bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A single cell transcriptional roadmap for cardiopharyngeal fate diversification

Wei Wang[1,#], Xiang Niu[2,3,#], Estelle Jullian[4], Robert G. Kelly[4], Rahul Satija[2,3,*] and Lionel Christiaen[1,*]

[1] Center for Developmental Genetics, Department of Biology, New York University, New York, NY, USA

[2] Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY, USA

[3] New York Genome Center, New York, NY, USA

[4] Aix-Marseille University, CNRS UMR 7288, Developmental Biology Institute of Marseille, Campus De Luminy Case 907, 13288 Marseille Cedex 9, France

[#]: contributed equally to the work

*: corresponding authors: lc121@nyu.edu (L.C.), rsatija@nygenome.org (R.S.)

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Summary

Dynamic gene expression programs determine multipotent cell states and fate choices during development. Multipotent progenitors for cardiomyocytes and branchiomeric head muscles populate the pharyngeal mesoderm of vertebrate embryos, but the mechanisms underlying cardiopharyngeal multipotency and heart vs. head muscle fate choices remain elusive. The tunicate *Ciona* emerged as a simple chordate model to study cardiopharyngeal development with unprecedented spatio-temporal resolution. We analyzed the transcriptome of single cardiopharyngeal lineage cells isolated at successive time points encompassing the transitions from multipotent progenitors to distinct first and second heart, and pharyngeal muscle precursors. We reconstructed the three cardiopharyngeal developmental trajectories, and characterized gene expression dynamics and regulatory states underlying each fate choice. Experimental perturbations and bulk transcriptome analyses revealed that ongoing FGF/MAPK signaling maintains cardiopharyngeal multipotency and promotes the pharyngeal muscle fate, whereas signal termination permits the deployment of a full pan-cardiac program and heart fate specification. We identified the *Dach1/2* homolog as a novel evolutionarily conserved second-heart-field-specific factor and demonstrate, through lineage tracing and CRISPR/Cas9 perturbations, that it operates downstream of Tbx1/10 to actively suppress the first heart lineage program. This data indicates that the regulatory state of multipotent cardiopharyngeal progenitors determines the first vs. second heart lineage choice, and that Tbx1/10 acts as a *bona fide* regulator of cardiopharyngeal multipotency.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Multicellular animals are composed of dozens to hundreds of distinct cell types that execute specialized functions. This diversity is generated progressively as early pluripotent cells and germ-layer-specific multipotent progenitors divide, and their daughter cells adopt distinct fates. These successive multipotent and fate-restricted states are largely defined by comprehensive gene and protein expression profiles that determine cellular competence and activities. In multicellular settings, extrinsic inputs resulting from cell-cell signaling, adhesion and cell-matrix interactions largely influence intrinsic networks governing global expression profiles. There are long-standing and widely shared interests in understanding how inherited intrinsic determinants and spatio-temporally defined extrinsic inputs coordinately determine the gene expression dynamics that underlie cells' progression through multipotent and fate-restricted states in development .

Recently, advances in single cell technologies, notably single cell RNA sequencing (scRNA-seq) have revolutionized our ability to characterize global cell transcriptomes with unprecedented spatial and temporal resolution (Moris et al., 2016; Tanay and Regev, 2017)(Moris et al., 2016; Tanay and Regev, 2017; Trapnell, 2015). Whereas previous studies relied on microdissection, laser microcapture, or fluorescence activated cell sorting (FACS) to physically isolated spatially-defined cell populations, scRNA-seq permits *a posteriori in silico* isolation of target cell populations, as long as complex tissues can be dissociated into single cell suspensions. Most strikingly, scRNA-seq approaches offer unique opportunities to study transcriptome dynamics with high (pseudo)temporal resolution, providing single cells can be obtained from successive stages along developmental trajectories of interest. In the past few years, scRNA-seq-based analyses have been extensively used to characterized developmental transitions in the hematopoietic system (Bendall et al., 2014; Moignard et al., 2015), during muscle differentiation (Trapnell et al., 2014), in early mesoderm development in the mouse (Scialdone et al., 2016), and during cardiac organogenesis in the mouse, and *in vitro* cardiomyocyte differentiation from human embryonic stem cells (DeLaughter et al., 2016; Li et al., 2016). The latter two studies provided insights into the transition between the earliest multipotent progenitors and initial cardiomyocyte differentiation.

The mammalian four-chambered heart is a complex organ with diverse cell types emerging primarily from *Mesp1*+ mesodermal progenitors in the early embryo (Devine et al., 2014; Lescroart et al., 2014; Saga et al., 1999). Genetic lineage tracing and clonal analyses indicated that distinct compartments arise from separate progenitor pools that have been referred to as

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the first and second heart fields (Buckingham et al., 2005; Kelly et al., 2001; Meilhac et al., 2003, 2004; Mosimann et al., 2015; Nevis et al., 2013; Waldo et al., 2001; Zhou et al., 2011). Lineage and clonal analyses also supported previous studies proposing that derivatives of the second heart field (i.e. the right ventricle and outflow tract) share a common origin with branchiomeric head muscles, in the cardiopharyngeal mesoderm of the early embryo (Diogo et al., 2015a; Gopalakrishnan et al., 2015; Harel et al., 2012; Lescroart et al., 2010, 2014, 2015; Mandal et al., 2017; Nathan et al., 2008; Tirosh-Finkel et al., 2006; Tzahor and Evans, 2011). The cellular and molecular characteristics of multipotent cardiopharyngeal progenitors, and the mechanisms underlying the early heart vs. pharyngeal/branchiomeric muscle fate choices remain largely elusive, and studies are partially hindered by the complexity of early vertebrate embryos (Chan et al., 2016; Mandal et al., 2017). Single cell approaches thus offer an opportunity to uncover the molecular underpinnings of early cardiopharyngeal development.

ScRNA-seq-based analyses of developmental progression provide insights into transcriptome dynamics, although the latter remains convolved with unknown clonal relationships between individual cells in vertebrate systems, where exact cell lineages vary between embryos and cell fate decisions are probabilistic. Therefore, reconstructed trajectories are abstract representations of possible developmental transitions, but not the actual path followed by defined successions of clonally related progenitors. Among model organisms, the nematode *Caenorhabditis elegans* and ascidians of the genera *Ciona*, *Halocynthia*, *Phallusia* and *Molgula* have extensively annotated genomes and develop with fixed and well-documented embryonic lineages (Brozovic et al., 2016; Hirano and Nishida, 1997; Nishida, 1987; Stolfi et al., 2014a; Sulston et al., 1983), thus offering opportunities to reconstruct developmental trajectories using scRNA-seq datasets with unequivocal clonal relationships between individual cells. Single cell analyses comprehensively characterized very early embryonic stages (Tintori et al., 2016), and the L2 larval stage in *C. elegans* (Cao, Junyue, Jonathan S. Packer, Vijay Ramani, Darren A. Cusanovich, Chau Huynh, Riza Daza, Xiaojie Qiu, et al., 2017), but there are technical difficulties in isolated healthy single cells from embryos. By contrast, single cell suspensions can be obtained from dissociated *Ciona* embryos throughout development, lineage-specific cell populations can be isolated by FACS, and profiled using -omics methods (Christiaen et al., 2008, 2009; Racioppi et al., 2014; Razy-Krajka et al., 2014). Moreover, *Ciona* emerged as an innovative chordate model to study conserved aspects of cardiopharyngeal development with unprecedented spatio-temporal resolution, and a well-established and invariant lineage (Kaplan et al., 2015; Wang et al., 2013). In 110-cell early gastrula embryos, a single bilateral pair of

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

blastomeres activate *Mesp* expression and form the only multipotent heart progenitors (Davidson et al., 2005; Hirano and Nishida, 1997; Satou et al., 2004). Following two divisions, two of their grand-daughter cells are induced by FGF/MAPK signaling to activate the conserved cardiac determinants, *Nk4/Nkx2-5*, *Gata4/5/6* and *Hand* (Christiaen et al., 2008; Davidson et al., 2006). These cells, called trunk ventral cells (TVCs), migrate towards the ventral midline before dividing twice asymmetrically and along the mediolateral axis to form median first and second heart precursors, and lateral pharyngeal muscle precursors (or Atrial siphon muscle founder cells, ASMFs; Figure 1A; (Stolfi et al., 2010)). We previously established that the *Tbx1* homolog, *Tbx1/10*, is activated in the common multipotent progenitors for the second heart and the pharyngeal muscle lineages, which specifically activate the *Ebf* and the universal muscle determinant Myogenic regulatory factors (*Mrf*)(Tolkin and Christiaen, 2016; Wang et al., 2013). Previous analyses showed that multipotent cardiopharyngeal progenitors exhibit multilineage transcriptional priming, whereby fate-specific determinants are transiently co-expressed before regulatory cross-antagonisms contribute to segregating the heart and pharyngeal muscle programs to their corresponding fate-restricted precursors (Kaplan et al., 2015; Razy-Krajka et al., 2014; Wang et al., 2013). However, the genome-wide extent of multilineage transcriptional priming and most initial regulatory events underlying the heart vs. pharyngeal muscle fate choices remain elusive.

Here, we combine single cell transcriptomics, genetic and chemical perturbations, and *in vivo* lineage tracing to comprehensively profile cardiopharyngeal multipotency and the transitions to early fate specification in the three main branches of the cardiopharyngeal lineage: the first and second cardiac, and pharyngeal muscle lineages. We found that extensive multilineage priming contributes substantially to the cardiopharyngeal multipotent state, and ongoing FGF/MAPK signaling prevents further deployment of the heart-specific program in multipotent progenitors. Continuous FGF/MAPK signaling in successive generations of multipotent and early pharyngeal muscle progenitors maintains a multipotent state and eventually promotes *Ebf* activation and commitment to a pharyngeal muscle identity, while reinforcing the antagonism with the heart program (see also (Razy-Krajka et al., 2017)). In the heart lineages, termination of FGF/MAPK signaling permits further deployment of a pan-cardiac program, which is composed of primed and *de novo*-activated markers, and is shared between the first and second heart lineages. Following establishment of a heart-specific identity, first heart precursors (FHPs) activate FHP-specific markers and later produce most

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mhc2+ cardiomyocytes in juvenile beating hearts. By contrast, second heart precursors activate an SHP-specific program marked by the expression of a *Dachshund/Dach1/2* homolog, which is activated by Tbx1/10 upon cessation of MAPK activity, and limits the propensity of SHP derivatives to form *Mhc2+* cardiomyocytes. These results indicate that, although the first and second heart lineages share a common pan-cardiac program, they emerge from successive generations of multipotent progenitors whose *Tbx1/10* expression status imparts a different potential to distinct heart progenitors**.**

Results

Single cell transcriptome profiling of cardiopharyngeal progenitors

To characterize the transcription profiles underlying the transition from multipotent cardiopharyngeal progenitors to distinct fate-restricted precursors, we performed scRNA-seq using the SMART-Seq2 protocol on cardiopharyngeal-lineage cells FACS-purified from synchronously developing embryos and larvae dissociated at successive time points (Figure 1A; Methods). After removing rare populations of non-cardiopharyngeal-lineage cells (Supplementary Methods), we obtained a total of 848 high-quality single cell transcriptomes, with an average of 4,415 expressed genes, from 5 time points covering early cardiopharyngeal development (Figures 1A, S1A). We clustered single cell transcriptomes from each time point independently, in order to identify the cell of origin according to a well-established lineage diagram (Figures 1A-C, S1B). The dataset comprises 27 multipotent progenitors cells (TVCs) from 12hpf embryos; 275 multipotent second trunk ventral cells (STVCs) and fate-restricted first heart precursors (FHPs) from 14hpf embryos; 114 cells from 16hpf pre-hatching larvae, when the population already comprises early atrial siphon muscle founder cells (ASMFs) and second heart precursors (SHPs), and FHPs; 144 and 288 cells from 18 hpf and 20hpf, respectively (Figures 1A, B; S1).

We focused our first analysis on the samples isolated from 20hpf larvae, where cardiopharyngeal-lineage cells have committed to either an ASM or a heart fate, and explored whether our single cell transcriptomes could uncover molecular heterogeneity among fate-committed progenitors. We used an unsupervised graph-based clustering strategy, implemented in the R package Seurat (Satija et al., 2015), to separate distinct groups of cells corresponding to first and second heart precursors, and atrial siphon muscle (ASM) precursors, which we identified using known markers, including *Ebf*, *Gata4/5/6* and *Tbx1/10* (Stolfi et al., 2010; Wang et al., 2013), and visualized individual clusters using t-distributed stochastic

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

neighbor embedding (t-SNE) (Figure 1C). Among cells expressing *Ebf* at high levels, we identified two known subpopulations, the inner and outer ASM precursors (iASMPs and oASMPs) based on differential expression of such markers as *Wnt2/4/5* and *Mhc3* (Figure S1B; (Razy-Krajka et al., 2014). Thus, our scRNA-seq clusters unequivocally identify defined fate-restricted precursor cell types isolated from swimming larvae.

Next, we sought to characterize comprehensive lists of marker genes that distinguish the first and second heart precursors, and the ASM founder cells. We conducted differential gene expression analysis as in (Satija et al., 2015) to identify (1) ASM/pharyngeal muscle vs. pan-cardiac specific markers and (2) first vs. second heart precursor-specific markers. The former analysis stems from previous observations that the heart vs. ASM fate choice is phenotypically striking, and coincides with most known differential gene expression in 20hpf swimming larvae (Stolfi et al., 2010; Wang et al., 2013; Razy-Krajka et al., 2014). Among the 86 top predicted pan-cardiac genes, we identified established cardiac determinants and markers, including *Gata4/5/6*, *Nk4/Nkx2-5*, *Hand* and *Bmp2/4* (Figure 1D, Table S1; Wang et al., 2013; Razy-Krajka et al., 2014), as well as novel candidate genes with heart-specific expression, and validated 16 of our predicted cardiac genes by fluorescent *in situ* hybridization (FISH) on 20hpf swimming larvae (Figures 1E, S1D; Tables S1, S2).

While the pan-cardiac vs. pharyngeal muscle segregation was the most prominent source of heterogeneity in the 20hpf data (Figures 1D, S1D), we also observed a second segregation between the first and second heart precursor populations, and differential expression analysis identified FHP- and SHP-specific markers. In addition to the single known discriminating marker *Tbx1/10*, we identified 28 novel FHP-specific (e.g. *Mmp21*, *Bmp5/6/7*, and *Wnt10*A) and 5 SHP-specific candidate markers (e.g. *Dach*, *Matn1/3/4*), and verified cell-specific gene expression using FISH (Figures 1D-E, S1; Table S2). In summary, our analyses of single cell cardiopharyngeal transcriptomes obtained from 20hpf swimming larvae characterized the specific molecular programs activated in fate-committed progenitors, including both shared ('pan-cardiac') and FHP- vs. SHP-specific signatures of cardiac precursors.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Reconstruction of temporal developmental profiles and identification of discrete regulatory states

Having characterized transcription profiles for fate-restricted first and second heart, and pharyngeal muscle precursors, we sought to characterize the gene expression dynamics establishing these distinct profiles as cells transition through multipotent cardiopharyngeal states. Single cell RNA-seq provides unique opportunities to decipher transcriptome dynamics by ordering cells on pseudotemporal developmental projection (Bendall et al., 2014; Trapnell et al., 2014). We took advantage of the established TVC lineage tree (Figure 1B; (Kaplan et al., 2015) to aggregate three subsets of cells, using annotated subpopulations corresponding to ASM, FHP and SHP trajectories (Figure 1B, C; S1). We used curated cardiopharyngeal-lineage-specific scRNA-seq data for cells isolated from 12hpf, 14hpf, 16hpf, 18hpf and 20hpf embryos and larvae. This time series encompasses the whole spectrum of progenitor subtypes leading to the emergence of distinct cell types from common multipotent cardiopharyngeal progenitors (Figure 1A, B). We first used unbiased clustering, to annotate each subpopulation as per the established lineage (Figure S1C). Single cells isolated from 12hpf embryos formed a homogenous population expressing known TVC markers, including *Gata4/5/6*, *Nk4/Nkx2-5* and *Bmp2/4* (Christiaen et al., 2008; Razy-Krajka et al., 2014; Satou et al., 2004). The 275 cells isolated from 14hpf embryos comprised 64 *Tbx1/10*-positive STVCs, and 211 *Tbx1/10*-negative FHPs, which expressed heart-specific genes (Figure S1C). The 114 cells isolated from 16hpf pre-hatching larvae consisted of three cell types: 58 ASMFs that activated *Ebf*, 31 *Tbx1/10*+ SHPs and 25 *Tbx1/10*- FHPs. The 144 cells from 18hpf consisted of 47 ASMs, 46 SHPs and 51 FHPs. Therefore, single cells obtained from 12, 14, 16, 18 and 20hpf samples captured all known subtypes and intermediate states during early cardiopharyngeal specification.

We then used trajectory-specific groups of annotated cells and applied diffusion maps, a non-linear dimensional reduction technique using cell-type-specific markers (Coifman and Lafon, 2006), to arrange our single cell data into two dimensions. We mapped each single cardiopharyngeal progenitor onto their developmental 'pseudotime' by fitting principal curves (Hastie and Stuetzle, 1989) on this diffusion/distance matrix, and computationally projecting each cell onto the principal curve. We repeated this strategy for all lineages, creating three unidirectional trajectories representing FHP, SHP, and ASM lineage specification (Figure 2A). By analyzing the distribution of cells along pseudotime axes, we observed a "strong"

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

correspondence (average PCC = 0.889) between developmental pseudotime and time point of origin: across each trajectory, the intermediate states were correctly ordered with ASM, FHP and SHP appearing late on pseudotime axis, while multipotent progenitors of TVC (at 12hpf) or STVC (at 14hpf) were ordered earlier on pseudotime axes (Figure 2A, B; S2A). This analysis also showed that equivalent cells collected at the same time point mapped to slightly different developmental pseudotimes, illustrating a unique strength of scRNA-seq and its potential to provide unprecedented insights into the gene expression dynamics underlying fate specification.

We next asked whether our *in silico* trajectories captured known lineage-specific expression dynamics of cardiopharyngeal regulators, such as *Hand-r*, *Tbx1/10* and *Ebf* (Figures 2C, S2B). Previous studies using FISH on embryos and larvae fixed at successive time points indicated that *Hand-r* is highly expressed in the TVCs, selectively maintained in the STVCs and ASMFs, and progressively eliminated from FHPs and SHPs following asymmetric divisions (Razy-Krajka et al., 2014). Accordingly, and in spite of biological and technical variability, *Hand-r* expression appeared steadily maintained along the ASM trajectory, whereas its expression declines sharply in the first and second heart precursor trajectories, albeit with a delay in the latter due to its STVC-specific expression (Figure 2C; (Razy-Krajka et al., 2014)). Similarly, analyses of *Tbx1/10* and *Ebf* expression dynamics along the SHP and ASM developmental trajectories captured their STVC- and ASMF-specific upregulation, respectively, whereas these genes were never expressed in the FHP trajectory (Figure S2B; (Razy-Krajka et al., 2014; Wang et al., 2013). These and further analyses of known cardiac determinants, *Nk4/Nkx2-5*, *Gata4/5/6* and *Hand* (Figure S2B), indicate that our *in silico* trajectories faithfully captured the molecular dynamics underlying fate specification across all three lineages.

In an attempt to visualize how cell-type-specific global expression profiles are established dynamically, we clustered ASM and pan-cardiac, FHP- and SHP- markers based on their changing expression levels along each cardiopharyngeal trajectory and generated a heatmap plotting the relative gene expression values (Figure 2D shows smoothed pseudotemporal profiles; see Figure S2C for a heatmap of the raw values). At first sight, the heatmap illustrates both the diversity of gene expression dynamics leading to cell-type-specific expression, as well as emerging broad patterns. For example, both first and second heart precursors ultimately express pan-cardiac genes at high levels, but these tend to become upregulated earlier along the FHP trajectory than the SHP trajectory. This mirrors a pattern previously observed for the

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cardiac-specific waves of *Gata4/5/6* expression (Wang et al., 2013), and suggests that the pan-cardiac program is activated independently in the first and second heart precursors, following divisions of the multipotent cardiopharyngeal progenitors. This convergence suggests that similar regulatory mechanisms control pan-cardiac gene expression in the first and second cardiac lineages (see below).

In addition, we observed that some ASM-specific markers were downregulated along cardiac trajectories, whereas some cardiac markers were down-regulated along the ASM trajectory (Figure 2D). This is the expected dynamics for "primed" marker genes, which are already expressed in multipotent progenitors and become restricted to their corresponding fate-restricted precursors following cell divisions (Graf and Enver, 2009; Nimmo et al., 2015). We previously reported this phenomenon for the ASM markers, *Hand-r*, *Ets1/2*, and *Rhod/f* and the heart markers *Hand*, *Gata4/5/6* and *Bmp2/4* (Razy-Krajka et al., 2014). Multilineage transcriptional priming is a hallmark of cardiopharyngeal progenitors multipotency, we thus aimed to characterize it globally; to this aim, we determined whether cell-type-specific markers were already expressed in multipotent cardiopharyngeal progenitors, the TVCs in 12hpf embryos. We used conservative expression cutoffs to classify cell-type-specific marker genes as "primed" (expressed in >50% of multipotent progenitor cells), or expressed "*de novo*" in fate-restricted precursors (expressed in <25% of the multipotent cells). We classified genes that we could not confidently assign to any of the other two categories as "ambiguous/undetermined", and excluded them from subsequent analyses. Applying these criteria to the dataset obtained from 18 and 20hpf cardiopharyngeal-lineage cells, we unequivocally classified 76% (3,982/5,208), 78% (87/111) and 88% (151/176) of the total transcriptome, pan-cardiac- and ASM- markers as primed or expressed *de novo* in fate-restricted progenitors, respectively (Figure 2E). Remarkably, "primed" genes represented 48% (73/151) and 68% (59/87) of the ASM- and pan-cardiac-specific markers, respectively, indicating that lineage-specific maintenance of primed genes is a major determinant of cell-type-specific transcriptomes in the cardiopharyngeal lineage. However, these values are significantly lower than for the whole transcriptome, since 88% (3,504/3,982) of the late-expressed genes were already active in the multipotent cardiopharyngeal progenitors (Figure 2E). Conversely, whereas only 12% (478/3,982) of the late transcriptome are activated *de novo* in fate-restricted precursors, *de novo*-expressed genes represent 52% (78/151) and 32% (28/87) of the ASM- and pan-cardiac-specific gene sets. This indicates that cell-type-specific gene *de novo* activation contributes significantly more to the deployment of ASM- vs.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

pan-cardiac-specific programs than expected by chance (Fisher's exact test, P= 4.2 x 10^{-7} and P=2.8 x 10^{-34} for the pan-cardiac- and ASM-specific gene sets, respectively; Figure 2E).

While developmental trajectories are built on the premise of a continuous process marked by gradual changes in gene expression levels, these changes occur preferentially in defined 'pseudotime' windows for multiple genes (Figures 2D, S2C). Abrupt transitions are consistent with switch-like biological discontinuities, such as cell divisions (e.g. (Razy-Krajka et al., 2017). Therefore, to augment our interpretation of the dataset, we explored the possibility that each trajectory could be represented as a stepwise progression through a series of defined regulatory states. This approach allows us to recapitulate the developmental continuum, while also identifying the most important developmental transitions, potentially occurring within a single cell cycle (Moris et al., 2016). We first determined the cell-to-cell cross-correlation structures along each lineage-specific pseudotime axis using dynamically expressed genes, we then subdivided the pseudotime space into discrete states through constrained hierarchical clustering (Grimm, 1987). In this way, we identified five states in the ASM trajectory and four states in each of the FHP and SHP trajectories, altogether comprising a total of 10 discrete regulatory states (TVC, STVC, FHP1, -2, and -3, SHP1 and -2, ASM1, -2, and -3; Figure 2F). In general, early regulatory states coincided well with cell identities defined initially by clustering on time-specific samples, but the analysis revealed the existence of further structure in the data. Smooth trajectories characterize the 'pseudotemporal' dynamics of any gene, while regulatory states describe key developmental transitions emerging from coordinated expression changes across many genes. For each of the three trajectories, we first explored the successive regulatory states leading to lineage commitment, and subsequently analyzed the underlying gene expression dynamics, in an attempt to identify genetic determinants of cell fate specification.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ebf is a key determinant of ASM regulatory states

To further explore how cells progress through successive regulatory states, we first focused on the ASM trajectory, for which the key regulators Hand-r, Tbx1/10 and Ebf have been identified and characterized (Razy-Krajka et al., 2014, 2017; Stolfi et al., 2010; Tolkin and Christiaen, 2016; Wang et al., 2013). The first two regulatory states corresponded squarely to two successive generations of multipotent cardiopharyngeal progenitors, the TVCs and STVCs (Figures 2F, S2D). Defined asymmetric cell divisions thus provide the biological basis for the first transitions. By contrast, transitions to the later regulatory states were independent of asymmetrical divisions and have only been partially described, although we surmise that they coincide with progressive ASM fate specification and commitment (Figure 2B,F, S2D ; (Razy-Krajka et al., 2014, 2017).

As described above, specific pseudotemporal patterns matched known expression dynamics, with *Tbx1/10* and *Ebf* expressions increasing sharply at the TVC-to-STVC and STVC-to-ASM1 transitions, respectively (Figures 3A, S2B). Among the cells isolated from 16hpf embryos, where ASMF are unequivocally identified as large lateral cells within the cardiopharyngeal mesoderm, a small number of *Ebf*+ ASM cells surprisingly clustered together with STVCs (Figure 2F, S2D). This suggests that, despite having upregulated a key determinant of pharyngeal muscle fate specification, these early ASMFs had not yet induced the full transcription program for ASM commitment, and were otherwise globally similar to their mother STVCs.

To test this possibility, we used both our newly defined ASM markers and a list of 159 ASM-specific candidate Ebf target genes obtained by microarray analyses of gain- and loss-of-function perturbations of Ebf (logFC>1, p-value<0.05; (Razy-Krajka et al., 2014), and we leveraged their pseudotemporal profiles on the ASM trajectory to calculate an inferred induction time ('turn-on' point) for each gene (Supplementary Methods; Figures 3A,B,D, S3). As expected, we observed that the onset of *Ebf* expression along the ASM trajectory preceded that of most of its ASM-specific candidate target genes, including *Myogenic regulatory factor* (*Mrf*) and *Myosin heavy chain 3* (Mhc3) (Figure 3A, B, D, S3). Notably, candidate Ebf target genes were predominantly activated at either of two distinct pseudotime points, which coincided with the last two ASM transitions (Figure 3B, C, S3A). This suggests that the major ASM1-to-ASM2 and ASM2-to-ASM3 transitions are defined, at least in part, by the activation of candidate Ebf target genes. While these observations are fully consistent with a key role for Ebf in ASM specification,

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

we did observe a small subset of genes whose expression was activated during the STVC-ASM1 transition (Figure 3B, S3B), and therefore may be driven by other regulators such as Tbx1.

We then used our published microarray (MA) data obtained following Ebf perturbations in the cardiopharyngeal lineage (Razy-Krajka et al., 2014), in order to further evaluate the extent of Ebf influence on the ASM transcriptome. Out of 11,082 genes for which we could directly compare scRNA-seq and MA data, 151 were annotated as ASM-specific candidates by scRNA-seq. This ASM candidate population was significantly enriched in genes predicted to be positively regulated by Ebf (42 genes, 26 % of ASM markers, a 19-fold enrichment, $P = 2.2 \times 10^{-43}$; hypergeometric test; Figures 3C, H, S3E-F). These genes were either upregulated by over-expression of a full length wild-type Ebf cDNA and/or inhibited by misexpression of a dominant repressor form of Ebf (Figures 3C, H, S3E-F; (Razy-Krajka et al., 2014)). Taken together, these analyses support the notion that Ebf controls the pharyngeal muscle program by activating a quarter or more of the ASM-specific genes.

We next asked whether Ebf also contributes to the negative regulation of pan-cardiac genes, the majority of which are primed in the multipotent progenitors, and subsequently downregulated in the ASM trajectory. While the expression patterns of these genes was generally anti-correlated with Ebf (Figure S3G), we observed distinct inactivation dynamics within the 'primed' cardiac program (Figure 3E-F). For example, *Hand* was gradually down-regulated, with an inferred 'turn-off' point in the ASM2 state, while *Nk4/Nkx2-5* and *Gata4/5/6* were more rapidly downregulated in the STVC state (Figure 3E). This demonstrates that, similar to *de novo*-expressed genes, distinct inhibition dynamics for pan-cardiac genes also contribute to successive transitions in regulatory states (Figures 3F, S3C). Whereas Ebf is unlikely to downregulate pan-cardiac genes prior to its onset, later down-regulation of primed genes and inhibition of *de novo* pan-cardiac gene activation may require Ebf. Consistent with the latter possibilities, pan-cardiac markers tended to be inhibited by over-expression of an Ebf cDNA and/or upregulated by misexpression of a dominant repressor form of Ebf as shown in our published MA data(Figure 3G, H, S3H-I; (Razy-Krajka et al., 2014)).

In summary, combining pseudotime and regulatory states analyses with perturbation data indicated that, along the ASM trajectory, the earliest cells to engage on a pharyngeal muscle specification path activate the key determinant Ebf, prior to deploying a later Ebf-dependent

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ASM-specific transcriptional program. Similarly, cells engaged on the ASM differentiation path begin to downregulate some of the primed pan-cardiac markers prior to the onset of *Ebf* expression, indicating that, whereas Ebf is in fact necessary to inhibit the heart-specific program in the ASM (Figure 3M-O; (Razy-Krajka et al., 2014; Stolfi et al., 2010, 2014b)), other regulatory inputs, such as Tbx1/10 (Wang et al., 2013), must act before and/or in parallel to Ebf in order to prevent deployment of the pan-cardiac program in the pharyngeal muscles.

De novo pan-cardiac gene activation drives early heart specification

Next, we sought to investigate the gene expression dynamics underlying early cardiac specification. To this aim, we applied diffusion map and cross-correlation analyses to order cells and identify successive regulatory states along the first and second heart lineages (Figure 2). The FHP and SHP pseudotime spaces each segregated into 4 discrete states (Figure 2F, S2D). In both trajectories, the first state principally comprised first generation multipotent cardiopharyngeal progenitors, the TVCs, and in the SHF trajectory, the second state consisted of second generation multipotent cardiopharyngeal progenitors, the STVCs. These observations are consistent with the established lineage tree, and indicate that -as is the case for the ASMs- asymmetric cell divisions condition transitions between multipotent states and the initiation of heart fate specification.

In order to study the gene expression dynamics underlying the transitions through successive cardiac regulatory states, we first focused on the pan-cardiac, and first- and second-heart-precursor-specific markers defined above (Table S1). Since each category was further parsed into "primed" vs. "*de novo*-expressed" (Figure 2), we suspected that the activation timing of *de novo*-expressed heart markers could explain major transitions along the cardiac trajectories. For instance, the primed pan-cardiac markers *Nk4/Nkx2-5* and *Gata4/5/6* remained expressed at high levels throughout the FHP trajectory, albeit with a transient downregulation in the early FHP1 state, as was previously documented for *Gata4/5/6* (Figure 4A; Wang et al., 2013), while the onset of the pan-cardiac marker *Slit* and FHP-specific marker *Mmp21* expression coincides with later transitions to the FHP2 and FHP3 states (Figure 4A-E, S4A). We observed similar patterns globally -- pan-cardiac markers tended to be induced earlier in the trajectory compared to FHP-specific markers, almost all of which were activated in the final FHP3 state. This demonstrates that FHP identity is established during a step-wise process, commencing with the induction of a pan-cardiac program, followed by a gene program which establishes an FHP-specific fate.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Next, we compared the pseudotemporal dynamics of the FHP and SHP fate specification. By definition, de novo 'pan-cardiac' genes must be induced along both trajectories. Accordingly, we observed a striking agreement between the ordered activation pattern of individual genes along each trajectory (Figure 4D), indicating a remarkably conserved developmental program. However, the onset of each gene was consistently delayed in the SHP trajectory, starting with the STVC-SHP1 transition, as these progenitors are born ~2h later than FHPs (Figures 2D, 4C, S4A). Therefore, the *de novo* pan-cardiac program is a tightly regulated and deployed in highly reproducible cascade, whose onset is induced independently at the birth of FHPs from TVCs, and at the birth of SHPs from STVCs.

To further analyze gene expression changes underlying transitions between regulatory states along the FHP trajectory, we analyzed the inactivation dynamics of primed ASM genes, and found that these genes, as well as TVC-specific genes that were not maintained in the ASM, were downregulated as early as the TVC-to-FHP1 transition (Figures 4G, S4B). To identify gene expression profiles underlying successive regulatory states, we performed a principal component analysis focused on cells assigned to the FHP trajectory and genes categorized as "TVC-specific", primed ASM markers, primed pan-cardiac markers, *de novo*-expressed pan-cardiac markers, and FHP-specific genes (Figure 4H, Table S1). PC1 correlated strongly with pseudotime (PCC=0.96; Figure 4H), allowing us to use the PC1 loading of each gene to calculate a contribution score for each gene category (Figure 4I). This analysis indicated that the TVC state is primarily determined by combined TVC-specific genes, and primed cardiac and ASM markers as expected. The TVC-to-FHP1 transition is marked by a sharp decline in the contribution of TVC-specific genes, accompanied by concomitant down-regulation of primed ASM genes, upregulation of primed pan-cardiac genes, and activation of *de novo*-expressed pan-cardiac genes. In this regard, the FHP1 state may be considered a "transition state" between a multipotent TVC state, and the FHP2 state, which is defined by the virtual absence of TVC-specific and primed ASM-specific transcripts, and high levels of both primed and *de novo*-expressed pan-cardiac markers. The FHP2 state may thus constitute an organ-level (heart)-specific program, whereas activation of cell-type/lineage-specific genes underlie the FHP2-to-FHP3 transition, and their expression helps define the FHP-specific state, FHP3.

These data suggest that early heart specification relies on the sustained expression of primed cardiac determinants, the down-regulation of primed ASM and candidate multipotency genes,

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and the initial activation of *de novo*-expressed pan-cardiac genes, which define an initial heart-specific regulatory state that later changes primarily activating an FHP-specific program.

Termination of FGF/MAPK signaling permits further deployment of the heart-specific program.

The previous analyses open the possibility that defined regulatory mechanisms coordinate the birth of FHPs, following asymmetric cell divisions, with the downregulation of TVC-expressed genes and the onset of *de novo*-expressed pan-cardiac genes. We recently found that the FGF/MAPK signaling pathway is active in the multipotent cardiopharyngeal progenitors (TVCs and STVCs), and in the early pharyngeal muscle precursors (ASMF), where it promotes the expression and successive activation of *Hand-r*, *Tbx1/10*, and *Ebf* (Razy-Krajka et al., 2017). By contrast, FGF/MAPK signaling does not persist in first and second heart precursors following asymmetric cell divisions. To test whether differential FGF/MAPK signaling functions as a molecular switch regulating the segregation of early ASM- and heart-specific transcriptional programs, we performed bulk RNA sequencing of FACS-purified cardiopharyngeal lineage cells isolated from embryos and larvae expressing either a dominant negative form the fibroblast growth factor receptor (dnFGFR), or a constitutively active form of M-Ras (caM-Ras), the sole Ras homolog in *Ciona* (Keduka et al., 2009), under the control of TVC-specific enhancers (Razy-Krajka et al., 2017). We then integrated the bulk RNA-seq and scRNA-seq datasets to test how pan-cardiac- and ASM-specific transcription programs respond to perturbations of FGF/MAPK signaling (Figure 4J-K, S4). We first focused on samples obtained from 18 and 20hpf larvae, and confirmed that, as was observed using FISH to monitor *Ebf* expression (Razy-Krajka et al., 2017), FGF/MAPK signaling opposed pan-cardiac gene expression, while promoting ASM-specific genes (Figure 4J-K). Taken together, these results demonstrate that differential FGF/MAPK signaling in the cardiopharyngeal mesoderm acts as a molecular switch that regulates the deployment of global 'pan-cardiac' and 'pharyngeal muscle' programs.

We found that *Nk4/Nk2-5*, *Gata4/5/6*, *Hand* and other primed pan-cardiac markers were inhibited by FGF/MAPK activity in 20hpf larvae (Figure 4J-K), which constitutes a possible conundrum since FGF/MAPK signaling is also active in the multipotent cardiopharyngeal progenitors (Razy-Krajka et al., 2017). To test whether the sensitivity of pan-cardiac genes to perturbations of FGF/MAPK signaling changed over time and/or whether genes are primed or expressed *de novo* in fate-restricted heart precursors, we repeated the bulk RNA-seq analyses

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

following dnFGFR or caM-Ras misexpression with cells isolated from 12, and 15 embryos and pre-hatching larvae. Remarkably, whereas both primed and *de novo*-expressed pan-cardiac genes were sensitive to perturbations of FGF/MAPK signaling at 20hpf, only the *de novo*-expressed genes were upregulated upon dnFGFR misexpression at 12 and 15hpf (Figure 4K). By contrast, primed pan-cardiac genes were generally not affected at these early time points, suggesting that their early expression is independent of FGF/MAPK signaling. Remarkably, similar analysis conducted using STVC- and FHP-specific markers, as defined by differential gene expression in our scRNA-seq data (Table S1), demonstrated that FGF/MAPK signaling inhibits FHP-specific gene expression, which is characterized by an abundance of *de novo* expression heart genes (Figure S4C-D). Taken together, these analyses indicated that FHP- and SHP-specific inhibition of FGF/MAPK signaling permits the activation of *de novo*-expressed pan-cardiac genes, and subsequent heart fate specification.

By contrast with the above effects on primed pan-cardiac genes, and consistent with the role for FGF/MAPK signaling in promoting *Hand-r* and *Tbx1/10* expression in the TVCs and STVCs (Davidson et al., 2006; Razy-Krajka et al., 2017), inhibition of FGF/MAPK signaling reduced the expression of primed ASM- and STVC-specific genes at 12 and 15hpf (Figures 4J, K, S4C, D; (Razy-Krajka et al., 2017). This data indicates that ongoing FGF/MAPK signaling in cardiopharyngeal progenitors promotes multipotency both by maintaining the primed pharyngeal muscle program and by inhibiting further deployment of the heart-specific program.

Heart cell-type specification: SHP vs. FHP fates

Besides the predominant roles of pan-cardiac gene expression in defining successive regulatory states within the first and second heart lineages, we identified FHP- and SHP-specific transcription programs (Figures 2D, 4D, 5A; Table S1).

To further explore the regulatory mechanisms govern heart cell-type specification, we mined the SHP trajectory, which is poorly characterized in *Ciona* and offers the potential to provide insights into the development and evolution of the vertebrate second heart field, since the ascidian and vertebrate SHFs share specific regulatory inputs from Nk4/Nkx2-5 and Tbx1/10 orthologs (Chen et al., 2009; Diogo et al., 2015; Grifone and Kelly, 2007; Liao et al., 2008; Nevis et al., 2013; Prall et al., 2007; Wang et al., 2013; Zhang et al., 2014).

A small number of genes were exclusively upregulated in the SHPs (Figures 1D, 2D, 5A; Table S1). Among these SHP-specific genes, the *dachshund* homolog *Dach*, was the only one to

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

encode a known transcription regulator (Chen et al., 1997; Davis et al., 1999, 2001; Heanue et al., 2002), which was upregulated as cells transitioned to an SHP-specific regulatory state (Figure 5B). Therefore, as was the case for *Ebf* in the ASM trajectory, the pseudotemporal dynamics of *Dach* expression is consistent with a role in specifying SHP identity.

Dach1 and -2 have not been formally implicated in second heart field (SHF) development in mammals. However, *Dach* homologs are part of a conserved regulatory network, termed "the retinal network" (Kumar, 2009), which comprises homologs of Six, EyA and Pax family transcription factors, and studies have uncovered a role for *Six* and *EyA* genes in the mouse SHF (Guo et al., 2011; Zhou et al., 2017). To investigate the function of *Ciona Dach* in SHP fate specification, we employed lineage-specific CRISPR/Cas9-mediated loss-of-function assays (Gandhi et al., 2017; Stolfi et al., 2014b). We reasoned that Dach could contribute to preventing SHP from acquired either an ASM and/or an FHP identity. To distinguish between these possibilities, we analyzed larvae expressing Cas9 and either control or *Dach*-specific sgRNAs by assaying the expression of *Ebf* and *Mmp21*, which specifically mark the ASMFs and FHPs, respectively (Figures 1C, E, S5A ; (Stolfi et al., 2010)). In either control or *Dach* CRISPR larvae, *Ebf* expression was specifically restricted to the lateral-most ASMFs, indicating that *Dach* is neither required for *Ebf* activation, nor for its exclusion from the SHPs (Figure S5A). By contrast, whereas *Mmp21* transcripts were observed only in the FHPs of 22hpf control larvae (Figure 5C,D), its expression domain expanded to include both FHPs and SHPs in larvae expressing the *Dach*-specific sgRNAs (Figure 5C,D). This data indicates that Dach function is required in the SHPs to prevent ectopic activation of the FHP-specific marker *Mmp21*.

Next, we sought to test whether *Dach* expression is sufficient to impose an SHP-like fate to cardiopharyngeal-lineage cells. We used a CRISPR-resistant form of the *Dach* cDNA ($Dach^{PAMmut}$), which we misexpressed using the TVC-specific *FoxF* enhancer (Beh et al., 2007). Co-expressing $Dach^{PAMmut}$ with Cas9 and *Dach*-specific sgRNAs was sufficient to restrict *Mmp21* expression to the FHPs, indicating that the observed effects of *Dach* sgRNAs are specifically due to a loss of *Dach* function in the cardiopharyngeal lineage. However, neither the rescue condition, nor *Dach* misexpression alone abolished *Mmp21* expression in the FHPs (Figure 5C,D), indicating that *Dach* is necessary but not sufficient to inhibit FHP-specific gene expression in the cardiopharyngeal lineage. Taken together, these analyses revealed an SHP-specific function for the *Ciona Dach* homolog in controlling a first- vs. second-heart-field-specific identity.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The first and second cardiac lineages share a common set of pan-cardiac transcripts, but differ by defined gene activities, thus opening the possibility that each lineage forms distinct cell types during heart development. For instance, although the beating *Ciona* heart is demonstrably simpler than its vertebrate counterpart, the single-compartment U-shaped adult heart is encased in a pericardial squamous epithelium, it contains two pacemakers at either ends of the tube, an epithelial row of undifferentiated cells, the contractile wall is composed of myoepithelial cells, and endothelial cells are missing altogether (Anderson and Christiaen, 2016; Davidson, 2007). In 2 to 3 day old post-metamorphic juveniles, the heart has not yet acquired its characteristic shape, but it already beats peristaltically, and double labeling with a B7.5-lineage-specific *Mesp>nls::lacZ* reporter and the cardiac-specific *myosin heavy chain 2* (*Mhc2*) marker showed that a β-galactosidase+; *Mhc2*- pericardium already encases *Mhc2*+ cardiomyocytes (Stolfi et al., 2010; Wang et al 2013). To test whether the first and second heart lineages contribute to forming different cells and/or regions of the beating juvenile heart, we performed lineage tracing using a nuclear form of the photoconvertible fluorescent reporter Kaede, which we expressed in the B7.5 lineage using the *Mesp>nls::kaede::nls* construct (Razy-Krajka et al., 2014). We photoconverted Kaede proteins specifically in the median FHPs of 16hpf larvae (Figure 5E.II), and raised them through metamorphosis. We imaged whole hearts at successive times points, and segmented and color-coded individual nuclei, which allowed us to distinguish FHP- from SHP-derivatives in differentiating hearts (Figures 5E). In ~65hpf juveniles, the heart typically consisted of 32 to 39 cells per half, arranged in a conical shape with the apex pointing towards the base of the juveniles (Figure 5E.IV). We observed that the FHP- and SHP-derived cells remained within largely separate domains, which mixed only to a small extent (Figure 5E). Specifically, FHP-derived cells occupied the more basal domain, near the apex, whereas SHP-derived cells form the more distal base of this conical heart.

We next asked whether FHP- and SHP-derived cells contributed equally to the population of *Mhc2*+ cardiomyocytes in juvenile hearts. We coelectroporated the B7.5 lineage marker *Mesp>nls::LacZ* with a construct containing three copies of a minimal STVC-specific enhancer from the *Tbx1/10* locus (*3XT12*; (Tolkin and Christiaen, 2016); Racioppi et al., in preparation) to express an histone H2B::mCherry fusion and label SHP-derived cells in juvenile hearts stained for *Mhc2* expression. In these juvenile hearts, most *Mhc2*+ cells were located near the FHP-derived apex of the conical heart, and few *Mhc2*+ cells coexpressed *3xT12*-driven H2B::mCherry, which is consistent with a predominantly FHP origin (Figure 5F, S5B). By

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

segmenting dual β-galactosidase- and H2B::mCherry-labeled, SHP-derived cells and assaying colocalization with *Mhc2* staining, we determined that only ~17% of SHP-derived cells express *Mhc2* in control juveniles (Figure 5F,G; S5B). This result indicates that distinct cardiac progenitors from the first and second heart fields produce spatially and qualitatively different parts of the ascidian beating heart, instead of converging on a common pan-cardiac progenitor type (Figure 5F,G; S5B).

Since SHPs specifically activate *Dach*, which is necessary to inhibit expression of the FHP marker *Mmp21*, we reasoned that Dach function could also limit the proportion of SHP-derived cells that acquire an *Mhc2+* cardiomyocyte identity. To test this possibility, we grew larvae expressing control or *Dach* CRISPR/Cas9 constructs alongside the *Mesp>nls::LacZ* and *3xT12>H2B::mCherry* reporters, and quantified the proportions of *Mhc2+* SHP-derived cells, which we found to increase from 17% in controls to 46% upon loss of *Dach* function. These results demonstrate that Dach inhibits the expression of *Mmp21* in early SHPs, and limits the potential of SHP derivatives to form *Mhc2+* cardiomyocytes during organogenesis and differentiation. Taken together, these experiments indicate that the earliest signs of FHP- vs. SHP-specific transcriptional divergence correspond to the establishment of distinct cardiac progenitor identities, which contribute to the formation of different cell types in separate domains of the juvenile heart.

Tbx1/10 acts as a dual regulator of cardiopharyngeal fates upstream of *Dach*

Dach was activated specifically in the SHPs and appeared necessary but not sufficient to inhibit the FHP-specific gene expression and later differentiation of *Mhc2+* cardiomyocytes. We thus reasoned that SHP-specific regulatory inputs govern *Dach* expression and the inhibition of the FHP program. SHPs emerge from *Tbx1/10+* multipotent progenitors, the STVCs, and transiently maintain *Tbx1/10* expression (Figures S2B, 5A; Wang et al., 2013). As mentioned above, *Tbx1* function is required for proper development of SHF-derivatives in vertebrates, which prompted us to test whether Tbx1/10 regulates *Dach* expression and later SHP-specific features.

We first used established sgRNA-expressing constructs to inhibit Tbx1/10 function by CRISPR/Cas9-mediated mutagenesis and assayed *Mmp21* expression in swimming larvae, and *Mhc2* expression in metamorphosing juveniles (Figure 5; (Gandhi et al., 2017; Tolkin and Christiaen, 2016)). Loss of Tbx1/10 function largely mimicked Dach inhibition while causing more pronounced phenotypes, with ectopic *Mmp21* activation in cells at the intermediate

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mediolateral position normally occupied by SHPs (Figure 5C,D), and the proportion of SHP-derived *Mhc2*+ cells increased to 75% in the Tbx1/10 CRISPR "background", compared to 17% in controls and 46% upon Dach inhibition (Figure 5F, G). These results indicate that *Tbx1/10*, which is specifically activated in the second multipotent cardiopharyngeal progenitors that produce the SHPs and ASMFs, but not the FHPs (Figure 1A; Wang et al., 2013), is necessary to oppose deployment of the FHP-specific program in the second heart lineage.

These results led us to test whether Tbx1/10 regulates *Dach* expression, as part of its role in promoting the SHP-specific program. We used CRISPR/Cas9 to inhibit Tbx1/10 function and assayed *Dach* expression in 18hpf larvae. Since *Dach* is also expressed in adjacent non-cardiopharyngeal tissues that obscured SHP-specific expression, we used nuclear and membrane markers to segment cells, and counted fluorescent dots, which provided an approximation of transcripts abundance in individual cells (Figure 6). These analyses revealed a marked downregulation of *Dach* expression in the SHPs following Tbx1/10 inhibition, which is consistent with a key role for Tbx1/10 in regulating *Dach* activation specifically in the second heart lineage (Figure 6A).

We previously showed that Tbx1/10 is necessary to activate *Ebf* and promote the pharyngeal muscle program in the lateral-most ASMFs (Wang et al., 2013), in a manner similar to Tbx1 function in branchiomeric myogenesis in vertebrates (Kelly et al., 2004; Kong et al., 2014). Since SHPs inhibit FGF/MAPK signaling, which remains active in the ASMFs and is required for *Ebf* activation in parallel to Tbx1/10 (Razy-Krajka et al., 2017), we reasoned that inhibition of FGF/MAPK may be necessary for SHP-specific activation of *Dach* by Tbx1/10. We first used the Mek inhibitor U0126 to block MAPK signaling prior to the division of STVCs, before after the onset of endogenous *Tbx1/10* expression, and assayed *Dach* expression in 18hpf larvae. We observed conspicuous ectopic activation in the lateral-most STVC-derived cells that normally form the *Ebf*+ ASMFs (Figure 6B). Since this treatment also inhibits *Ebf* expression (Razy-Krajka et al., 2017), inhibition FGF/MAPK signaling in Tbx1/10+ cardiopharyngeal derivatives appears sufficient to activate *Dach* expression and promote the SHP identity.

To further test whether a Tbx1/10+, MAPK- logic governs *Dach* expression in the cardiopharyngeal mesoderm, we used gain- and loss-of-function assays by over-expressing a *Tbx1/10* cDNA using the TVC-specific *FoxF* enhancer, in combination with U0126-mediated Mek inhibition. Since *FoxF>Tbx1/10* causes precoce misexpression in the first generation of multipotent progenitors, we analyzed *Dach* expression in individual cardiopharyngeal cells following the first asymmetric division, in 14hpf pre-hatching larvae. Remarkably, only the

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

combined misexpression of *Tbx1/10* and inhibition of MAPK signaling caused substantial upregulation of *Dach* in both the small median and large lateral lateral cells that normally correspond to the *Dach*-negative FHPs and STVCs, respectively (Figure 6C, D).

In summary, we showed that the combination of *Tbx1/10* expression in the second multipotent progenitors (STVCs), with FGF/MAPK inhibition suffice to specify a second heart lineage progenitor identity, in part by upregulating the novel SHF regulator *Dach*. The latter contributes to opposing the alternative first heart lineage program, marked in part by early *Mmp21* expression and a higher propensity to produce *Mhc2*+ cardiomyocytes in beating juvenile hearts.

Conserved expression of a *Dach* homolog in the mouse SHF

Finally, we asked whether *Dach* homologs mark second heart field derivatives in the mouse. Recent single cell RNA-seq analysis of early mesodermal lineages identified a population of pharyngeal mesoderm marked by high levels of *Tbx1* expression, and remarkably similar levels of *Dach1* expression (Figure 7A; (Scialdone et al., 2016)). To complement these observations, we perform multicolor immunohistochemical staining using an anti-Dach1 antibody, and assayed co-expression with the cardiac and second heart field markers Nkx2.5 and Islet1 between embryonic day (E) 7.5 and E9.5 (Figure 7B, C; S6). At E7.5, Dach1 is broadly expressed in the pharyngeal mesoderm, where it colocalizes with Islet1, and with nascent Nkx2.5 in a subset of cells (Figure 7B). Dach1 remains broadly expressed at E8.5 (Figure S6), but by E9.5, its expression becomes restricted to a defined population of outflow tract cells, which also express Islet1, and is coexpressed with Islet1 in second heart field cells in the dorsal pericardial wall. Notably, by this stage, Dach1 expression was excluded from the Nkx2.5+ ventricle and absent from the Islet1+ skeletal muscle progenitor cells in the core mesoderm of the first and second pharyngeal arches (Figure 7C), in a manner reminiscent of *Dach* exclusion from the ASMFs in *Ciona* (Figure 1E). The latter pattern indicates that, like Islet1, Dach1 expression is specifically maintained in the second heart field and distal outflow tract, but excluded from other cardiopharyngeal mesoderm derivatives. These observations suggest that expression of *Dach* homologs in second heart field progenitor cell is thus conserved in chordates, although the gene expression dynamics leading to SHF-specific expression may differ since *Dach* is specifically upregulated in the *Ciona* SHP, whereas Dach1 appears to be broadly expressed at E7.5 and progressively restricted to the mouse SHF by E9.5.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

Here, we have presented an extensive analysis of the transcriptome dynamics underlying multipotency and early fate specification in the cardiopharyngeal mesoderm of a simple chordate. By profiling the transcriptome of single cardiopharyngeal lineage cells isolated from embryos and larvae dissociated at defined time points, we unequivocally recovered cell identities and globally characterized cell-specific gene expression. We showed that the pharyngeal muscle vs. pan-cardiac split accounts from most differential gene expression between cell types. We also uncovered subsets of first- and second-heart-precursor-specific transcripts. Because cardiopharyngeal development proceeds according to a fixed and known lineage in *Ciona*, our time series dataset allowed us to reconstruct the three main developmental trajectories whereby multipotent cardiopharyngeal progenitors (the TVCs) produce either first or second heart precursors, or atrial siphon muscle (ASM) precursors. This specificity of the *Ciona* model highlights the advantage of reconstructing developmental trajectories using a known and fixed lineage. This prevents unknown clonal relationships between individual cells to convolve the analysis of global gene expression changes. Indeed, most trajectory reconstructions using complex vertebrate systems produce abstract models that do not necessarily represent the actual developmental path that each individual progenitor follows in the embryo (Moris et al., 2016). By contrast, we can confidently ascertain that the progeny of every cardiopharyngeal progenitor in every *Ciona* embryo follows either one of the three trajectories that we modeled to produce first or second heart precursors, or pharyngeal muscle founder cells.

Based on unequivocal clonal relationships between cells, we characterized the transcriptome dynamics along each trajectory. We showed that multilineage transcriptional priming affects large proportions of cell-type-specific markers, although *de novo* gene activation accounts for greater fractions of cell-type-specific transcriptomes than expected by chance, highlighting the importance of both transcriptional control and post-transcriptional regulation in the heart vs. pharyngeal muscle fate choice.

Our analyses of transcriptome dynamics, and the impact of defined regulators (Ebf, FGF/MAPK signaling, Tbx1/10 and Dach) operating at distinct times and places, lead us to propose a provisional mechanistic model to explain the transition from multipotent cardiopharyngeal progenitors to distinct first and second heart precursors, and pharyngeal muscle precursors. First, following Mesp- and FGF/MAPK-dependent induction (Beh et al.,

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2007; Christiaen et al., 2008; Davidson et al., 2005, 2006; Satou et al., 2004; Woznica et al., 2012), the multipotent cardiopharyngeal progenitors, the TVCs, maintain detectable levels of FGF/MAPK signaling, which is required for the maintenance of TVC-specific and primed ASM programs (Figure 7D; (Razy-Krajka et al., 2017)). The TVC-specific program contributes to regulating cell behavior, including collective cell migration and oriented asymmetric cell divisions (Christiaen et al., 2008), and possibly also specific aspects of cardiopharyngeal multipotency. By contrast, past the initial induction, the primed pan-cardiac program no longer requires FGF/MAPK signaling. Instead, continuous FGF/MAPK signaling prevents the precocious activation of *de novo* pan-cardiac markers. The molecular mechanisms for this MAPK-mediated inhibition of cardiac fate specification remain elusive, although they may be conserved with vertebrates ((Hutson et al., 2010; Marques et al., 2008; Tirosh-Finkel et al., 2010; van Wijk et al., 2009); see discussion in (Razy-Krajka et al., 2017)). We propose that seemingly conflicting reports on the role of FGF/MAPK signaling in early cardiac specification in vertebrates can be reconciled simply by considering that FGF/MAPK inputs are necessary to induce multipotent cardiac progenitors (Abu-Issa et al., 2002; Barron et al., 2000; Brand, 2003; Reifers et al., 2000; Zaffran and Frasch, 2002), whereas termination of FGF/MAPK signaling is required for subsequent commitment to a heart fate and differentiation (Hutson et al., 2010; Marques et al., 2008; Tirosh-Finkel et al., 2010; Watanabe et al., 2012; van Wijk et al., 2009), in a manner analogous to the different sequential requirements for canonical Wnt signaling during cardiogenesis (Tzahor, 2007).

The first oriented and asymmetric division of multipotent progenitors gives birth to small median first heart precursors, the FHPs, and large lateral second multipotent progenitors, the STVCs. Unknown mechanisms cause FGF/MAPK to be maintained only in the large lateral STVCs where, together with the primed ASM regulator Hand-r, it contributes to activating *Tbx1/10* expression (Razy-Krajka et al., 2017). Whereas ongoing FGF/MAPK signaling also continues to prevent activation of *de novo* pan-cardiac markers, the newly activated Tbx1/10+ STVC program presumably antagonizes both the TVC-specific program, and certain aspects of the primed pan-cardiac program (Figure 7D, e.g. Tbx1/10 inhibits *Gata4/5/6* expression, (Wang et al., 2013). In the newborn FHPs, termination of FGF/MAPK signaling impedes the maintenance of TVC-specific and primed ASM gene expression, while permitting *de novo* expression of pan-cardiac marker genes, which we surmise are activated by the primed and

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

universally conserved pan-cardiac regulators Nk4/Nkx2-5, Gata4/5/6 and Hand (e.g. (Ang et al., 2016; Luna-Zurita et al., 2016; Tsuchihashi et al., 2011).

Following the second oriented asymmetric division of STVCs, FGF/MAPK signaling is further restricted to the lateral large pharyngeal muscle progenitors (the ASMFs) where, together with Hand-r and Tbx1/10, it activates *Ebf* expression (Razy-Krajka et al., 2017). Whereas ongoing FGF/MAPK signaling presumably continues to inhibit the deployment of the cardiac program in the ASMFs, *Ebf* expression rises sharply, becomes auto-regulative, and reinforces the antagonism with the heart program, thus promoting commitment to a pharyngeal muscle identity upstream of *Mrf* (Razy-Krajka et al., 2014, 2017; Stolfi et al., 2010; Tolkin and Christiaen, 2016). *Ebf* homologs have been implicated in branchiomeric muscle development in chicken (El-Magd et al., 2014a, 2014b) and *Xenopus* (Green and Vetter, 2011) embryos, suggesting that their roles in branchiomeric myogenesis are conserved in chordates.

As is the case in the first heart lineage, termination of FGF/MAPK signaling in the SHPs promotes the activation of *de novo* pan-cardiac markers, presumably also downstream of primed pan-cardiac regulators, albeit with a delay compared to the FHPs (Figures 2D, 4D). In the latter, deployment of the cardiac program leads to (1) upregulation of primed pan-cardiac markers, possibly as a result of positive feedbacks, (2) elimination of primed pharyngeal muscle markers, and (3) activation of FHP-specific markers, such as *Mmp21*, presumably downstream of a general pan-cardiac program, itself driven by both primed and *de novo*-activated regulators.

In the second heart lineage, deployment of the pan-cardiac program does not lead to the activation of FHP-specific markers. Instead, SHPs deploy an alternative cardiac cell-type-/lineage-specific program, marked by the expression of *Dach*. This *Dachshund/Dach1/2* homolog is necessary to inhibit the expression of FHP-specific markers, and limit the propensity of SHP-derived heart cells to form *Mhc2+* cardiomyocytes in beating juvenile hearts (Figures 5, 7D). We hypothesize that, together with downregulation of FGF/MAPK signaling, activation of the SHP-specific program requires inputs from primed pan-cardiac regulators, in addition to regulatory inputs from the STVC-specific determinant Tbx1/10 (Figures 6, 7D). Since Tbx1/10 also promotes the activation of the pharyngeal muscle determinants Ebf and Mrf (Tolkin and Christiaen, 2016; Wang et al., 2013), it emerges as a *bona fide* regulator of cardiopharyngeal multipotency, and one could regard the *Tbx1/10+* STVC regulatory state as a hallmark of multipotent cardiopharyngeal progenitors.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In vertebrates, Tbx1 homologs also play dual roles in branchiomeric myogenesis and second heart field development (Chen et al., 2009; Kelly and Papaioannou, 2007; Kelly et al., 2004; Kong et al., 2014; Liao et al., 2008; Vitelli et al., 2002; Zhang et al., 2006)). We showed that mouse Dach1 is co-expressed with Isl1 at E9.5 in in the distal outflow tract and second heart field cells in the dorsal pericardial wall. However, the dynamics leading to SHF-specific expression differs from that observed in *Ciona*, inasmuch as Dach1 is broadly expressed in the cardiopharyngeal mesoderm et E7.5 and E8.5, before being downregulated in both the differentiating ventricles and branchiomeric mesodermal core (Figure 7). Therefore, whereas Tbx1 could in principle contribute to Dach1 expression, we must invoke other mechanisms accounting for its early widespread expression and selective maintenance in the SHF. Notably, together with *Pax*, *Six* and *Eya* family genes, *Dach* homologs form a conserved "retinal network" (Davis and Rebay, 2017; Kumar, 2009). In the mouse, *Six* and *Eya* homologs have been implicated in early SHF development (Guo et al., 2011; Zhou et al., 2017), opening the possibility that conserved elements of the retinal network contribute to SHF-specific gene expression.

In summary, we have used single cell RNA-seq in a highly tractable model for early cardiopharyngeal development, and characterized the transcriptome features and dynamics underlying cardiopharyngeal multipotency and early fate specification. We characterized essential features of cardiopharyngeal multipotency, such as the extent of multilineage transcriptional priming, the roles of regulatory inputs from FGF/MAPK signaling and Ebf in the early heart vs. pharyngeal muscle fate choice. We show that first and second lineages of fate restricted heart progenitors activate a shared pan-cardiac program upon inhibition of MAPK signaling, but differ by small subsets of lineage-specific genes. We show that the second heart lineage specifically activates a *Dach1/2* homolog downstream of Tbx1/10, which inhibits the first-heart-lineage-specific program. These results indicate that, by analogy with neuronal and muscle fate diversification by temporal patterning in *Drosophila* (Dubois et al., 2016; Isshiki et al., 2001; Li et al., 2013), the first and second heart precursors share a common cardiac identity but differ because they emerge from successive multipotent progenitors before and after the onset of *Tbx1/10* expression, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Authors' contributions

W.W. performed the *Ciona* experiments. E.J. performed the mouse experiments. X.N. performed computational analyses. W.W., X.N., R.K., R.S., and L.C. designed the experiments and analyses. W.W., X.N., R.S. and L.C. wrote the paper.

Acknowledgements

We are grateful to Florian Razy-Krajka for discussions and sharing reagents prior to publication. We thank Ashley Powers for help processing the single cell samples in the early phase of this study, Christopher Hafemeister and Andrew Butler for discussion and help on computational analyses. This project was funded by NIH/NHLBI R01 award HL108643 to L.C., trans-Atlantic network of excellence award 15CVD01 from the Leducq Foundation to R.K. and L.C., and an NIH New Innovator Award (DP2-HG-009623) to R.S.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1

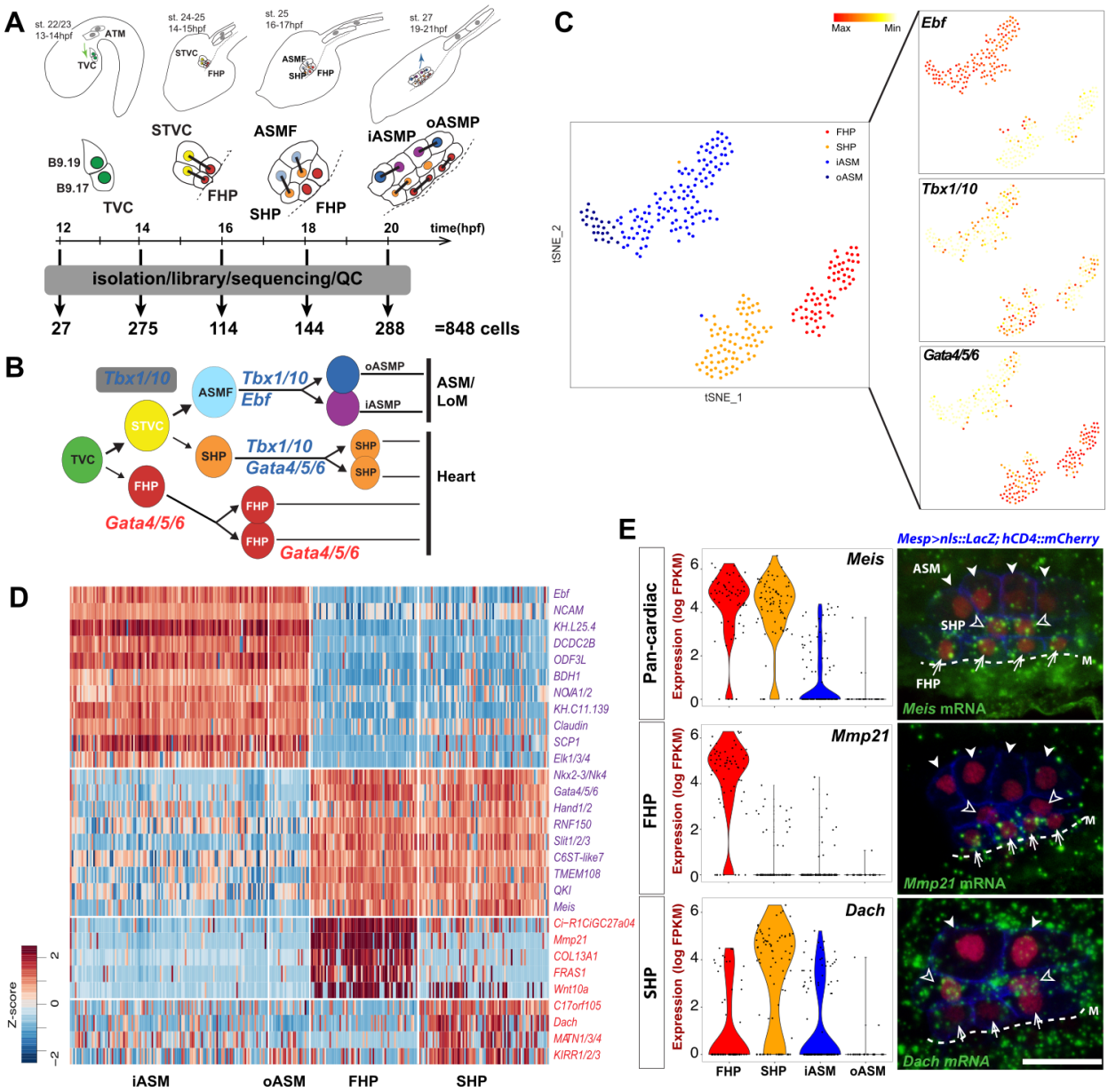


Figure 1. Single cell RNA-seq and clustering identifies cardiopharyngeal progenitor subtypes and novel cell-type-specific markers. (A) Schematic representation of early cardiopharyngeal development in *Ciona robusta* and sampling stages for scRNA-seq. Abbreviations: st., stage according to (Hotta et al., 2007); hpf, hours post-fertilization; TVC, trunk ventral cell; STVC, second trunk ventral cell; FHP, first heart precursor; ASMF, atrial siphon muscle founder cells; SHP, second heart precursor; iASMP, inner atrial siphon muscle precursor; oASMP, outer atrial siphon muscle precursor, according to (Razy-Krajka et al., 2014); QC, quality control. Dotted line: midline, cardiopharyngeal lineage cells are shown for only one side of the animals. (B) Established TVC lineage tree illustrating early cardiopharyngeal specification events. Known cell-type-specific marker genes are indicated

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

on corresponding branches. Abbreviations as in (A); LoM, longitudinal muscles. (C) The t-distributed Stochastic Neighbor Embedding (t-SNE) plots of 20 hpf scRNA-seq data depicts separation of distinct progenitor subtypes: FHP (red), SHP (orange), iASM (blue) and oASM(dark blue) cell clusters are shown. Indicated marker gene expression levels color-coded and shown on corresponding clusters. (D) Expression heatmap of 20 hpf single cell transcriptomes showing top predicted differentially expressed marker genes across different cell types. Blue names: previously known ASM and heart markers, red names: newly discovered markers. (E) Violin plots and FISH validations of newly identified cell-type-specific markers. Solid arrowheads, ASM; open arrowheads, SHPs; arrows, FHPs; M, midline (dotted line). Scale bar = 20μm.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

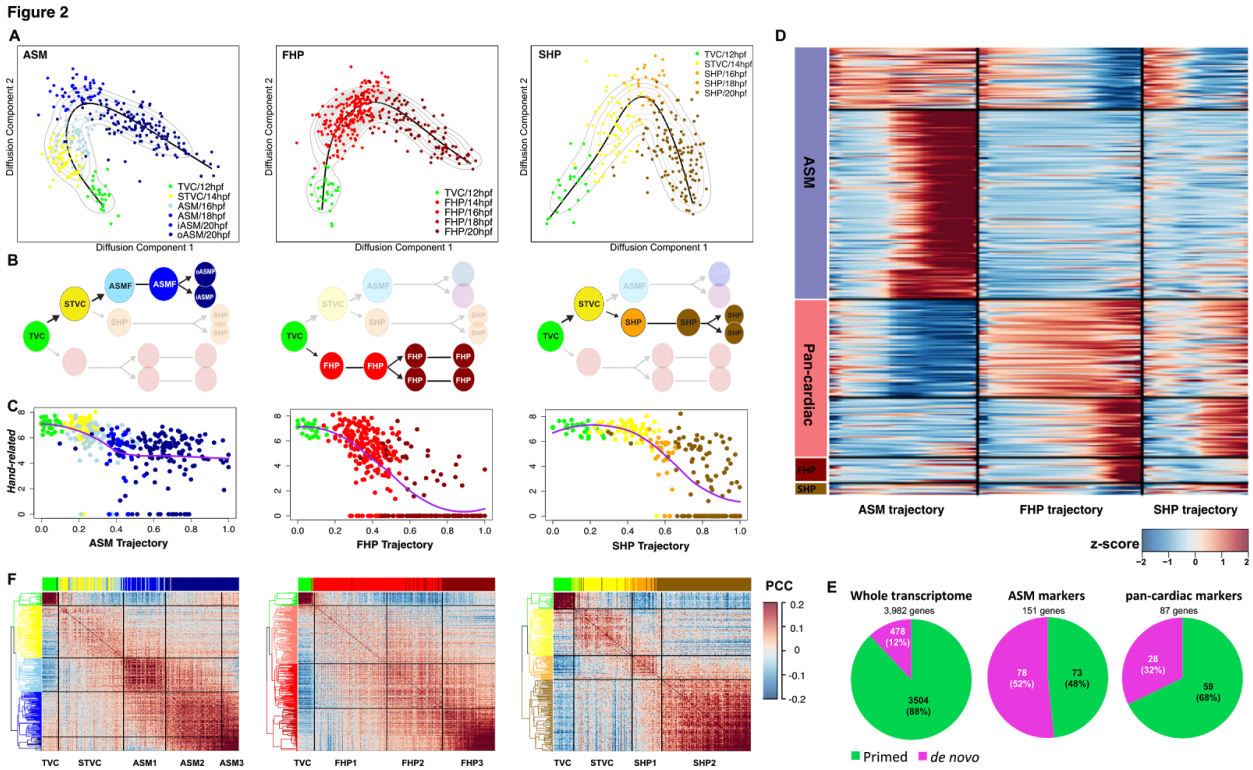
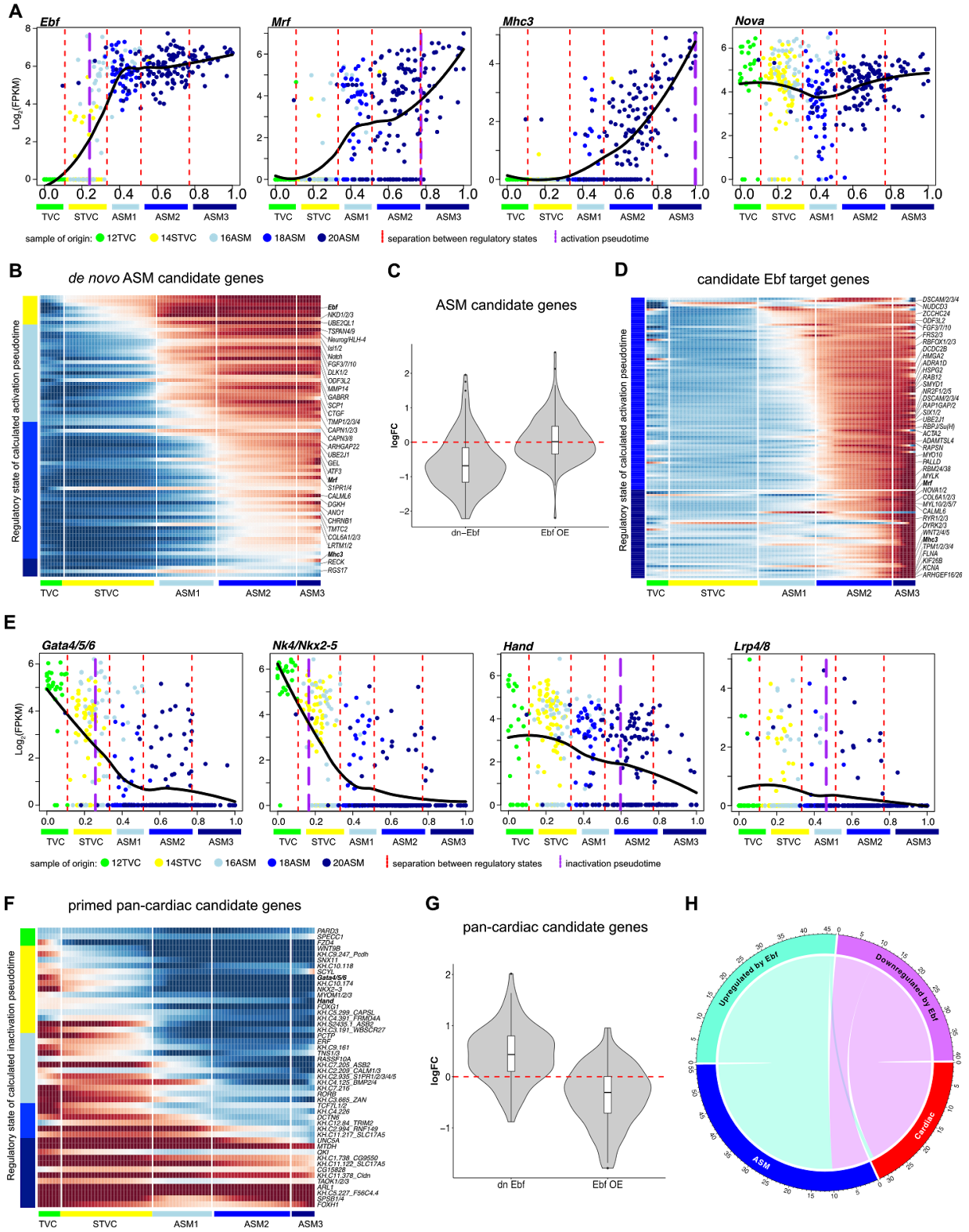


Figure 2. Pseudotemporal reconstruction of cardiopharyngeal developmental trajectories. (A) 2D diffusion maps single cell transcriptomes for three developmental trajectories. Black lines indicate principal curves and light gray contours indicate single cell density distribution along principal curve. Color codes correspond to assigned cell identities following clustering at each time point. Hpf, hours post-fertilization. (B) Cartoon showing three unidirectional cardiopharyngeal progenitor trajectories. (C) *Hand-related/Hand-r* expression on three trajectories. The purple lines show smoothed expression profiles. Y-axis: $\log_2$(FPKM). X-axis, normalized pseudotime point along the principal curve (shown in (A)). Colors codes as in (A). (D) Heatmap of smoothed temporal expression pattern of indicated cell type specific marker genes on three different trajectories. Expressions are grouped by k-means clustering. (E) Pie-charts showing the proportion of primed vs. *de novo* genes in three categories: Whole transcriptome, ASM Markers and Pan-cardiac Markers. (F) Cross-correlation heatmaps showing potential regulatory states along three trajectories. The left side dendrograms indicate the results of constrained hierarchical clustering. The top bars indicate the sample of origin for each cell ordered along the pseudotime axes. Color codes for top bars as in (A).

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3

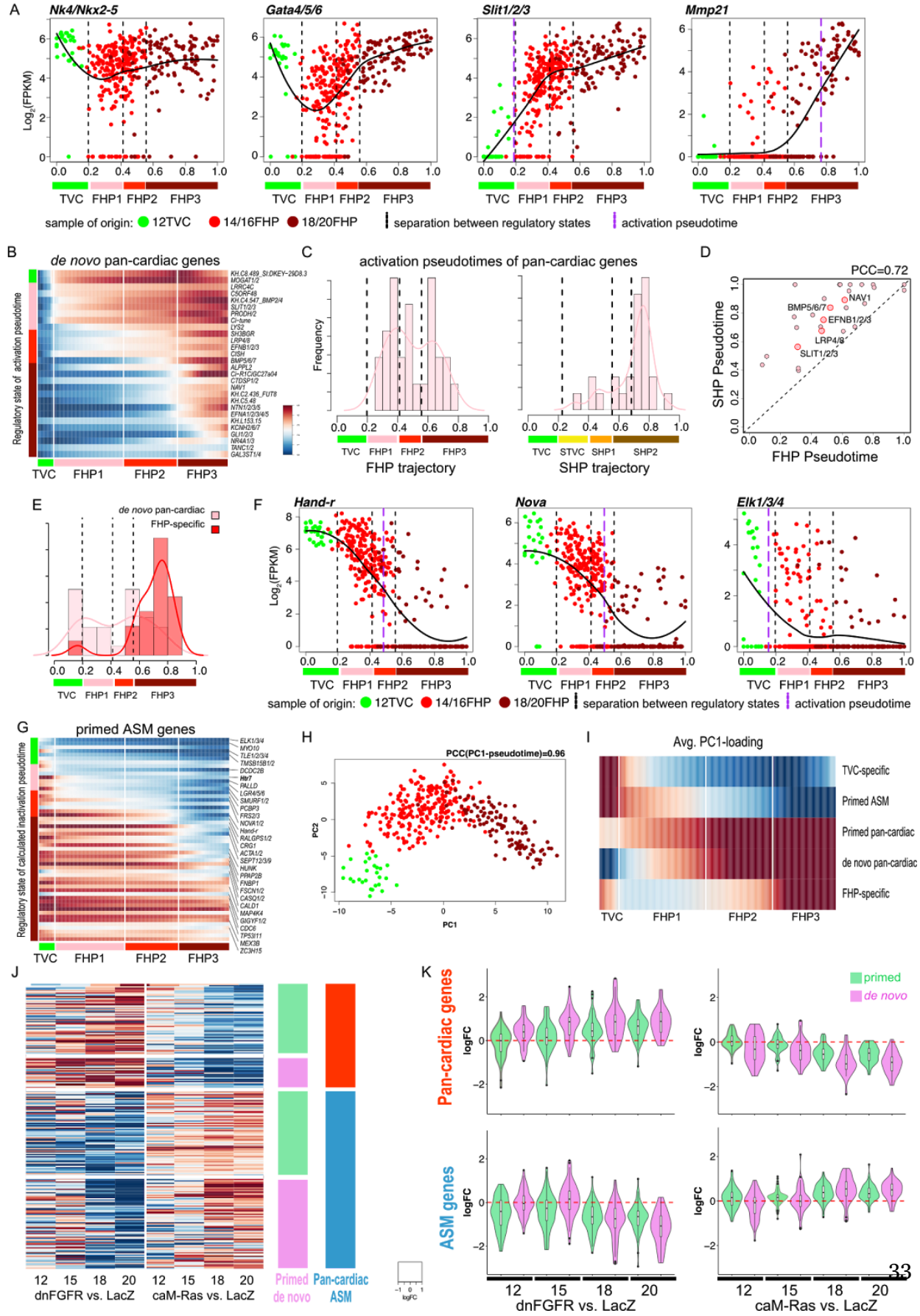


bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3. Gene expression dynamics and transcriptional regulation of ASM fate specification. (A) Pseudotemporal expression profiles of indicated genes along the ASM trajectory. X-axis: normalized pseudotime as defined in Figure 2A. Red dashed lines indicate the transitions between predicted regulatory states, indicated and color-coded below, and purple dashed lines indicate calculated activation pseudotime. Dots colors refer to the sample of origin as indicated below and in Figure 2A. (B) Heatmap of smoothed pseudotemporal expression profiles showing successive activation candidate *de novo* ASM genes. The white vertical lines mark transitions between indicated regulatory states along the ASM trajectory. Colored bars on the left indicate the regulatory state of calculate activation pseudotime, showing that most ASM candidates turn on in the ASM1 (light blue) and ASM2 (navy blue) states. (C) Violin plots showing the log fold change of all candidate ASM-specific genes in response to defined Ebf perturbations, a dominant-negative (dn-Ebf) and Ebf precocious and misexpression (Ebf-OE) as described in (Razy-Krajka et al., 2014). (D) Heatmap of smoothed pseudotemporal expression profiles for candidate ASM-specific Ebf target genes defined in (Razy-Krajka et al., 2014), and showing two main activation waves of activation in regulatory states ASM2 and ASM3 (left colored bars). The white vertical lines mark transitions between indicated regulatory states. (E) Gene plots showing the progressive depletion of the indicated primed pan-cardiac genes along the ASM trajectory. Red dashed lines indicate transitions between indicated regulatory states and purple dashed lines indicate predicted inactivation pseudotime. (F) Heatmap of smoothed pseudotemporal expression profiles showing progressive depletion of primed pan-cardiac genes along the ASM trajectory, with calculated pseudotime of inactivation mapped onto discrete regulatory states indicated on the left colored bars. (G) Violin plots showing the log fold changes of pan-cardiac genes corresponding to Ebf perturbation. (H) Chord diagram showing mutual enrichment of ASM vs. Cardiac genes among candidate target genes activated or inhibited by Ebf, respectively. Note that Ebf is predicted to downregulate a few ASM candidate genes, which are primed and quickly downregulated after the onset of Ebf (e.g. *Hand-r*, (Razy-Krajka et al., 2014).

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4

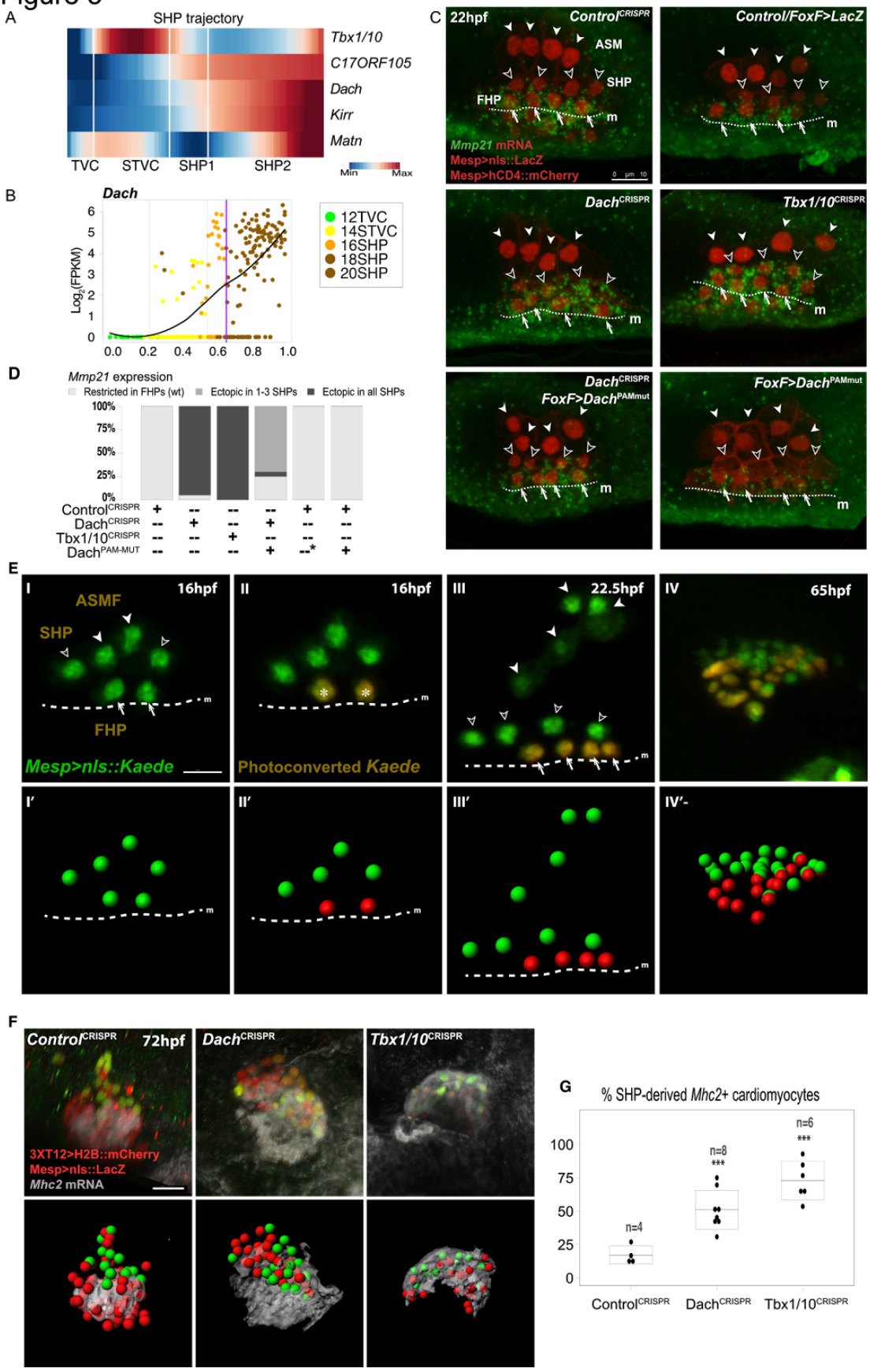


bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4. Gene expression dynamics underlying heart fate specification. (A) Pseudotemporal expression plots showing the activation of the representative pan-cardiac genes on FHP trajectory. The black dashed lines and color bars indicate the regulatory states and the purple dashed lines indicate predicted induction time. (B) Heatmap of smoothed gene expression showing the successive activations of *de novo*-expressed pan-cardiac genes. White vertical lines indicate transitions between indicated regulatory states on FHP trajectory. Left-side color bar indicates relative expression level. (C) Density plots showing the number of *de novo* pan-cardiac genes with calculated activation pseudotime in binned pseudotime windows along FHP (right) and SHP (left) trajectories. The black dashed lines indicate the regulatory states. (D) Activation pseudotimes for *de novo*-expressed pan-cardiac genes along the FHP and SHP pseudotime axes. Note the conserved order (high Pearson's Correlation Coefficient, PCC) in which genes turn on in two independent trajectories, which only share the TVC state. (E) Density plots showing the number of *de novo* pan-cardiac and FHP-specific genes with calculated activation pseudotime in binned pseudotime windows along FHP trajectory. FHP-specific genes tend to be activated later than pan-cardiac markers. (F) Pseudotemporal expression plots showing the depletion of the representative primed ASM genes on FHP trajectory. The black dashed lines indicate the regulatory states and the purple dashed lines indicate predicted inactivation time. (G) Heatmap of smoothed gene expression showing progressive depletion of primed ASM genes on FHP trajectory. White vertical lines indicate predicted regulatory states on FHP trajectory. (H) Principal component analysis showing that PC1 correlates strongly with pseudotime. (I) Average PC1-loading scores per indicated gene category, mapped onto the FHP trajectory. (J) Heatmap showing log fold changes of primed/*de novo* ASM and pan-cardiac genes in dnFGFR vs. LacZ and caM-Ras vs. LacZ pairwise comparisons from FACS-purified samples obtained from 12, 15, 18 and 20 hpf larvae (hpf: hours post-fertilization). (K) Violin plots showing log fold changes as in indicated conditions and time points relative to LacZ controls, and parsed by primed or *de novo*-expressed, pan-cardiac or ASM genes.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5



bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5. *Dach* promotes early SHP specification by inhibiting FHP-specific gene expression.

(A)Heatmap of smoothed gene expression showing Tbx1/10 activation and SHP-specific genes on the SHP trajectory. White vertical lines indicate predicted SHP regulatory states. Dach is the only known transcription regulator identified as SHP-specific. (B) Gene plots showing Dach temporal expression pattern on the SHP trajectory. Black dashed lines indicate transitions between SHP regulatory states and the purple dashed line indicates predicted induction time. (C) Dach and Tbx1/10 are required to prevent expression of the FHP specific gene *Mmp21* in the SHPs. Conditions indicated on top-right corner of each panel. Mmp21 mRNAs are visualized by whole mount fluorescent *in situ* hybridization (green signal). Nuclei of TVC-derived cells progeny are marked by Mesp>nls::LacZ and revealed by anti beta-galactosidase antibody (red). Mesp driven hCD4::mCherry accumulates in the cell membrane and revealed by anti mCherry antibody (red). Open arrowheads, ASMPs; solid arrowheads, SHPs; arrows. Dotted line: ventral midline. Anterior to the left. Scale bar, 10 μm. (D) Proportions of larvae showing indicated phenotypes, as illustarted in (C) for indiacted experimental conditions. Wt, wild-type, indicates normal *Mmp21* expression restricted to the FHPs. Dach^{PAMmut}, Dach cDNA with wobble base mutation in the PAMs top render the rescue construct resistant to CRISPR. *, The Dach^{PAMmut} control was electroporated with a neutral *FoxF>Venus* construct. (E) Photoconversion and lineage tracing of TVC progeny. Nucleus of live B7.5 lineage cells are labelled with *Mesp> nls::Kaede::nls* (green). Nuclear Kaede photoconverted from green (I) to red (II) specifically in the FHPs in a 16hpf larva. The same animal is shown at successive time points (indicated as hours post-fertilization, hpf). I' to IV' show segmented nuclei. Dotted line: midline. (F) *Dach* and *Tbx1/10* is required to limit the proportion of SHP-derived cells forming *Mhc2+* cardiomyocytes in juveniles. Grey: *Mhc2* mRNA visualized by *in situ* hybridization. SHP-derived cells are labelled with *3XT12>H2B::mCherry* (green), B7.5 lineage cells are labelled with *Mesp>nls::LacZ* (red). (G) Corresponding boxplots with the proportions of *Mhc2+* cells among the 3xT12>H2B::mCherry+ SHP-derived cells in juvenile hearts. Scale bar, 10μm

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 6

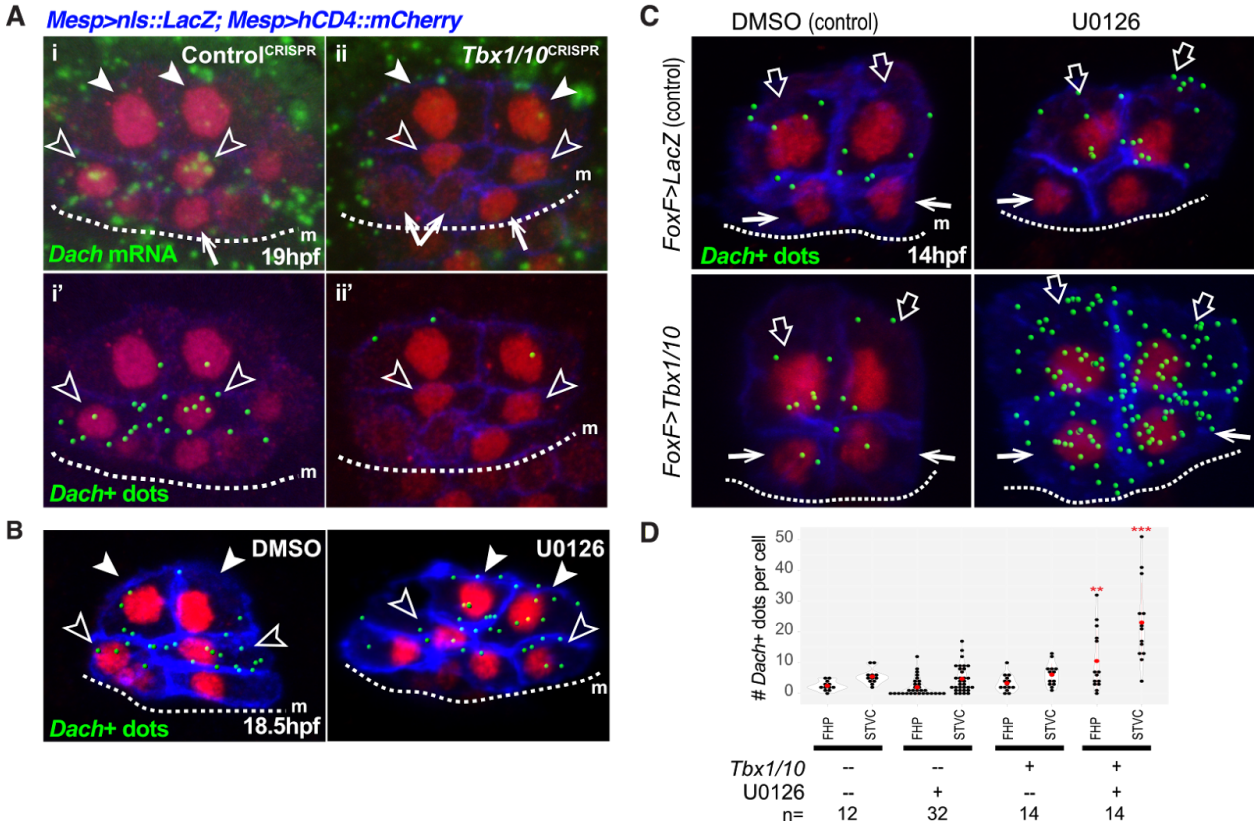


Figure 6. Tbx1/10 and MAPK inhibition promote *Dach* activation in the second heart lineage. (A) Tbx1/10 is required for *Dach* expression in SHP. Dach expression is visualized by *in situ* hybridization. Nuclei of B7.5 lineage cells are labelled by *Mesp>nls::LacZ* and revealed with an anti beta-galactosidase antibody (red). Mesp driven hCD4::mCherry accumulates at the cell membrane as revealed by anti mCherry antibody (red). Solid arrowheads, ASMFs, open arrowheads, SHPs; arrows, FHPs. Anterior to the left, stages indicated in hpf. Scale bar, 10 µm. Upper panels show raw confocal data, and lower panels show segmented FISH signal superimposed onto cell outlines. The green dots reveal the signal from *Dach in situ* hybridization. (i) and (i'), endogenous expression of Dach in SHPs in CRISPR^{Control}. (ii) and (ii'), endogenous *Dach* expression is inhibited upon CRISPR/Cas9-mediated loss of Tbx1/10. Experiment performed in biological replicates. For each replicate, confocal stacks were acquired for 10 larvae in each condition. None of the 20 Tbx1/10 CRISPR larvae showed *Dach* expression in SHPs. (B) FGF/MAPK signaling negatively regulates *Dach* expression in Tbx1/10+ ASMFs. Representative Imaris processed confocal stacks showing *Dach* expression in 18.5hpf larvae, following 3.5 hours treatments with DMSO (control) or U0126 (MEK inhibitor) as indicated. Blocking MEK activity causes ectopic *Dach* expression in the ASMFs (solid arrowheads), in additions to its endogenous expression in the SHPs (open arrowheads). Experiments were performed in biological replicates, and confocal stacks acquired for 10 animals in each condition. All U0126 treated larvae showed ectopic *Dach* expression in ASMFs. (C) Overexpression of Tbx1/10 and blocking of FGF/MAPK induce precocious expression of *Dach* in all of B7.5 lineage cells at 14hpf. Imaris-processed images showing *Dach* expression in

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the TVC-derived cells of 14hpf embryos. Open arrows: STVCs, arrows: FHPs; dotted line: midline (m). (D) Corresponding violin plots representing the counts of *Dach*+ dots per cell. Red dots indicate the mean value in each condition. Vertical red lines indicate mean +/- SD. *** $p<0.05$; ** $p<0.1$. (Student's t-test)

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

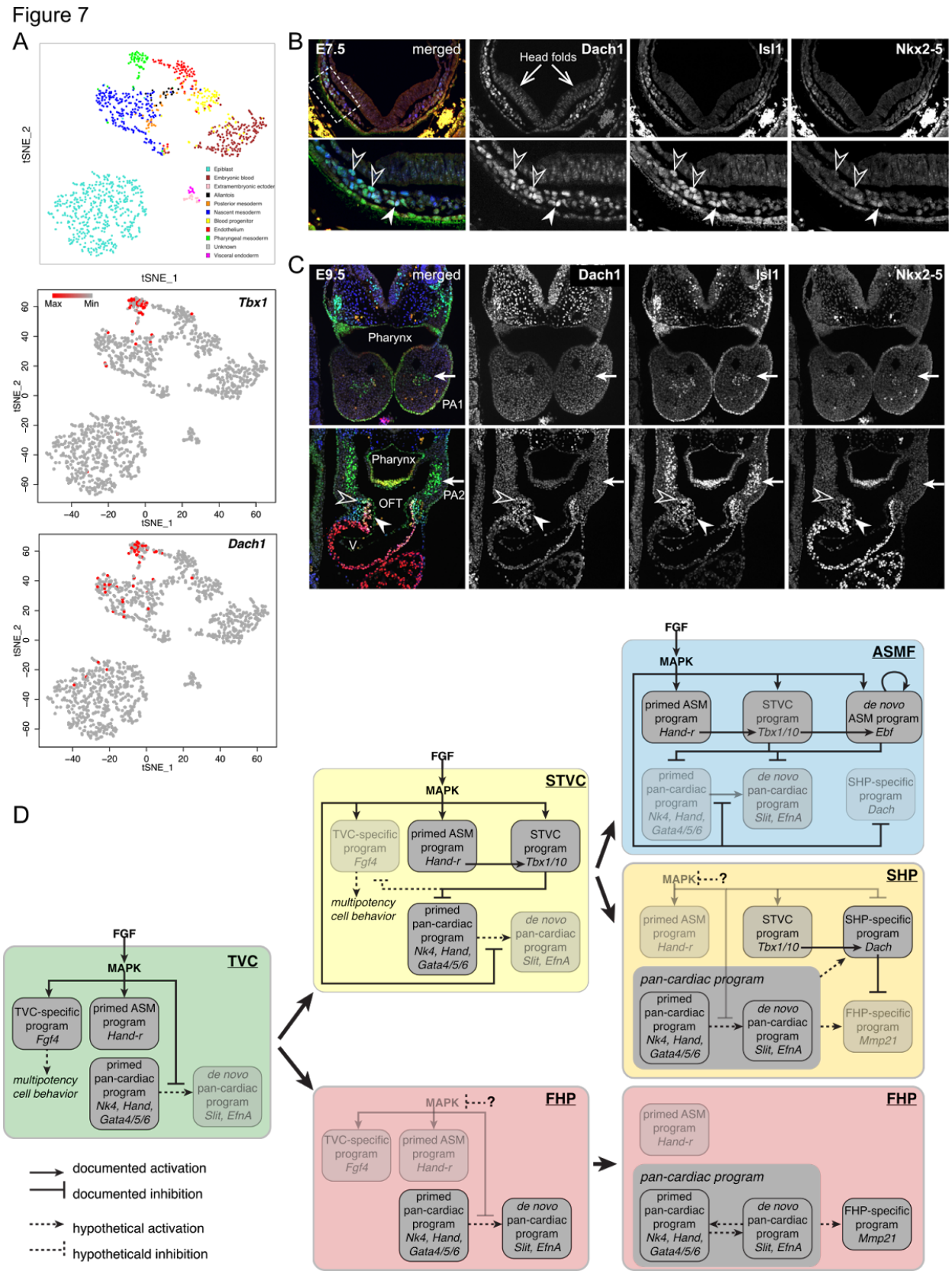


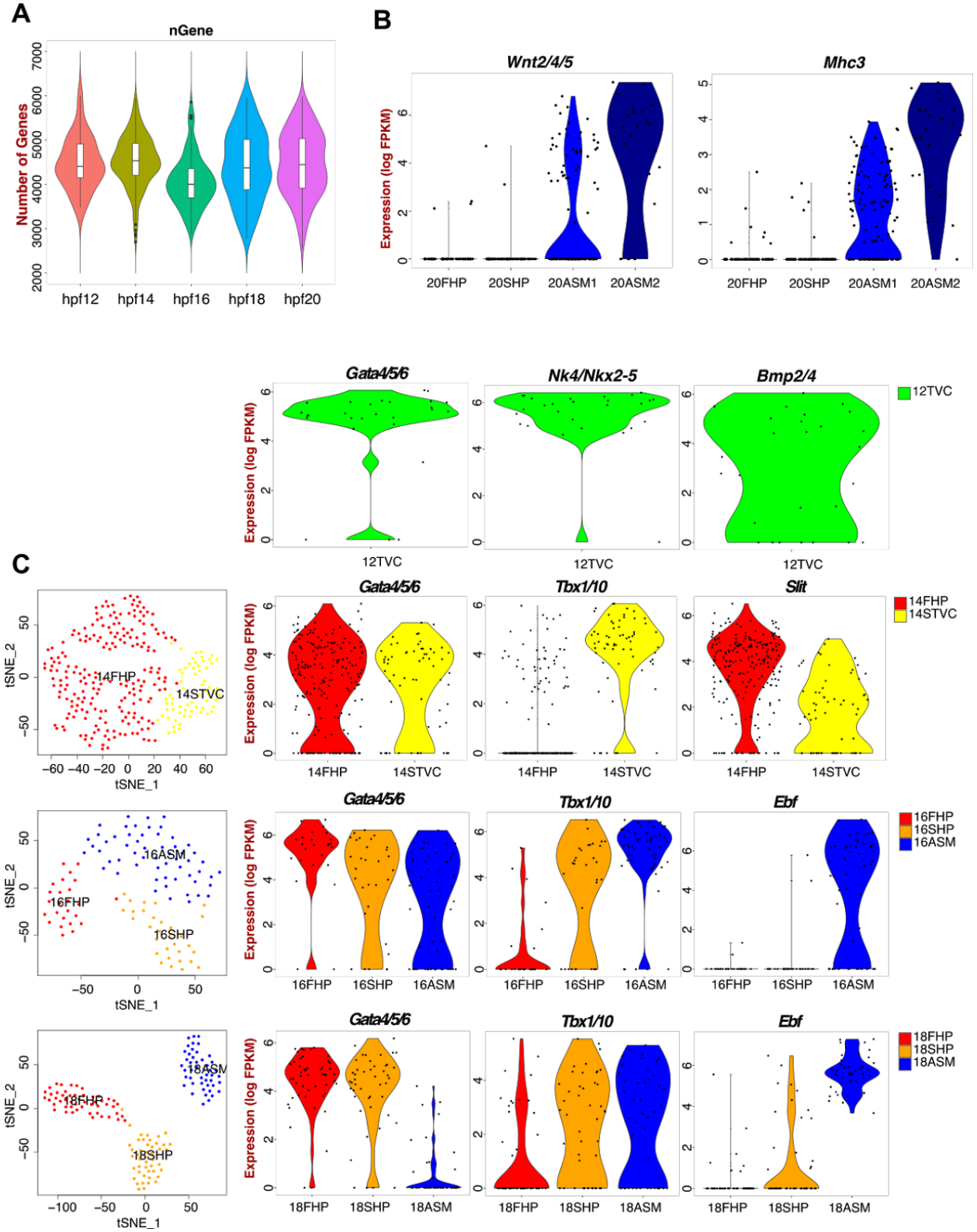
Figure 7. Dach1 expression in the mouse and summary model. (A) t-SNE plots of

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mouse scRNA-seq data after (Scialdone et al., 2016), with expression pattern of *Tbx1* and *Dach1*. Cluster identities are as determined in the original publication, with the pharyngeal mesoderm shown in green. (B) Expression patterns of Dach1, Islet1 and Nkx2.5 proteins in E7.5 embryos as indicated. Bottom panels are close-up views of the region boxed with a dotted line. Arrows: head folds, Open arrowheads: double Dach1+, Islet1+ pharyngeal mesoderm cells, solid arrowhead: nascent Nkx2.5 expression in a Dach1+, Islet1+ cell. (C) Expression patterns of Dach1, Islet1 and Nkx2.5 proteins in E9.5 embryos. Arrows: Islet1+ head muscle progenitor cells in the mesodermal core of the first (PA1, top) and second (PA2, bottom) pharyngeal arches, showing absence of Dach1 and Nkx2.5 expression. Open arrowhead: Dach1+, Islet1+, Nkx2-5- second heart field cells in the dorsal pericardial wall, solid arrowhead: Triple Nkx2.5+, Dach1+, Islet1+ second heart field-derived cells in the outflow tract (OFT). Note the Nkx2.5+, Dach1-, Islet1- cells in the ventricle (V). (D) Summary model showing the maintenance and progressive restriction of FGF/MAPK signaling in the multipotent progenitors (TVCs, trunk ventral cells, and STVCs, second trunk ventral cells) and atrial siphon muscle founder cells (ASMFs). Inhibition of MAPK activity permits the deployment of *de novo*-expressed pan-cardiac genes in both cardiac lineages (FHP, first heart precursors, and SHP, second heart precursors). FHPs specifically activate genes like *Mmp21*, and later produce most *Mhc2*+ cardiomyocytes, whereas SHPs descend from *Tbx1/10*+ multipotent progenitors, and thus activate *Dach*, which contributes to inhibiting the FHP-specific program. See discussion and (Razy-Krajka et al., 2017) for details.

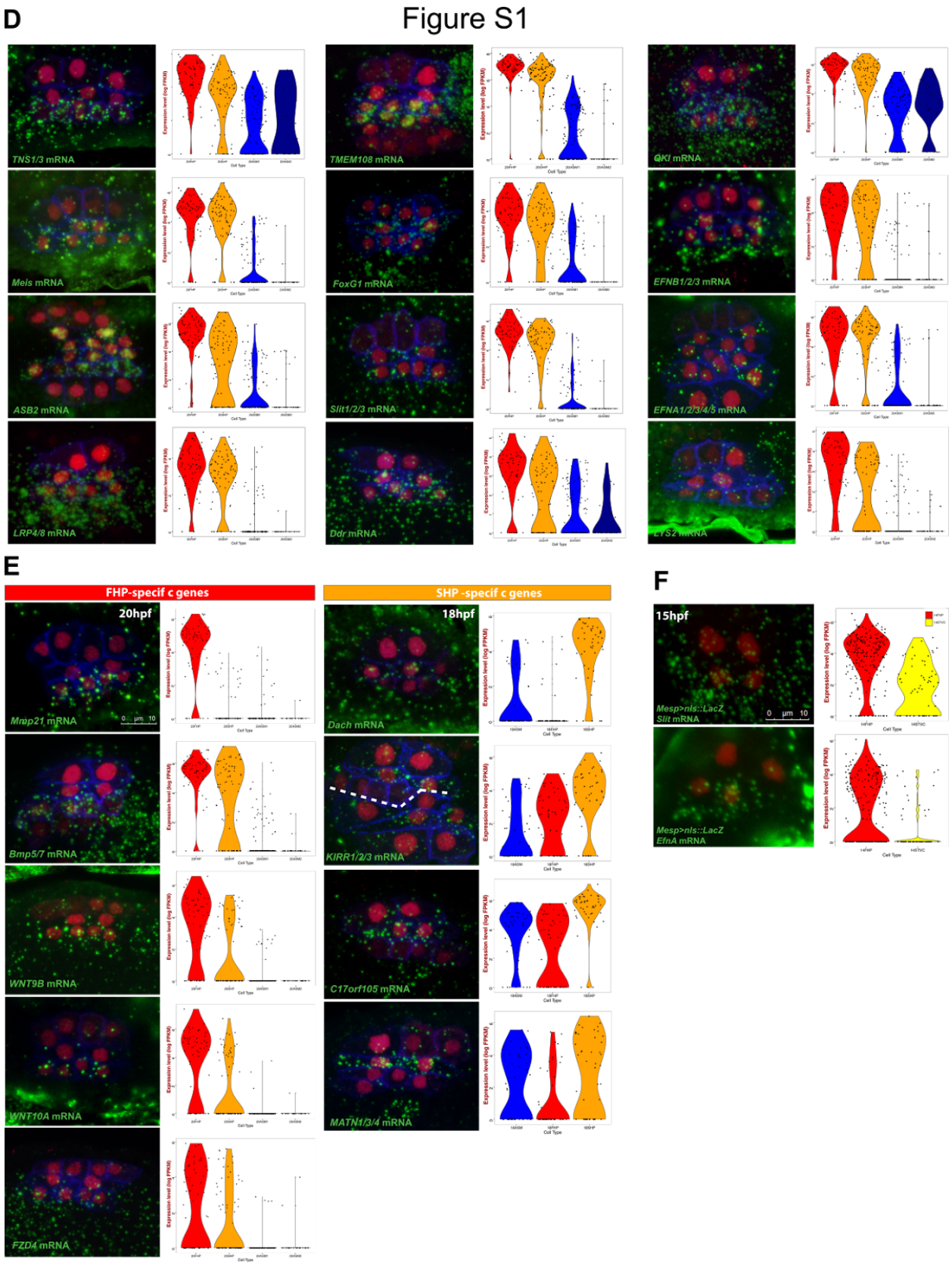
bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S1



bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S1



bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

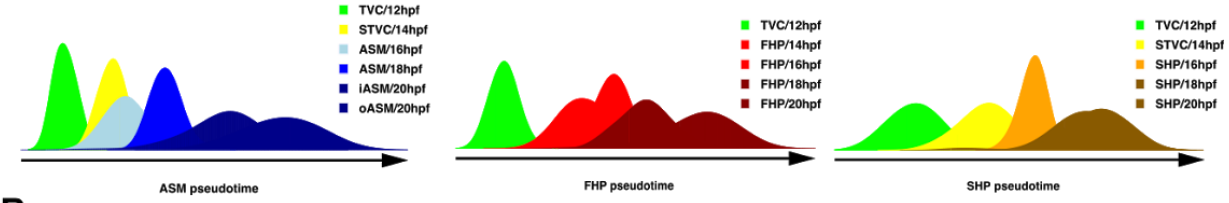
Figure S1

(A) Number of genes detected by scRNA-seq in samples obtained from larvae dissociated at different time points (12, 14, 16, 18, 20 hpf).
(B) Violin plots showing expression of known oASMP markers in 20hpf scRNA-seq dataset.
(C) t-SNE plots showing clusters of individual transcriptomes from 14, 16, and 18hpf samples (left), and violin plots showing expression of key marker genes in defined clusters. See lineage tree Figure 1A for cell identities. (t-SNE is not shown for 12 hpf data because it only contains homogeneous TVC cells).
(D) Violin plots and FISH validation for indicated predicted pan-cardiac genes.
(E) Violin plots and FISH validation for indicated predicted FHP and SHP specific genes.
(F) Violin plots and FISH validation for indicated predicted early FHP and STVC genes from 14hpf scRNA-seq data.

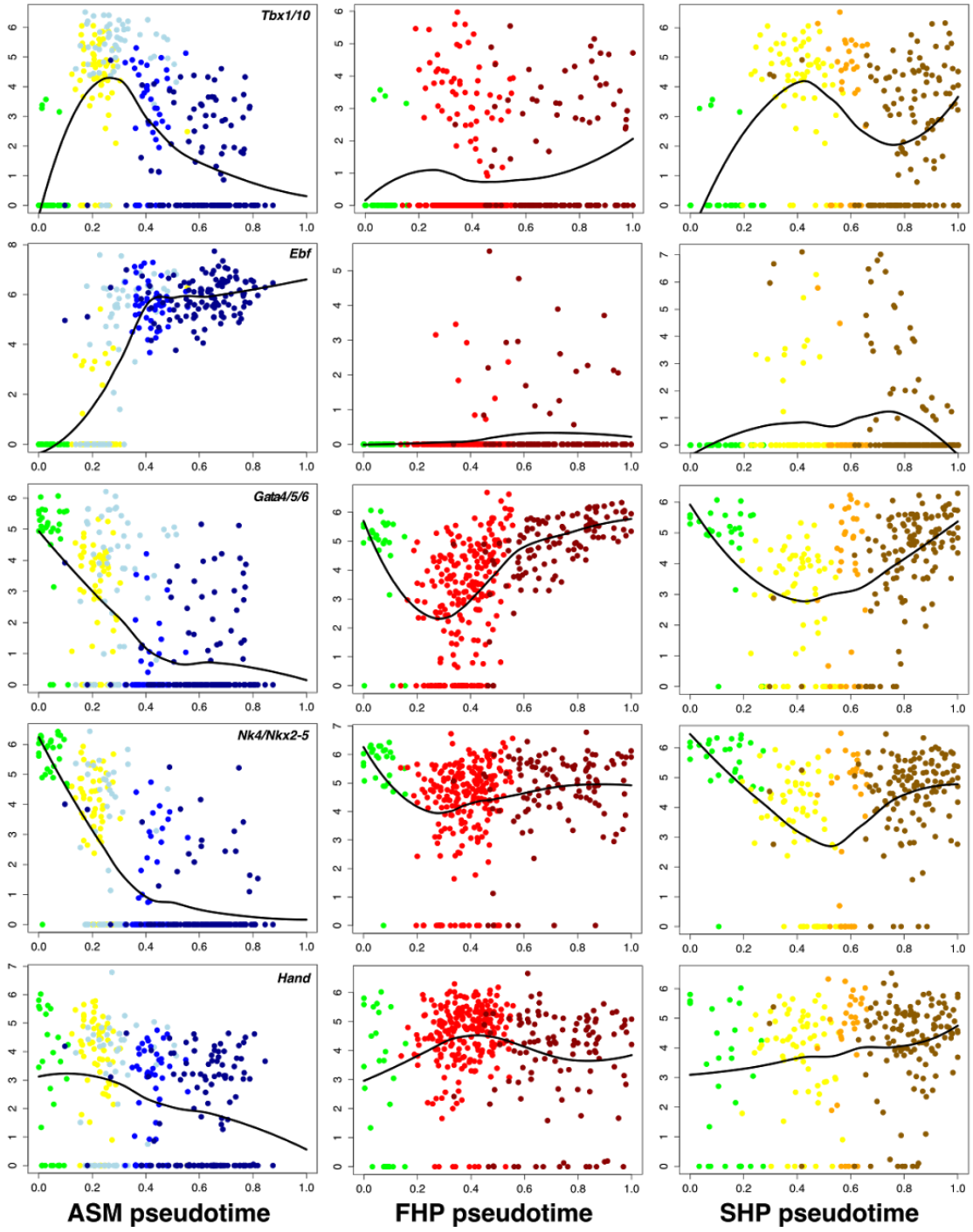
bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S2

A



B



bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S2

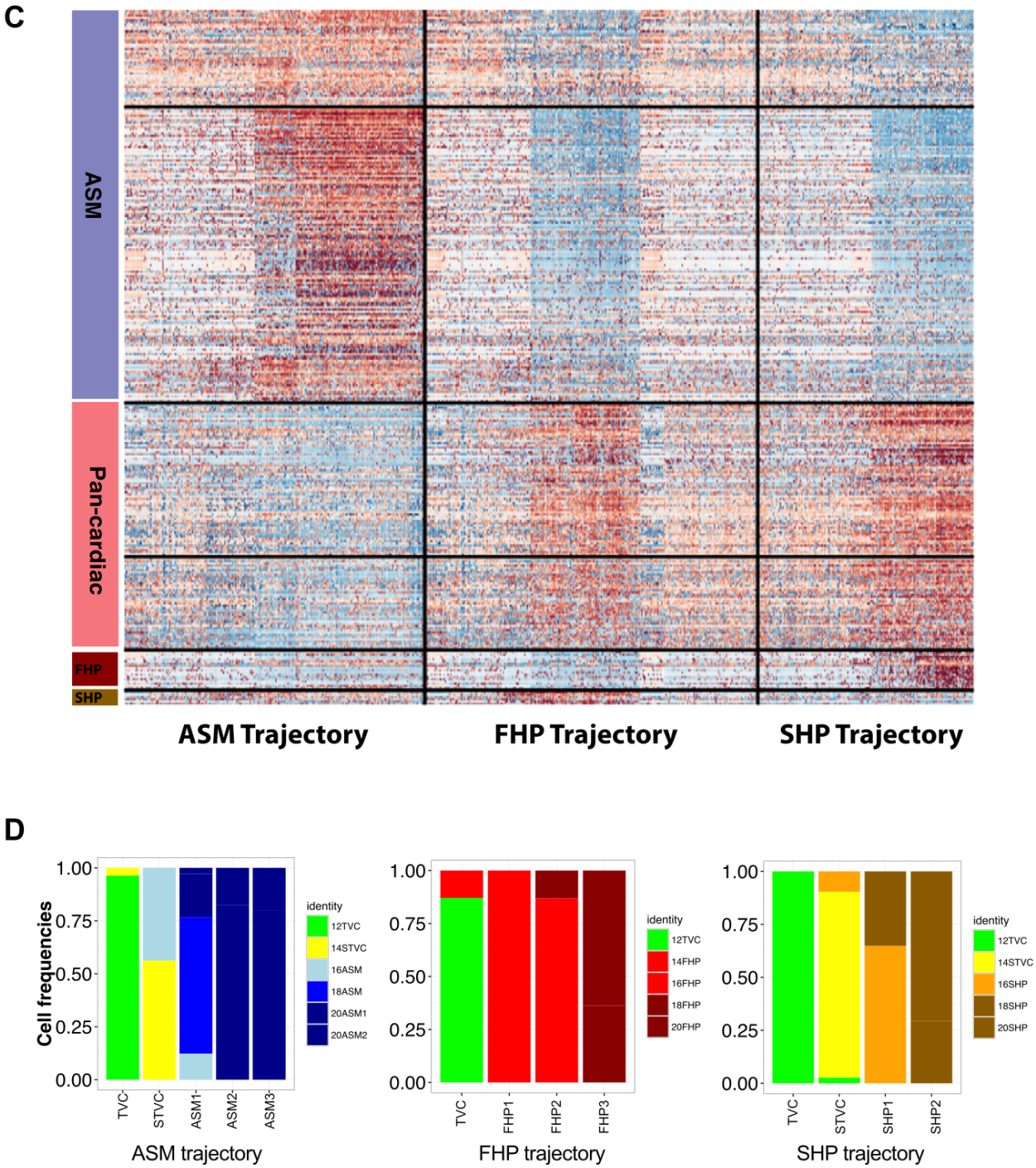


Figure S2

(A) Distribution of defined cell types along the ASM, FHP and SHP trajectories. This shows a good agreement between the time series and developmental progression (see main text for details).

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(B) Pseudotemporal expression profiles of indicated known marker genes along the ASM, FHP and SHP trajectories. Y-axes: expression in log FPKM. Lines indicate the smoothed expression as plotted on the main heatmap in Figure 2C.

(C) Raw scaled expression heatmap corresponding to Figure 2C.

(D) Bar plots showing relative cell type composition for each regulatory states identified on ASM, FHP and SHP trajectories. Note that numbers of 16ASM cells clustering with the 'STVC' state in the ASM trajectory, and indicating that these cells retain most STVC characteristics and have not yet activated the ASM-specific program (see text for details).

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S3.

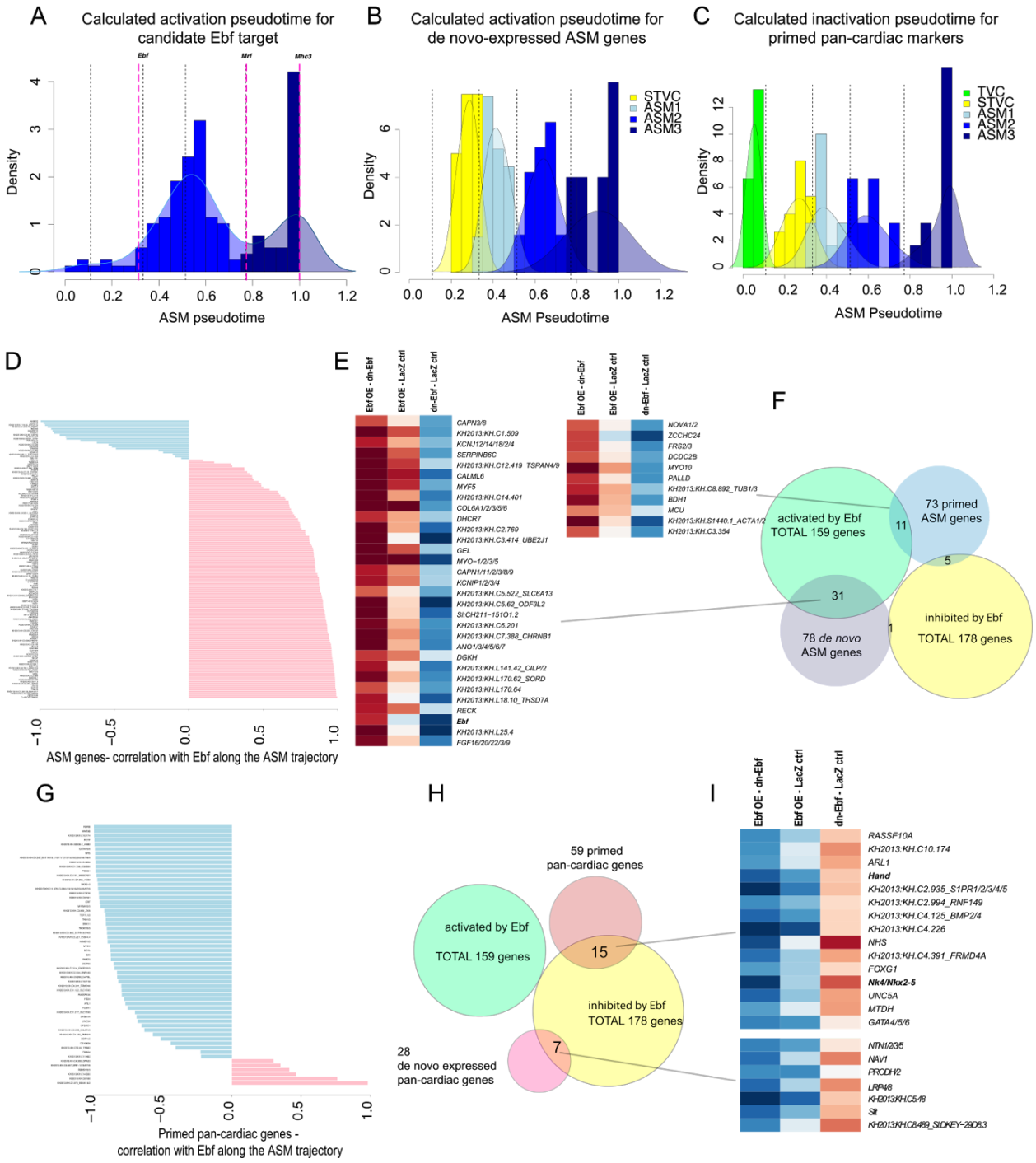


Figure S3

(A) Histogram with density lines showing predicted activation pseudotime of candidate Ebf target genes with corresponding ASM regulatory states (black dashed lines). Purple lines indicate predicted Ebf, Mrf and Mhc3 induction time respectively.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(B) Histogram with density lines showing predicted induction time of *de novo* ASM genes with corresponding ASM regulatory states (black dashed lines).
(C) Histogram with density lines showing predicted inactivation pseudotime of primed cardiac genes with corresponding ASM regulatory states (black dashed lines).
(D) Correlation between candidate ASM markers and Ebf. Genes are ranked with negative to positive correlation from top to bottom.
(E) Heatmaps showing log fold changes of ASM and pan cardiac genes in Ebf over-expression (OE) vs. dominant negative Ebf (dn-Ebf), Ebf OE vs. LacZ (control) and dn-Ebf vs. LacZ conditions.
(F) Corresponding Venn diagram showing the mutual enrichment in candidate ASM genes and candidate targets activated by Ebf.
(G) Anti-correlation between Ebf expression and primed pan-cardiac markers along the ASM trajectory. Genes ranked with low to high correlation from top to bottom.
(H) Corresponding Venn diagram showing the mutual enrichment in candidate pan-cardiac genes and candidate targets inhibited by Ebf.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S4

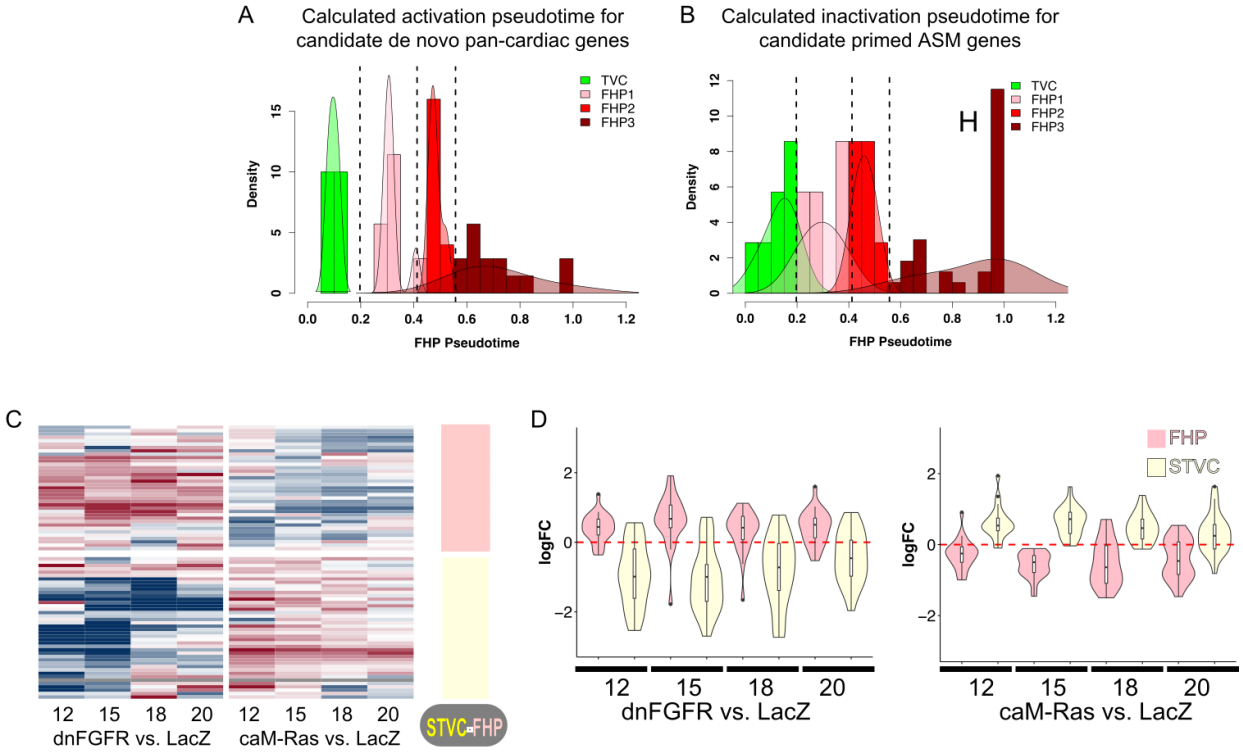


Figure S4

(A) Histogram with density lines showing predicted activation pseudotime of *de novo* cardiac genes along the FHP trajectory, with corresponding regulatory states (black dashed lines).

(B) Histogram with density lines showing predicted inactivation pseudotime of primed ASM genes with corresponding FHP regulatory states (black dashed lines).

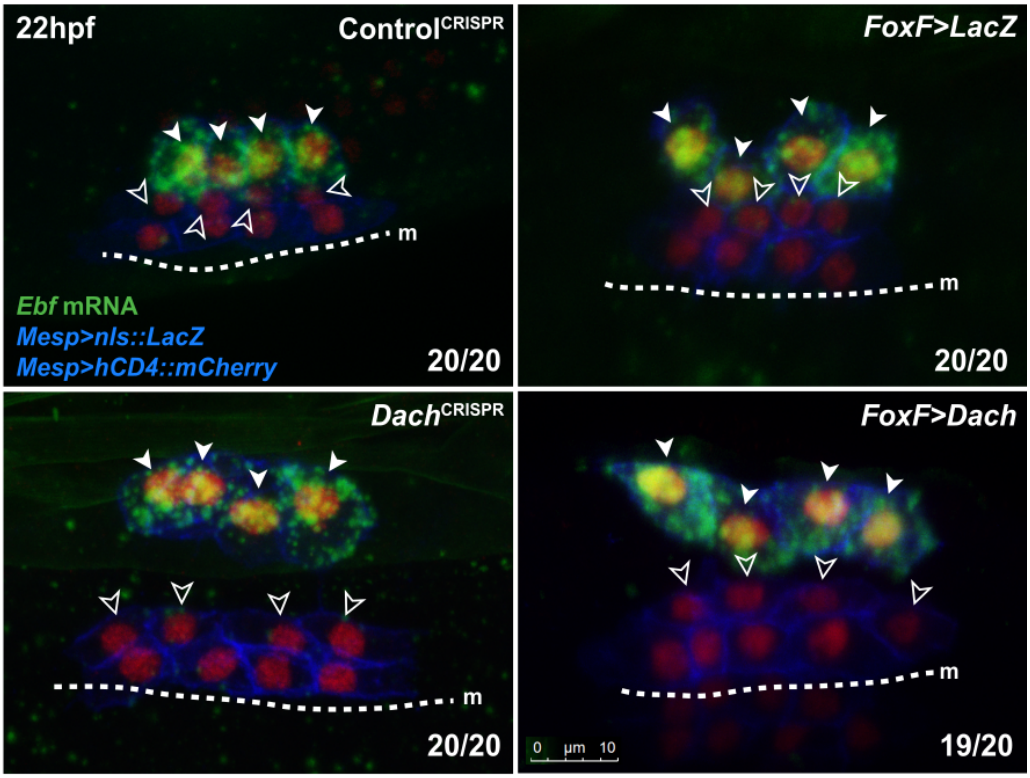
(C) Heatmap showing log fold changes of early FHP and STVC genes from 14 hpf scRNA-seq data in dnFGFR vs LacZ and caM-Ras vs LacZ conditions (12, 15, 18, 20 hpf).

(D) Violin plots showing log fold changes for defined categories of genes (FHP and STVC markers identified in the 14hpf scRNA-seq data), as in Figure 4K.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S5

A



B

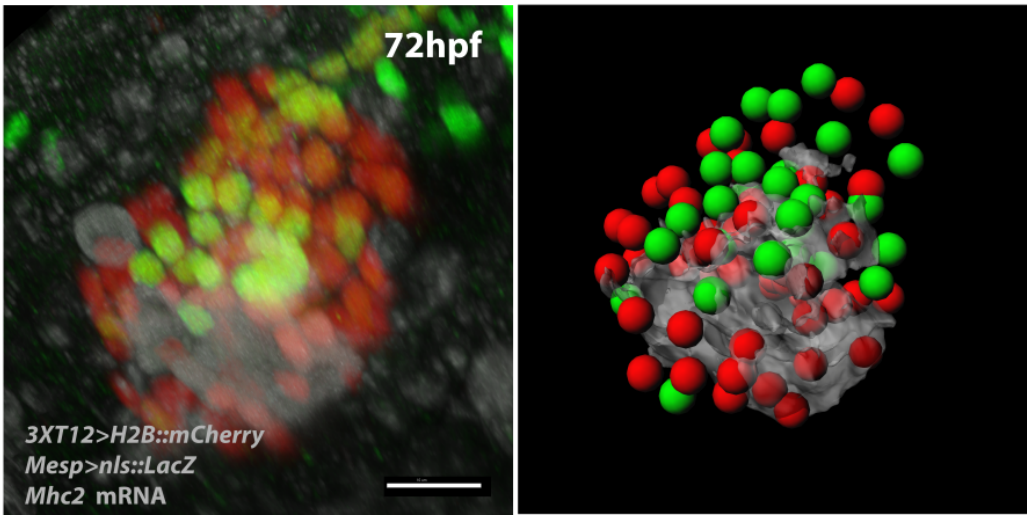


Figure S5

(A) *Ebf* expression specifically in the ASMPs at 22hpf is not altered by perturbations of Dach function. Top panels: controls, bottom panels: CRISPR/Cas9-mediated loss-of-function

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(left) and FoxF TVC enhancer-driven misexpression. Numbers indicate observed/total scored. Solid arrowheads: ASMPs, open arrowheads: SHPs, dotted line: midline (m).

(B) Raw confocal data (left) and segmented image (right) showing Mhc2 expression (grey) primarily in FHP-derived cells marked by the B7.5 lineage marker Mesp>nls::LacZ (red), but not by the STVC-specific marker 3xT12>H2B::mCherry (green), which marks the SHP-derived cells. Scale bar, 10μm.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

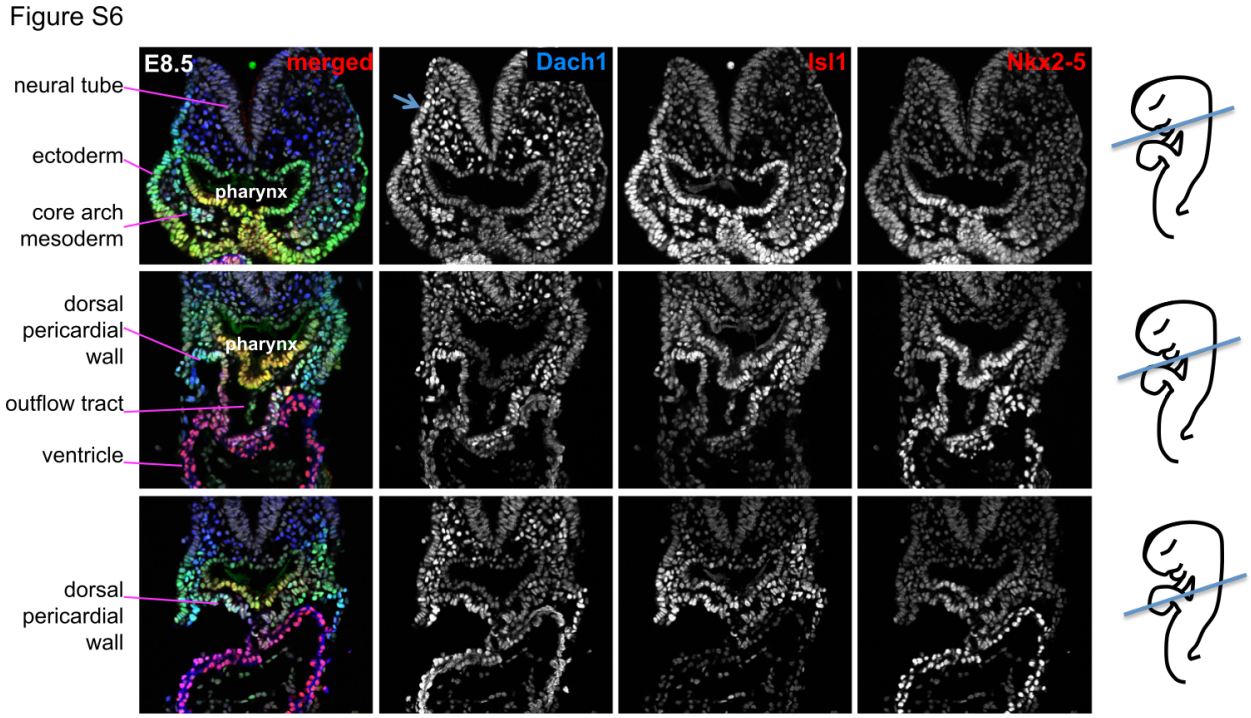


Figure S6

Dach1, Isl1 and Nkx2-5 protein expression detected by immunofluorescence in E8.5 mouse embryos at distinct levels in the pharyngeal mesoderm and developing heart. Note the broad Dach1 expression, which overlaps with Isl1 in second heart field cells in the dorsal pericardial wall.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Abu-Issa, R., Smyth, G., Smoak, I., Yamamura, K.-I., and Meyers, E.N. (2002). Fgf8 is required for pharyngeal arch and cardiovascular development in the mouse. Development *129*, 4613–4625.

Anderson, H.E., and Christiaen, L. (2016). Ciona as a Simple Chordate Model for Heart Development and Regeneration. J Cardiovasc Dev Dis *3*.

Ang, Y.-S., Rivas, R.N., Ribeiro, A.J.S., Srivas, R., Rivera, J., Stone, N.R., Pratt, K., Mohamed, T.M.A., Fu, J.-D., Spencer, C.I., et al. (2016). Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. Cell *167*, 1734–1749.e22.

Barron, M., Gao, M., and Lough, J. (2000). Requirement for BMP and FGF signaling during cardiogenic induction in non-precardiac mesoderm is specific, transient, and cooperative. Dev. Dyn. *218*, 383–393.

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.

Bendall, S.C., Davis, K.L., Amir, E.-A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014). Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. Cell *157*, 714–725.

Brand, T. (2003). Heart development: molecular insights into cardiac specification and early morphogenesis. Dev. Biol. *258*, 1–19.

Brozovic, M., Martin, C., Dantec, C., Dauga, D., Mendez, M., Simion, P., Percher, M., Laporte, B., Scornavacca, C., Di Gregorio, A., et al. (2016). ANISEED 2015: a digital framework for the comparative developmental biology of ascidians. Nucleic Acids Res. *44*, D808–D818.

Buckingham, M., Meilhac, S., and Zaffran, S. (2005). Building the mammalian heart from two sources of myocardial cells. Nat. Rev. Genet. *6*, 826–835.

Cao, Junyue, Jonathan S. Packer, Vijay Ramani, Darren A. Cusanovich, Chau Huynh, Riza Daza, Xiaojie Qiu, et al. (2017). Comprehensive Single Cell Transcriptional Profiling of a Multicellular Organism by Combinatorial Indexing. bioRxiv.

Chan, S.S.-K., Hagen, H.R., Swanson, S.A., Stewart, R., Boll, K.A., Aho, J., Thomson, J.A., and Kyba, M. (2016). Development of Bipotent Cardiac/Skeletal Myogenic Progenitors from MESP1+ Mesoderm. Stem Cell Reports *6*, 26–34.

Chen, L., Fulcoli, F.G., Tang, S., and Baldini, A. (2009). Tbx1 regulates proliferation and differentiation of multipotent heart progenitors. Circ. Res. *105*, 842–851.

Chen, R., Amoui, M., Zhang, Z., and Mardon, G. (1997). Dachshund and Eyes Absent Proteins Form a Complex and Function Synergistically to Induce Ectopic Eye Development in Drosophila. Cell *91*, 893–903.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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. (2009). 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.

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

Davidson, B. (2007). Ciona intestinalis as a model for cardiac development. Semin. Cell Dev. Biol. *18*, 16–26.

Davidson, B., Shi, W., and Levine, M. (2005). Uncoupling heart cell specification and migration in the simple chordate Ciona intestinalis. Development *132*, 4811–4818.

Davidson, B., Shi, W., Beh, J., Christiaen, L., and Levine, M. (2006). FGF signaling delineates the cardiac progenitor field in the simple chordate, Ciona intestinalis. Genes Dev. *20*, 2728–2738.

Davis, T.L., and Rebay, I. (2017). Master regulators in development: Views from the Drosophila retinal determination and mammalian pluripotency gene networks. Dev. Biol. *421*, 93–107.

Davis, R.J., Shen, W., Heanue, T.A., and Mardon, G. (1999). Mouse Dach , a homologue of Drosophila dachshund , is expressed in the developing retina, brain and limbs. Dev. Genes Evol. *209*, 526–536.

Davis, R.J., Shen, W., Sandler, Y.I., Heanue, T.A., and Mardon, G. (2001). Characterization of mouse Dach2, a homologue of Drosophila dachshund. Mech. Dev. *102*, 169–179.

DeLaughter, D.M., Bick, A.G., Wakimoto, H., McKean, D., Gorham, J.M., Kathiriya, I.S., Hinson, J.T., Homsy, J., Gray, J., Pu, W., et al. (2016). Single-Cell Resolution of Temporal Gene Expression during Heart Development. Dev. Cell *39*, 480–490.

Devine, W.P., Wythe, J.D., George, M., Koshiba-Takeuchi, K., and Bruneau, B.G. (2014). Early patterning and specification of cardiac progenitors in gastrulating mesoderm. Elife *3*.

Diogo, R., Kelly, R.G., Christiaen, L., Levine, M., Ziermann, J.M., Molnar, J.L., Noden, D.M., and Tzahor, E. (2015). A new heart for a new head in vertebrate cardiopharyngeal evolution. Nature *520*, 466–473.

Dubois, L., Frendo, J.-L., Chanut-Delalande, H., Crozatier, M., and Vincent, A. (2016). Genetic dissection of the Transcription Factor code controlling serial specification of muscle identities in Drosophila. Elife *5*.

El-Magd, M.A., Saleh, A.A., El-Aziz, R.M.A., and Salama, M.F. (2014a). The effect of RA on the chick Ebf1-3 genes expression in somites and pharyngeal arches. Dev. Genes Evol. *224*, 245–253.

El-Magd, M.A., Saleh, A.A., Farrag, F., Abd El-Aziz, R.M., Ali, H.A., and Salama, M.F. (2014b). Regulation of chick Ebf1-3 gene expression in the pharyngeal arches, cranial sensory ganglia and

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

placodes. Cells Tissues Organs *199*, 278–293.

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.

Gopalakrishnan, S., Comai, G., Sambasivan, R., Francou, A., Kelly, R.G., and Tajbakhsh, S. (2015). A Cranial Mesoderm Origin for Esophagus Striated Muscles. Dev. Cell *34*, 694–704.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Green, Y.S., and Vetter, M.L. (2011). EBF proteins participate in transcriptional regulation of Xenopus muscle development. Dev. Biol. *358*, 240–250.

Grifone, R., and Kelly, R.G. (2007). Heartening news for head muscle development. Trends Genet. *23*, 365–369.

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.

Guo, C., Sun, Y., Zhou, B., Adam, R.M., Li, X., Pu, W.T., Morrow, B.E., Moon, A., and Li, X. (2011). A Tbx1-Six1/Eya1-Fgf8 genetic pathway controls mammalian cardiovascular and craniofacial morphogenesis. J. Clin. Invest. *121*, 1585–1595.

Harel, I., Maezawa, Y., Avraham, R., Rinon, A., Ma, H.-Y., Cross, J.W., Leviatan, N., Hegesh, J., Roy, A., Jacob-Hirsch, J., et al. (2012). Pharyngeal mesoderm regulatory network controls cardiac and head muscle morphogenesis. Proc. Natl. Acad. Sci. U. S. A. *109*, 18839–18844.

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

Heanue, T.A., Davis, R.J., Rowitch, D.H., Kispert, A., McMahon, A.P., Mardon, G., and Tabin, C.J. (2002). Dach1, a vertebrate homologue of Drosophila dachshund, is expressed in the developing eye and ear of both chick and mouse and is regulated independently of Pax and Eya genes. Mech. Dev. *111*, 75–87.

Hirano, T., and Nishida, H. (1997). Developmental fates of larval tissues after metamorphosis in ascidian Halocynthia roretzi. I. Origin of mesodermal tissues of the juvenile. Dev. Biol. *192*, 199–210.

Hotta, K., Mitsuhara, K., Takahashi, H., Inaba, K., Oka, K., Gojobori, T., and Ikeo, K. (2007). A web-based interactive developmental table for the ascidian Ciona intestinalis, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. Dev. Dyn. *236*, 1790–1805.

Hutson, M.R., Zeng, X.L., Kim, A.J., Antoon, E., Harward, S., and Kirby, M.L. (2010). Arterial pole progenitors interpret opposing FGF/BMP signals to proliferate or differentiate. Development *137*, 3001–3011.

Isshiki, T., Pearson, B., Holbrook, S., and Doe, C.Q. (2001). Drosophila neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. Cell

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

106, 511–521.

Kaplan, N., Razy-Krajka, F., and Christiaen, L. (2015). Regulation and evolution of cardiopharyngeal cell identity and behavior: insights from simple chordates. Curr. Opin. Genet. Dev. *32*, 119–128.

Keduka, E., Kaiho, A., Hamada, M., Watanabe-Takano, H., Takano, K., Ogasawara, M., Satou, Y., Satoh, N., and Endo, T. (2009). M-Ras evolved independently of R-Ras and its neural function is conserved between mammalian and ascidian, which lacks classical Ras. Gene *429*, 49–58.

Kelly, R.G., and Papaioannou, V.E. (2007). Visualization of outflow tract development in the absence of Tbx1 using an FgF10 enhancer trap transgene. Dev. Dyn. *236*, 821–828.

Kelly, R.G., Brown, N.A., and Buckingham, M.E. (2001). The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. Dev. Cell *1*, 435–440.

Kelly, R.G., Jerome-Majewska, L.A., and Papaioannou, V.E. (2004). The del22q11.2 candidate gene Tbx1 regulates branchiomeric myogenesis. Hum. Mol. Genet. *13*, 2829–2840.

Kong, P., Racedo, S.E., Macchiarulo, S., Hu, Z., Carpenter, C., Guo, T., Wang, T., Zheng, D., and Morrow, B.E. (2014). Tbx1 is required autonomously for cell survival and fate in the pharyngeal core mesoderm to form the muscles of mastication. Hum. Mol. Genet. *23*, 4215–4231.

Kumar, J.P. (2009). The molecular circuitry governing retinal determination. Biochim. Biophys. Acta *1789*, 306–314.

Lescroart, F., Kelly, R.G., Le Garrec, J.-F., Nicolas, J.-F., Meilhac, S.M., and Buckingham, M. (2010). Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. Development *137*, 3269–3279.

Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolosse, A., Auer, H., Achouri, Y., Dubois, C., Bondue, A., et al. (2014). Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. Nat. Cell Biol. *16*, 829–840.

Lescroart, F., Hamou, W., Francou, A., Théveniau-Ruissy, M., Kelly, R.G., and Buckingham, M. (2015). Clonal analysis reveals a common origin between nonsomite-derived neck muscles and heart myocardium. Proc. Natl. Acad. Sci. U. S. A. *112*, 1446–1451.

Li, G., Xu, A., Sim, S., Priest, J.R., Tian, X., Khan, T., Quertermous, T., Zhou, B., Tsao, P.S., Quake, S.R., et al. (2016). Transcriptomic Profiling Maps Anatomically Patterned Subpopulations among Single Embryonic Cardiac Cells. Dev. Cell *39*, 491–507.

Li, X., Erclik, T., Bertet, C., Chen, Z., Voutev, R., Venkatesh, S., Morante, J., Celik, A., and Desplan, C. (2013). Temporal patterning of Drosophila medulla neuroblasts controls neural fates. Nature *498*, 456–462.

Liao, J., Aggarwal, V.S., Nowotschin, S., Bondarev, A., Lipner, S., and Morrow, B.E. (2008). Identification of downstream genetic pathways of Tbx1 in the second heart field. Dev. Biol. *316*, 524–537.

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Luna-Zurita, L., Stirnimann, C.U., Glatt, S., Kaynak, B.L., Thomas, S., Baudin, F., Samee, M.A.H., He, D., Small, E.M., Mileikovsky, M., et al. (2016). Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. Cell *164*, 999–1014.

Mandal, A., Holowiecki, A., Song, Y.C., and Waxman, J.S. (2017). Wnt signaling balances specification of the cardiac and pharyngeal muscle fields. Mech. Dev. *143*, 32–41.

Marques, S.R., Lee, Y., Poss, K.D., and Yelon, D. (2008). Reiterative roles for FGF signaling in the establishment of size and proportion of the zebrafish heart. Dev. Biol. *321*, 397–406.

Meilhac, S.M., Kelly, R.G., Rocancourt, D., Eloy-Trinquet, S., Nicolas, J.-F., and Buckingham, M.E. (2003). A retrospective clonal analysis of the myocardium reveals two phases of clonal growth in the developing mouse heart. Development *130*, 3877–3889.

Meilhac, S.M., Esner, M., Kelly, R.G., Nicolas, J.-F., and Buckingham, M.E. (2004). The clonal origin of myocardial cells in different regions of the embryonic mouse heart. Dev. Cell *6*, 685–698.

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.

Moris, N., Pina, C., and Arias, A.M. (2016). Transition states and cell fate decisions in epigenetic landscapes. Nat. Rev. Genet. *17*, 693–703.

Mosimann, C., Panáková, D., Werdich, A.A., Musso, G., Burger, A., Lawson, K.L., Carr, L.A., Nevis, K.R., Sabeh, M.K., Zhou, Y., et al. (2015). Chamber identity programs drive early functional partitioning of the heart. Nat. Commun. *6*, 8146.

Nathan, E., Monovich, A., Tirosh-Finkel, L., Harrelson, Z., Rousso, T., Rinon, A., Harel, I., Evans, S.M., and Tzahor, E. (2008). The contribution of Islet1-expressing splanchnic mesoderm cells to distinct branchiomeric muscles reveals significant heterogeneity in head muscle development. Development *135*, 647–657.

Nevis, K., Obregon, P., Walsh, C., Guner-Ataman, B., Burns, C.G., and Burns, C.E. (2013). Tbx1 is required for second heart field proliferation in zebrafish. Dev. Dyn. *242*, 550–559.

Nimmo, R.A., May, G.E., and Enver, T. (2015). Primed and ready: understanding lineage commitment through single cell analysis. Trends Cell Biol. *25*, 459–467.

Nishida, H. (1987). Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. Dev. Biol. *121*, 526–541.

Prall, O.W.J., Menon, M.K., Solloway, M.J., Watanabe, Y., Zaffran, S., Bajolle, F., Biben, C., McBride, J.J., Robertson, B.R., Chaulet, H., et al. (2007). An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. Cell *128*, 947–959.

Racioppi, C., Kamal, A.K., Razy-Krajka, F., Gambardella, G., Zanetti, L., di Bernardo, D., Sanges,

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

R., Christiaen, L.A., and Ristoratore, F. (2014). Fibroblast growth factor signalling controls nervous system patterning and pigment cell formation in Ciona intestinalis. Nat. Commun. *5*, 4830.

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.

Razy-Krajka, F., Gravez, B., and Christiaen, L. (2017). An FGF-driven feed-forward circuit for spatiotemporal patterning of the cardiopharyngeal mesoderm in a simple chordate (bioRxiv).

Reifers, F., Walsh, E.C., Léger, S., Stainier, D.Y., and Brand, M. (2000). Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (fgf8/acerebellar). Development *127*, 225–235.

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

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.

Satou, Y., Imai, K.S., and Satoh, N. (2004). The ascidian Mesp gene specifies heart precursor cells. Development *131*, 2533–2541.

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., Gainous, T.B., Young, J.J., Mori, A., Levine, M., and Christiaen, L. (2010). Early chordate origins of the vertebrate second heart field. Science *329*, 565–568.

Stolfi, A., Lowe, E.K., Racioppi, C., Ristoratore, F., Brown, C.T., Swalla, B.J., and Christiaen, L. (2014a). Divergent mechanisms regulate conserved cardiopharyngeal development and gene expression in distantly related ascidians. Elife *3*, e03728.

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

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119.

Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. Nature *541*, 331–338.

Tintori, S.C., Osborne Nishimura, E., Golden, P., Lieb, J.D., and Goldstein, B. (2016). A Transcriptional Lineage of the Early C. elegans Embryo. Dev. Cell *38*, 430–444.

Tirosh-Finkel, L., Elhanany, H., Rinon, A., and Tzahor, E. (2006). Mesoderm progenitor cells of common origin contribute to the head musculature and the cardiac outflow tract. Development

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

133, 1943–1953.

Tirosh-Finkel, L., Zeisel, A., Brodt-Ivenshitz, M., Shamai, A., Yao, Z., Seger, R., Domany, E., and Tzahor, E. (2010). BMP-mediated inhibition of FGF signaling promotes cardiomyocyte differentiation of anterior heart field progenitors. Development *137*, 2989–3000.

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. (2015). Defining cell types and states with single-cell genomics. Genome Res. *25*, 1491–1498.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Tsuchihashi, T., Maeda, J., Shin, C.H., Ivey, K.N., Black, B.L., Olson, E.N., Yamagishi, H., and Srivastava, D. (2011). Hand2 function in second heart field progenitors is essential for cardiogenesis. Dev. Biol. *351*, 62–69.

Tzahor, E. (2007). Wnt/beta-catenin signaling and cardiogenesis: timing does matter. Dev. Cell *13*, 10–13.

Tzahor, E., and Evans, S.M. (2011). Pharyngeal mesoderm development during embryogenesis: implications for both heart and head myogenesis. Cardiovasc. Res. *91*, 196–202.

Vitelli, F., Morishima, M., Taddei, I., Lindsay, E.A., and Baldini, A. (2002). Tbx1 mutation causes multiple cardiovascular defects and disrupts neural crest and cranial nerve migratory pathways. Hum. Mol. Genet. *11*, 915–922.

Waldo, K.L., Kumiski, D.H., Wallis, K.T., Stadt, H.A., Hutson, M.R., Platt, D.H., and Kirby, M.L. (2001). Conotruncal myocardium arises from a secondary heart field. Development *128*, 3179–3188.

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.

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

van Wijk, B., van den Berg, G., Abu-Issa, R., Barnett, P., van der Velden, S., Schmidt, M., Ruijter, J.M., Kirby, M.L., Moorman, A.F.M., and van den Hoff, M.J.B. (2009). Epicardium and myocardium separate from a common precursor pool by crosstalk between bone morphogenetic protein- and fibroblast growth factor-signaling pathways. Circ. Res. *105*, 431–441.

Woznica, A., Haeussler, M., Starobinska, E., Jemmett, J., Li, Y., Mount, D., and Davidson, B. (2012). Initial deployment of the cardiogenic gene regulatory network in the basal chordate,

bioRxiv preprint doi: https://doi.org/10.1101/150235; this version posted June 15, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ciona intestinalis. Dev. Biol. *368*, 127–139.

Zaffran, S., and Frasch, M. (2002). Early signals in cardiac development. Circ. Res. *91*, 457–469.

Zhang, L., Nomura-Kitabayashi, A., Sultana, N., Cai, W., Cai, X., Moon, A.M., and Cai, C.-L. (2014). Mesodermal Nkx2.5 is necessary and sufficient for early second heart field development. Dev. Biol. *390*, 68–79.

Zhang, Z., Huynh, T., and Baldini, A. (2006). Mesodermal expression of Tbx1 is necessary and sufficient for pharyngeal arch and cardiac outflow tract development. Development *133*, 3587–3595.

Zhou, Y., Cashman, T.J., Nevis, K.R., Obregon, P., Carney, S.A., Liu, Y., Gu, A., Mosimann, C., Sondalle, S., Peterson, R.E., et al. (2011). Latent TGF-β binding protein 3 identifies a second heart field in zebrafish. Nature *474*, 645–648.

Zhou, Z., Wang, J., Guo, C., Chang, W., Zhuang, J., Zhu, P., and Li, X. (2017). Temporally Distinct Six2-Positive Second Heart Field Progenitors Regulate Mammalian Heart Development and Disease. Cell Rep. *18*, 1019–1032.