CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to Lionel Christiaen (lc121@nyu.edu) and Rahul Satija (rsatija@nygenome.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

All animal care and experiments were carried out in accord with current NIH guidelines. *Ciona robusta* adults were purchased from M-Rep (San Diego, CA), maintained in artificial seawater with constant illumination and used for experiment within one week after arrival. CD1 mice were crossed to generate embryos that were staged using the day of the copulation plug as embryonic day (E) 0.5.

METHOD DETAILS

Isolation of gametes, fertilization, dechorionation, electroporation, and development

Ciona eggs and sperm were collected from at least two individual adult animals and kept separately in filtered artificial seawater (FASW) until fertilization. Eggs were mixed with activated sperm and incubated in FASW at room temperature for 5 min. Chorion and surrounding follicle cells were chemically removed with a pronase solution (FASW, 7.5mg/L sodium thioglycolate, 0.05% pronase, 0.042N NaOH), as described (Christiaen et al., 2009a). Fertilized and dechorionated eggs were electroporated as described (Christiaen et al., 2009b), and cultured in FASW in agarose coated plastic Petri dishes at 18°C. The amount of fluorescent reporter DNA (*Mesp>nls::lacZ*, *Mesp>hCD4::mCherry*, *Mesp>tagRFP*, *MyoD905>EGFP* and *Hand-r>tagBFP*) for electroporation was typically 50 μg, except for *3XT12>H2B::mCherry* is 30 µg.

Dissociation and FACS

Sample dissociation and FACS were performed essentially as described (Christiaen et al., 2009c; Razy-Krajka et al., 2014). Embryos and larvae were harvested at 12, 14, 16, 18 and 20hpf in 5ml borosilicate glass tubes (Fisher Scientific, Waltham, MA. Cat.No. 14-961-26) and washed with 2ml calcium- and magnesium-free artificial seawater (CMF-ASW: 449 mM NaCl, 33 mM Na_2SO4, 9 mM KCl, 2.15 mM $NaHCO_3$, 10 mM Tris-Cl pH 8.2, 2.5 mM EGTA). Embryos and larvae were dissociated in 2ml 0.2% trypsin (w/v, Sigma, T- 4799) ASW by pipetting with glass Pasteur pipettes. The dissociation was stopped by adding 2ml filtered ice cold 0.05% BSA CMF-ASW. The dissociated cells were passed through 40µm cell-strainer and collected in 5ml polystyrene round-bottom tube (Corning Life Sciences, Oneonta, New York. REF 352235). Cells were collected by centrifugation at 800g for 3 min at 4°C, followed by two rounds of washing with ice cold 0.05% BSA CMF-ASW. Cell suspensions were filtered again through a 40µm cell-strainer and stored on ice. Following dissociation, cell suspensions were used for sorting within 1 hour.

The B7.5 lineage cells were labeled by *Mesp>tagRFP* reporter. Contaminating B-line mesenchyme cells were counter-selected using *MyoD905>EGFP* as described (Christiaen et al., 2008; Razy-Krajka et al., 2014). The TVC-specific *Hand-r>tagBFP* reporter was used in a 3-color FACS scheme for positive co-selection of TVC-derived cells, in order to minimize the effects of mosaicism. Dissociated cell were loaded in the BD FACS Aria™ cell sorter. 488 nm laser, FITC filter was used for EGFP; 407 nm laser, 561 nm laser, DsRed filter was used for tagRFP and Pacific BlueTM filter was used for tagBFP. The nozzle size was 100 μm. tagRFP +, tagBFP + and EGFP – cells were collected for downstream RNA sequencing analysis.

Fluorescent *In Situ* Hybridization-Immunohistochemistry (FISH-IHC) in *Ciona* Larvae

IFSH-IHC were performed essentially as described (Christiaen et al., 2009d; Razy-Krajka et al., 2014; Wang et al., 2013). Embryos were harvested and fixed at desired developmental stages for 2 h in 4% MEM-PFA and stored in 75% ethanol at −20°C. Antisense RNA probes were synthesized using either Gateway gene collections or amplified fragments of desired genes as templates (Table S2). *In vitro* antisense RNA synthesis was performed using T7 RNA Polymerase (Roche, Cat. No. 10881767001) and DIG RNA Labeling Mix (Roche, Cat. No. 11277073910). Anti-Digoxigenin-POD Fab fragment (Roche, IN) was first used to detect the hybridized probes, then the signal were revealed using the Tyramide Signal Amplification (TSA) with Fluorescein TSA Plus Evaluation Kits (Perkin Elmer, MA). Anti–β-galactosidase monoclonal mouse antibody (Promega) was co-incubated with anti-mCherry polyclonal rabbit antibody (Bio Vision, Cat. No. 5993-100) for immunodetection of *Mesp>nls::lacZ* and *Mesp>hCD4::mCherry* products respectively. Goat anti-mouse secondary antibodies coupled with AlexaFluor-555 and AlexaFluor-633 were used to detect β-galactosidase-bound mouse antibodies and mCherry-bound rabbit antibodies after the TSA reaction. FISH samples were mounted in ProLong® Gold Antifade Mountant (ThermoFisher Scientific, Waltham, MA. Catalog number P36930).

Multicolor Immunohistochemical Staining in Mouse Embryo

After dissection, embryos were fixed for 30 minutes (E8.5 and E9.5) to 1 hour (E7.5 in decidua) in 4% paraformaldehyde, dehydrated and embedded in paraffin prior to sectioning at 10µm. Immunofluorescence was performed using standard protocols. Briefly, after rehydration, sections were treated for 15 minutes with antigen unmasking solution (H-3300, Vector laboratories). Slides were washed twice in 1xPBS Tween (0.05%) and incubated for one hour in TNB buffer (0.1M Tris-HCl pH 7.5, 0.15M NaCl, 0.5% Blocking reagent (Roche 11096176001)). Sections were incubated with primary antibodies for 36 hours in TNB using the following dilutions: Dach1 (1/100, Proteintech 10914-1-AP), Nkx2-5 (1/100, Santa Cruz sc-8697), Islet1 (1/100, DSHB 39.4D5 and 40.2D6). After three 5 minutes washes in 1xPBS Tween (0.05%) sections were incubated with secondary antibodies for one hour using Alexa 488, 568 and 647 (1/500, Invitrogen). Sections were counterstained with Hoechst (Sigma 33258), mounted using Fluoromount (Southern Biotech 0100-001) and imaged using a Zeiss Axio Imager Z1 with an Apotome module.

CRISPR/Cas9-Mediated Gene Knock-Down

Two guide RNAs targeting the third and the fifth exon of *Dach* with Fusi Sore (http://crispor.tefor.net) as 63 (sgDach1: AAAAGATTAAGCATCGCCC) and 64 (sgDach2: GAGCATTGCCATTGACGTG) respectively were designed to knock-down *Dach* expression in TVC progeny (Gandhi et al., 2017). Two guide RNAs described by Tolkin et al (Tolkin and Christiaen, 2016). (sgTbx1.6 TGCGGCTTCGGCTCCGTGG; sgTbx1.8 AACGAAAGATTGGTGGCCG), were used to knock-down Tbx1/10 expression. The guide RNAs were subcloned into the expression cassette driven by U6 promoter (Stolfi et al., 2014). For each gene, two guide RNAs were used as combination with 25 μg of each expression plasmid. 30 μg of *Mesp>nls::Cas9::nls* plasmid were co-electroporated with guide RNA expression plasmids for B7.5 lineage-specific CRISPR/Cas9-mediated mutagenesis. Rescue of the *Dach* loss-of-function was achieved by TVC-specific overexpression of $Dach^{PAMmut}$ driven by a *FoxF* enhancer (Beh et al., 2007). The point mutagenesis (G303A and C852A) were introduced to the PAMs of sgDach1 and sgDach2 using optimized QuikChange™ Site-directed Mutagenesis protocol in lab. Two pairs of mutagenesis primers were designed using PrimerX website (http://www.bioinformatics.org/primerx/cgi-bin/DNA_1.cgi) as sgDACH_1_G303A_F: GATTAAGCATCGCCCCAGTCGTGTGCAACGTTG; sgDACH_1_G303A_R: CAACGTTGCACACGACTGGGGCGATGCTTAATC and sgDACH_2_C852A_F: CGTCGGGAATTCCACCCACGTCAATG; sgDACH_2_C852A_R: CATTGACGTGGGTGGAATTCCCGACG. The PCR mixture was prepared as 20 µl 5X HF buffer, 71 µl H2O, 3 µl 10 mM dNTP, 1.75 µl WT Dach plasmid at 15ng/µl, 2 µl 125 ng/µl top mutagenesis primer, 2 µl 125 ng/µl bottom mutagenesis primer, 1 µl Phusion® High-Fidelity DNA Polymerase (NEB M0530). PCR mixture were evenly distributed into 8 PCR tube-strip and PCR was performed with a denaturation at 95 °C for 4 min, followed by 18 cycles of (95 °C for 30 s, 50 to 72 °C (gradient) for 1 min, and 72 °C for 1 min) and a final extension at 72° for 5 min. The PCR products were pooled and 4 µl of DpnI was added directly to the tube, followed by the incubation at 37 °C for 2 hours. After the purification using QIAquick PCR Purification Kit (QIAGEN), the eluate are transformed into TOP10 cells. The successful mutagenesis was confirmed by sequencing.

Photoconversion and Lineage Tracing

Photoconversion and lineage tracing were performed as described (Razy-Krajka et al., 2014). Fertilized eggs were electroporated with 50 μg Mesp>nls:Kaede:nls to label the B7.5 lineage. Embryos were raised on agarose coated plastic Petri dishes in ASW at 18°C and transferred individually into Nunc™ MicroWell™ 96-Well Optical-Bottom plates (ThermoFisher Scientific, Waltham, MA. Supplier No. 164588) at 15hpf. Photoconversions were performed using the HC PL FLUOTAR 203/0.50 objective on Leica Microsystems inverted TCS SP8 X confocal microscope, by shedding 405 nm UV light on ROI continuously for 2 min. Stack scanning of whole TVC lineage were documented at 16, 22.5, 40, 48 and 65 hpf.

Confocal Microscopy

Images were acquired with an inverted Leica TCS SP8 X confocal microscope, using HC PL APO 63×/1.30 objective. Z-stacks were acquired with 1 µm z-steps. Maximum projections were processed with maximum projection tools from the LEICA software LAS-AF.

Image Processing and Quantification

Confocal Z-stacks were processed using Imaris x64 8.4.1 (BitPlane). A Region containing the TVC progeny was segmented first. For nuclei detection, the expected nucleus diameter was set at 2.5µm, Nucleus Threshold(Absolute Intensity) were calculated automatically by Imaris. The cell segmentation was carried out using "Detect Cell Boundary from Cell Membrane" function, the "Cell Smallest Diameter" was set as 5µm. The transcripts signal within the cell boundary was detected using "Vesicles Detection" function, the estimated diameter of vesicles was set as 1.44µm.

Bulk RNA-seq Library Preparation and Sequencing

200 to 800 cells were directly sorted in 100*µ*l lysis buffer included in RNAqueous®-Micro Total RNA Isolation Kit (Ambion). The total RNA extraction was performed following the manufacturer's instruction. The quality and quantity of total RNA was checked using Agilent RNA 6000 Pico Kit (Agilent) with Agilent 2100 Bioanalyzer. RNA samples with RNA Integrity Number (RIN)>8 were kept for downstream cDNA synthesis. 250-2000 pg of total RNA was loaded as template for cDNA synthesis using the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech) with template switching technology. RNA-Seq Libraries were prepared and barcoded using Ovation® Ultralow System V2 1–16 (NuGen). Up to 6 barcoded samples were pooled in one lane of the flow cell and sequenced by Illumina Hi-Seq 2500. One direction and 50bp length reads were obtained from all the bulk RNA-seq libraries.

Single Cell RNA-seq Library Preparation and Sequencing

Reverse transcription and cDNA amplification were carried out using modified Smart-seq2 protocol (Picelli et al., 2013). Single cells were sorted by FACS as described above into 96-well plates and collected in 3.4 µl RT buffer (0.5 µl 10µM 3' RT Primer (5' - AAG CAG TGG TAT CAA CGC AGA GTA C T30 VN - 3'), 0.5µl 10µM dNTP Mix, 0.5 µl 4U/µl RNase Inhibitor, 1 µl Maxima RT Buffer, 0,9 µl nuclease-free water) in each well. Plates were stored at -80°C or processed immediately. Plates were incubated at 72°C for 3min and chilled on the ice to denature the template RNA. 2µl RT reaction mixture (0.5µl 10um TSO primer (5′-AGACGTGTGCTCTTCCGATCTNNNNNrGrGrG-3′), 0.925µl 5M Betaine, 0.4 100mM MgCl2, 0.125µl 40U/ RNAase inhibitor, 0.05µl 200 U/µl Maxima H Minus Reverse Transcriptase) were added to each well. Reverse transcription was carried out by incubating the plate at 42°C for 90 min, followed by 10 cycles of (50 °C for 2 min, 42 °C for 2 min) and heat inactivation at 70° for 15 min. 7µl PCR amplification mixture (0.25µl 10 µM PCR primer (5′AGACGTGTGCTCTTCCGATCT-3′), 6.25µl KAPA HIFI ReadyMix, 0.5µl nuclease-free water) were added to each well. PCR amplification was carried out with a denaturation at 98 °C for 3 min, followed by 21 cycles of (98 °C for 15 s, 67 °C for 20 s, and 72 °C for 6 min) and a final extension at 72° for 5 min. PCR products were purified by adding 10µl (0.8×) Agencourt AMPureXP SPRI beads (Beckman-Coulter) to each well, followed by 5min incubation and two

rounds of wash using 100µl freshly prepared 70% Ethanol at room temperature. Purified cDNA were eluted in 20µl TE. The concentration of amplified cDNA was measured across the entire plate using Picogreen assay. The concentration of amplified cDNA was in a 0.5–2 ng/µl range. Fragment size distribution was checked for randomly selected wells with High-Sensitivity Bioanalyzer Chip (Agilent), the expected size average should be ~2 kb. For each sample, the amplified cDNA were normalized to a working concentration ranging from 0.1 to 0.2 ng/µl with TE buffer. 1.25 µl of diluted cDNA from each well were used for library preparation. Single cell libraries were prepared using the Nextera XT DNA Sample Kit (Illumina) according to manufacturer's instructions. After library amplification, 2.5µl from each well were pooled into a single 1.5-ml microcentrifuge tube, purified using Agencourt AMPure XP beads and eluted with 30µl TE buffer. 1µl purified library was used to measure the fragment size distribution using the Agilent HS DNA BioAnalyzer chip and another 1µl of the purified library was loaded into Qubit fluorometer to estimate library concentration according to the manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq 2500 sequencer to obtain paired-end 50bp reads.

QUANTIFICATION AND STATISTICAL ANALYSIS

Read alignment and generation of gene expression matrix

For each demultiplexed bulk and single cell RNA-seq library, sequencing reads were mapped to *Ciona* genome (Joined-scaffold (KH), http://ghost.zool.kyotou.ac.jp/datas/JoinedScaffold.zip) using TopHat 2.0.12 (Kim et al., 2013; Trapnell et al., 2012) with parameter: --no-coverage-search. Cufflinks 2.2.0 (Trapnell et al., 2012) was used to calculate the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) using KH gene models (ver.2013, http://ghost.zool.kyoto-u.ac.jp/datas/KH.KHGene.2013.gff3.zip) with default parameters. The FPKM values of all the single cell RNA-seq libraries were pooled to generate the gene expression matrix such that each row corresponds to a gene and each column corresponds to a single cell library. Specifically for the bulk RNA-seq data, we took the log-transformed FPKM values and calculated the log-fold-change by comparing the log2(FPKM) between the perturbed conditions and their corresponding lacZ controls.

Preprocessing and batch effect removal

We adopted multiple quality control criteria to filter out low quality single cell transcriptomes. First, we only retained single cells that had more than 2,000 and less than 6,000 expressed genes, and genes that were detected in more than 3 cells. 1,182 out of 1,796 single cells and 14,864 out of 15,287 genes were retained. We used total reads and overall read mapping rates from TopHat output files to assess the quality of scRNA-seq. Cells with mapping rates less than 30% and total reads more than 2-million were removed (see Supplementary Note). 1,138 out of 1,182 cells passed the quality controls and were retained for downstream analyses.

Batch effects were identified by principal component analysis (PCA) using all detected genes. Principal components 2, 5 and 7 were dominated either by ribosomal genes or unannotated genes that showed strong expressions only in certain batches (Supplementary Note). These PC's

were considered as batch effects created by sequencing and library preparation. For each gene *j*, its expression level y_j was fitted by a linear mixed model of the total sum of latent batch effects (x_i) and its real biological expression level (ε_j) as the formula $y_j = \sum_i a_i x_i + \varepsilon_j$, where a_i denotes coefficient of batch effect x_i . In our case, PCA rotation matrices of PC2, PC5 and PC7 served as batch effects and were regressed out by the above model (Supplementary Note).

Contaminating subpopulations were discovered upon clustering. 59 *Twist1*+ mesenchymal cells and 198 cells without previously identified lineage markers were detected in cluster 8, 9, 11 and 12 (Supplementary Note). These contaminating non-cardiopharyngeal lineage cells were removed before downstream analysis. 881 out of 1,138 single cells were retained for clustering and trajectory analysis.

Identification of variable genes and dimensional reduction analysis.
All scRNA-seq analyses were performed on each time point data individually. Downstream analysis followed the procedures of 'Seurat' R package v1.2 ((Satija et al., 2015); http://satijalab.org/seurat).

We first identified the set of genes that was most variable in 12, 14 and 20hpf single-cell data. We calculated the mean and dispersion (variance/mean) for each gene across all single cells, and placed genes into 20 bins based on their average expression. Within each bin, we then z-normalized the dispersion measure of all genes within the bin to identify genes whose expression values were highly variable even when compared to genes with similar average expression. We used a z-score cutoff of 2 for dispersion and average expression cutoff of 4 to identify highly variable genes. We then used those highly variable genes as input to the PCA to identify the primary data structures in 12, 14 and 20hpf data. For intermediate stages 16hpf and 18hpf, because of the known cell type similarity, we used cell type specific markers from 14hpf and 20hpf as input to PCA to obtain more robust dimensional reduction.

We extended the results of PCA analysis globally by projecting the PCA rotation matrix across the entire transcriptome. This additional projection allows us to identify other genes with strong PCA loadings that may not be included in our variable gene list. Statistically significant PCs were identified using a permutation test and independently confirmed using a modified resampling procedure ((Chung and Storey, 2015)) encoded in Seurat's 'jackStraw' function. Significant and biological meaningful PCs were retained for clustering and visualization. To visualize single cell data, we projected individual cells based on their PC scores onto a single two-dimensional map using t-distributed Stochastic Neighbor Embedding (t-SNE) (Maaten and Hinton, 2008).

Single cell clustering and differential gene expression
Clustering of single cells was performed using the weighted shared nearest neighbor (SNN) graph-based clustering method (Villani et al., 2017). To validate the legitimacy of clusters, we used 'ValidateClusters' function in Seurat, where we selected top 30 genes from significant PC's

as defined above and utilized them to build a linear kernel SVM. The predictive accuracy of the SVM was assessed by repeated 5-fold cross validation. The accuracy cutoffs of 0.8 and 0.85 were used, and the merging of clusters was done based on the minimal connectivity from the SNN graph with a threshold of 0.001. Subsequently, TVC, STVC, FHP, SHP and ASM cells were identified from each time point data based on both known and newly identified markers (Supplementary Table 2). Specifically, for 12hpf data, no significant PC was identified due to homogeneous TVC population. In 18hpf data, we also identified a cluster of 33 cells that expressed noticeable degree of both cardiac and ASM markers (Supplementary Note). We inferred this cluster of cells is possibly due to insufficient tissue-disassociation or sorting and sequencing errors. We removed these cells from further analysis.

To find markers differentially expressed among clusters, we used the same approach as in (Macosko et al., 2015). We used the binary classifier with ROC curve that was incorporated in Seurat's 'find.markers' function with parameters: test.use = 'roc', thresh.use = 1 and min.pct = 0.5, which selects genes that are expressed in more than 50% of single cells in the given cluster and with average expression larger than 1 logFPKM for differential expression analysis. The selected genes were ranked based on AUCs from 0 to 1. The higher the AUC or power value the more differentially expressed the gene is for the given cluster. An AUC of 0.5 has limited predictive power.

Single cell trajectory and transition state

We retrieved scRNA-seq data for each of the FHP, SHP, ASM trajectories by subsetting the master Seurat object containing all the single cell data. We adopted nonlinear dimensional reduction technique of diffusion map (Coifman and Lafon, 2006; Haghverdi et al., 2015; Moignard et al., 2015), which reduces dimensionality though a random walking process, to identify developmental trajectory. We used the markers (power>0.3) of all cell types in each trajectory to calculate a cell-to-cell pairwise Euclidean distance matrix and used this matrix as input to diffusion map ('diffuse' function from 'diffusionMap' package). We retained only the first two diffusion map components as developmental trajectory for pseudotime analysis. Every cell was assigned to a pseudotime coordinate by fitting a principal curve (Hastie and Stuetzle, 1989) to the first two diffusion map components. Pseudotime was determined by the unit-speed arc-length parameterization of each cell on principal curve and normalized to [0,1] range.

After identification of pseudotime, we selected genes that are dynamically expressed across the pseudotime using 'aic' function in the 'locfit' package. Genes expressed in more than 50% of single cells with mean expression level greater than 2 log(FPKM) were considered as expressed in each cell type. For every expressed gene, we built two local polynomial models: a null model with degree 0 that assumes the gene expression stays constant along pseudotime and an alternative model with degree 2 that assumes gene expression changes along pseudotime (Scialdone et al., 2016). We evaluated these two model using Akaike Information Criterion (AIC) to calculate the AIC score differences as AIC(degree=2) – AIC(degree = 0). Genes with AIC score differences lower than -5 were considered to favor the alternative model to be dynamically expressed in pseudotime space.

We then used these dynamic genes to subdivide the pseudotime space into distinct regulatory states separated by discrete transitions. First, we built a cell-to-cell cross-correlation matrix based on dynamic gene expressions for each trajectory. Constrained hierarchical clustering tree which maintains pseudotime ordering was built by CONISS algorithm (Grimm, 1987) using 'chclust' function from 'rioja' package based on the cross-correlation matrix. The tree was cut to split each trajectory into 4 or 5 clusters empirically as the transition states along pseudotime.

Primed and *de novo* gene expression

Before defining primed and *de novo* genes, we first unbiasedly clustered the temporal gene expression patterns of both pan cardiac and ASM markers across all three trajectories. With k=2 of 'kmeans' clustering, we identified two groups of gene expression patterns in both cardiac and ASM genes that mimicked the primed and *de novo* patterns. Then we performed more stringent selection to define primed and *de novo* cardiac/ASM marker genes. We first identified all pan-cardiac and ASM markers (power>0.5) using the 18hpf and 20hpf datasets. Then we defined primed genes as genes that were expressed in more than 50% of single cells in both the multipotent progenitors (12hpf TVCs) and the fate restricted cells (18/20hpf FHPs/SHPs/iASMs/oASMs). Similarly, we defined *de novo*-expressed genes as expressed in less than 25% of single cells in the multipotent progenitors (12hpf TVCs) but expressed in more than 50% of single cells in the fate restricted cells (18/20hpf FHPs/SHPs/iASMs/oASMs). The genes expressed between 25%-50% of single cells were classified as ambiguous. The progenitor genes were defined as genes that were expressed in more than 50% of single cells in 12hpf TVCs but less than 25% of single cells in any of 18/20hpf FHP/SHP/iASM/oASM clusters. FHP/SHP specific markers were defined as genes that only belonged to FHP/SHP group but not to pan-cardiac or ASM gene set.

To visualize the smoothed pseudotime expression pattern and predict the induction time of a given gene, we smoothed the expression profiles along the pseudotime axis using local polynomial fit ('loess' function) with degree of smoothing equals to 0.75. Gene induction time was predicted based on the smoothed pseudotime expression profile using a logistic regression model. Gene expressions were first normalized to [0,1] range. Normalized expression values that were smaller than 0.5 were considered in 'off' state and bigger or equal to 0.5 were considered in 'on' state. We used this binary state notation and pseudotime coordinates to train a logistic model ('glm' function with family=binomial(link='logit')) to predict the on/off state of a given gene. The induction time was determined as the closest pseudotime coordinate to 0.5. Genes were then subdivided into two groups, either turning on or turning off, and sorted by their induction pseudotime.

To quantify the relative contribution of multipotent progenitor, primed ASM/cardiac, *de novo* ASM/cardiac and FHP/SHP specific genes, we performed principal component analysis using these groups of genes defined using above criteria on FHP and SHP trajectories separately. We observed that principal component 1 (PC1) strongly correlated with the defined pseudotime (PCCs>0.9). This allowed us to used PC1 loadings of each gene to calculate the contribution of

each group as a scaled score using the formula: $G = {X_g}^T * PCA^{rot}_{[\bullet,1]}$. The X_g represents scaled expression matrix of group g where each column is a single cell and each row is a gene from group g. The $PCA^{rot}_{[\bullet,1]}$ represents the PC1 variable loadings which is the first column of the loading matrix. The G is the scaled score of gene group g. For heatmap visualization, we smoothed this score along the pseudotime axis using local polynomial fit ('loess' function) with degree of smoothing equals to 0.75.

Mouse single cell RNA-seq data

Mouse single cell count matrix and clustering results were retrieved from (Scialdone et al., 2016) http://gastrulation.stemcells.cam.ac.uk/scialdone2016. The matrix was then log transformed and variable genes were identified as previously described. We performed PCA analysis on the identified variable genes and PC1 to PC6 were used for two-dimension tSNE plot construction. The visualization was based on the clustering results of the original paper.

DATA AND SOFTWARE AVAILABILITY

The accession number for the expression data reported in this paper is GEO: GSE99846. The expression data and the code/Rmarkdown files for the analyses reported in this paper are available at https://github.com/stevexniu/single-cell-ciona.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit Anti-mCherry Polyclonal Antibody, Unconjugated	**BioVision, Inc**	**5993-100**
Monoclonal Antibody that Recognizes E. coli β-galactosidase	**Promega**	**Z3781**
Bacterial and Virus Strains		

Biological Samples		
Chemicals, Peptides, and Recombinant Proteins		
Falcon® 5mL Round Bottom Polystyrene Test Tube, with Cell Strainer Snap Cap	**Corning Life Science**	**352235**
Critical Commercial Assays		
Deposited Data		
Single Cell RNA-seq FASTQ files	**This Paper**	**GEOxxx**
Bulk RNA-seq FASTQ files	**This Paper**	**GEOxxx**
Gene Expression Matrices (Single Cell RNA-seq and Bulk RNA-seq)	**This Paper**	https://github.com/stevexniu/single-cell-ciona.

Microarray Gene Expression Matrix	(Razy-Krajka et al., 2014)	**GEO/GSE54746**
Mouse Single Cell Gene Expression Matrix	(Scialdone et al., 2016)	**http://gastrulation.stemcells.cam.ac.uk/scialdone2016**
Experimental Models: Cell Lines		
Experimental Models: Organisms/Strains		
Ciona robusta		
Oligonucleotides		
Recombinant DNA		

Software and Algorithms		
Tophat v.2.0.12	(Kim et al., 2013; Trapnell et al., 2012)	
Cufflinks v.2.2.0	(Trapnell et al., 2012)	
Seurat v1.2	(Satija et al., 2015)	
Other		

Methods Reference

Beh, J., Shi, W., Levine, M., Davidson, B., and Christiaen, L. (2007). FoxF is essential for FGF-induced migration of heart progenitor cells in the ascidian Ciona intestinalis. Development *134*, 3297–3305.

Christiaen, L., Davidson, B., Kawashima, T., Powell, W., Nolla, H., Vranizan, K., and Levine, M. (2008). The transcription/migration interface in heart precursors of Ciona intestinalis. Science *320*, 1349–1352.

Christiaen, L., Wagner, E., Shi, W., and Levine, M. (2009a). Isolation of sea squirt (Ciona) gametes, fertilization, dechorionation, and development. Cold Spring Harb. Protoc. *2009*,

db.prot5344.

Christiaen, L., Wagner, E., Shi, W., and Levine, M. (2009b). Electroporation of transgenic DNAs in the sea squirt Ciona. Cold Spring Harb. Protoc. *2009*, db.prot5345.

Christiaen, L., Wagner, E., Shi, W., and Levine, M. (2009c). Isolation of individual cells and tissues from electroporated sea squirt (Ciona) embryos by fluorescence-activated cell sorting (FACS). Cold Spring Harb. Protoc. *2009*, db.prot5349.

Christiaen, L., Wagner, E., Shi, W., and Levine, M. (2009d). Whole-mount in situ hybridization on sea squirt (Ciona intestinalis) embryos. Cold Spring Harb. Protoc. *2009*, db.prot5348.

Chung, N.C., and Storey, J.D. (2015). Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics *31*, 545–554.

Coifman, R.R., and Lafon, S. (2006). Diffusion maps. Appl. Comput. Harmon. Anal. *21*, 5–30.

Gandhi, S., Haeussler, M., Razy-Krajka, F., Christiaen, L., and Stolfi, A. (2017). Evaluation and rational design of guide RNAs for efficient CRISPR/Cas9-mediated mutagenesis in Ciona. Dev. Biol. *425*, 8–20.

Grimm, E.C. (1987). CONISS: a FORTRAN 77 program for stratigraphically constrained cluster analysis by the method of incremental sum of squares. Comput. Geosci. *13*, 13–35.

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

Hastie, T., and Stuetzle, W. (1989). Principal Curves. J. Am. Stat. Assoc. *84*, 502–516.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

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

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

Razy-Krajka, F., Lam, K., Wang, W., Stolfi, A., Joly, M., Bonneau, R., and Christiaen, L. (2014). Collier/OLF/EBF-dependent transcriptional dynamics control pharyngeal muscle specification

from primed cardiopharyngeal progenitors. Dev. Cell *29*, 263–276.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N.K., Macaulay, I.C., Marioni, J.C., and Göttgens, B. (2016). Resolving early mesoderm diversification through single-cell expression profiling. Nature *535*, 289–293.

Stolfi, A., Gandhi, S., Salek, F., and Christiaen, L. (2014). Tissue-specific genome editing in Ciona embryos by CRISPR/Cas9. Development *141*, 4115–4120.

Tolkin, T., and Christiaen, L. (2016). Rewiring of an ancestral Tbx1/10-Ebf-Mrf network for pharyngeal muscle specification in distinct embryonic lineages. Development *143*, 3852–3862.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*.

Wang, W., Razy-Krajka, F., Siu, E., Ketcham, A., and Christiaen, L. (2013). NK4 antagonizes Tbx1/10 to promote cardiac versus pharyngeal muscle fate in the ascidian second heart field. PLoS Biol. *11*, e1001725.