Combinatorial chromatin dynamics foster accurate cardiopharyngeal fate choices

Claudia Racioppi*, Keira A Wiechecki, and Lionel Christiaen*

Center for Developmental Genetics, Department of Biology, New York University, New York, NY, USA
* Correspondence to : cr1636@nyu.edu (C.R.), lc121@nyu.edu (L.C.)

ABSTRACT

In embryos, lineage-specific profiles of chromatin accessibility control gene expression by modulating transcription, and thus impact multipotent progenitor states and subsequent fate choices. Subsets of cardiac and pharyngeal/head muscles share a common origin in the cardiopharyngeal mesoderm, but the chromatin landscapes that govern multipotent progenitors' competence and early fate choices remain largely elusive. Here, we leveraged the simplicity of the chordate model *Ciona* to profile chromatin accessibility through stereotyped transitions from naive *Mesp*+ mesoderm to distinct fate-restricted heart and pharyngeal muscle precursors. An FGF-Foxf pathway acts in multipotent progenitors to establish cardiopharyngeal-specific patterns of accessibility, which govern later heart vs. pharyngeal muscle-specific expression profiles, demonstrating extensive spatiotemporal decoupling between early cardiopharyngeal enhancer accessibility and late cell-type-specific activity. Combinations of *cis*-regulatory elements with distinct chromatin accessibility profiles are required to activate of *Ebf* and *Tbx1/10*, two key determinants of cardiopharyngeal fate choices. We propose that this higher order combinatorial logic increases the repertoire of regulatory inputs that control gene expression, through either accessibility and/or activity, thus fostering spatially and temporally accurate fate choices.

KEY WORDS: Heart, Pharyngeal Muscle, Head Muscle, Chordate, Transcription, Enhancer, ATAC-seq, CRISPR/Cas9

How a species' genome encodes its diverse and specific biological features has fascinated generations of biologists, and answers regarding the genetic control of body plan, organ, tissue and cell type formation have emerged from steady progress in developmental biology. Cell types arise as cells divide and the progeny of pluripotent embryonic stem cells progress through multipotent and fate-restricted states. The ontogeny of diverse terminal cell identities involves differential expression of hundreds to thousands of genes, whose dynamic activities are orchestrated by complex gene regulatory networks, whereby DNA-binding proteins and co-factors act upon specific *cis*-regulatory elements to control gene expression[1]. Technical and conceptual revolutions in genome biology have extensively characterized the chromatin dynamics that govern the function of these *cis*-regulatory elements[2]. Specifically, as the nuclear genome is packaged in nucleosomes, DNA-binding transcription factors have to compete with histones to interact with *cis*-regulatory elements and control gene expression. Thus, identifying changing landscapes of accessible chromatin governing the transition from multipotent to fate restricted progenitors offers privileged insights into the genomic code for progressive cell type specification.

Dynamic chromatin states underlying cardiomyocyte differentiation have been extensively profiled[3,4] and chromatin state regulation is essential for heart development[5–7]. However, different parts of the heart originate from separate first and second progenitor fields, including those referred to as cardiopharyngeal because they also produce branchiomeric head muscles[8,9]. Bulk and single cell transcription profiling[10,11] have begun to illuminate gene expression changes underlying cardiopharyngeal fate choices, but the corresponding chromatin dynamics remains largely elusive. Moreover, these analyses are partially hindered by the complexity of early vertebrate embryos. By contrast, the tunicate *Ciona* emerged as a simple and powerful chordate model to study early cardiopharyngeal development at high

spatio-temporal resolution[8,12] (Fig. 1A). Building on previous extensive transcription profiles[13,14], including from single cells[15], here we characterize the genome-wide chromatin accessibility dynamics underlying cardiopharyngeal fate specification. We identify regulatory inputs that govern *cis*-regulatory element accessibility and activity, as well as cell-type-specific enhancers for key cardiopharyngeal determinants, and propose models connecting combinatorial chromatin dynamics to cell-specific gene expression and fate choices.

RESULTS

A reference accessome for cardiopharyngeal development

To characterize the chromatin landscape underlying early cardiopharyngeal development, we used the assay for transposon-accessible chromatin (ATAC-seq[16]) on lineage-specific samples isolated at successive time points, and following defined perturbations[15,17] (Fig. 1A). We obtained ~4,000 FACS-purified cells per replicate, collecting samples encompassing early cardiopharyngeal fate transitions, B-line mesenchymal cells and whole embryo cell suspensions (Table S1). We obtained ~500 million unique ATAC-seq reads, with fragment-size distributions showing the characteristic ~150 bp periodicity and patterns of mono-, di- and tri-nucleosomal fragments[16], which were absent in the genomic DNA control (Fig. S1). We identified ATAC-seq peaks using MACS2[18], and generated a combined atlas of 56,090 unique and non-overlapping accessible regions covering 9.25% of the *C. robusta* genome, which we used as our reference "accessome" (Fig. S1). General metrics including peak numbers, size, GC content and genomic distribution were comparable to consensus peaksets reported in similar studies of chromatin accessibility in developmental contexts[19–23] (Supplementary text; Fig. S1).

Next, we annotated the reference accessome by associating accessible regions with other genomic features. In *Ciona*, the transcripts of approximately half of the protein-coding genes undergo spliced-leader (SL) *trans*-splicing, causing the 5' end of mRNAs to differ from the transcription start site (TSS)[24,25]. Using annotated TSSs[26–28], RNA-seq datasets[15], and our ATAC-seq data (Fig. S2), we determined that promoter regions and 5' untranslated regions (5'UTR) were over-represented in the accessome (P<0.01, binomial test; Fig. S1), and we detected nucleosome footprints immediately upstream of TSSs , consistent with a tendency for constitutive accessibility[29–31] (Fig. S2). By contrast, intronic and intergenic regions were significantly under-represented in our reference accessome, compared to the whole genome, although they were the most abundant elements (32.8% and 20.8%, respectively; Fig. S1), suggesting context-specific accessibility, as expected for tissue-specific *cis*-regulatory elements[32]. We associated annotated genes with ATAC-seq peaks located within 10 kb of the TSS or transcription termination site (TTS) (Fig. S3)[33], thus assigning median values of 11 peaks per gene, and 3 genes per peak, owing to the compact *Ciona* genome (Fig. S3). Notably, active regulatory genes encoding transcription factors (TF) and signaling molecules (SM) were associated with significantly more peaks than other expressed genes (P<0.001, binomial test; Fig. S3). This high peak density surrounding regulatory genes is reminiscent of previously described super-enhancers[34] and Clusters of Open *Cis*-Regulatory Elements (CORE) surrounding developmental regulators[35–37].

Cardiopharyngeal accessibility profiles are established in multipotent progenitors

Using this annotated reference accessome, we investigated lineage-specific and dynamic patterns of chromatin accessibility during fate decisions. We observed the greatest contrast in accessibility between the B7.5 and B-line mesenchyme lineages, whereas biological replicates correlated most highly (Spearman's ρ>0.93), indicating reproducible detection of extensive lineage-specific accessibility (Fig. 1B). Within the B7.5 lineage, correlation analysis suggested that most changes occur between 6 and 10 hpf, during the transition from naive *Mesp*+ mesoderm to multipotent cardiopharyngeal progenitors (aka trunk ventral cells, TVCs). Higher correlation between multipotent progenitors and mixed heart and pharyngeal muscle precursors, obtained from 18 hpf larvae, suggested more stable accessibility profiles

during and immediately following early cardiopharyngeal fate choices (Fig. 1B). Consistent with correlation analyses, most significant temporal changes in accessibility occurred during the transition from naive *Mesp*+ mesoderm to multipotent progenitors (5,450 peaks, FDR<0.05; Fig. 1C). Specifically, about two thirds (64.7%, 3,525/5,450) of these regions showed reduced accessibility at 10 hpf, in multipotent progenitors, compared to 6 hpf naive *Mesp*+ mesoderm (Fig. 1C). Conversely, 1,252 regions appeared to open between 6 and 10 hpf or later (Fig. 1C), and 38.8% (486/1,252) of these regions were more accessible in the B7.5-lineage compared to the mesenchyme (Fig. 1C). Moreover, the subset of regions opening between 6 and 10 hpf or later was enriched in genomic elements associated with cardiopharyngeal markers, including primed pan-cardiac and pharyngeal muscle markers, while elements flanking tail muscle markers (anterior tail muscle, aka ATMs) as well as multipotent progenitor genes that are downregulated after 12 hpf, were predominantly closing between 6 and 10 to 18 hpf (Fig. 1D). Taken together, these observations suggest that cardiopharyngeal accessibility profiles are established specifically in the B7.5 lineage, upon induction of multipotent progenitors, and persist in fate-restricted cells.

To further analyze changes in accessibility associated with multipotent progenitor induction, we performed ATAC-seq on B7.5 lineage cells isolated at 10 hpf following defined perturbations of FGF-MAPK signaling, a constitutively active form of Mek ($Mek^{S216D,S220E}$), which converts all B7.5 lineage cells into multipotent cardiopharyngeal progenitors or a dominant negative Fgf receptor ($Fgfr^{DN}$), which blocks induction and transforms all cells into ATMs[17,38] (Fig. 1A). We repurposed DESeq2[39] to analyze differential accessibility using the reference accessome, and identified 2,728 and 2,491 regions that either closed or opened following inhibition and activation of FGF-MAPK signaling, respectively (Figs. 1E,G, S4). Using peak-to-gene annotations, we cross-referenced ATAC-seq with expression microarray data obtained from B7.5 lineage cells expressing the same $Fgfr^{DN}$[13], and observed a positive correlation between changes in differential accessibility and differential gene expression at 10 hpf (Spearman's $\rho = 0.47$; Figs. 1E, S5). Specifically, 48% of differentially FGF-MAPK-regulated genes were associated with at least one element showing consistent differential accessibility, including 260 candidate FGF-MAPK-regulated TVC markers associated with 557 regions predicted to open specifically in multipotent cardiopharyngeal progenitors at 10 hpf (Table S2, Fig. S5). Conversely, the majority of ATAC-seq peaks associated with FGF-MAPK-inhibited tail muscles markers were also more accessible upon inhibition of FGF signaling (Table S2; Fig. S4). Taken together, these observations indicate that cardiopharyngeal accessibility profiles are established in the multipotent progenitors by opening regions associated with genes upregulated upon induction by FGF-MAPK signaling.

These results open the possibility that differentially accessible elements act as tissue-specific enhancers. Consistent with this hypothesis, FGF-MAPK-dependent and cardiopharyngeal-specific elements were predominantly found in intronic or intergenic regions (48% and 37%, respectively, FDR<0.05, Hypergeometric test), whereas tissue-specific peaks associated to tail muscle markers were enriched in promoters, TSS and 5'UTR (Hypergeometric test, FDR<0.05, 57%, 23% and 15%, respectively) (Fig. S5). Previously characterized enhancers for TVC-specific genes *Lgr4/5/6*, *Rhodf*, *Foxf*, *Unc5*, *Rgs21, Ddr, Asb2* and *Gata4/5/6*[13,40–42] showed ATAC-seq patterns consistent with cardiopharyngeal-specific accessibility (Fig. 1F, Table S3). We thus leveraged differential accessibility profiles to identify novel enhancers for cardiopharyngeal gene expression. We focused on a locus containing the conserved cardiac determinant and TVC marker, *Nk4/Nkx2-5*[43], and two tail-muscle specific *Myosin regulatory light chain* (*Mrlc*) genes[44–46], with associated elements showing the predicted TVC- and ATM-specific accessibility patterns, respectively (Fig. 1G). Reporter gene expression assays showed that a DNA fragment containing differentially accessible elements located in the *Nk4/Nkx2-5* intron (*KhC8.2200* and *.2001*) was sufficient to drive *GFP* expression in cardiopharyngeal progenitors (Fig. 1I); while B7.5 lineage-specific CRISPR/Cas9-induced deletions of these elements reduced or eliminated *Nk4/Nkx2-5* expression,

demonstrating its role as a *bona fide* cardiopharyngeal enhancer (Fig. 1J). Extending these analyses to other loci, including *Fgf4*, *Fzd4*, *Foxg-r*, *Fbln*, *Eph1, Ncaph, Zan, Tmtc2, Hand* and *Smurf1/2*, 10 out 15 candidate cardiopharyngeal enhancers drove reporter expression specifically in multipotent progenitors (Fig. S6, Table S4), and B7.5-lineage-specific CRISPR/Cas9-mediated mutagenesis targeting differentially accessible elements reduced TVC-specific expression of the neighbouring genes *Fgf4*, *Smurf1/2* and *Fbln* (Fig. S7, Table S5). Conversely, candidate atrial tail muscle (aka ATM)-specific elements activated reporter gene expression in the tail muscles, including ATM cells, but not in the cardiopharyngeal progenitors, and were located near tail muscle markers (Fig. S4, Table S6). Collectively, these findings indicate that genomic elements that open specifically in multipotent progenitors upon FGF-MAPK induction act as transcriptional enhancers of cardiopharyngeal gene expression, illustrating the predictive power of our differential accessibility analysis for functional annotations.

A Foxf-dependent code for cardiopharyngeal accessibility

Next, we harnessed chromatin accessibility patterns predictive of cardiopharyngeal enhancer activity to identify enriched sequence motifs, and thus candidate regulators of chromatin accessibility and gene expression. To this aim, we repurposed chromVAR[47], which was developed to analyze sequence motifs associated with cell-type-specific accessibility in single cell ATAC-seq data (Fig. 2A). Naive *Mesp*+ mesoderm-specific elements, which closed between 6 and 10 hpf, were enriched in motifs for Homeodomain, T-box and Ets families of transcription factors (TF), consistent with documented roles for Lhx3/4, Tbx6 and Ets homologs in B7.5 blastomeres[38,48,49] (Fig. 2A). Candidate tail muscle-specific elements, which opened upon $Fgfr^{DN}$ misexpression, were similar to naive *Mesp*+ mesoderm, and enriched in motifs for the basic helix-loop-helix (bHLH) family of TFs, which includes Mesp and Mrf/MyoD, a conserved muscle-specific transcription regulator that promotes tail muscle differentiation[13,14,50,51]. By contrast, motifs for GATA, Forkhead and nuclear receptor families of TFs were enriched among candidate cardiopharyngeal-specific elements, revealing a typical mesendodermal signature for early cardiopharyngeal progenitors[52].

Combined with motif enrichment analyses, temporal gene expression profiles[14] identified candidate *trans*-acting regulators of cardiopharyngeal-specific accessibility and/or activity (Fig. 2B). For example, besides consistent *Mrf/MyoD*, *Lhx3/4*, *Ets1/2*, and *Tbx6* expression and motif enrichment associated with tail muscle- or naive *Mesp*+ mesoderm specific accessibility, the potential association between K50 Paired homeodomain proteins - Otx, Pitx and Gsc - and founder cell-specific accessibility is consistent with and suggests a possible role for early *Otx* expression in B7.5 blastomeres[53] (Fig. 2B). Cardiopharyngeal-specific enrichment for Fox/Forkhead and GATA motifs pointed to several known factors, including *Foxf*, one of the first gene activated in multipotent progenitors upon induction by FGF-MAPK, prior to *Gata4/5/6*[13,42,54]. GATA and Forkhead proteins are founding members of a group of TFs known as pioneers, which can bind their target sites in closed chromatin and promote accessibility[55,56]. Protein sequence alignments indicated the presence of key residues, conserved between the DNA binding domains of Foxf and the classic pioneer FOXA, which mimic linker histone H1 in its ability to displace DNA-bound nucleosomes[57] (Fig. S8). Finally, the *Foxf* enhancer was accessible in naive *Mesp*+ founder cells, suggesting that it can be activated without further requirement for pioneering activity, unlike the intronic *Gata4/5/6* enhancer (Fig. S9). Consistent with a role for Fox and GATA proteins in opening and activating the *Nk4/Nkx2-5* enhancer, we found putative cognate binding sites in the newly identified element in a region of >65% sequence identity with the closely related species, *C. savignyi* (Fig. S10). Taken together, these analyses identified a putative code for cardiopharyngeal-specific accessibility and enhancer activity, which comprise motifs for candidate DNA binding factors of the Forkhead, GATA, and nuclear receptor families, and identified *Foxf* as candidate determinant of cardiopharyngeal accessibility.

To test a role for Foxf in establishing cardiopharyngeal accessibility and gene expression profiles, we used reagents for B7.5 lineage-specific loss of function by CRISPR/Cas9-mediated mutagenesis (Foxf^{CRISPR,58}), and performed ATAC- and RNA-seq on FACS-purified cells isolated from tailbud embryos at 10 hpf. RNA-seq confirmed that CRISPR/Cas9 mutagenesis inhibited *Foxf* itself, and other TVC-expressed genes, including effectors of collective cell migration such as *Ddr*, consistent with previous microarray data[13,41] (Fig. 2C, Fig. S9). Out of 52 differentially expressed genes (Table S7), seven down-regulated genes were previously annotated as primed pan-cardiac markers[15], including *Hand*, *Gata4/5/6* and *Fzd4*. Down-regulated genes also included primed pharyngeal muscle markers, such as *Rhod/f*[3] (Fig. S9), suggesting that Foxf promotes the onset of both the cardiac and pharyngeal muscle programs in multipotent progenitors, a feature known as multilineage transcriptional priming[14,15].

Consistent with the effects of Foxf mutagenesis on gene expression, regions closed in Foxf^{CRISPR} samples included known cardiopharyngeal enhancers for *Gata4/5/6* and *Ddr*[13,40,41] (Fig. 2D, Fig. S11), newly identified enhancers for *Eph1, Smurf1/2* and *Fzd4*, and a novel enhancer of *Hand* expression (*Hand_KhC14.805 - .807*) (Figs. 2E-F, S12, Table S8). These differentially accessible elements contain several, evolutionary conserved, putative Fox binding sites (Figs. S12, S13). Notably, 98% of the regions that closed upon Foxf inhibition and were located near a down-regulated gene, also showed the ATAC-seq profile for cardiopharyngeal-specific accessibility (Figs. 2D, S9). Conversely, 22% (600/2,728) of the predicted cardiopharyngeal-specific elements were closed upon Foxf inhibition, and gene set enrichment analysis further indicated that *Foxf* loss-of-function generally decreased the accessibility of cardiopharyngeal-specific elements (Fig. S9). Finally, 51% of the Foxf-dependent elements associated with candidate Foxf targets were closed in 6 hpf founder cells and appear to open specifically in the cardiopharyngeal progenitors by 10 hpf (Fig. S14; Table S9). This dynamics is consistent with a requirement for Foxf activity following its activation specifically in the TVCs, immediately after cell division of the naive *Mesp*+ progenitors. Taken together, these results indicate that, in newborn multipotent progenitors, FGF-MAPK signaling upregulates *Foxf*[3,42], which is then required to open a substantial fraction of cardiopharyngeal-specific elements for gene expression in multipotent progenitors, including for such essential determinants as *Gata4/5/6* and *Hand*.

Chromatin accessibility in late heart vs. pharyngeal muscle precursors

Besides controlling coherent chromatin opening, enhancer activity and gene expression in multipotent cardiopharyngeal progenitors, FGF-Foxf inputs also appeared to open regions associated with later de novo-expressed heart and pharyngeal muscle markers[14,15,43,59] (Fig. 1E; Table S10). Accessibility patterns were also better correlated between 10 and 18 hpf (Fig. 1B), suggesting a decoupling between early accessibility and late heart- vs. pharyngeal muscle-specific expression in late fate-restricted precursors. To identify accessibility patterns underlying the heart vs. pharyngeal muscle fate choices, we compared bulk RNA-seq[15] and ATAC-seq datasets obtained from cardiopharyngeal lineage cells isolated from 18 hpf larvae, following the same defined perturbations of FGF-MAPK signaling[38,59,60] (Figs. 1A, 3A, S15, S16). Among cardiac and pharyngeal muscle markers, we identified 35 FGF-MAPK-regulated genes associated with one or more element showing consistent differential accessibility (Figs. 3A, S21, Table S11), suggesting that, at least for a subset of cardiopharyngeal marker genes, FGF-MAPK-dependent changes in gene expression follow corresponding changes in chromatin accessibility in early heart and pharyngeal muscle precursors.

Gene-level inspection of differential accessibility associated with either inhibition or activation of gene expression revealed that only a fraction of associated elements were either closing or opening upon perturbation of FGF-MAPK signaling (Fig. S16). For example, the first heart lineage marker *Matrix metalloproteinase 21/Mmp21*[15] was associated with multiple elements located upstream and in introns, but only some of these elements were differentially accessible following either gain or loss of FGF-MAPK

function (Fig. S16), and a 3kb fragment containing the upstream differentially accessible element sufficed to enhance transcription throughout the cardiopharyngeal lineage, but not specifically in the first heart precursors (Fig. S16), suggesting that cardiac-specific accessibility determines tissue-specific gene expression.

Remarkably, the vast majority (91%, 356/391, Table S12) of differentially expressed genes was not associated with differentially accessible elements (Figs. 3A, S16). Specifically, out of 30 de novo-expressed pan-cardiac genes that were also differentially expressed upon FGF-MAPK perturbation at 18hpf, only 8 (27%±8%, SE) were associated with one differentially accessible element following perturbation of FGF-MAPK signaling (Fig. 3B). This suggested that most differential gene expression in early heart and pharyngeal muscle precursors arise from differential *cis*-regulatory activity of elements that are otherwise accessible throughout the cardiopharyngeal mesoderm. In keeping with this hypothesis, accessible regions associated with de novo expressed pan-cardiac markers tended to open between 6 and 10 hpf, in a pattern consistent with FGF- and Foxf-dependent cardiopharyngeal-specific accessibility (Fig. 3C, Fig. S17). These observations suggested that *cis*-regulatory elements controlling cell-type-specific de novo gene expression open in multipotent progenitors prior to becoming active in fate-restricted precursors. Such decoupling between enhancer accessibility and activity has been observed in other developmental contexts, including early cardiogenesis in mammals[3,4].

As a proof of principle, we analyzed the *Lrp4/8* locus, which harbors two intronic elements (*KhC4.137* and *KhC4.144*) that opened upon TVC induction in an FGF- and Foxf-dependent manner, prior to *Lrp4/8* activation in cardiac progenitors[15], and were not differentially accessible at 18 hpf (Fig. 3C,D). Reporter gene expression and CRISPR/Cas9-mediated mutagenesis assays followed by FISH indicated that *KhC4.137* is both necessary and sufficient to activate gene expression in heart precursors (Fig. 3E,F), showing that it acts as a *bona fide* enhancer, and demonstrating a specific case of decoupling between early and broad accessibility and late, cell-type-specific, activity.

To identify candidate regulators of late accessibility and activity, we parsed accessible elements associated with de novo expressed heart and pharyngeal muscle markers into pre-accessible/primed or de novo accessible elements, and discovered sequence motifs enriched in each category (Fig. 4A, Table S13). As expected, Forkhead motifs were overrepresented in primed elements associated with both de novo-expressed cardiac and pharyngeal muscle genes. Putative binding sites for SMAD and K50 Paired homeodomain proteins - such as Otx,Pitx, Crx and Gsc (Fig. S18)- were enriched among pre-accessible elements associated with cardiac markers, and found in the primed elements regulating de novo expression of *Lrp4/8*, suggesting a specific role in transcriptional activation, consistent with conserved roles for Pitx2 and BMP-SMAD signaling during heart development[61,62] (Fig. S18). Motifs for known regulators of cardiac development, including Meis[3,63], were over-represented among de novo accessible elements associated to cardiac markers, suggesting roles in establishing accessibility and/or regulating enhancer activity (Fig. S18). Notably, GATA motifs were enriched among both primed and de novo accessible elements associated with cardiac markers, consistent with conserved roles for GATA factors as pioneer factors, and during cardiac development[64]. Among motifs enriched in accessible elements associated with de novo-expressed pharyngeal muscle markers, the presence of ETS-, TBX-, bHLH/Mrf-, CSL- and EBF-family motifs is consistent with established roles for FGF-MAPK, Hand-r, Tbx1/10, Mrf, Ebf and Notch signaling in pharyngeal muscle specification[14,17,43,51,65,66]. Of note, the enrichment of EBF motifs among de novo accessible elements associated with de novo expressed pharyngeal muscle markers is reminiscent of the ability of EBF-family factors to interact with nucleosome-bound cognate sites, suggestive of a pioneering activity in committed pharyngeal muscle precursors[16]. In summary, this analysis identified distinct combinations of established and putative *trans*-acting factors differentially controlling the accessibility and/or activity of *cis*-regulatory elements that govern heart- vs. pharyngeal-muscle-specific gene expression.

Combinatorial *cis*-regulatory control of cardiopharyngeal determinants

The above analyses focused on one-to-one associations between accessible elements and neighboring genes to uncover candidate *trans*-acting inputs controlling gene expression through defined elements. However, most genes are associated with multiple accessible regions, especially developmental control genes (Fig. S3), suggesting that individual genes may respond to a variety of regulatory inputs mediated through separate *cis*-regulatory elements. For instance, the loci of many de novo-expressed heart and pharyngeal muscle markers contained both primed cardiopharyngeal-specific and de novo cell-type-specific accessible elements (Fig. S18). As a proof of concept, we focused on *Tbx1/10* and *Ebf*, two established determinants of cardiopharyngeal fates[14,15,17,43,65]. Both loci contained multiple accessible regions, including elements open since the naive *Mesp*+ mesoderm state (e.g. *Ebf_KhL24.35/36*, Figs. 4B, S19), cardiopharyngeal-lineage-specific elements that open prior to gene activation (e.g. *Ebf_KhL24.34*, Fig. 4A) and elements that open de novo in fate-restricted pharyngeal muscle precursors, where the gene is activated (e.g. *Ebf_KhL24.37*, Figs. 4A, S19). Previous reporter gene expression assays identified the latter element as a minimal, although weak, enhancer active in pharyngeal muscle precursors[43]. CRISPR/Cas9-mediated mutagenesis assays followed by FISH indicated that each one of these elements is necessary for proper activation of *Ebf* in pharyngeal muscle progenitors (Figs. 4E-G and S19).

Consistent with the established roles of Hand-r, Tbx1/10 and Ets-mediated FGF-MAPK signaling in activating *Ebf*, the primed cardiopharyngeal-specific element (*KhL24.34*) contained E-box and Fox motifs, and the more distal de novo accessible minimal enhancer (*KhL24.37*) also contained putative Ets, E-box and RORγ binding sites, whereas the constitutively accessible elements (*KhC24.35* and *.36*) contained primarily CREB and T-box binding sites (Fig. 4E, S19). Of note, *Ebf* expression is maintained by auto-regulation[17], which requires separate intronic elements that open later and harbor putative Ebf binding sites (Fig. S21). This indicates that *Ebf* expression in pharyngeal muscle precursors is controlled by a combination of *trans*-acting inputs mediated by distinct elements with variable dynamics of accessibility. *Tbx1/10* showed a similar logic, whereby a constitutively accessible upstream element (*KhC7.909*) acts as an enhancer of cardiopharyngeal expression[17], whose activity also requires a primed cardiopharyngeal-specific intronic element (*KhC7.914*) (Fig. S20). Taken together, these results indicate that multiple elements are necessary for proper activation of cell fate determinants, such as *Tbx1/10* and *Ebf*, in a manner reminiscent of super- and shadow enhancers[67], but here, individual regulatory elements show distinct motifs composition and dynamics of chromatin accessibility. Thus, the sequence diversity and modular organization permits distinct chromatin dynamics, thereby increasing the repertoire of *trans*-acting inputs controlling gene expression via both enhancer accessibility and activity.

DISCUSSION

We characterized the accessible genome of the tunicate *Ciona*, with a special focus on the cardiopharyngeal lineage that produce heart and pharyngeal muscles. As seen in other systems, less than 10% of the *Ciona* genome is accessible, and distributed across thousands of short regions, most of which are stably accessible across time and lineages, especially promoter regions. By contrast, developmentally regulated regions either closed upon induction of multipotent progenitors, or opened specifically in the cardiopharyngeal lineage in response to FGF-MAPK signaling and Foxf activity. The latter elements were predominantly found in intergenic and intronic regions, and near cardiopharyngeal markers, consistent with their function as transcriptional enhancers. Similarly to other Forkhead factors[55], *Ciona* Foxf is required to open cardiopharyngeal elements for either immediate or later activation in multipotent or fate-restricted progenitors, respectively. Notably Foxf homologs play deeply conserved roles in visceral muscles specification[68–70], including during heart development in mammals[71]. Moreover, GATA motifs are over-represented among cardiopharyngeal-specific elements, also consistent with a conserved role for GATA homologs in heart development[72–77]. As combinations of Fox and GATA inputs play well-established

roles in early endoderm specification[78], we speculate that cardiopharyngeal regulatory programs were built upon an ancestral endomesodermal chromatin landscape during Olfactores evolution.

Finally, the majority of cell-type-specific markers expressed de novo are associated with "primed accessible" elements, as observed in numerous systems including cardiac differentiation of embryonic stem cells[3,4], and consistent with the role of pioneer factors in establishing competence for subsequent activation[55]. In the case of *Tbx1/10* and *Ebf*, spatially and temporally accurate activation is essential to permit the emergence of first and second cardiac, and pharyngeal muscle lineages[14,15,17,43]. We propose that, whereas the activity of multiple elements with similar spatio-temporal transcriptional outputs permits precise and robust gene activation[67,79], the modular organization of the *cis*-regulatory system increases the repertoire of regulatory inputs, acting through both accessibility and activity, to control cell-specific gene expression. These multi-level combinatorial inputs achieve exquisite spatio-temporal control, while permitting strong activation, thus ensuring both transcriptional precision and accuracy for developmental fate choices (Fig. S21).

ACKNOWLEDGEMENTS

We are grateful to Dayanne M. Castro for comments and support with ATAC-seq data analysis. We thank Karen Lam, Emily Miraldi and Richard Bonneau for help processing ATAC-seq data in the early phase of this study. We thank Emily Huang for TVC-specific enhancers validation and Wei Wang for discussions and sharing single cell and bulk RNA-seq data prior to publication. We thank Tatjana Sauka-Spengler for advice with ATAC-seq. We thank members of the Christiaen lab for discussions. C.R. has been supported by a long-term fellowship ALTF 1608-2014 from EMBO. This work was funded by awards R01 HL108643 from NIH/NHLBI, R01 HD096770 from NIH/NICHD, and 15CVD01 from the Leducq Foundation to L.C.

AUTHORS CONTRIBUTIONS

C.R. and L.C. obtained funding. C.R., K.W. and L.C. designed the experiments and analyses, C.R. performed the experiments, C.R. and K.W. performed the computational analyses, C.R., K.W. and L.C. wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

Fig. 1

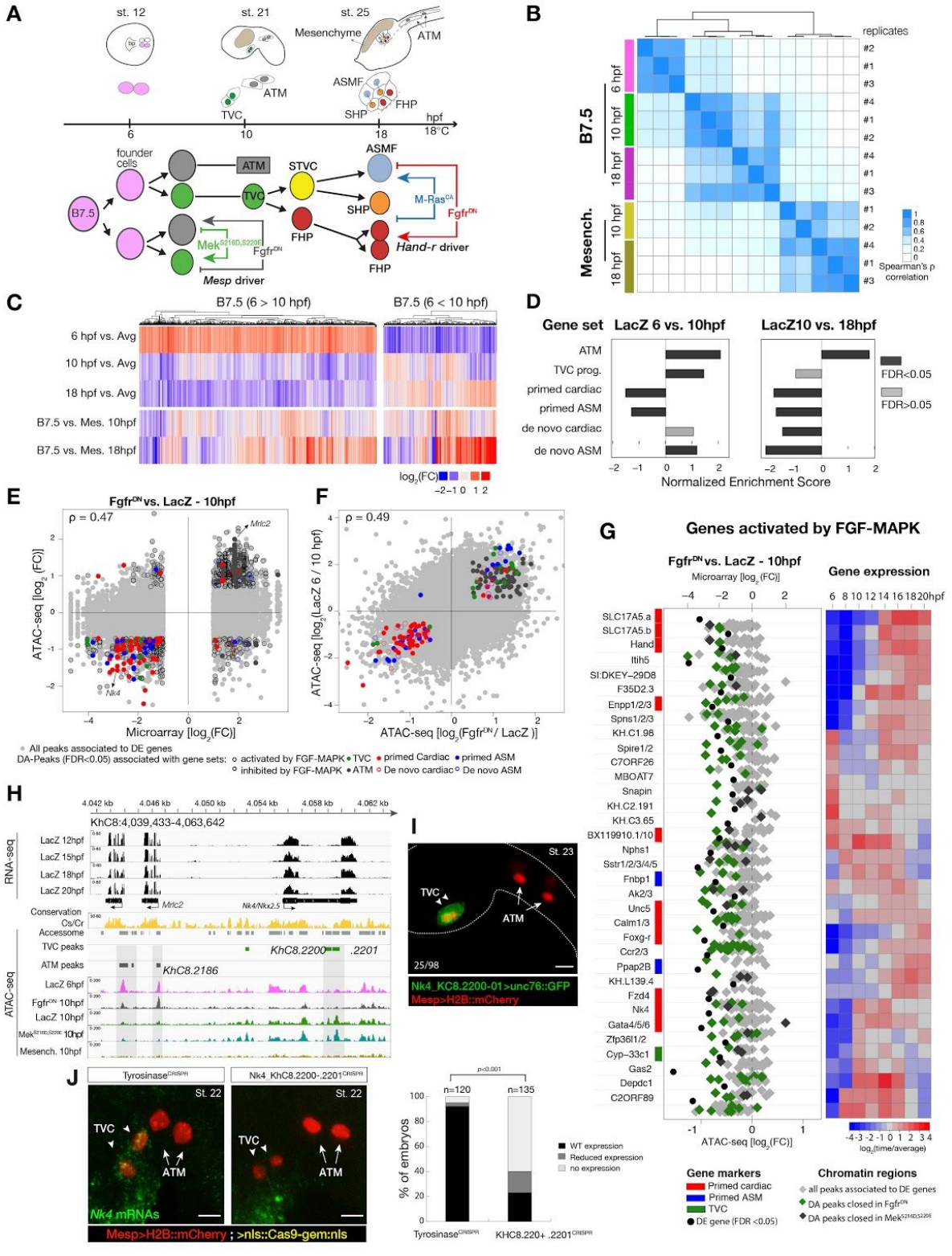

Fig. 1. Profiling chromatin accessibility dynamics during early cardiopharyngeal cell development. (**A**) Embryos, larvae and lineage diagram showing B7.5 blastomeres, their cardiopharyngeal progeny, and the main stages sampled for ATAC-seq. Stages (St.) according to [80] with hours post fertilization (hpf). (**B**) Spearman correlation of RPKM (reads per kb per million mapped reads) values in 14,178 regions changing accessibility over time or between mesenchyme and B7.5. (**C**) Temporal changes in chromatin accessibility for 5,450 peaks. "Open peaks at 6 hpf": 3,691 regions more accessible at 6 than 10 hpf in control samples. "Open peaks at 10 hpf": 1,759 regions more accessible at 10 than 6 hpf. The accessibility of these peaks is shown for 6, 10, and 18 hpf vs. the average (avg) accessibility in the control cells. Cell-type specific chromatin accessibility is shown in the comparison of B7.5 vs. mesenchyme (mes.) at 10 and 18 hpf. (**D**) Gene Set Enrichment Analysis (GSEA) normalized enrichment score of defined gene sets in peaks ranked by difference in accessibility between time points as indicated (see Supplementary Text). (**E-F**) Correlations between differentially expressed (DE) genes and differentially accessible (DA) peaks (E) and between two ATAC-seq samples (F). Colored dots are significantly changing in both comparisons. In (E), ϱ is the Spearman correlation of expression and accessibility for DA peaks associated to DE genes. (**G**) Relationship between expression and accessibility of DE genes associated to DA peaks. Microarray $\log_2$(fold change (FC)) values are shown as black dots. ATAC-seq $\log_2$(FC) values are shown as diamonds. (**H**) A 24 kb region on chromosome 8 displaying expression (RNA-seq) and chromatin accessibility (ATAC-seq; normalized by total sequencing depth). Gray shaded boxes show validated ATM-specific promoters and a newly identified TVC-specific enhancer in *Nk4* intron. (**I**) Enhancer driven *in vivo* reporter expression (green) of tested ATAC-seq peaks (*KhC8.2200* and *.2201*). TVCs marked with *Mesp>H2B::mCherry* (red). Numbers indicate observed/total of halves embryos scored. (**J**) Endogenous expression of *Nk4* visualized by *in situ* (green) in *Tyrosinase*^{CRISPR} and upon CRISPR/Cas9-induced deletions of TVC-specific peaks. Nuclei of B7.5 lineage cells are labelled by Mesp>nls::LacZ and revealed with an anti beta-galactosidase antibody (red). Experiment performed in biological replicates. Scale bar, 20 μm. Statistical analysis using a Fisher exact test, total numbers of embryos are shown.

Fig. 2

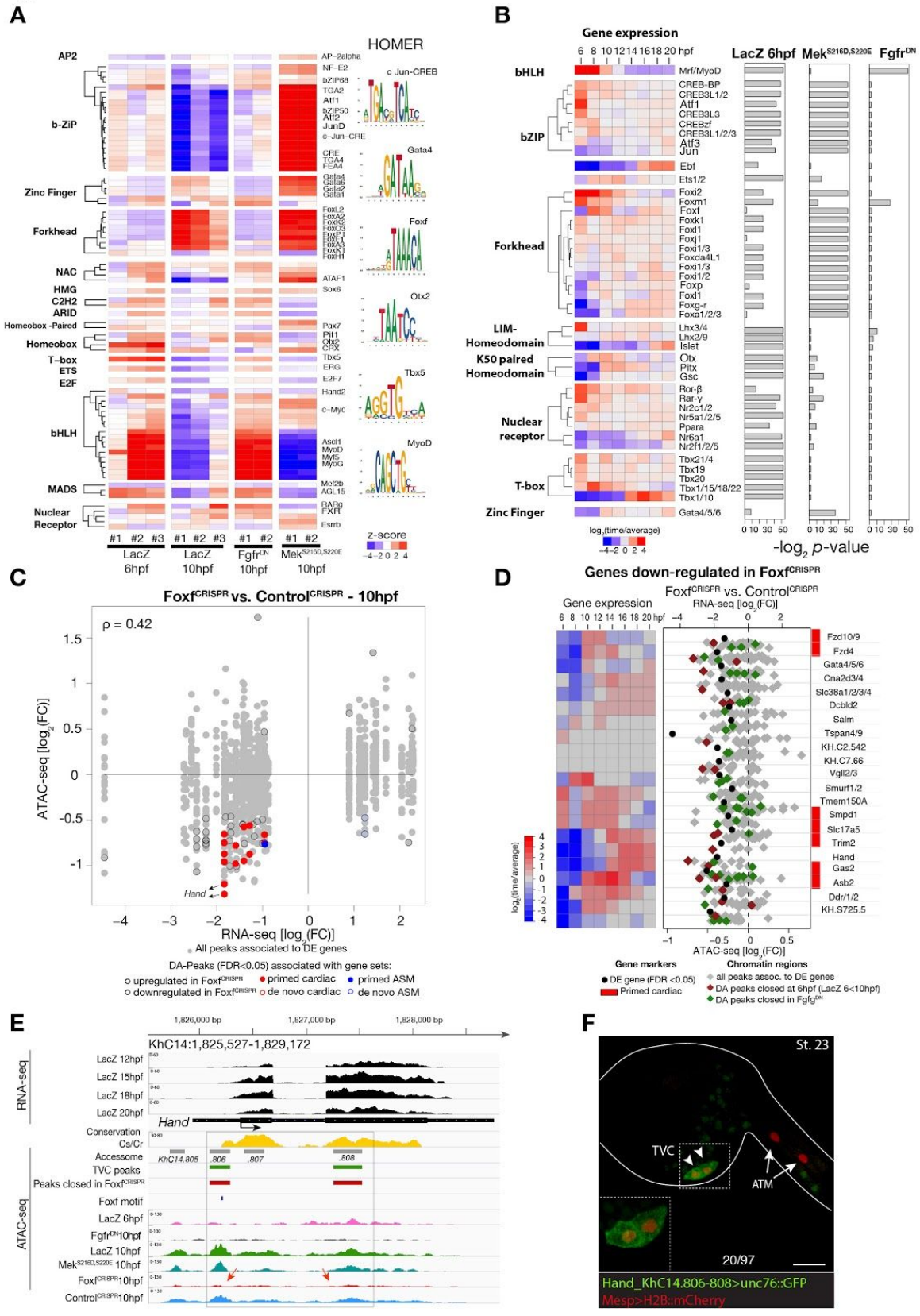

Fig. 2. Foxf is required for cardiopharyngeal-specific chromatin accessibility. (**A**) Deviation z-scores from ChromVAR[47]. Motifs were taken from the HOMER Known Motifs database[81]. Deviations were computed for FGF signaling-dependent peaks at 10 hpf and B7.5 replicates at 6 and 10 hpf. Only DA motifs (FDR < 0.01, |z| > 2 in any replicate) are shown. (**B**) Microarray gene expression of transcription factors over time compared to enrichment of corresponding HOMER motifs in condition specific peak sets. $\log_2$ *p*-values are shown for a hypergeometric test for enrichment of each motif in each peak set. (**C**) Differential expression from bulk RNA-seq (DE) vs. chromatin accessibility from ATAC-seq (DA). ϱ is the Spearman correlation of expression and accessibility for DA peaks associated to DE genes.(**D**) Relationship between expression and accessibility of DE genes associated to DA peaks as in Fig. 1G. (**E**) A 3.6 kb region on chromosome 14 displaying expression profiles of RNA-seq and chromatin accessibility profiles of ATAC-seq normalized tag count. Foxf core binding site (GTAAACA) is displayed as blue line. The boxed region indicates a newly identified TVC-specific enhancer in *Hand* locus. Red arrow indicates a TVC-specific enhancer showing closed chromatin in Foxf^{CRISPR} ATAC-seq. (**F**) Enhancer driven *in vivo* reporter expression (green) of tested ATAC-seq peaks. TVCs marked with Mesp>H2B::mCherry (red). Numbers indicate observed/total of halves embryos scored. Experiment performed in biological replicates. Scale bar, 30 μm.

Fig. 3

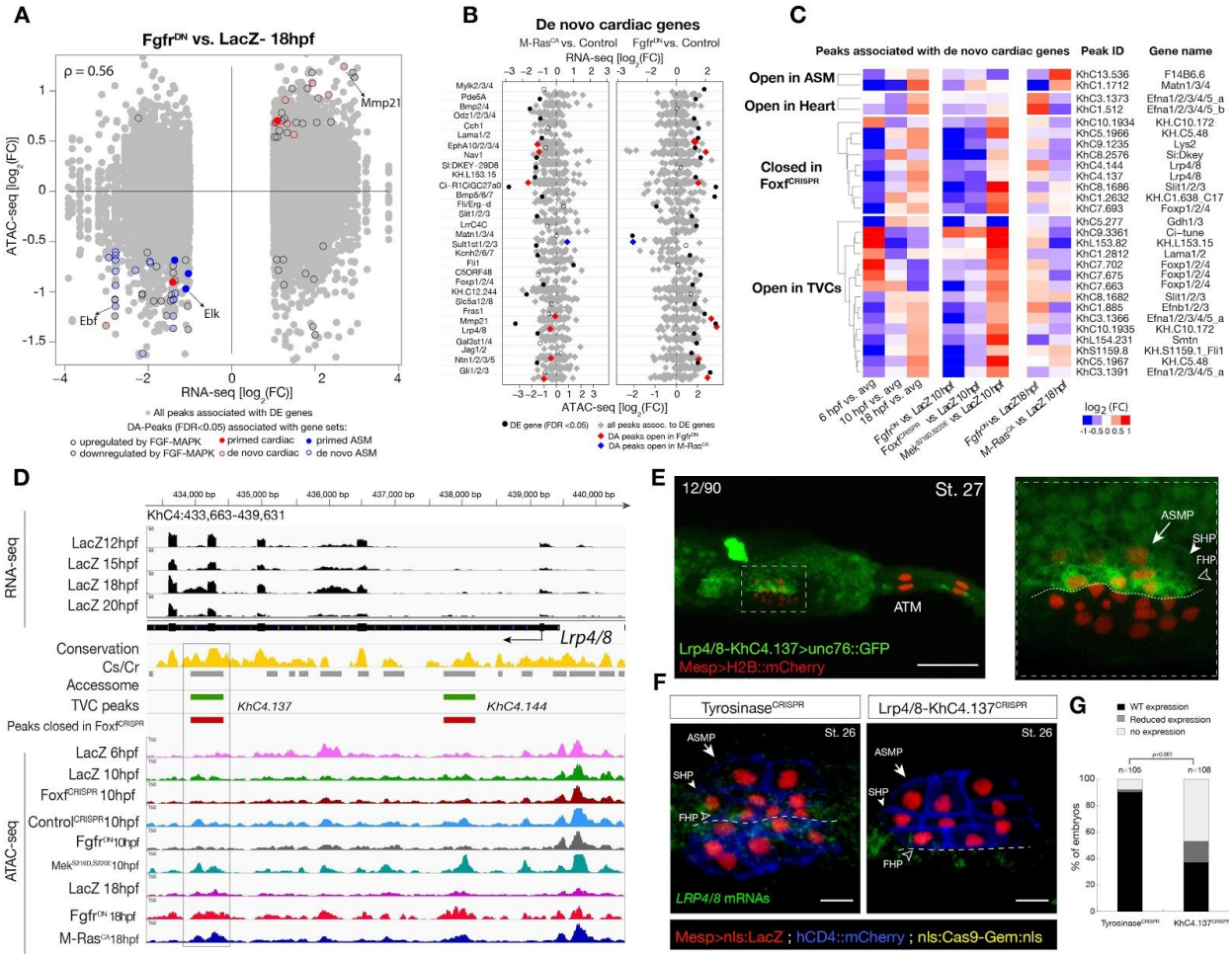

Fig. 3 Cardiopharyngeal lineage-specific accessibility profiles and decoupling between enhancer accessibility and activity for de novo expressed genes. (**A**) Differentially expressed (DE) genes vs differentially accessible (DA) peaks. ρ is the Spearman correlation of expression and accessibility for DA peaks associated to DE genes. (**B**) Relationship between accessibility and expression of de novo cardiac genes as in Fig. 1G. DE genes in either condition are shown as black dots. Non-significant genes are shown as hollow circles. (**C**) ATAC-seq peaks associated to de novo expressed pan-cardiac genes. The accessibility of these peaks is shown for 6, 10 and 18 hpf vs. the average (avg) accessibility in the controls (LacZ) and upon FGF-MAPK perturbations at either 10 or 18hpf. Peaks were classified as "Open in ASM" (less accessible in Fgfr^{DN} vs. M-Ras^{CA} or LacZ at 18hpf), "Open in Heart" (less accessible in M-Ras^{CA} vs. Fgfr^{DN} or LacZ at 18hpf), "Closed in Fofx^{CRISPR}" (less accessible in Fofx^{CRISPR} vs. Control^{CRISPR}), or "Open in TVC" (less accessible in Fgfr^{DN} vs. Mek^{S216D,S220E} or LacZ at 10 hpf). (**D**) A 6 kb region on chromosome 4 displaying expression profiles of RNA-seq and chromatin accessibility profiles of ATAC-seq normalized tag count. Peak ID refers to elements tested for reporter assay *in vivo*. The newly identified enhancer in *Lrp4/8* locus is in the boxed region. (**E**) Enhancer driven *in vivo* reporter expression (green) of tested "*KhC4.137*" peak. TVCs marked with Mesp>H2B::mCherry (red). Numbers indicate observed/total of halves embryos scored. Zoom on cardiopharyngeal cell lineage (panel on the right). (**F**) Endogenous expression of *Lrp4/8* visualized by *in situ* (green) in Tyrosinase^{CRISPR} and upon CRISPR/Cas9-induced

deletion of ATAC-seq peaks. 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 (Blue). Experiment performed in biological replicates. Scale bar, 10 μm. (**G**) Statistical analysis using a Fisher exact test ($p < 0.001$), total numbers of embryos are shown.

Fig. 4

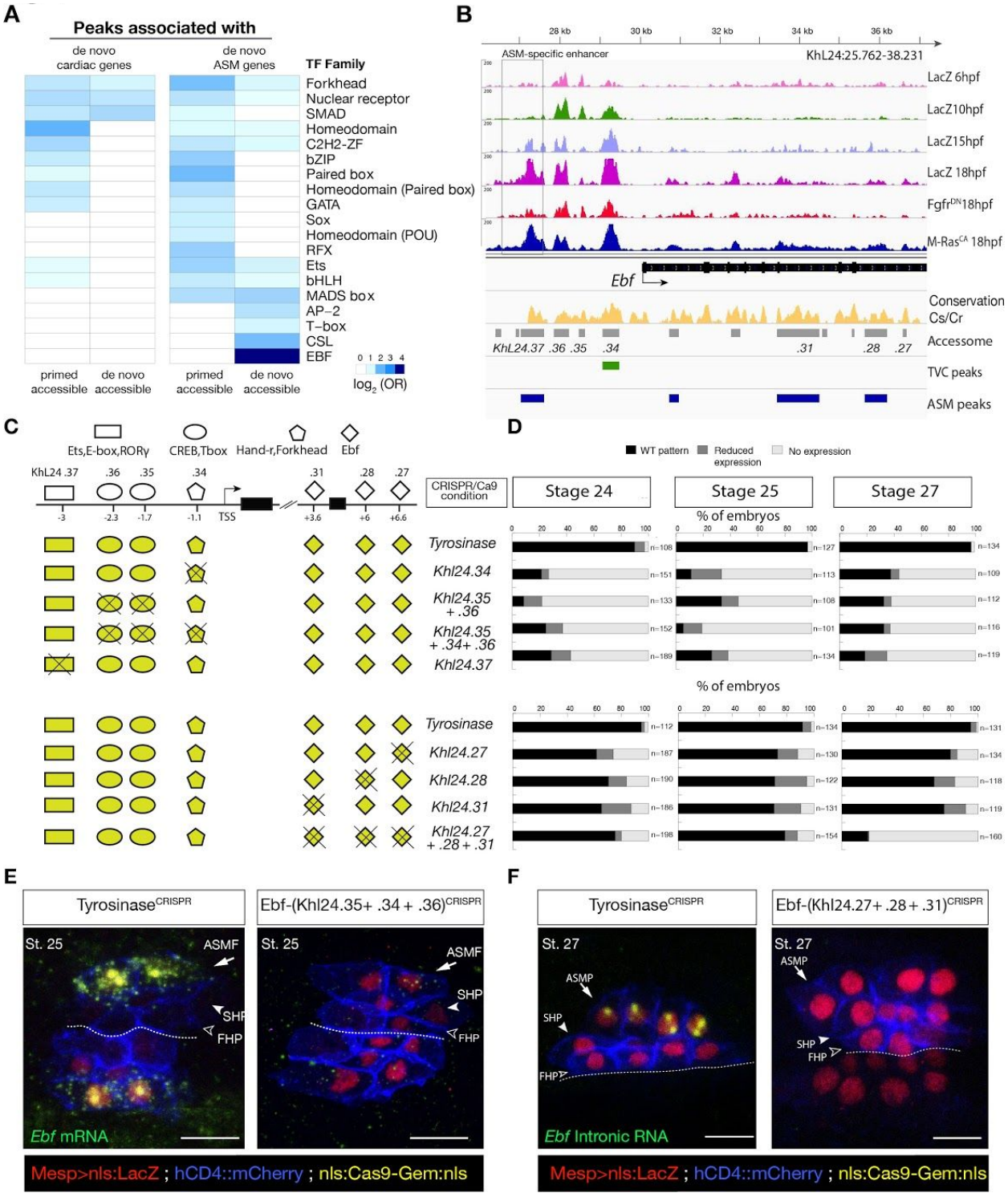

Fig. 4 Combinations of *cis*-regulatory elements with distinct chromatin accessibility profiles are required for *Ebf* transcription in pharyngeal-muscle precursors. (**A**) TF binding motif families enriched in peaks associated to de novo cardiac and pharyngeal expressed genes parsed based on primed

and de novo accessibility (see Material and Methods). $\log_2$ odds ratios (see Materials and Methods) are shown for motif families significantly enriched (one-tailed hypergeometric test, FDR < 0.05) in the indicated peak classes. (**B**) A 12 kb region of the scaffold L24 displaying expression profiles of RNA-seq and chromatin accessibility profiles of ATAC-seq (normalized tag count) in the *Ebf* locus. (**C**) Schematic representation of *Ebf* regulatory elements targeted for CRISPR/Cas9-mediated deletions. Shapes represent binding sites located in the regulatory elements. (**D**) Proportions of larvae halves showing the indicated *Ebf* transcription patterns, in indicated experimental conditions; all the treatments were significant versus *Tyrosinase* (Fisher exact test, $p < 0.001$). (**E-F**) Endogenous expression of *Ebf* visualized by *in situ* (green) in Tyrosinase^{CRISPR} and upon CRISPR/Cas9-induced deletion of ATAC-seq peaks as indicated, at stage 25 (E) and 27 (F)[80]. Nuclei of B7.5 lineage cells are labelled by Mesp>nls::LacZ and revealed with an anti beta-galactosidase antibody (red). Scale bar = 10 µm.

REFERENCES

1. Davidson, E. H. *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*. (Elsevier, 2010).
2. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* (2019). doi:10.1038/s41576-018-0089-8
3. Paige, S. L. *et al.* A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* **151**, 221–232 (2012).
4. Wamstad, J. A. *et al.* Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206–220 (2012).
5. Rosa-Garrido, M., Karbassi, E., Monte, E. & Vondriska, T. M. Regulation of Chromatin Structure in the Cardiovascular System. *Circ. J.* **77**, 1389–1398 (2013).
6. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–223 (2013).
7. He, A. *et al.* Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat. Commun.* **5**, 4907 (2014).
8. Diogo, R. *et al.* A new heart for a new head in vertebrate cardiopharyngeal evolution. *Nature* **520**, 466–473 (2015).
9. Lescroart, F. *et al.* Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. *Development* **137**, 3269–3279 (2010).
10. Lescroart, F. *et al.* Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nat. Cell Biol.* **16**, 829–840 (2014).
11. Lescroart, F. *et al.* Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science* **359**, 1177–1181 (2018).
12. Kaplan, N., Razy-Krajka, F. & Christiaen, L. Regulation and evolution of cardiopharyngeal cell identity and behavior: insights from simple chordates. *Curr. Opin. Genet. Dev.* **32**, 119–128 (2015).
13. Christiaen, L. *et al.* The transcription/migration interface in heart precursors of Ciona intestinalis. *Science* **320**, 1349–1352 (2008).
14. Razy-Krajka, F. *et al.* Collier/OLF/EBF-dependent transcriptional dynamics control pharyngeal muscle specification from primed cardiopharyngeal progenitors. *Dev. Cell* **29**, 263–276 (2014).
15. Wang, W. *et al.* A single cell transcriptional roadmap for cardiopharyngeal fate diversification. *bioRxiv* 150235 (2017). doi:10.1101/150235
16. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
17. Razy-Krajka, F. *et al.* An FGF-driven feed-forward circuit patterns the cardiopharyngeal mesoderm in space and time. *Elife* **7**, (2018).
18. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
19. Hockman, D. *et al.* A genome-wide assessment of the ancestral neural crest gene regulatory network. (2018). doi:10.1101/275875
20. Daugherty, A. C. *et al.* Chromatin accessibility dynamics reveal novel functional enhancers in C. elegans. *Genome Res.* (2017). doi:10.1101/gr.226233.117
21. Janes, J. *et al.* Chromatin accessibility is dynamically regulated acrossC. elegansdevelopment and ageing. (2018). doi:10.1101/279158
22. Li, L., Zhu, Q., He, X., Sinha, S. & Halfon, M. S. Large-scale analysis of transcriptional cis-regulatory modules reveals both common features and distinct subclasses. *Genome Biol.* **8**, R101 (2007).
23. Madgwick, A. *et al.* Evolution of the embryonic cis-regulatory landscapes between divergent Phallusia and Ciona ascidians. (2018). doi:10.1101/371005
24. Ganot, P., Kallesøe, T., Reinhardt, R., Chourrout, D. & Thompson, E. M. Spliced-leader RNA trans splicing in a chordate, Oikopleura dioica, with a compact genome. *Mol. Cell. Biol.* **24**, 7795–7805 (2004).
25. Hastings, M. L. *et al.* An LKB1 AT-AC intron mutation causes Peutz-Jeghers syndrome via splicing at noncanonical cryptic splice sites. *Nat. Struct. Mol. Biol.* **12**, 54–59 (2005).

26. Vandenberghe, A. E., Meedel, T. H. & Hastings, K. E. mRNA 5'-leader trans-splicing in the chordates. *Genes Dev.* **15**, 294–303 (2001).
27. Satou, Y., Hamaguchi, M., Takeuchi, K., Hastings, K. E. M. & Satoh, N. Genomic overview of mRNA 5'-leader trans-splicing in the ascidian Ciona intestinalis. *Nucleic Acids Res.* **34**, 3378–3388 (2006).
28. Yokomori, R. *et al.* Genome-wide identification and characterization of transcription start sites and promoters in the tunicate Ciona intestinalis. *Genome Res.* **26**, 140–150 (2016).
29. Mavrich, T. N. *et al.* A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18**, 1073–1083 (2008).
30. Mavrich, T. N. *et al.* Nucleosome organization in the Drosophila genome. *Nature* **453**, 358–362 (2008).
31. Mayran, A. *et al.* Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat. Genet.* **50**, 259–269 (2018).
32. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170–1187 (2016).
33. Brozovic, M. *et al.* ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. *Nucleic Acids Res.* **46**, D718–D725 (2018).
34. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
35. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2015).
36. Gaulton, K. J. *et al.* A map of open chromatin in human pancreatic islets. *Nat. Genet.* **42**, 255–259 (2010).
37. Khan, A., Mathelier, A. & Zhang, X. Super-enhancers are transcriptionally more active and cell-type-specific than stretch enhancers. (2018). doi:10.1101/310839
38. Davidson, B., Shi, W., Beh, J., Christiaen, L. & Levine, M. FGF signaling delineates the cardiac progenitor field in the simple chordate, Ciona intestinalis. *Genes Dev.* **20**, 2728–2738 (2006).
39. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
40. Woznica, A. *et al.* Initial deployment of the cardiogenic gene regulatory network in the basal chordate, Ciona intestinalis. *Dev. Biol.* **368**, 127–139 (2012).
41. Bernadskaya, Y. Y., Brahmbhatt, S., Gline, S. E., Wang, W. & Christiaen, L. Dual functions of Discoidin domain receptor coordinate cell-matrix adhesion and collective polarity in migratory cardiopharyngeal progenitors. (2017). doi:10.1101/154880
42. Beh, J., Shi, W., Levine, M., Davidson, B. & Christiaen, L. FoxF is essential for FGF-induced migration of heart progenitor cells in the ascidian Ciona intestinalis. **3305**, 3297–3305 (2007).
43. Wang, W., Razy-Krajka, F., Siu, E., Ketcham, A. & Christiaen, L. NK4 antagonizes Tbx1/10 to promote cardiac versus pharyngeal muscle fate in the ascidian second heart field. *PLoS Biol.* **11**, e1001725 (2013).
44. Kusakabe, T., Yoshida, R., Ikeda, Y. & Tsuda, M. Computational discovery of DNA motifs associated with cell type-specific gene expression in Ciona. *Dev. Biol.* **276**, 563–580 (2004).
45. Satou, Y. *et al.* Gene expression profiles in Ciona intestinalis tailbud embryos. *Development* **128**, 2893–2904 (2001).
46. Sierro, N. *et al.* DBTGR: a database of tunicate promoters and their regulatory elements. *Nucleic Acids Res.* **34**, D552–5 (2006).
47. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
48. Satou, Y., Imai, K. S. & Satoh, N. Early embryonic expression of a LIM-homeobox gene Cs-lhx3 is downstream of beta-catenin and responsible for the endoderm differentiation in Ciona savignyi embryos. *Development* **128**, 3559–3570 (2001).
49. Davidson, B., Shi, W. & Levine, M. Uncoupling heart cell specification and migration in the simple chordate Ciona intestinalis. *Development* **132**, 4811–4818 (2005).
50. Meedel, T. H., Chang, P. & Yasuo, H. Muscle development in Ciona intestinalis requires the b-HLH myogenic regulatory factor gene Ci-MRF. *Dev. Biol.* **302**, 333–344 (2007).
51. Tolkin, T. & Christiaen, L. Rewiring of an ancestral Tbx1/10-Ebf-Mrf network for pharyngeal muscle

specification in distinct embryonic lineages. (2016). doi:10.1101/039289
52. Cusanovich, D. A. *et al.* The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018).
53. Hudson, C. A conserved role for the MEK signalling pathway in neural tissue specification and posteriorisation in the invertebrate chordate, the ascidian Ciona intestinalis. *Development* **130**, 147–159 (2003).
54. Ragkousi, K., Beh, J., Sweeney, S., Starobinska, E. & Davidson, B. A single GATA factor plays discrete, lineage specific roles in ascidian heart development. *Dev. Biol.* **352**, 154–163 (2011).
55. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
56. Cirillo, L. A. *et al.* Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. **9**, 279–289 (2002).
57. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420 (1993).
58. Gandhi, S., Haeussler, M., Razy-Krajka, F., Christiaen, L. & Stolfi, A. Evaluation and rational design of guide RNAs for efficient CRISPR/Cas9-mediated mutagenesis in Ciona. *Dev. Biol.* **425**, 8–20 (2017).
59. Razy-Krajka, F. *et al.* An FGF-driven feed-forward circuit patterns the cardiopharyngeal mesoderm in space and time. *Elife* **7**, (2018).
60. Keduka, E. *et al.* 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 (2009).
61. Schultheiss, T. M., Burch, J. B. & Lassar, A. B. A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev.* **11**, 451–462 (1997).
62. Nowotschin, S. *et al.* Tbx1 affects asymmetric cardiac morphogenesis by regulating Pitx2 in the secondary heart field. *Development* **133**, 1565–1573 (2006).
63. Desjardins, C. A. & Naya, F. J. The Function of the MEF2 Family of Transcription Factors in Cardiac Development, Cardiogenomics, and Direct Reprogramming. *J Cardiovasc Dev Dis* **3**, (2016).
64. Pikkarainen, S., Tokola, H., Kerkelä, R. & Ruskoaho, H. GATA transcription factors in the developing and adult heart. *Cardiovasc. Res.* **63**, 196–207 (2004).
65. Stolfi, A. *et al.* Early chordate origins of the vertebrate second heart field. *Science* **329**, 565–568 (2010).
66. Stolfi, A., Gandhi, S., Salek, F. & Christiaen, L. Tissue-specific genome editing in Ciona embryos by CRISPR/Cas9. *Development* **141**, 4115–4120 (2014).
67. Lagha, M., Bothma, J. P. & Levine, M. Mechanisms of transcriptional precision in animal development. *Trends Genet.* **28**, 409–416 (2012).
68. Scimone, M. L. *et al.* foxF-1 Controls Specification of Non-body Wall Muscle and Phagocytic Cells in Planarians. *Curr. Biol.* **28**, 3787–3801.e6 (2018).
69. Jakobsen, J. S. *et al.* Temporal ChIP-on-chip reveals Biniou as a universal regulator of the visceral muscle transcriptional network. *Genes Dev.* **21**, 2448–2460 (2007).
70. Zaffran, S., Küchler, A., Lee, H. H. & Frasch, M. biniou (FoxF), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in Drosophila. *Genes Dev.* **15**, 2900–2915 (2001).
71. Hoffmann, A. D. *et al.* Foxf genes integrate tbx5 and hedgehog pathways in the second heart field for cardiac septation. *PLoS Genet.* **10**, e1004604 (2014).
72. Qian, L. & Bodmer, R. Partial loss of GATA factor Pannier impairs adult heart function in Drosophila. *Hum. Mol. Genet.* **18**, 3153–3163 (2009).
73. Sorrentino, R. P., Gajewski, K. M. & Schulz, R. A. GATA factors in Drosophila heart and blood cell development. *Semin. Cell Dev. Biol.* **16**, 107–116 (2005).
74. Reiter, J. F. *et al.* Gata5 is required for the development of the heart and endoderm in zebrafish. *Genes Dev.* **13**, 2983–2995 (1999).
75. Holtzinger, A. & Evans, T. Gata5 and Gata6 are functionally redundant in zebrafish for specification of cardiomyocytes. *Dev. Biol.* **312**, 613–622 (2007).
76. Zhao, R. *et al.* Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. *Dev. Biol.* **317**, 614–619 (2008).

77. Molkentin, J. D., Lin, Q., Duncan, S. A. & Olson, E. N. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev.* **11**, 1061–1072 (1997).
78. Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
79. Bentovim, L., Harden, T. T. & DePace, A. H. Transcriptional precision and accuracy in development: from measurements to models and mechanisms. *Development* **144**, 3855–3866 (2017).
80. Hotta, K. *et al.* 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 (2007).
81. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).