, L. *et al.* Tbx5-Hedgehog Molecular Networks Are Essential in the Second Heart Field for Atrial Septation. *Dev. Cell* **23**, 280–291 (2012).

37. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science (80-.).* **370**, (2020).

38. Aevermann, B. *et al.* A machine learning method for the discovery of minimum marker gene combinations for cell type identification from single-cell RNA sequencing. *Genome Res.* **31**, 1767–1780 (2021).

39. Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science (80-.).* **362**, (2018).

40. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).

41. Wu, T. *et al.* PRDM16 Is a Compact Myocardium-Enriched Transcription Factor Required to Maintain Compact Myocardial Cardiomyocyte Identity in Left Ventricle. *Circulation* **145**, 586–602 (2022).

42. Koibuchi, N. & Chin, M. T. CHF1/Hey2 plays a pivotal role in left ventricular maturation through suppression of ectopic atrial gene expression. *Circ. Res.* **100**, 850–855 (2007).

43. Christoffels, V. M., Keijser, A. G. M., Houweling, A. C., Clout, D. E. W. & Moorman, A. F. M. Patterning the embryonic heart: Identification of five mouse Iroquois homeobox genes in the developing heart. *Dev. Biol.* **224**, 263–274 (2000).

44. Kim, K. H., Rosen, A., Bruneau, B. G., Hui, C. C. & Backx, P. H. Iroquois homeodomain transcription factors in heart development and function. *Circulation Research* **110**, 1513–1524 (2012).

45. Anderson, R. H., Mori, S., Spicer, D. E., Sanchez-Quintana, D. & Jensen, B. The Anatomy, Development, and Evolution of the Atrioventricular Conduction Axis. doi:10.3390/jcdd5030044

46. Itoh, A. *et al.* Active stiffening of mitral valve leaflets in the beating heart. *Am. J. Physiol. - Hear. Circ. Physiol.* **296**, H1766 (2009).

47. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

48. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).

49. Oyler-Yaniv, A. *et al.* A Tunable Diffusion-Consumption Mechanism of Cytokine Propagation Enables Plasticity in Cell-to-Cell Communication in the Immune System. *Immunity* **46**, 609–620 (2017).

50. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329-337.e4 (2019).

51. Huang, H. *et al.* CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat. Genet.* **53**, 1064–1074 (2021).

52. Moffitt, J. R. *et al.* High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11046–11051 (2016).

53. Zhang, M. *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021).

54. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

55. D'Amato, G., Luxán, G. & de la Pompa, J. L. Notch signalling in ventricular chamber development and cardiomyopathy. *FEBS J.* **283**, 4223–4237 (2016).

56. Jongbloets, B. C. & Jeroen Pasterkamp, R. Semaphorin signalling during development. *Dev.* **141**, 3292–3297 (2014).

57. Alto, L. T. & Terman, J. R. Semaphorins and their signaling mechanisms. *Methods Mol. Biol.* **1493**, 1–25 (2017).

58. Towbin, J. A., Lorts, A. & Jefferies, J. L. Left ventricular non-compaction cardiomyopathy. in *The Lancet* **386**, 813–825 (Lancet Publishing Group, 2015).

59. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1–20 (2021).

60. Toyofuku, T. *et al.* Repulsive and attractive semaphorins cooperate to direct the navigation of cardiac neural crest cells. *Dev. Biol.* **321**, 251–262 (2008).

61. Bagnard, D., Thomasset, N., Lohrum, M., Püschel, A. W. & Bolz, J. Spatial distributions of guidance molecules regulate chemorepulsion and chemoattraction of growth cones. *J. Neurosci.* **20**, 1030–1035 (2000).

62. Suto, F. *et al.* Interactions between Plexin-A2, Plexin-A4, and Semaphorin 6A Control Lamina-Restricted Projection of Hippocampal Mossy Fibers. *Neuron* **53**, 535–547 (2007).

63. Haklai-Topper, L., Mlechkovich, G., Savariego, D., Gokhman, I. & Yaron, A. Cis interaction between Semaphorin6A and Plexin-A4 modulates the repulsive response to Sema6A. *EMBO J.* **29**, 2635–2645 (2010).

64. Renaud, J. *et al.* Plexin-A2 and its ligand, Sema6A, control nucleus-centrosome coupling in migrating granule cells. *Nat. Neurosci.* **11**, 440–449 (2008).

65. Lian, X. *et al.* Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1848-57 (2012).

66. Chirikian, O. *et al.* CRISPR/Cas9-based targeting of fluorescent reporters to human iPSCs to isolate atrial and ventricular-specific cardiomyocytes. *Sci. Rep.* **11**, 3026 (2021).

67. Scuderi, G. J. & Butcher, J. Naturally Engineered Maturation of Cardiomyocytes. *Front. Cell Dev. Biol.* **5**, 50 (2017).

68. Mikryukov, A. A. *et al.* BMP10 Signaling Promotes the Development of Endocardial Cells from Human Pluripotent Stem Cell-Derived Cardiovascular

Progenitors. *Cell Stem Cell* **28**, 96-111.e7 (2021).

69. Faber, J. W., D'Silva, A., Christoffels, V. M. & Jensen, B. Lack of morphometric evidence for ventricular compaction in humans. *Journal of Cardiology* **78**, 397–405 (2021).

70. Lian, X. *et al.* Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1848-57 (2012).

71. Ma, X. *et al.* Deterministically patterned biomimetic human iPSC-derived hepatic model via rapid 3D bioprinting. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2206–2211 (2016).