Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects

T. Yvanka de Soysa1,2,3, Sanjeev S. Ranade1,3, Satoshi Okawa4,5, Srikanth Ravichandran4, Yu Huang1,3, Hazel T. Salunga1,3, Amelia Schricker1,3, Antonio del Sol4,6,7, Casey A. Gifford1,3,8 & Deepak Srivastava1,3,8,9

Organogenesis involves integration of diverse cell types; dysregulation of cell-type-specific gene networks results in birth defects, which affect 5% of live births. Congenital heart defects are the most common malformations, and result from disruption of discrete subsets of cardiac progenitor cells1, but the transcriptional changes in individual progenitors that lead to organ-level defects remain unknown. Here we used single-cell RNA sequencing to interrogate early cardiac progenitor cells as they become specified during normal and abnormal cardiogenesis, revealing how dysregulation of specific cellular subpopulations has catastrophic consequences. A network-based computational method for single-cell RNA-sequencing analysis that predicts lineage-specifying transcription factors2,3 identified Hand2 as a specifier of outflow tract cells but not right ventricular cells, despite the failure of right ventricular formation in Hand2-null mice4. Temporal single-cell-transcriptome analysis of Hand2-null embryos revealed failure of outflow tract myocardium specification, whereas right ventricular myocardium was specified but failed to properly differentiate and migrate. Loss of Hand2 also led to dysregulation of retinoic acid signalling and disruption of anterior–posterior patterning of cardiac progenitors. This work reveals transcriptional determinants that specify fate and differentiation in individual cardiac progenitor cells, and exposes mechanisms of disrupted cardiac development at single-cell resolution, providing a framework for investigating congenital heart defects.

The heart develops from diverse cell lineages specified from two pools of mesodermally derived cardiac progenitor cells (CPCs), the first and second heart fields (FHF and SHF, respectively) and multipotent neural crest cells5. Although genetic analyses are revealing mutations that may contribute to congenital heart defects6, identification of specific cell types and the progenitors from which they are derived that are affected by such mutations has remained a challenge. To address this, we identified transcriptional features of cardiac cell specification and morphogenesis by sequencing over 36,000 individual cells collected from the cardiogenic region of mouse embryos at three developmental stages: (1) as CPCs begin to differentiate in the late cardiac crescent at embryonic day (E)7.75; (2) as the FHF forms a linear heart tube and the SHF migrates into the anterior and posterior poles of the tube at E8.25; and (3) as the heart tube loops and incorporates the SHF-derived outflow tract (OFT), right ventricle (RV), sinus venosus (SV) and atrial cells with the FHF-derived left ventricle (LV), atrial and atrioventricular canal (AVC) cells at E9.25 (Fig. 1a–c, Extended Data Fig. 1a–c, Supplementary Table 1). Among these, the transcriptomes of 21,366 mesoderm and neural crest cells were partitioned into seven broadly defined populations6; multipotent Isl1+ progenitors, endothelial or endocardial cells, epicardium, myocardium, neural crest-derived mesenchyme, paraxial mesoderm and lateral plate mesoderm (Fig. 1d, e, Supplementary Table 2). Several genes with de novo mutations identified through whole-exome sequencing of congenital-heart-defect probands and parents5 exhibited expression specific to or enriched in one of these populations (Extended Data Fig. 1d–f).

Within each broad population, further transcriptome analyses revealed distinct cell types characteristic of unique progenitor pools (Extended Data Fig. 2), which we validated and spatially resolved by in situ hybridization of specific marker genes (Extended Data Fig. 3). Three subpopulations of the endothelial or endocardial lineage emerged: haematoendothelial progenitors, specified endothelial or endocardial cells, and endocardial cells initiating an endothelial-to-mesenchymal transition program typical of valve development (Extended Data Fig. 4a, b). The Isl1+ multipotent progenitor population segregated into the anterior and posterior domains of the SHF (the anterior heart field (AHF) and posterior SHF (pSHF), respectively), and the branchiomyocline muscle progenitors that share a common origin with SHF cells in the cardiopharyngeal mesoderm (Extended Data Fig. 4c, d). Moreover, we identified transcriptomes representative of ventricular, atrial, SV, AVC and OFT myocardium (Extended Data Fig. 5a, b). This dataset therefore represents a catalogue of cardiac-cell states that arise during embryonic development, providing a foundation for studying the transcriptional dynamics that underlie specification of multiple cardiac subtypes and to uncover subtype-specific genes (Supplementary Table 2).

Although the LV and RV perform similar functions, they arise from distinct progenitors, and dysregulation of genes uniquely enriched in each population leads to chamber-specific congenital malformations6,7. Focusing on ventricular progenitors identified LV-fated cells, an early, less differentiated RV population, and a more differentiated pool of RV cells (Extended Data Fig. 5c, d). Several genes were enriched in only one chamber, consistent with the distinct origins and physiology of the ventricles (Supplementary Table 2). Notably, phospholamban (Pln), a critical regulator of calcium handling, and cholecystokinin (Cck), an intestinal hormone, were predominantly expressed in RV cells, which we confirmed with in situ hybridization (Extended Data Fig. 5d–g).

Lineage tracing of Cck-expressing cells and their progeny using a constitutive Cck-IRES-Cre transgenic mouse crossed with a floxed TdToma reporter mouse established their specific contribution to the RV, particularly the trabecular myocardium and interventricular septum postnatally (Extended Data Fig. 5f–h).

A myocardial population that we could not ascribe to a known cell type comprised a large proportion of E7.75 cells and expressed genes of multiple subtypes. We used pseudotemporal ordering to test the hypothesis that this cluster represents a mixed population of early myocardium progenitors (EMP)6 (Extended Data Fig. 6–c). This analysis placed the majority of EMP cells at the start of the pseudotime trajectory, whereas EMP cells from later time points were represented in all other myocardial cell states, consistent with our hypothesis. Additionally, the early RV cells segregated with the OFT state (state d),...
differentiated LV–RV state (state f) and an intermediate state (state e) that expressed lower levels of differentiated ventricle genes, such as Nppa, and higher levels of Fgf8, supporting their designation as early RV cells (Extended Data Fig. 6b, d, f, Supplementary Table 2).

The large number of early CPCs in our dataset compared with previous single-cell CPC analyses8–11, allowed us to characterize the heterogeneity of this compartment. AHF and pSHF populations captured at E7.75 and E8.25 segregated into nine subpopulations (Fig. 2a, b, c, Supplementary Table 2). Populations A–C were of pSHF origin, whereas E–I were of AHF origin. Clusters A and I were derived from E7.75 and the majority of cells in each population co-expressed the left–right asymmetric genes Nodal, Lefty2 and Pitx2, indicating differences among transcriptomes of left–right asymmetric cells of the AHF and pSHF (Supplementary Table 2). Population D had enriched expression of FHF, LV and sarcomeric genes with minimal Isil expression, features indicative of early FHF progenitors, whereas population F appears to represent the earliest OFT cells (Fig. 2b). Pseudotime analysis suggests that pSHF clusters A and B, AHF clusters E and F, and AHF clusters H and I, represent sequential stages of three differentiating populations (Fig. 2c–e). By contrast, pSHF cluster C encompasses a continuum of cells that are differentiating into myocardium (Fig. 2h, d).

As CPCs form the cardiac crescent, their cardiac gene expression is driven by fibroblast growth factor (FGF) and bone morphogenetic protein (BMP) signals secreted from the adjacent anterior endoderm12. Analysis of definitive endoderm cells captured at E7.75 revealed enriched expression of cardiac inducers Bmp2, Bmp4 or Fgf8 in specific subpopulations, suggesting CPCs receive combinatorial inductive signals from distinct endodermal domains (Extended Data Fig. 7, Supplementary Table 2). One endodermal cluster (cluster 5) expressed genes characteristic of mesendoderm, while another (cluster 1) may represent cardiogenic anterior visceral endoderm, revealing transcriptomes of these specific endodermal populations. Notably, the gene Wnt5a was enriched in multiple clusters that co-expressed Bmp4, Bmp2 and Fgf8 (Extended Data Fig. 7b). It is possible that the spatial positioning of these endodermal populations relative to the adjacent cardiac mesoderm determine myocardium subtype specification.

---

Fig. 1 Single-cell RNA sequencing reveals heterogeneity of cardiogenic regions during early embryonic development. a, Representative images of mouse embryos used for cell collection at E7.75, E8.25 and E9.25, with microdissected regions indicated, in frontal view (top) and right sagittal view (bottom). Scale bars, 200 μm. Single-cell experiments were repeated with n = 5 biologically independent embryos at E7.75 and n = 2 biologically independent embryos at E8.25 and E9.25; similar results were obtained for embryos collected at the same developmental stage. b, Spatial organization of captured cardiac cell populations at each stage: frontal view at E7.75 and E8.25; right sagittal view at E9.25. Darker shaded region on left side of SHF at E7.75 indicates left–right asymmetric stage. c, SHF, AHF and pSHF populations captured at E7.75 and the majority of cells in each population co-expressed the left–right asymmetric genes Nodal, Lefty2 and Pitx2, indicating differences among transcriptomes of left–right asymmetric cells of the AHF and pSHF. d, UMAP plot of all captured mesodermal and neural crest populations coloured by cluster identity (top) and embryonic stage of collection (bottom). Total cells, n = 21,366; E7.75, n = 4,326; E8.25, n = 5,664; E9.25, n = 11,376. e, Expression heat map of five marker genes of broadly defined populations. Statistics for differential gene-expression tests were applied to n = 21,366 cells. Data are shown for 100 cells subsampled from each population. Scale indicates z-scored expression values. All genes represented have a Bonferroni correction adjusted P < 1 × 10^-4 (two-sided Wilcoxon rank-sum test). HF, head folds; CC, cardiac crescent; HT, heart tube; PA, pharyngeal arches; NC, neural crest-derived mesenchyme cells; A, atria; PEO, proepicardial organ containing epicardial cells. MP, multipotent progenitors; EC, endocardium or endothelial cells; PM, paraxial mesoderm; LPM, lateral plate mesoderm.

---

---
The molecular regulators governing the progressive specification decisions of cardiac progenitors into distinct sublineages are largely unknown. We applied a Boolean network-based lineage-specifier prediction method\(^1\) to AHF, RV and OFT cells captured at E7.75 and E8.25, and identified \textit{Irx4} and \textit{Plagl1} as specifiers of RV myocardium cells, and \textit{Hand2}, \textit{Tead2} and \textit{Arid3b} as OFT myocardium cell-fate determinants. \textit{Irx4} is an established specifier of ventricular identity\(^1\) and \textit{Arid3b} is important for SHF progenitor deployment in the OFT\(^1\)\(^3\), confirming the validity of the analysis. \textit{Hand2} expression was enriched in the OFT myocardium (Extended Data Fig. 5b); however, the prediction that \textit{Hand2} is a lineage specifier for the OFT, but not RV myocardium, was not expected, as its global deletion causes severe RV hypoplasia\(^4\). This phenotype is recapitulated upon specific deletion of \textit{Hand2} in the SHF, underscoring its requirement in this progenitor compartment, although the mechanism has remained unknown\(^1\).\(^4\).

To resolve the discrepancy of the predicted lineage-specifying function of \textit{Hand2} in the OFT but not the RV and the morphologic loss-of-function consequence, we analysed single-cell transcriptomes of wild-type and \textit{Hand2}-null cardiac progenitors captured at E7.75 and E8.25 (Fig. 3a–c, Supplementary Table 3). AHF, OFT and RV precursors in \textit{Hand2}-null embryos were transcriptionally dysregulated as early as E7.75 (Fig. 3d–g, Extended Data Fig. 8a–c), well before any morphologic defect. For instance, the \textit{Rgs5} gene, which we identified as a marker of AHF and OFT, was downregulated in \textit{Hand2}-null AHF and OFT cells at E7.75 (Fig. 3d, h, Extended Data Fig. 8b, d). The gene encoding the chromatin remodeler \textit{Smyd1} was also downregulated in \textit{Hand2}-null AHF cells at E8.25, consistent with the observation that \textit{Smyd1} loss phenotypically mimics \textit{Hand2} loss\(^5\) (Extended Data Fig. 8e).

SHF cells undergo a binary decision to adopt an anterior or posterior lineage identity\(^6\); we investigated whether this decision was disrupted on loss of \textit{Hand2}. \textit{Crabp1} and \textit{Crabp2} are opposing regulators of retinoic acid signalling, which is involved in posteriorization of SHF progenitors, resulting in an atrial-like fate\(^1\)\(^9\). \textit{Crabp1}, normally highly expressed in the AHF, sequesters retinoic acid to facilitate its catabolism, whereas \textit{Crabp2}, which promotes retinoic acid nuclear transport and subsequent transcriptional activation\(^1\)\(^7\), is low in AHF cells. In \textit{Hand2}-null AHF cells, \textit{Crabp1} was downregulated, whereas \textit{Crabp2} was upregulated; this would be expected to cause ectopic retinoic acid signalling and AHF posteriorization (Fig. 3d). \textit{Hand2}-null embryos expressing a \textit{LacZ} reporter driven by a retinoic acid response element (RARE–\textit{LacZ}) revealed ectopic \textit{LacZ} expression in the AHF progenitors extending into the cardiac OFT region (Fig. 3i, Extended Data Fig. 8d). Consistent with this, \textit{Hoxa1} and \textit{Hoxb1}, established transcriptional targets of retinoic acid\(^1\)\(^8\) and pSHF markers\(^1\)\(^6\), were upregulated in \textit{Hand2}-null AHF cells at E9.25 (Fig. 3g, Extended Data Fig. 8d, f). \textit{Upp1}, a gene typically expressed in pSHF derivatives, was also ectopically expressed in \textit{Hand2}-null OFT and RV cells (Fig. 3e, f, j, Extended Data Fig. 8d). Notably, the proportions of AHF and pSHF cells captured at E7.75 were higher in \textit{Hand2}-null mutants than in wild-type embryos, probably
In addition to this posteriorized gene signature, the AHF and its derivatives displayed broader dysregulation. In comparison, wild-type and Hand2-null embryos showed disrupted specification, whereas RV-fated cells were appropriately specified but had differentiation defects, consistent with the lineage-specifier analysis. In agreement, cardiac transcriptomes from E9.25 indicated the presence of an RV population that comprised Hand2-null cells in comparable numbers to wild-type cells (Extended Data Fig. 9b–e), despite the absence of the RV chamber in Hand2-null mutants. Furthermore, Hand2-null AHF cells at E9.25 failed to upregulate differentiation genes and continued to express progenitor markers consistently with a differentiation defect (Extended Data Fig. 9f).

In situ hybridization demonstrated that ifx4Cck RV cells were present in Hand2-null embryos but were located in the AHF area behind the LV at E8.25 and in the area of the OFT at E9.25 (Fig. 4e, Extended Data Fig. 10a, b). Moreover, in Hand2-null embryos at E8.5, Sema3c-positive RV cells accumulated in the pharyngeal mesoderm.
In this study, we used single-cell RNA sequencing to reveal mechanisms of normal organogenesis and how regulatory defects in discrete cell subsets can lead to morphologic developmental defects. The ability to acquire quantitative and spatial resolution of large numbers of individual transcriptomes enabled us to dissect functional deficits such as improper differentiation and migration from disrupted fate specification in distinct cell types in response to loss of a single transcription factor, Hand2. Single-cell transcriptomics therefore offers a powerful strategy to more effectively determine the precise mechanisms and cell types that underlie phenotypic presentation of developmental defects associated with genetic variation. Precise dissection of transcriptional landscapes during organogenesis is a prerequisite for defining preventative approaches for birth defects and potential postnatal intervention for ongoing sequelae of human malformations.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1144-x.

Received: 9 July 2018; Accepted: 19 June 2019; Published online 24 July 2019.

1. Srivastava, D. Making or breaking the heart: from lineage determination to morphogenesis. Cell 126, 1037–1048 (2006).
2. Okawa, S., Nicklas, S., Zickenrott, S., Schwamborn, J. C. & del Sol, A. A generalized gene-regulatory network model of stem cell differentiation for predicting lineage specifiers. Stem Cell Reports 7, 307–315 (2016).
3. Okawa, S. & del Sol, A. A computational strategy for predicting lineage specifiers in stem cell subpopulations. Stem Cell Res. 15, 427–434 (2015).
4. Srivastava, D. et al. Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. Nat. Genet. 16, 154–160 (1997).
5. Jin, S. C. et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. Nat. Genet. 49, 1593–1601 (2017).
6. 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).
7. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods 14, 979–982 (2017).
8. Lesca, R. et al. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. Science 359, 1177–1181 (2018).
9. DeLaughter, D. M. et al. Single-cell resolution of temporal gene expression during heart development. Dev. Cell 39, 480–490 (2016).
10. Li, G. et al. Transcriptomic profiling maps anatomically patterned subpopulations among single embryonic cardiac cells. Dev. Cell 39, 491–507 (2016).
11. Jia, G. et al. Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. Nat. Commun. 9, 4877 (2018).
12. Nascone, N. & Mercola, M. An inductive role for the endoderm in Xenopus cardiogenesis. Development 121, 515–523 (1995).
13. Uribe, V. et al. Arid3b is essential for second heart field cell deployment and heart patterning. Development 141, 4168–4181 (2014).
14. Tsuchihashi, T. et al. Hand2 function in second heart field progenitors is essential for cardiogenesis. Dev. Biol. 351, 62–69 (2011).
15. Gottlieb, P. D. et al. Bop encodes a muscle-restricting protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. Nat. Genet. 31, 25–32 (2002).
16. Xaver-Neto, J. et al. A retinoic acid-inducible transgenic marker of sino-atrial development in the mouse heart. Development 126, 2677–2687 (1999).
17. Napoli, J. L. Cellular retinoid binding-proteins, CRBP, CRABP, FABP5: effects on retinoid metabolism, function and related diseases. Pharmacol. Ther. 173, 19–33 (2017).
18. Stefanovic, S. & Zaffran, S. Mechanisms of retinoic acid signaling during cardiogenesis. Mech. Dev. 143, 9–19 (2017).
19. Epstein, J. A., Aghajanian, H. & Singh, M. K. Semaphorin signaling in cardiovascular development. Cell Metab. 21, 163–173 (2015).
20. Sinha, T. et al. Loss of Wnt5a disrupts second heart field cell deployment and may contribute to OFT malformations in DiGeorge syndrome. Hum. Mol. Genet. 24, 1704–1716 (2015).
21. Dupays, L., Kotecha, S., Angst, B. & Mohun, T. J. Tbx2 misexpression impairs deployment of second heart field derived progenitor cells to the arterial pole of the embryonic heart. Dev. Biol. 333, 121–131 (2009).
22. George, R. M. & Firulli, A. B. Hand factors in cardiac development. Anat. Rec. 302, 101–107 (2019).
**METHODS**

Animal models. Animal studies were conducted in strict compliance with all relevant ethical regulations in the animal use protocols, UCSC animal use guidelines and the NIH Guide for the Care and Use of Laboratory Animals. All protocols concerning animal use were approved by the Institutional Animal Care and Use Committee (IACUC) at UCSC and were accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). Transcripts were captured from wild-type and Hand2-null embryos from intercrossed C57BL/6 mice heterozygous for the Hand2-null allele. The sexes of all embryos used for capture of single-cell transcriptomes are listed in Supplementary Table 1a. Lineage tracing of Cck-expressing cells was performed using Cck-Ires-Cre (JAX stock no. 012706) and Ai4 (JAX stock no. 007914) mice (JAX stock no. 008477).

Timed matings between male (8–10 weeks of age) and female (6–8 weeks of age) mice were set up for which noon on the day of plug detection was considered E0.5. Pregnant females were identified by echocardiography performed at E6.5 and euthanized to harvest embryos at E7.75, E8.25 and E9.25 for single-cell RNA sequencing and at E7.75, E8.25, E8.5 and E9.25 for whole-mount and section-based in situ hybridization experiments. Transcriptomes from at least two embryos were collected per embryonic stage, per genotype. The sample sizes of embryos used for single-cell transcriptome analysis at each time point was chosen to obtain cell numbers comparable to estimated cell numbers in the cardiogenic region at each embryonic stage. Embryos were developmentally matched at each time point by somite count (4, 8, and 21 somites for E7.75, E8.25 and E9.25, respectively.)

Randomization was not implemented—experiments were determined by genotype; that is, Hand2 wild-type embryos were compared to Hand2-null embryos. Covariates were not relevant to the analysis of the Hand2-null phenotype as the developmental defect is highly penetrant regardless of embryo sex, and the developmental stages analysed were before the onset of overt heart failure. Investigators were not blinded to allocation of embryos during experiments. Blinding was not possible for the wild-type and Hand2-null embryo comparisons owing to the need to match somite counts to control for developmental timing.

**Embryo dissection and single-cell library generation.** The entire cardiogenic region was dissected at each time point, including the SHF region that lies behind the cardiac crescent and heart tube, as well as the first and second pharyngeal arches at E9.25. Owing to the small size of embryos at these stages, some surrounding tissue (indicated in Fig. 1a) encompassing the posterior lateral plate mesoderm (E7.75), head folds (E7.75 and E8.25) and endoderm, was microdissected to ensure complete retrieval of cardiac populations. Embryos were dissected in cold PBS (Life Technologies, 141900250), de-yolked and placed in PBS with 1% FBS (ThermoFisher Scientific, 10439016) solution on ice until dissociation (approximately 3 h). Yolk-sac DNA was extracted (QuickExtract DNA Extraction Solution, 44438016) and used for genotyping to distinguish wild-type from Hand2-null embryos before further microdissection of cardiac regions at each stage. Dissected cardiac tissue was incubated in 200 µl TrypLE (ThermoFisher Scientific, 12563029) for 5 min, triturated with a 200-µl pipette tip, and incubated for an additional 5 min. The TrypLE solution was quenched with 600 µl PBS with 1% FBS. Cells were filtered through a 70-μm cell strainer (BD Falcon, 08-771-2), centrifuged at 150g for 3 min, and resuspended in 35 µl PBS with 1% FBS. Single-cell droplet libraries from this suspension were generated by the 10X Genomics Chromium controller according to the manufacturer’s instructions in the Chromium Single Cell 3′ Reagent Kit v2 User Guide. The cell capture efficiency of the Chromium controller is ~57%, thus, we loaded all cells dissected from embryos without pre-counting, to minimize cell loss and maximize the number of captured single cells. Additional components used for library preparation include the Chromium Single Cell 3′ Library and Gel Bead Kit v2 (PN-120237) and the Chromium Single Cell 3′ Chip kit v2 (PN-120236).

**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 17 Multiplex Kit (PN-120262). Final libraries were sequenced on the NextSeq 500 and HiSeq 4000. Somite-matched wild-type and Hand2-null replicate libraries from each litter were pooled and sequenced in the same lane. Sequencing parameters were selected according to the Chromium Single Cell v2 specifications. All libraries were sequenced to a mean read depth of at least 50,000 total aligned reads per cell.

**Processing of raw sequencing reads.** Raw sequencing reads were processed using the Cell Ranger v.2.2.0 pipeline from 10X Genomics. In brief, reads were demultiplexed, aligned to the mouse mm10 genome and UMI counts were quantified per gene using a gene barcode matrix in the barcodeR package. In the barcodeR pipeline only analysis, wild-type–Hand2-null E7.75–E8.25 and wild-type–Hand2-null E9.25 analyses) were aggregated and normalized to the same sequencing depth, resulting in a combined gene-barcode matrix of all samples.

**Cell filtering and cell-type clustering analysis.** We sequenced the transcriptomes of 36,777 cells captured from wild-type and 37,149 cells captured from Hand2-null embryos in total. Further filtering and clustering analyses of these cells were performed with the Seurat v.2.2 R package, as described in the tutorials (http://satijalab.org/seurat). For each aggregated dataset (wild-type only, wild-type–Hand2-null E7.75–E8.25, wild-type–Hand2-null E9.25), cells were normalized for genes expressed per cell and total expression, then multiplied by a scale factor of 10,000 and log-transformed. Cells that were of low quality or represented doubles were excluded from our analyses—this was achieved by filtering out cells with greater than 8,000 and fewer than 1,500 genes in Seurat. We then performed a linear regression on all genes to eliminate technical variability due to the number of genes detected, embryonic time point, embryo replicate and stage of the cell cycle (ScaleData function). For the Hand2-null analyses, we also regressed out the Hand2 gene to eliminate its contribution to cell clustering. Highly variable genes in the dataset were computed and used as input for principal component analysis. Significant principal components were used for downstream graph-based, semi-supervised clustering into distinct populations (FindClusters function) and uniform manifold approximation and projection (UMAP) dimensionality reduction was used to project these populations in two dimensions. For clustering, the resolution parameter, which indirectly controls the number of clusters, was approximated based on the number of cells according to Seurat guidelines; a vector of resolution parameters was passed to the FindClusters function and the optimal resolution that established discernible clusters with distinct marker gene expression was selected. One or two cell clusters would emerge that expressed marker genes representing multiple populations; these combined cells with low UMI and gene counts that escaped the first filtering step. These cells were removed from the analyses. To identify marker genes, the clusters were compared pairwise for differential gene expression using the Wilcoxon rank-sum test for single-cell gene expression (FindAllMarkers function, min.pct = 0.25, min.diff.pct = 0.1, return.thresh (p-value cut-off) = 1 × 10⁻⁴. To assign identities to these subpopulations, we cross-referenced their marker genes with known cardiac subtype markers and in situ hybridization data from the literature (Supplementary Table 2). We also validated several of these marker genes by fluorescence in situ hybridization (Extended Data Fig. 3). We removed blood, endoderm and ectoderm-derived clusters based on their expression of known blood markers such as the haemoglobin genes, endoderm markers such as Epcam and Fossa2, and ectoderm markers such as Pous3f1 and Sox2 (Extended Data Fig. 1c), retaining the cells of mesodermal or neural crest identity. The clustering approach was then repeated for these retained mesodermal and neural crest cells, beginning with the regression of technical variables, identification of highly variable genes, principal component analysis, graph-based clustering, UMAP projection and marker analysis. Similarly, all re-clustering analyses (Extended Data Fig. 2) were processed as described above. For computing differentially expressed genes across wild-type and Hand2-null embryos, all UMI ≥ 10 (normalized UMI < 0.1) cells in clusters A and I (Fig. 2a) and for the Hand2-null analyses, the FindMarkers function was used on wild-type and Hand2-null cells from each cluster with the following parameters: Wilcoxon rank-sum test, min. diff.pct = 0.1, min.pct = 0.25, logFC.threshold = 0.2. Additionally, the max.cells.per.ident argument was used to ensure that equivalent numbers of cells were considered in the analysis. For each population analysed, the number of cells (n) given to this argument was set to the population or genotype that had the lower cell number. An adjusted P value (Bonferroni correction) cut-off < 1 × 10⁻⁴ was used to identify differentially expressed genes.

**Prediction of cell-fate determinants.** Cell-fate determinants for OFF and RV myocardium from the AHF were predicted using a modified version of the method that we previously developed. This procedure was performed on E7.75 and E8.25 wild-type OFF, RV and AHF cells. In brief, 100 cells were randomly selected from the AHF and from the OFF and RV daughter populations and the normalized ratio difference (NRD) was computed for all combinations of these 100 cells, yielding 10,000 parent–daughter cell combinations. The NRD was calculated for all pairs of differentially expressed transcription factors between OFF, RV and AHF cells and averaged over the 10,000 cell combinations. We minimized the impact of zero inflation by relying on the highly stringent adjusted (Bonferroni) P value cut-off (P < 1 × 10⁻⁴) and by filtering out very lowly expressed genes in defining differentially expressed transcription factors (average expression ≥ 0.5 normalized UMI). We filtered out genes that were detected in less than 25% of cells in populations being compared, and only tested genes that show a minimum difference of 0.1 in the fraction of detection between the two groups. Transcription factor pairs for which the mean NRD was more than 0.05 in one lineage direction but less than 0.01 in the other lineage direction were selected. Finally, the transcription factor pairs that were present in the transcription factor component of the GRN were kept as the final candidate cell fate determinants.

**Cell trajectory analysis.** Pseudotime analyses were performed using the Monocle 2 package, as described in the tutorials (http://cole-trapnell-lab.github.io/...
monocle-release\textsuperscript{17}. Differentially expressed genes, as determined in Seurat using the FindAllMarkers function, between the myocardium, CPC and wild-type–Hand2-null AHF, OFT and RV cells were used as input for temporal ordering of these cells along the differentiation trajectory.

In situ hybridization experiments. Each in situ hybridization experiment was replicated at least twice for identifying spatial expression of genes and three times for quantification of in situ signal for differentially expressed genes in the Hand2-null analysis. For whole-mount experiments: de-yolked whole embryos were fixed in a 4% formaldehyde solution (ThermoFisher Scientific, 28006) overnight at 4°C followed by 2× PBST washes and 5-min incubations in a dehydration series of 25%, 50%, 75% and 100% methanol (Fisher Scientific, A454-1). At this point embryos were stored in 100% methanol at −20°C until the in situ protocol was initiated. Yolk sac DNA was used for genotyping. The whole-mount in situ assay was adapted from the protocol formulated for whole-mount zebrafish embryos\textsuperscript{27} using the RNAscope Multiplex Fluorescent Reagent Kit v.2 (Advanced Cell Diagnostics, 323100), with minor modifications (the air-drying step was excluded in our protocol). Protease Plus was used for embryo permeabilization, and the 0.2× SSCST wash step between reagent incubations was reduced to 3×8 min). Whole-mount embryos were imaged in 0.1% PBST using the Leica M165 fluorescence dissecting scope (FC/PLANAP0 1.0×10450028; camera DFC3000G; ET GFP 10447408; ET nCHER10450019; acquisition software LAS v.4.6). Quantification of transcript signal from whole-mount wild-type and Hand2-null embryos was performed using ImageJ\textsuperscript{1} v.1.51m9. The mean grey value and integrated density of a defined area, which was kept consistent between wild-type and mutant embryos, as well as the background fluorescence level for the same defined area per embryo was measured. The corrected total fluorescence for each gene was calculated using the following formula: Integrated density = (Area × Background fluorescence). Corrected total fluorescence values were then log<sub>10</sub>-transformed before t-tests were conducted to satisfy the prerequisite assumptions of normality.

For in situ hybridization experiments performed on embryonic and postnatal day 1 (P1) heart sections: embryos and P1 hearts were washed 3× PBS after overnight fixation in 4% formaldehyde and stored in 70% ethanol (VWR, 89125-08) for up to one month. For cardiac tissue samples, 506051; C130080G10Rik, 316791; Wnt5a, 481911-C2; Actc1, 441441-C3; Tbx1, 506411; Tbx5, 480301-C3; and Vangl2, 506241; Tbx5, 480301-C3; were used. Tissue slices were serially sectioned at 5-μm intervals, mounted on slides and stored at room temperature until initiation of the RNAscope protocol for paraffin embedded sections (User manual catalogue number 322452-USM). Sections were imaged with a Zeiss Axiol Observer Z1 inverted epifluorescence microscope (Carl Zeiss Microscopy) with Zeiss Axiocam MRm and Pco.edge cSMOS (PCO Imaging) monochrome cameras run by Zeiss Zen imaging software. For three-dimensional expression reconstruction, embryos were embedded in low melting agarose and imaged with a Zeiss lightsheet Z.1 selective plane illumination microscope (Carl Zeiss Microscopy) with 888-nm and 561-nm lasers, tandem Pco.edge cSMOS cameras (PCO Imaging) and Zeiss Zen imaging software. Three-dimensional reconstructions of multi-view images were performed using Bitplane Imaris software v.9.0.2 (Andor Technology).

Catalog numbers for RNAscope probes used in this study: Cxcr4, 402271-C3; Citd1, 432471; Hand1, 429651-C2; Irx4, 504831; Pln, 506241; Rgs5, 430181; Sema3c, 441441-C3; Tgfl1, 506411; Tbx5, 519581-C2 and Upp1, 504841-C2; Tbx1, 481911-C2; Trnt2, 480301-C3; Hoxd, 541861; Lzct, 313451; Tbx18, 515221-C2; Mash2l2, 456901; Wnt5a, 316791; Bmp4, 410301-C2; Smad5, 136245I06648, 505031; C130980G10Rik, 506051; Tbx5, 481861-C4; Actc1, 510361-C2; Ccna2d2, 449221-C2; Tbx5, 448991-C2; Sho2x, 554291-C3; TdTomato, 317041-C2. Statistics and reproducibility. Standard statistical analyses were performed using GraphPad Prism 8. The number of replicates, statistical test used and test result are described in the figure legends. Differentially expressed genes were determined by a Bonferroni correction adjusted $P < 1 \times 10^{-4}$ (two-sided Wilcoxon rank-sum test). The level of significance in all graphs is represented as follow: *P < 0.05 and **P < 0.01. For fluorescence quantification, corrected total fluorescence values were log-transformed before t-tests were conducted to satisfy the prerequisite assumptions of normality. For all quantitative analyses, the mean ± s.e.m. is reported. No experimental samples were excluded from the statistical analyses. Sample size was not pre-determined through power calculations, and no randomization or investigator blinding approaches were implemented during the experiments and data analyses. When representative results are presented to indicate expression patterns of genes in wild-type embryos, at least two independent embryos were analysed.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability
All source data, including sequencing reads and single-cell expression matrices have been deposited in NCBI’s Gene Expression Omnibus and are accessible through GEO series accession number GSE126128. Data underlying each figure are available as Source Data, in the Supplementary Information and on the UCSC cell browser at https://mouse-cardiac.cells.ucsc.edu. Users can use the cell browser to explore the data, view expression of genes of interest in each UMAP plot and download datasets for custom analysis.

Code availability
All analyses were performed using standard protocols with previously described R packages\textsuperscript{27}. The R scripts are available upon request.

Acknowledgements
We thank B. Bruneau and members of the Srivastava laboratory for helpful discussion and feedback and C. Trapnell for guidance on single-cell RNA-sequencing analysis. We acknowledge the Gladstone Histology and Light Microscopy Core, the Gladstone Genomics Core and the Gladstone Bioinformatics Core for their technical expertise and the Gladstone Animal Facility for support with mouse colony maintenance. We thank T. Marsh, J. Deb Nath, H. Yin, F. Chanut, B. Taylor, T. Roberts and Q. Maki for their assistance with imaging, literature review, editing and graphics. We thank M. Haessler and M. Speir for formatting and hosting the processed datasets on the UCSC cell browser. T.Y.d.S. was supported by the UCSF Chancellor’s Fellowship, Genentech Foundation Fellowship, Discovery Fellows Program and Phi Beta Kappa Graduate Scholarship. C.A.G. is a HHMI fellow of the Damon Runyon Cancer Research Foundation (DRG-2206-14). S.S.R. is a Winslow Fellow. D.S. is supported by the National Heart Lung and Blood Institute (R01 HL057181, FO1 HL098707, UM1 HL098179, UM1 HL128761), the California Institute for Regenerative Medicine (DSC2-00908), the Roddenberry Foundation, the L.K. Whittier Foundation, and the Younger Family Fund. S.O. is supported by an FNR CORE grant (C15/1BM/10397420) and S.R. is supported by the University of Luxembourg IRP Grant (R-AGR-3227-11). This work was also supported by NIH/NICRR grant C06 RR018928 to the Gladstone Institutes.

Author contributions
T.Y.d.S., C.A.G. and D.S. conceived the study, interpreted the data and wrote the manuscript. T.Y.d.S. prepared chromosome libraries, performed in situ hybridization experiments and analysed data with Seurat and Monocle. T.Y.d.S. and S.S.R. dissected and processed embryos for single-cell library preparation and in situ hybridization experiments and conducted whole-mount and section imaging. T.Y.d.S. and A.S. performed genotyping of mice. S.O. and S.R. performed, and A.d.S. conducted the computational modelling for cell-fate determinant predictions. Y.H. and H.T.S. identified pregnant female mice by echocardiography.

Competing interests
D.S. is a co-founder and member of the board of directors of Tenaya Therapeutics and has equity in Tenaya Therapeutics. The other authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1414-x.

Correspondence and requests for materials should be addressed to C.A.G. or D.S.

Reprints and permissions information is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | Genes associated with congenital heart defects are enriched in specific cardiac populations. a, b, UMAP plot of all captured cell populations coloured by cluster (a) and embryonic stage of collection (b). c, UMAP feature plot showing expression of marker genes used to identify and remove endoderm (Epcam), ectoderm (Sox2) and blood (Hbb-y) cell populations. Statistics for differential gene expression tests were applied to \( n = 36,777 \) cells.
d, UMAP plot of all mesodermal and neural crest populations captured at E7.75, E8.25 and E9.25 coloured by cluster identity as in Fig. 1d. e, Expression of Flt4 and Upp1 in endocardium or endothelium population. f, Expression of Rrad, Ank3 and Prkaa2 in subpopulations of the myocardium. Statistics for differential gene expression tests were applied to \( n = 21,366 \) cells. All genes represented in c, e, f, have a Bonferroni correction adjusted \( P < 1 \times 10^{-4} \) (two-sided Wilcoxon rank-sum test).
Extended Data Fig. 2 | Focused analyses of cardiac populations. Schema of progressive subdivisions of broadly clustered cell populations from Fig. 1d that are discussed in the manuscript.
Extended Data Fig. 3 | Spatial validation of marker gene expression by in situ hybridization. a, Ventral view of Tdgf1 and Tnnt2 expression in the cardiac crescent (CC), right lateral views of Wnt5a and Bmp4 in the OFT, Mab21l2 and Shox2 in the SV, and Hoxa1 and Hoxb1 in the pSHF by in situ hybridization at the indicated embryonic stages, which informed assignment of population identities in Extended Data Figs. 4c, 5a. b, Expression of Tbx1 in the SHF and pharyngeal arches (PA) and of novel unannotated gene 3632451O06Rik in the OFT at E8.25 and E9.25. Scale bars indicate 200 μm unless otherwise noted. c, Expression of Rgs5 and Isl1 in the SHF, Hand2 in the SHF, OFT and RV, and Tbx18 in the proepicardial organ (PEO) of E9.25 embryos. d, In situ hybridization of mRNA expression of Isl1, Fgf8 and Hoxb1 at E8.25 and Nr2f2 at E9.25 in right lateral histologic sections. Scale bars, 50 μm. n = 2 independent embryos per gene for all panels.
Extended Data Fig. 4 | Heterogeneity in endocardium or endothelium and multipotent progenitor populations. 

**a**, UMAP plot of reclustered endocardium or endothelium population coloured by cluster and embryonic stage of collection. **b**, Violin plot of markers indicating distinct subpopulations of endocardium or endothelial cells. Summary statistics reported in violin plots: the centre white line represents median gene expression and the central black rectangle spans the first quartile to the third quartile of the data distribution. The whiskers above or below the box indicate values that are $1.5 \times$ the interquartile range above the third quartile or below the first quartile. Statistics for differential gene expression tests were applied to $n = 2,199$ cells. **c**, UMAP plot of reclustered multipotent progenitor populations coloured by cluster and embryonic stage of collection. **d**, Heat map showing curated list of marker genes that identify pSHF, AHF and branchiomeric muscle progenitors. Scale indicates $z$-scored expression values. Statistics for differential gene expression tests were applied to $n = 5,376$ cells. HE, haematoendothelial progenitors; EndMT, endothelial–mesenchymal transition cells. All genes represented in **b**, **d**, have a Bonferroni correction adjusted $P < 1 \times 10^{-4}$ (two-sided Wilcoxon rank-sum test).
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | Focused analyses of myocardial populations and spatial validation of RV markers. a, UMAP plot of reclustered ‘myocardium’ population coloured by cluster and embryonic stage of collection. b, Heat map of highly and uniquely expressed genes in myocardial subpopulations. Scale indicates $z$-scored expression values. Statistics for differential gene expression tests were applied to $n = 6,474$ cells. c, UMAP plot of reclustered ventricle populations coloured by cluster and embryonic stage of collection. d, Heat map showing curated list of genes that identify LV, RV and early RV cells. Scale indicates $z$-scored expression values. Statistics for differential gene expression tests were applied to $n = 1,976$ cells. All genes represented in b, d, have a Bonferroni correction adjusted $P < 1 \times 10^{-4}$ (two-sided Wilcoxon rank-sum test). e, mRNA expression of LV marker *Hand1* (green) and *Pln* (red) in frontal view of the E9.5 heart showing enrichment of *Pln* in the RV region by whole-mount in situ hybridization. $n = 2$ independent embryos per gene. Scale bar, 200 μm. f, Breeding scheme for lineage-tracing *Cck* expressing cells. g, mRNA expression of endogenous *Cck* and *TdTomato* driven by *Cck-cre* transgene at E9.25 in right oblique view of the heart. $n = 2$ independent embryos per gene. Scale bars, 200 μm. h, Expression of *TdTomato* in whole-mount and sectioned P1 heart from *Ai14 × Cck-cre* lineage-traced mice showing location of progeny of *Cck*-expressing cells. Left panels show bright-field view (top) or TdTomato (bottom) of whole-mount P1 heart; right panels show sections of TdTomato and DAPI (top) or TdTomato alone (bottom) in P1 heart section. $n = 2$ independent embryos. Scale bars, 100 μm. A, atria; V, ventricle.
Extended Data Fig. 6 | Pseudotemporal ordering of myocardium populations. a–d, Pseudotime trajectory of myocardium populations coloured by pseudotime value (a), cluster identity (b), embryonic stage of collection (c) and cell state (d). Pseudotime trajectory analysis was applied to \( n = 6,474 \) cells. e, Percentage of cells in each state that were captured at E7.75, E8.25 or E9.25. f, Violin plots showing expression of Nppa and Fgf8 in state e and state f from pseudotime trajectory in d. Statistics for differential gene expression tests were applied to \( n = 455 \) cells from each state. Bonferroni correction adjusted \( P < 1 \times 10^{-4} \) (two-sided Wilcoxon rank-sum test). Summary statistics reported in violin plots: the centre white line represents median gene expression and the central black rectangle spans the first quartile to the third quartile of the data distribution. The whiskers above or below the box indicate values that are \( 1.5 \times \) the interquartile range above the third quartile or below the first quartile.
Extended Data Fig. 7 | Endoderm populations adjacent to the cardiac crescent. a, UMAP plot of endoderm populations captured at E7.75 coloured by cluster. b, Dot plot highlighting expression patterns of known and novel endodermal secreted factors, Fgf8, Bmp4, Bmp2 and Wnt5a. The size of the dot indicates the percentage of cells expressing that gene within a cluster (% exp) and the colour indicates the average expression level of that gene within a cluster. c, Expression heat map of the top-ten marker genes of each endodermal population and secreted factors from b. Scale indicates z-scored expression values. Statistics for differential gene expression tests were applied to n = 915 cells. All genes represented have a Bonferroni correction adjusted P < 1 × 10⁻⁴ (two-sided Wilcoxon rank-sum test).
Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | Transcriptional perturbation in Hand2-null embryos. a, Heat map of marker genes of populations from Fig. 3a. Statistics for differential gene expression tests were applied to \( n = 13,185 \) cells. b, c, Heat map of differentially expressed genes between wild-type and Hand2-null OFT (b) and RV (c) cells captured at E7.75 and E8.25. Statistics were applied to \( n = 253 \) OFT cells per genotype at E7.75, \( n = 276 \) OFT cells per genotype at E8.25, \( n = 132 \) RV cells per genotype at E7.75, and \( n = 331 \) RV cells per genotype at E8.25. Scale indicates \( z \)-scored expression values. d, Quantification of fluorescence signal for indicated genes in f and Fig. 3h–j. \( n = 3 \) independent embryos per genotype. The mean ± s.e.m. is indicated. Two-tailed \( t \)-test. *\( P < 0.05 \) and **\( P < 0.01 \). e, Violin plots showing expression of Smyd1 and Sema3c in wild-type and Hand2-null E8.25 AHF cells. f, Expression of Hoxb1 in wild-type and Hand2-null embryos at E9.25 by whole-mount in situ hybridization in right lateral view. Arrowheads indicate expanded anterior Hoxb1 expression in Hand2-null embryos. Scale bars, 500 \( \mu \text{m} \). \( n = 3 \) independent experiments with similar results. g, Proportion of wild-type and Hand2-null cells from each population captured at E7.75. \( n = 5 \) wild-type embryos and \( n = 3 \) Hand2-null embryos. The mean ± s.e.m. is indicated. Two-tailed \( t \)-test. *\( P < 0.05 \) and **\( P < 0.01 \). h, Violin plots showing expression of Nppa and Nppb in E8.25 wild-type and Hand2-null RV cells. All genes in a–c, e, h have a Bonferroni correction adjusted \( P < 1 \times 10^{-4} \) (two-sided Wilcoxon rank-sum test). Violin plot summary statistics: centre white line represents median gene expression and central black rectangle spans the first to third quartile of the data distribution. The whiskers indicate values that are 1.5 \( \times \) the interquartile range above the third quartile or below the first quartile.
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | RV cells are present in Hand2-null embryos at E9.25. a, Violin plots showing expression of Nppa and Nppb in RV states 1 and 2 from Fig. 4c. Statistics for differential gene expression tests were applied to \( n = 251 \) cells from each state. b, c, UMAP plot of subset of cardiac populations captured at E9.25 coloured by cluster (b) and genotype (c). Curated list of highly and uniquely enriched genes in cardiac populations at E9.25. Scale indicates z-scored expression values. e, UMAP feature plot showing expression domains of Irx4, Cited1 and Cck, indicating presence of LV and RV at E9.25. Statistics for differential gene expression tests for d and e were applied to \( n = 5,211 \) cells. f, Violin plots of genes differentially expressed in wild-type versus Hand2-null AHF cells captured at E9.25. Isl1 is shown to indicate equivalent expression, and thus progenitor identity, in wild-type and Hand2-null cells. ns, not significant. All other genes represented in a, d–f, have a Bonferroni correction adjusted \( P < 1 \times 10^{-4} \) (two-sided Wilcoxon rank-sum test). Violin plot summary statistics: centre white line represents median gene expression and the central black rectangle spans the first quartile to the third quartile of the data distribution. The whiskers indicate values that are \( 1.5 \times \) the interquartile range above the third quartile or below the first quartile.
Extended Data Fig. 10 | RV cell migration is impaired in Hand2-null embryos. a, b, Whole-mount in situ hybridization for Irx4 and Cck in right lateral view (a) and transverse sections (b) at E9.25, indicating presence of RV cells in Hand2 mutants (arrowheads). n = 2 independent experiments with similar results. Scale bars, 200 μm. c, Quantification of Sema3c fluorescence signal in Fig. 4f. n = 3 replicate embryos per genotype. The mean ± s.e.m. indicated. Two-tailed t-test: *P < 0.05. d, e, Violin plots of Wnt5a and Tbx2 expression in wild-type and Hand2-null RV cells at E7.75 and E8.25, respectively (d), and Hand1 and Hand2 expression in LV and OFT cells at E9.25 (e). All genes in d and e have a Bonferroni correction adjusted P < 1 × 10^{-4} (two-sided Wilcoxon rank-sum test). Violin plot summary statistics: centre white line represents median gene expression and central black rectangle spans the first to third quartile of the data distribution. Whiskers indicate values that are 1.5× the interquartile range above the third quartile or below the first quartile. f, In situ hybridization for Hand1 in wild-type and Hand2-null embryos at E9.25 in frontal view. n = 3 independent experiments with similar results. g, Quantification of Hand1 fluorescent signal in the OFT. n = 3 replicated embryos per genotype. The mean ± s.e.m. is indicated. Two-tailed t-test. *P < 0.05. h, GO biological process terms of differentially expressed genes in wild-type and Hand2-null AHF (n = 406 cells per genotype), OFT (n = 362 cells per genotype) or RV (n = 227 cells per genotype) cells at E9.25, as determined with DAVID v.6.8. Significant functional enrichment was statistically determined using a modified Fisher’s exact test (EASE score) followed by Benjamini–Hochberg correction for multiple comparisons, with 0.01 as a P-value cut-off.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

- Transcriptomes were sequenced using the NextSeq 500 and HiSeq 4000 and corresponding Illumina commercial software. Whole-mount embryo images with fluorescence were acquired with LAS v4.6 acquisition software and section images/lightsheet images were acquired with Zeiss Zen imaging software.

Data analysis

- Cellranger (v2.2.0), R (3.5.1), Seurat v2, Monocle v2, ImageJ v1.51m9, GraphPad Prism 8, Bitplane Imaris software v.9.0.2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All source data, including sequencing reads and single-cell expression matrices have been deposited in NCBI’s Gene Expression Omnibus and are accessible through GEO series accession number GSE126128. Data underlying each figure are available in the Source Data Files, Supplementary Information and on the UCSC cell browser at https://mouse-cardiac.cells.ucsc.edu. Users can use the UCSC cell browser to explore the data, view expression of genes of interest in each UMAP plot and download datasets for custom analysis.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

**Sample size**  
For single cell RNAseq experiments, the cell capture efficiency of the Chromium technology is ~57%. Thus we loaded all cells dissected from embryos without pre-counting, to minimize cell loss and maximize the number of captured single cells. The sample sizes of embryos used for single-cell transcriptome analysis at each time point was chosen to obtain cell numbers comparable to estimated cell numbers in the cardiogenic region at each embryonic stage. For RNA in situ hybridization experiments, at least 2 embryos per probe, per genotype were tested. For WT vs Hand2-null comparisons, at least 3 independent embryos per genotype were tested.

**Data exclusions**  
All exclusion criteria were pre-established. We captured cells of the endoderm, ectoderm and mesoderm germ layers - but we focused primarily on mesodermal lineages, ectodermal neural crest cells and endoderm cells captured at E7.75 in this study. Of the analyzed cells, those that were of low quality or represented doublets were excluded from our analyses -this was achieved by filtering out cells with greater than 8000 and fewer than 1500 genes in Seurat. After clustering and UMAP projection, one or two cell clusters would emerge that expressed markers representing multiple populations; these were cells that had low transcript (nUMI) and gene (nGene) counts that escaped the first filtering step. These cells were also removed from the analyses.

**Replication**  
Cells from at least 2 somite-matched biological replicate embryos per genotype were captured at E7.75, E8.25 and E9.25. Expression and enrichment of marker genes in clusters derived from the single cell RNA-seq data were validated and reproducible by in situ hybridization experiments (at least n=2). For each in situ experiment comparing gene expression in wild type and Hand2-null embryos, at least 3 embryos per genotype (WT vs Hand2null) were assayed. All attempts at replication were successful.

**Randomization**  
Experimental groups were determined by genotype i.e., Hand2 wild type embryos were compared to Hand2-null embryos. Covariates were not relevant to the analysis of the Hand2-null phenotype as the developmental defect is highly penetrant regardless of embryo sex, and the developmental stages analyzed were prior to the onset of overt heart failure.

**Blinding**  
Investigators were not blinded to allocation of embryos during experiments. Blinding was not possible for the WT and Hand2-null embryo comparisons due to the need to match embryos with the same somite count to control for developmental timing.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

**Materials & experimental systems**

- n/a Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology
- Animals and other organisms
- Human research participants
- Clinical data

**Methods**

- n/a Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

**Laboratory animals**

Transcriptomes were captured from wildtype (WT) and Hand2-null embryos from intercrossed C57BL/6 mice heterozygous for the Hand2-null allele. The sexes and ages of all embryos used for capture of single-cell transcriptomes are listed in Supplementary Table 1a. Lineage tracing of Cck expressing cells was performed using Cck-ires-Cre mice (JAX stock #012706) and Ai14 mice (JAX stock #007914). Validation of ectopic RA signaling in the Hand2 mutant was done by crossing the mutant line to RARE- hsp68LacZ mice (JAX stock #008477). All animals used for timed matings were aged 6-8 weeks (female) or 8-10 weeks (male) of age.

**Wild animals**

The study did not involve wild animals
| Field-collected samples | The study did not involve samples collected from the field |
|-------------------------|---------------------------------------------------------|
| Ethics oversight        | All protocols concerning animal use were approved by the Institutional Animal Care and Use Committee (IACUC) at UCSF and were accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). |

Note that full information on the approval of the study protocol must also be provided in the manuscript.