Supplementary Materials for

**GATA4/5/6 family transcription factors are conserved determinants of cardiac versus pharyngeal mesoderm fate**

Mengyi Song, Xuefei Yuan, Claudia Racioppi, Meaghan Leslie, Nathan Stutt, Anastasiia Aleksandrova, Lionel Christiaen, Michael D. Wilson*, Ian C. Scott*

*Corresponding author. Email: michael.wilson@sickkids.ca (M.D.W.); ian.scott@sickkids.ca (I.C.S.)

Published 11 March 2022, *Sci. Adv.* 8, eabg0834 (2022)
DOI: 10.1126/sciadv.abg0834

**The PDF file includes:**
- Supplementary Text
- Figs. S1 to S8
- Tables S1 to S6
- Legends for data S1 to S5
- References

**Other Supplementary Material for this manuscript includes the following:**
- Data S1 to S5
Supplementary Text

The expression of canonical pharyngeal genes within the forced-directed graph and pharyngeal sub-populations at 13 hpf

As Tbx1 and Ebf are classical pharyngeal regulators, and their orthologs play important roles in pharyngeal lineage specification in Ciona and mouse (18), we analyzed the expression of tbx1 and Ebf genes on the forced-directed graph. While tbx1 expression is relatively broad, it is enriched in the pharyngeal trajectory. Expression of ebf genes is observed at the branching tips of pharyngeal lineages (Fig. S1S). To better understand pharyngeal sub-populations, we next examined the expression of tbx1, tcf21, and ebf genes in sub-clustering analysis. In agreement with expression patterns seen in the forced-directed graph (Fig. S1S), tbx1 was expressed in many cardiac and pharyngeal clusters, most strongly in the SHF progenitors (sub-cluster S9), first pharyngeal arch progenitors (sub-cluster S5) and putative pharyngeal arch two (sub-cluster S2). In contrast, the expression of tcf21 was still quite sparse at early somite stages. Although the expression and roles of ebf genes have been primarily associated with neural lineages in zebrafish (79–81), we observed the expression of ebf1a, ebf2 and ebf3a in some cranial-pharyngeal sub-clusters (Fig. S3D), suggesting they may play conserved roles in early pharyngeal specification.

Mouse and zebrafish cardiac and pharyngeal marker gene comparisons

To identify genes that are expressed in both mouse and zebrafish pharyngeal-related cells, we performed overlaps of marker genes in the pharyngeal mesoderm (sub)clusters and the zebrafish orthologs of the mouse paraxial mesoderm (PM) and branchiomeric muscles (BM) marker genes identified through scRNA-seq (36). As the markers for each sub-cluster are identified through comparison with the rest of the cells, which are enriched for pharyngeal mesoderm, broadly expressed markers could be underrepresented. Thus, we overlapped the zebrafish orthologs of the mouse PM and BM markers with marker genes from a) two pharyngeal clusters in 13 hpf WT (Fig. S7A), b) four pharyngeal clusters in 13 hpf WT+KD (Fig. S7B), c) six pharyngeal sub-clusters, respectively (Fig. S7C)

GFP+/-specific ATAC-seq peaks

GREAT functional enrichment analyses showed that the GFP+ specific accessible regions were highly enriched for various mesoderm (vascular endothelial growth factor signaling pathway: FDR=3.07E-27; embryonic heart tube development: FDR=1.93E-26; endothelium development: FDR=1.47E-19) and endoderm (endoderm development: FDR=1.22E-15; digestive
tract development: FDR=1.77E-15) development pathways (Fig. S6D), while the GFP- specific regions were strongly enriched for many brain and neural development-related processes (hindbrain development: FDR=5.50E-33; forebrain development: FDR=1.36E-28; peripheral nervous system development: FDR=8.14E-25; Fig. S6D). This is consistent with the fact that the GFP+/- specific accessible chromatin regions contain many lineage-specific regulatory elements.

Our two motif enrichment analyses both revealed that GATA motifs showed the highest enrichment within GFP+ specific peaks (Fig. S6C, E). This suggests that Gata factors play essential roles in regulating the development of the gata5::GFP+ populations, which is highly consistent with our cell composition analysis of the single-cell data. Similarly, motifs most enriched in the GFP- specific regions are the SOX factor motifs (Fig. S6C, E), consistent with the important role that SOX factors play in brain development.

**Independent alleles of tbx1 DARs**

2-6 independent alleles were raised for each tbx1 closed DARs (Fig. S8A-G). We observed consistent GFP activity in the developing heart. 4 independent alleles were isolated for the Tg(tbx1open::GFP) background. Each allele showed pharyngeal muscle activity marked by tcf21. One allele displayed additional mosaic GFP activity in the heart, which is likely due to positional effect (Fig. S8H)
Supplementary Methods
Single-cell mRNA-seq data alignment, quality control, and clustering analysis

Cell Ranger (version 2.0.0) was used for initial alignment, filtering non-cellular barcodes, and UMI counting to generate cell gene-count matrices with default parameters. Reads were aligned to the Ensembl Zv10 Release 89 reference genome and counted with the Ensembl transcriptome (Release 89) with a gfp sequence manually added. Sequencing saturation was estimated in the Cell Ranger pipeline by downsampling sequencing depth in mean reads per cell and assessing library complexities subsequently. The cell gene-count matrices generated in the Cell Ranger pipeline were filtered to exclude low-quality cells (low number of genes detected, high percentage of mitochondria gene reads) or potential cell doublets (high number of UMI counts). Different thresholds were used for samples from different stages, as cell size, RNA amounts, and RNA capture efficiencies varied between samples. Our filtering thresholds were 6 hpf (genes > 1000, UMI < 25000, mitochondria percentage < 4%), 8 hpf (genes > 1000, UMI < 20000, mitochondria percentage < 4%), 10 hpf (genes > 500, UMI < 12000, mitochondria percentage < 4%), WT and Gata5/6 KD 13 hpf (genes > 1000, UMI < 30000, mitochondria percentage < 4%) which were more stringent than that used in the previous studies of zebrafish embryos (26, 28). The filtered gene-count matrices were used as input of Seurat (version 3.0.0) for dimensional reduction, cell clustering, and marker gene identification of each cluster. In the Seurat pipeline, SCTransform (82) was selected for data normalization, and the effects of mitochondria gene counts and cell cycles were regressed out before identifying variable genes according to package instruction.

Only genes that were identified as variable genes in SCTransform and were detected in large than 10% of cells in a cluster were tested as signature genes. MAST (83) was used and genes with log|\(\text{FC}\)| > 0.25 and \(p_{\text{val}} < 0.01\) were identified as signature genes. Signature genes that were well studied and annotated (known marker genes) or best represented the transcriptome differences between clusters were selected for heatmap plotting. The 4 WT samples were analyzed separately (Fig. S1E-P). For comparing 13 hpf WT and Gata5/6 KD samples, these two samples were merged directly for clustering analysis without special integration.

To verify the SHF-like cluster, marker genes of the SHF-like sub-cluster were first converted to their mouse orthologs using gprofiler2. GO enrichment was performed on the mouse orthologs (Binomial test, Data S2) (84, 85). To compare the genes in the SHF-like cluster with a mouse SHF dataset (36), the zebrafish orthologs (\(n = 221\)) of the mouse anterior SHF genes
(E7.75 and E8.25, n = 212) determined by the authors through scRNA-seq were first obtained using gprofiler2 and intersected with marker genes identified from each sub-cluster separately.

**Pseudotime analysis**

After clustering analysis within each sample, the 4 WT samples were merged with cluster identity kept and then normalized as one dataset using the SCTransform (82) method. Clusters of each lineage were extracted from this merged dataset based on marker gene expression and fate-mapping knowledge (Table S4) and analyzed using the dynverse package (86, 87) for building up a pseudotime trajectory. *gata5* and *gata6* expression in each trajectory was plotted with ggplot2. To compare the *gata5* expression dynamics between different lineage trajectories, the pseudotime at each time point was divided into 10 quantile bins and the mean *gata5* expression of each bin was calculated and plotted in a heatmap (Fig. 1E).

**Visualization of the developmental progression of gata5GFP+ cells**

STITCH (26) was implemented to generate a forced-layout visualization as previously described. In short, the filtered cell x gene counts matrix from all WT time points were merged and normalized as the input together with the developmental timepoint ‘ID’ for each cell. Variable genes were first identified for each time point. STITCH next constructed a single-cell graph by computing k-nearest neighbors for each time point and identifying neighboring cells in pairs of adjacent time points to stitch the graph with a pre-defined order (6 hpf, 8 hpf, 10 hpf, 13 hpf). Up to 200 nearest neighbors were identified and up to 20 nearest neighbors were retained for each cell (nodes) with edges linking them in between. The single-cell graph was then visualized using ForceAtlas2 layout in Gephi (https://gephi.org/). Nodes are colored based on either timepoints or major branches. The latter required a lineage tag for each cell obtained from the clustering information in Seurat.

**ATAC-seq read mapping and differential peak identification**

Raw reads were preprocessed by FastQC (version 0.11.2, https://github.com/s-andrews/FastQC) (88) and Trimmomatic (version 0.32) (89) (LEADING:20 TRAILING:20 SLIDINGWINDOW:5:25 MINLEN:36) before being aligned to GRCz11 zebrafish genome assembly by BWA (version 0.7.8) under aln model (90). Fastq files from the two batches of sequencing were aligned individually and the two alignment outputs for the same libraries (bam files) were merged using SAMtools merge (version 1.2) (91). Reads mapped to the non-
mitochondrial genome and with mapping quality scores > 30 were kept for downstream analysis using SAMtools (version 1.2) (91) (samtools view -b -q 30). Initial peak calling of each library were conducted by MACS2 (version 2.7.9) (92) (--fe-cutoff 4 -nomodel --nolambda --gsize 1.4e9). Around 100,000 peaks were identified in each library. Peaks showing enriched signals in one population versus the other (DARs or GFP+/− specific peaks) were identified using DiffBind package (version 2.6.6) (93) (edgeR, FDR < 0.05, fold change > 1). We used DiffBind to conduct three pairwise comparisons: 1) WT GFP+ samples versus WT GFP- samples to identify GFP+/− specific peaks; 2) WT GFP+ samples versus Gata5/6 KD GFP+ samples to identify DARs in GFP+ cells; 3) WT GFP- cells versus Gata5/6 KD GFP- samples to identify DARs in GFP- cells. DiffBind first generated a consensus peak set that represent an overall set of open chromatin regions from the samples included in each comparison. Then it counted if the normalized sequences mapped to each interval in the consensus peak set were significantly different in the compared samples. The coordinates of differential and shared peaks identified from each comparison were shown in Data S2. We used the output from DiffBind for all downstream analyses. deepTools (version 2.2.3) (94) was used to plot the ATAC-seq signals centered around the peaks (plotHeatmap and plotProfile functions), as well as to generate the merged, normalized coverage files shown in the genome browser tracks (bamCoverage –binSize 1 –normalizeUsing CPM).

Enrichment analyses of ATAC-seq

Homer (version 4.11) (95) findMotifs.pl was used to identify overrepresented motifs within differential peak sets. The default homer motif database was used for the motif scan and the corresponding shared peaks were used as the background. Fold of enrichment was calculated as the % of targets sequences with the given motif divided by the % of background sequences with the same motif. CentriMO (version 5.3.3) (96) was used to identify motifs enriched at the center of ATAC-seq peaks. To run CentriMO, the middle point of the ATAC-seq peaks were extracted from each peak and extended 250 bp to each side. These regions with the uniform length were used as the input for CentriMO. The JASPAR 2020 CORE vertebrate non-redundant motif database (97) was used for the motif scan and the corresponding shared peaks were used as the background. Fisher E-value was used to determine the statistical significance (Fisher E-value < 0.05).

GREAT (version 3.0.0) (98) online tool was used for functional enrichment analysis of the different peak sets. In order to run GREAT, the genomic coordinates of the peaks were first
converted from GRCz11 assembly to danRer7 (Zv9) using liftOver (default setting) (99). The whole genome was used as background and the basal plus extension rule was used for associating genomic regions with genes. In GREAT, each gene is assigned a basal regulatory domain of 5 kb upstream and 1kb downstream of the TSS (regardless of other nearby genes). The gene regulatory domain is extended in both directions to the nearest gene's basal domain but no more than 1 Mb in one direction. Functional terms showing a binomial FDR < 0.05, hypergeometric FDR < 0.05 and a minimal region-based fold of enrichment of 2 were ranked by their binomial FDR and selected for plotting. The full motif and GREAT functional enrichment results are shown in Data S4.

To identify GFP+ open DARs (n = 372) near conserved pharyngeal genes (Fig. S7A-C), distanceToNearest function (GenomicRanges (100)) was used to associate the open DARs to their nearest protein-coding gene (annotation: Ensembl 101 (101), GRCz11) and 8 open DARs were identified (Table S6). All peaks (n = 560) associated with the conserved pharyngeal genes were identified with a similar approach using the consensus peak set (n = 114, 987) from WT and Gata5/6KD GFP+ datasets. Hypergeometric test was performed to test if open chromatin regions near conserved pharyngeal genes were enriched for open DARs.
Figure S1

A

\[ \text{tgBAC(gata5:GFP)} \]

B

\[ \text{tgBAC(gata5:GFP)} \]

C

\[ \text{tgBAC(gata5:GFP)} \]

D

\[ \text{tgBAC(gata5:GFP)} \]
Figure S1

WT 13 hpf

Before filtering: n = 1640
After filtering: n = 1605

Frequency

UMI counts (10^4)

n < 300000

Frequency

n > 1000

genes (10^4)

n < 4%

Frequency

mitochondrial gene (%)

Cluster ID

geta1a
cldng
gfl1aa
taf1
fxp1a
hnf1bb
acy3.2
pxk2a
thbs3b
chtnc1a
twist1b
tbx1
junba
dhrs3a
m21a
rpolms2b
foxa2
dap1b
prdx5
nkd2.7
hand2
cdx4
cdf
emilin2b
col15a1b
pkrax2
foxo2a
pitu3
gpx1a
sna1a
eve1
cyp26a1
met2a
tnn.2
pocdi
met2b
bgl2
redg9
lmsb2
barbila
etv2
che
fl1b
kdr
onec2f3
angpt3
link2
hhbe1.3
sec61g
serp1
hef1b
ctsl
cat5a.1
krtcap2
hyor1
inhbae
cldhi
cyt1
myh9a
kr4
myod1
tnn.1
myt5
ofmp2a

Scaled expression

Anterior

Anterior mesoderm

Posterior mesoderm

Cluster 1: erythroid progenitors
Cluster 2: pronephric progenitors
Cluster 3: twist/b+ cranial-pharyngeal mesoderm
Cluster 4: atrium/fin progenitors
Cluster 5: dorsal/anterior endoderm
Cluster 6: putative mesothelium progenitors
Cluster 7: col15a1b+ cranial-pharyngeal mesoderm
Cluster 8: tail bud
Cluster 9: cardiac mesoderm
Cluster 10: presomatic mesoderm
Cluster 11: vascular/myeloid progenitors
Cluster 12: liver/pancreas endoderm
Cluster 13: hatching gland 1
Cluster 14: hatching gland 2
Cluster 15: epidermis
Cluster 16: somitic muscle progenitors
Figure S1

T

Pseudotime trajectory

comp_2

comp_1

Pseudotime

0 20 40 60 80

0 20 40 60 80

0 20 40 60 80

0 20 40 60 80

0 20 40 60 80

WT

6 hpf

8 hpf

10 hpf

13 hpf

anterior endoderm

posterior endoderm

cardiac

arium/off bud

vascular

pronephric

presomitic

erythroid

twist/fb+ cranial-pharyngeal

hatching gland 1

gata6 expression

Pseudotime

0 20 40 60 80

0 20 40 60 80

0 20 40 60 80

0 20 40 60 80

0 20 40 60 80

gata6 expression

Pseudotime

0 20 40 60 80

0 20 40 60 80

0 20 40 60 80

0 20 40 60 80

0 20 40 60 80

gata6 expression

Pseudotime
Fig. S1 Analysis of scRNA-seq on gata5:GFP+ cells from 6 to 13 hpf
(A-D) RNA ISH images of gfp transcript in TgBAC(gata5:GFP) embryos, and gata5 and gata6 expression in WT embryos at 6 hpf (A), 8 hpf (B), 10 hpf (C), and 13 hpf (D). Arrowheads indicate similar expression patterns of gfp, gata5, and gata6. A: anterior, P: posterior. (E, H, K, N) Histograms showing the distribution of UMI counts, numbers of genes detected, and percent of mitochondrial genes detected in the 6 hpf (E), 8 hpf (H), 10 hpf (K), and 13 hpf (N) WT sample before and after filtering. Cyan shaded areas indicate cells that were filtered out to remove low-quality cells and potential doublets. (F, I, L, O) UMAP visualization of the 6 hpf (F), 8 hpf (I), 10 hpf (L), and 13 hpf (O) datasets. WT sample colored by cluster IDs (6 hpf: 879 cells, 8 hpf: 1465 cells, 10 hpf: 1407 cells, 13 hpf: 1605 cells). (G, J, M, P) Heatmap showing the expression of top marker genes for each cluster in the 6 hpf (G), 8 hpf (J), 10 hpf (M), and 13 hpf (P) WT samples. (Q) Force-directed graph showing the connection between all single cells from the four WT samples (6, 8, 10 and 13 hpf). Cells are colored based on their developmental origins (germ layers). Expression of marker gene for each lineage (R) and canonical pharyngeal mesoderm (S) within the forced-directed graph. (T) Pseudotime developmental trajectories of the major lineages visualized in reduced dimensions (component_1 and component_2) and expression dynamics of gata5 and gata6 along each trajectory (x-axis, pseudotime; y-axis, normalized log expression level). In the gene expression scatter plots, lines show smoothed conditional means after local polynomial regression fitting (LOESS method) and shaded areas indicate standard errors.
Figure S2

A

WT + Gata5/6 MO 13 hpf

Before filtering: n = 3550

After filtering: n = 3445

Frequency

UMI counts (10^5)

n < 30000

n > 1000

n < 4%

genes (10^5)

mitochondrial gene (%)

B

Cluster ID

12 13 14 15 16 17 18

C

Population fold change (cardiac cluster removed)

Color by clusters

Reduced

Expanded

Log10(% in Gata5/6MO/% in WT)

% in WT

1% 4% 7% 10%

Scaled expression

-2 -1 0 1 2
Fig. S2 Analysis of scRNA-seq of gata5:GFP+ cells from 13 hpf WT and Gata5/6 KD embryos

(A) Histograms showing the distribution of UMI counts, numbers of genes detected, and percent of mitochondrial genes detected in WT and Gata5/6 MO combined samples at 13 hpf before and after filtering. Cyan shaded areas indicate cells that were filtered out to remove low-quality cells and potential doublets. 

(B) Heatmap showing the expression of top marker genes for each cluster in the 13 hpf WT and Gata5/6 MO combined samples (n = 3448). Five marker genes were plotted for most clusters except the last four clusters (somitic muscle, epidermis and two hatching gland clusters)

(C) Cell composition changes of each cluster between Gata5/6 KD and WT samples after the cardiac cluster was removed. Asterisks indicate significant differences (Fisher’s exact test, Bonferroni correction, adjusted p-value < 0.05). Dot sizes show the percentage of each cluster within the whole WT population before cardiac cluster removal (% total 13 hpf WT cells). Note that removing the cardiac cluster resulted in no significant effect on trends of cell composition changes for most clusters except that the reduction for liver endoderm became statistically significant due to less multiple test correction.
Figure S3

D
Fig. S3 Sub-clustering of the cardiac and pharyngeal lineages at 13 hpf
(A) Heatmap showing the expression of top marker genes for each cardiac and pharyngeal cluster after sub-clustering, with WT and MO cells plotted separately. S1: posterior pharyngeal mesoderm, S2: cranial-pharyngeal mesoderm, S3: cranial-pharyngeal mesoderm, S4: cardiac progenitors, S5: first pharyngeal arch progenitors, S6: cranial-pharyngeal mesoderm, S7: cranial-pharyngeal mesoderm. (B) Sankey diagrams showing relationships between cardiac and pharyngeal sub-clustered populations (Figure 3, after sub-clustering) and cardiac and pharyngeal clusters identified in all 13 hpf WT+Gata5/6KD data (Figure 2, before sub-clustering). (C) Venn diagram showing the overlap of markers genes in the SHF-like sub-cluster (S9, n = 62) and the zebrafish orthologs (n = 221) of the mouse anterior SHF genes (n = 212) identified through scRNA-seq (36). (D) Expression levels of canonical pharyngeal mesoderm genes in 13 hpf scRNA-seq cardiac-pharyngeal sub-clusters.
Figure S4

A
Marker gene for each sub-cluster 13 hpf
- S2: cthrc1a
- S3: kazak2
- S6: foxd1
- S7: sxx3b
- S2: cthrc1a
- S3: kazak2
- S6: foxd1
- S7: sxx3b

WT          Gata5/6 KD

B
- S4: Pan-cardiac / S9: SHF 13 hpf
- nolx2.5 fgg8a nolx2.5 fgg8a
- cyp28c1 fgg8a cyp28c1

C
-fgg8a/nolx2.5 signal distribution

D
- S5: Pharyngeal arch one 13 hpf
- irx1b
- cyp28c1

E
- S9: SHF 13 hpf

F
- S4: Pan-cardiac
- S5: Pharyngeal arch one 13 hpf

G
- Cardiac and pharyngeal mesoderm spatial map 13 hpf

H
Cardiomyocytes and pectoral fin progenitors 24 hpf
- tbx5a
- gate5+/+, gate6+/
- gate5-/-, gate6-/-
- WT, Gata5/6 KD

I
- SHF and pharyngeal 13 hpf
- tbx1
- gata5+, gata6+/
- gata5-, gata6-/-
**Fig. S4 Gene expression analysis on cardiac and pharyngeal subtypes**

(A) RNA ISH hybridization for markers of each sub-cluster in WT embryos and embryos injected with Gata5/6 morpholino (Gata5/6 KD). Arrowheads indicate cranial-pharyngeal or head mesoderm-related expression domains of these genes that potentially overlap gata5:GFP+ marked regions. A: anterior, P: posterior. (B) Confocal images of nkx2.5 (S4: pan-cardiac, pseudocolored green) and fgf8a (S9: SHF, pseudocolored magenta) expression domain at 13 hpf as indicated by RNAscope double-fluorescent RNA *in situ* hybridization. The boxed region is shown at a higher magnification and with DAPI staining. A: anterior, P: posterior, M: medial, L: lateral. Scale bar: 20 μm. (C) Quantifications of fgf8a and nkx2.5 expression signals based on percentage of positive cells (top) or pixels (bottom). Solid lines: smooth lines; dash lines: average moving trendlines, period = 2. (D) Double FISH against irx1b and cyp26c1 (S5: pharyngeal arch one) together with GFP immunostaining on 13 hpf TgBAC(gata5:EGFP) embryos. (E) RNA ISH against wnt11r at 13 hpf. (F) Double FISH against nkx2.5 (S4: pan-cardiac) and cyp26c1 (S5: pharyngeal arch one) at 13 hpf. Scale bar: 20 μm. The white arrowhead indicates foci double positive for nkx2.5 and cyp26c1. (G) Schematic representation of the spatial organization of cardiac and pharyngeal sub-clusters at 13 hpf. r2: rhombomere 2; r4: rhombomere 4; r6: rhombomere 6. Dotted line boxes: clusters whose spatial organization was not confirmed by high-resolution gene expression assays. (H) RNA ISH against tbx5a at 24 hpf. Arrowheads indicate loss of tbx5a expression in the heart. Asterisks represent reduced tbx5a expression in pectoral fin bud progenitors. (I) RNA ISH against tbx1 (S2, S5, S9). Arrowheads show that tbx1 expression in anterior lateral mesoderm is largely unaffected upon loss of gata5/6. All scale bars represent 100 μm unless otherwise specified.
Figure S5

I

mef2cb expression pattern at 13 hpf

| Genotype | n |
|----------|---|
| g5 +/+ g6 +/+ | 4 |
| g5 +/+ g6 +/− | 3 |
| g5 +/− g6 +/+ | 2 |
| g5 +/− g6 +/− | 1 |
| g5 +/− g6 −/− | 0 |

J

aldh1a2 expression pattern at 13 hpf

| Genotype | n |
|----------|---|
| g5 +/+ g6 +/+ | 4 |
| g5 +/+ g6 +/− | 3 |
| g5 +/− g6 +/+ | 2 |
| g5 +/− g6 +/− | 1 |
| g5 +/− g6 −/− | 0 |
Figure S5

K

**tbx20 expression pattern at 13 hpf**

| Genotype   | n=2 | n=3 | n=9 | n=21 | n=2 | n=3 | n=6 | n=8 | n=3 |
|------------|-----|-----|-----|------|-----|-----|-----|-----|-----|
| g5 +/-     | 4   | 3   | 2   | 1    | 1   | 0   |
| g6 +/-     |     |     |     |      |     |     |

Cardiac

\[ gata5/6 \text{ gene dosage} \]

L

**hand2 expression pattern at 13 hpf**

| Genotype   | n=11 | n=15 | n=15 | n=33 | n=3 | n=6 | n=10 | n=15 | n=4 |
|------------|------|------|------|------|-----|-----|------|------|-----|
| g5 +/-     | 4    | 3    | 2    | 1    | 0   |
| g6 +/-     |     |     |     |      |     |     |

Cardiac

\[ gata5/6 \text{ gene dosage} \]
### Figure S5

#### M

| Gene | Time | Condition |
|------|------|-----------|
| nkd2.5 | 11.5 hpf | WT |
| nkd2.5 | 12 hpf | WT |
| nkd2.5 | 13 hpf | WT |
| nkd2.5 | 14 hpf | WT |
| nkd2.5 | 24 hpf | WT |
| nkd2.5 | 28 hpf | WT |

| Gene | Time | Condition |
|------|------|-----------|
| nkd2.5 | 20/20 | Gata5/6MO |
| nkd2.5 | 18/18 | Gata5/6MO |
| nkd2.5 | 20/20 | Gata5/6MO |
| nkd2.5 | 19/19 | Gata5/6MO |
| nkd2.5 | 42/42 | Gata5/6MO |
| nkd2.5 | 22/22 | Gata5/6MO |

#### N

| Gene | Time | Condition |
|------|------|-----------|
| tbx1 | 11.5 hpf | WT |
| tbx1 | 12 hpf | WT |
| tbx1 | 19/19 | WT |

| Gene | Time | Condition |
|------|------|-----------|
| tbx1 | 20/20 | Gata5/6MO |
| tbx1 | 18/18 | Gata5/6MO |
Figure S5

Q. gata4 expression in WT VS Morphants 10 and 13 hpf

P. gata4 in WT VS Mutants 13 hpf

Q: S1: Posterior pharyngeal arches 13 hpf
- aldh1a2
- gata4+/-, gata5+/-
- gata4-/-, gata5+/-

R: Expression domain size (A.U.)
- n.s.
- ***

S: S5: Pharyngeal arch one 13 hpf
- lrx1b
- gata4+/-, gata5+/-
- gata4-/-, gata5+/-

T: Expression domain size (A.U.)
- n.s.
- ***

U: GFP, gata4, gata4dGFP

V: WT, gata5KD, gata5GFP+
- Peaks in WT GFP+ VS GSK6D GFP+
- Refseq Genes

Legend:
- WT
- gata5KD
- gata5GFP+
- Scale chr20
- 53,000,000 | 53,090,000 | 53,100,000 | 53,160,000
Fig. S5 Characterization of gata4, gata5, and gata6 null mutants

(A) Schematic representation of the gata5 and gata6 null-alleles generated through CRISPR/Cas9-mediated genome editing. A 29bp deletion was generated in exon2 of gata5. An 11bp deletion was introduced in exon2 of gata6. Asterisks indicate the premature stop-codon positions, at amino acid 233 in the first zinc finger domain of the Gata5 protein and amino acid 332 in the first zinc finger domain of the Gata6 protein. GATA-N: N-terminal GATA-type transcription activation domain; ZNF-GATA: Zinc finger DNA binding domain. (B) Bright-field images of Tg(myl7:EGFP) at 48 hpf in a gata5 null mutant, a gata6 null mutant, a gata5/gata6 double mutant and a sibling control from incrossing gata5/gata6 compound heterozygous mutants. (C) Confocal images of gata5 and gata6 single mutants and their WT siblings in Tg(myl7:EGFP) and in Tg(gata1a;dsRED) backgrounds at 48 hpf. Yellow arrowheads indicate normal blood distribution in WT, blood pooling in gata5-/-, and impaired circulation in gata6-/- embryos. V: ventricle, A: atrium. (D) Bright-field images of the pectoral fins in WT, Gata5/6 knockdown, and gata5; gata6 compound homozygous mutant backgrounds at 72 hpf. Black arrowheads indicate the pectoral fins. (E) RNA ISH against known cardiac progenitor genes (nkx2.5, tbx20, mef2cb, and hand2) in gata5/6 double mutants and WT siblings. (F) RNA ISH against known cardiac genes (myl7, tbx20, hand2) in WT embryos and embryos injected with Gata5/6 morpholinos (Gata5/6 KD). (G-L) Gene expression analysis of a marker gene for the cardiac (S4: nkx2.5 (G), mef2cb (I), tbx20 (K) and hand2 (L)) or pharyngeal (S1: irx1b (H), S1: aldh1a2 (J)) lineage in nine different genotypes that were obtained from incrossing gata5+/-; gata6+/- double heterozygous mutants. Embryos are grouped into different categories according to their gene expression patterns. Representative embryo patterns are shown in the top (G, I, K) or bottom (H, J, L) panels. For each genotype, embryos that fall into different categories are counted and shown in the bar plots. As the gata5/6 dosage goes down, the cardiac progenitor population decreases with a concurrent expansion of the pharyngeal population. (M) RNA ISH against nkx2.5 in WT and Gata5/6 KD embryos from early segmentation (11.5 hpf) to pharyngula (28 hpf) stages. (N) RNA ISH against tbx1 in WT and Gata5/6 KD embryos at 11.5 and 12 hpf. (O) RNA ISH against gata4 in WT and Gata5/6 KD embryos at 10 and 13 hpf. (P) RNA ISH against gata4 in in gata5/6 double mutants and WT siblings at 13 hpf. (Q) RNA ISH against the posterior pharyngeal arch progenitor (S1) marker aldh1a2 in gata4+/-, gata5+/-; gata4+/-, gata5-/-; gata4-/-, gata5+/-; and gata4-/-, gata5-/- embryos at 13 hpf. (R) Quantification of aldh1a2 expression intensity and expression domain sizes of four genotypes accessed in (Q). (S) RNA ISH against the pharyngeal arch one progenitor (S5) marker aldh1a2 in gata4+/-, gata5+/-; gata4+/-, gata5-/-; gata4-/-, gata5+/-; and gata4-/-, gata5-/- embryos at 13 hpf. (T)
Quantification of *irx1b* expression intensity and expression domain sizes of four genotypes accessed in (S). Only lateral staining was quantified. T-test was used to determine statistical significance. ***: p-value < 0.001, **: p-value < 0.01, n.s: p-value > 0.05. (U) FISH against *gata4* transcripts and immunostaining against GFP at 13 hpf in the *TgBAC(gata5:GFP)* background. (V) Genome browser view of the ATAC-seq signals and peaks at the *gata4* loci. Grey sticks: shared peaks, orange sticks: closed DARs in GFP+ cells. All scale bars represent 100 μm.
Figure S6

A

B

C

D

E

F

GFP+/specific
(n=8323)

GFP- specific
(n=7279)

WT

G4

WT

G4

GFP+ specific

WT VS

WT

GFP+ specific

-0.5kb peak center

0.5kb peak center

-0.5kb peak center

0.5kb peak center

Interactive panel

GTAT

GFP+ specific

WT gata5GFP

Gata6 KD gata5GFP

WT gata5GFP

Gata6 KD gata5GFP

-8000 -6000 -4000 -2000 0 2000 4000 6000 Set Size

GWG

D

WT

GFP+ specific

regulation of vascular endothelial growth factor receptor signaling pathway
response to growth factor stimulus
embryonic heart tube development
embryonic pattern specification
embryonic camera-type eye development
embryonic visceral organ morphogenesis
Transcription factor, T-box
Transcription factor, T-box, conserved site
Homeobox KN domain

Fold of enrichment
+ 2.01
+ 2.5
+ 3.0
+ 3.5

-10(log10(BinomFDRQ))

GO Biological Process

GO Molecular Function

InterPro

Wiki Pathways

E

GFP+ closed DMRs

GATA2

GATA3

Zic2

Zic3

WT GFP+ specific

-2000 -1000 0 1000 2000 3000 4000 5000 6000 7000 8000 Prob.

WT GFP+ specific

GATA2

GATA3

Zic2

Zic3

WT GFP+ specific

Sox13

Sox4

Sox5

Sox6

Sox9

Sox11

Enrichment Score

1.5
2.0
2.5

F

WT GFP+ specific

GATA2

GATA3

Zic2

Zic3

WT GFP+ specific

Sox13

Sox4

Sox5

Sox6

Sox9

Sox11

Enrichment Score

1.5
2.0
2.5

3.0

AP-2γ
Fig. S6 ATAC-seq in *gata5*:GFP+ and *gata5*:GFP- cells from WT and Gata5/6 KD embryos at 8 hpf.

(A) Heatmap showing the GFP+-/- specific peaks identified in WT Tg (*gata5*:GFP) embryos. The read intensity within 3 kb of the peak center was plotted for each peak. (B) Aggregate plots showing the ATAC-seq signals in GFP+-/- specific peaks identified in WT Tg (*gata5*:GFP) embryos. (C) UpSet plot showing the overlap between GFP+-/- specific peaks and the closed/open DARs. (D) Barplot showing the top 10 most enriched terms (or all enriched terms if the total number < 10) obtained from GFP+-/- specific peaks using GREAT analysis. Terms from four categories were plotted. (E) Probabilities of the top five enriched motifs (or all enriched motifs if the total number < 5; significant threshold: Fisher E-value < 0.05) within the GFP+-/- specific peaks calculated by CentriMo. Each curve shows the probability of the best match to a given motif occurring at a given position in the input sequences. Solid lines represent probabilities calculated from query peak sets (GFP+-/- specific peaks, or closed/open DARs) while dash lines show that from the background sequences (the corresponding shared peaks). Fisher E-value was used to determine the statistical significance (Fisher E-value < 0.05) and showed for each plotted motif. (F) Motif enrichments identified by Homer within WT GFP+ specific and WT GFP-specific peaks. The top 12 motifs are plotted.
Figure S7

B

marker genes of WT+Gata5/6KD 13 hpf cluster1: Cranial-pharyngeal mesoderm (n = 110)

zebrafish orthologs (n = 79) of mouse PM genes (n = 75) Soysa et al. 2019

66 13 97

twist1a twist1b meox1
colec12 tbx1 fsta
tbx1 dedx2 six1a sfxb
foca1e foxc1e

marker genes of WT+Gata5/6KD 13 hpf cluster2: Cranial-pharyngeal mesoderm (n = 40)

zebrafish orthologs (n = 56) of mouse BM genes (n = 39) Soysa et al. 2019

50 6 104

tbx1 fsta foxd2
sdx1a sdx1b
cdh11

marker genes of WT+Gata5/6KD 13 hpf cluster5: Cranial-pharyngeal mesoderm (n = 136)

zebrafish orthologs (n = 79) of mouse PM genes (n = 75) Soysa et al. 2019

77 2 38

cdh11 mycn

twist1a cole12 foxd2
sdx1a sdx1b
foca1e foxc1e

marker genes of WT+Gata5/6KD 13 hpf cluster13: Cranial-pharyngeal mesoderm (n = 75)

zebrafish orthologs (n = 79) of marker genes of WT+Gata5/6KD 13 hpf cluster13: Cranial-pharyngeal mesoderm (n = 75) Soysa et al. 2019

74 5 70

twist1a tbx1 sdx1a
sdx1b foxc1e

marker genes of WT+Gata5/6KD 13 hpf cluster13: Cranial-pharyngeal mesoderm (n = 75)

zebrafish orthologs (n = 56) of mouse BM genes (n = 39) Soysa et al. 2019

51 5 70

ebf3a tbx1 sdx1a
sdx1b nppb
Figure S7

C. zebrafish orthologs (n = 79) of mouse PM genes (n = 75)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 1: posterior pharyngeal mesoderm
(n = 22)
cdh11

zebrafish orthologs (n = 79) of mouse PM genes (n = 75)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 2: Cranial-pharyngeal mesoderm
(n = 68)
twist1a
twist1b
tbx1
fstaa
foxc1b
mex1
cole12

zebrafish orthologs (n = 56) of mouse BM genes (n = 39)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 2: Cranial-pharyngeal mesoderm
(n = 68)
cdh11

zebrafish orthologs (n = 56) of mouse BM genes (n = 39)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 3: Cranial-pharyngeal mesoderm
(n = 62)
twist1a
twist1b
fstaa
mex1
cole12
inx3a
six1b
foxd2
cdh11
tcf15

zebrafish orthologs (n = 56) of mouse BM genes (n = 39)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 3: Cranial-pharyngeal mesoderm
(n = 62)
fsta
cdh11

zebrafish orthologs (n = 56) of mouse BM genes (n = 39)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 5: first pharyngeal arch prog
(n = 58)
six1a
six1b

zebrafish orthologs (n = 56) of mouse BM genes (n = 39)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 5: first pharyngeal arch prog
(n = 58)
six1a
six1b
tbx1
ebf3a
rgp2b

zebrafish orthologs (n = 56) of mouse BM genes (n = 39)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 6: Cranial-pharyngeal mesoderm
(n = 96)
twist1a
tcf15
inx5b
foxd2
lxpa1
colec12
foxc1a
foxc1b

zebrafish orthologs (n = 56) of mouse BM genes (n = 39)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 6: Cranial-pharyngeal mesoderm
(n = 96)
six1a
ramp2
foxd2
efb3a
lxpa1

zebrafish orthologs (n = 56) of mouse BM genes (n = 39)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 7: Cranial-pharyngeal mesoderm
(n = 97)
twist1a
six1a
foxd2

zebrafish orthologs (n = 56) of mouse BM genes (n = 39)
Soyza et al. 2019

marker genes of WT+Gata5/6KD 13 hpf
subcluster 7: Cranial-pharyngeal mesoderm
(n = 97)
ebf3a
six1a
foxd2
Fig. S7 Mouse and zebrafish cardiac and pharyngeal marker gene comparisons
Venn diagrams showing the overlap of the zebrafish orthologs (PM: n = 79; BM: n = 56) of the mouse PM (n = 75) and BM genes (n = 39) identified through scRNA-seq (36) and markers genes in (A) the 13 hpf WT-only pharyngeal mesoderm clusters (Cluster3, n = 146 and Cluster7, n = 170), (B) the 13 hpf WT+Gata5/6KD pharyngeal mesoderm clusters (Cluster1, n = 110; Cluster2, n = 40; Cluster5, n = 136; and Cluster13, n = 45), and (C) the 13 hpf WT+Gata5/6KD pharyngeal mesoderm sub-clusters (S1, n = 22; S2, n = 68; S3, n = 62; S5, n = 58; S6, n = 96; and S7, n = 97), respectively. PM: paraxial mesoderm, BM: branchiomeric muscles.
Fig. S8 Activity of independent alleles of tbx1 DARs
Fluorescent images of F2 transgenic lines at 50-60 hpf, generated using tbx1 closed DAR sequences: (A) closed_1<sup>hsc171</sup>, (B) closed_2<sup>hsc173</sup>, and (C-G) closed_3<sup>hsc175-179</sup>. (H) Confocal (lower) images of Tg(tbx1<sub>open</sub>:GFP, tcf21:dsRed) at 72 hpf from up to bottom: alleles hsc183, hsc181, hsc184. Confocal images of tbx1 closed DAR transgenics: Tg(tbx1<sub>closed1</sub>:GFP) (I), Tg(tbx1<sub>closed1</sub>:GFP) (K), Tg(tbx1<sub>closed1</sub>:GFP) (M) at 13 hpf. Real-time activity of tbx1 closed DARs illustrated by gfp RNA ISH in Tg(tbx1<sub>closed1</sub>:GFP) (J) at 13 and 18 hpf, Tg(tbx1<sub>closed1</sub>:GFP) (L) at 10 and 13 hpf, Tg(tbx1<sub>closed1</sub>:GFP) (N) at 10 and 13 hpf. Arrowheads indicate the position of heart-forming regions. All scale bars represent 100 μm.
| Transgenic Line | Reference |
|----------------|-----------|
| TgBAC (gata5:GFP)pd25 | Kikuchi et al., 2011 (103) |
| TgBAC (nkx2.5:ZsYellow)kb7 | Zhou et al., 2011 (104) |
| Tg(gata1a:dsRed)sd2 | Traver et al., 2003 (105) |
| Tg(acta1:EGFP)zd13 | Higashijima et al., 1997 (106) |
| Tg(myl7:EGFP)twu34 | Huang et al., 2003 (107) |
| TgBAC(tcf21:NSL-EGFP)pd41 | Wang et al., 2011 (108) |
| TgBAC(tcf21:dsRed)pd37 | Wang et al., 2011 (108) |
| Tg(kdrl:EGFP)s843 | Jin et al., 2005 (109) |
| sih (tnnt2ab109) | Sehnert et al., 2002 (110) |
| gata4wcm6/gata5wcm8 | Sam et al., 2020 (3) |

Table S1 Previously established transgenic and mutant lines used in this study
### Table S2 gRNA design for targeting *gata5* and *gata6*

Target sequence in red: PAM sequence. Whole primer in Red: T7 promoter (18 bp), in blue: target sequence without PAM, in green: overlap sequence with Oligo2

|          | *gata5*-gRNA                                                                 |          | *gata6*-gRNA                                                                 |
|----------|-------------------------------------------------------------------------------|----------|-------------------------------------------------------------------------------|
| **Target** | ggccgcgagttgtgaactg cgg                                                        | **Target** | gcggtcgttcacagcg ggg                                                          |
| **sequence** |                                                                                 | **sequence** |                                                                                 |
| **Whole**  | TTAATACGACTCACTATAGggccgcgagttgtgaactgGTTTTAGAGCTAGAAATAGC                   | **primer** | TTAATACGACTCACTATAGgcggtcgttcacagcgGTTTTAGAGCTAGAAATAGC                      |
| Gene  | Reference                                      |
|-------|-----------------------------------------------|
| myl7  | Yelon et al., 1999 (111)                      |
| nkx2.5| Chen and Fishman, 1996 (112)                  |
| tbx20 | Szeto et al., 2002 (113)                      |
| mef2cb| Lazic and Scott, 2011 (114)                   |
| hand2 | Schoenebeck et al., 2007 (115)                |
| gata4 | Reiter et al., 1999 (9)                       |
| gata5 | Reiter et al., 1999 (9)                       |
| gata6 | Reiter et al., 1999 (9)                       |
| gfp   | Yuan et al., 2018 (48)                        |
| fgf8a | Reifers et al., 1998 (116)                    |
| tbx5a | Ruvinsky et al., 2000 (117)                   |

Table S3 Previously established probes used in this study
| Location                | WT_6hpf          | WT_8hpf          | WT_10hpf        | WT_13hpf        |
|------------------------|------------------|------------------|----------------|----------------|
| **anterior endoderm**   | c6_dorsal-endo   | c4_dorsal-endo   | c1_anterior-endo| c5_anterior-endo|
| **posterior endoderm**  |                  | c1_ventral-endo  | c6_posterior-endo, c10_posterior-endo |
| **cardiac**             |                  |                  | c7_cardiac      | c9_cardiac      |
| **atrium pectoral fin**|                  |                  | c2_anterior-meso| c4_atrium-fin-progenitors |
| **vascular**            |                  |                  | c2_anterior-meso| c11_vascular    |
| **erythroid**           |                  |                  | c5_posterior-meso| c1_erythroid |
| **presomitic**          |                  |                  | c3_posterior-meso, c5_posterior-meso | c10_presomitic |
| **pronephric**          |                  |                  | c3_posterior-meso | c2_pronephric |
| **twist1b pharyngeal**  | c2_dorsal-meso   | c5_cranial-pharyngeal, c4_cranial-pharyngeal | c3_cranial-pharyngeal |

**Legend:**
- c6_dorsal-endoderm
- c4_dorsal-endoderm
- c1_anterior-endoderm
- c5_anterior-endoderm
- c1_ventral-endoderm
- c6_posterior-endoderm
- c10_posterior-endoderm
- c2_lateral-meso
- c2_anterior-meso
- c3_ventral-meso
- c3_posterior-meso
- c5_posterior-meso
- c7_cardiac
- c9_cardiac
- c1_ventral-lateral-meso
- c2_lateral-meso
- c2_anterior-meso
- c3_ventral-meso
- c3_posterior-meso
- c4_atrium-fin-progenitors
- c5_posterior-meso
- c6_posterior-endoderm
- c11_vascular
- c10_presomitic
- c2_pronephric
- c3_cranial-pharyngeal
|                |          | c6 Cranial-Pharyngeal | c9 Cranial-Pharyngeal |
|----------------|----------|-----------------------|-----------------------|
| *coll15a1b*    | pharyngeal | c2 dorsal-meso        | c5 Cranial-Pharyngeal|
| hatching       | gland 1   | c3 axial-meso         | c6 Cranial-Pharyngeal|
| gland 2        | c3 axial-meso | c7 hatching-gland   | c8 hatching-gland     |
|                |          |                       |                       |

**Table S4 WT Time-series single-cell trajectories**
For each lineage (each row), the cluster number (c: cluster) together with the cluster annotation is shown for each time point.
| Sub-cluster | Gene    | Reference                                      |
|------------|---------|------------------------------------------------|
| S1         | aldh1a2 | Grandel et al., 2002 (118)                     |
| S2         | efn2b   | Piotrowski and Nusslein-Volhard et al., 2000 (119) |
| S3         | tenm3   | Cheung et al., 2019 (120)                      |
| S5         | cyp26c1 | Gu et al., 2005 (35)                           |
| S6         | foxd1   | Thisse (http://zfin.org) (121)                 |

Table S5 Genes used for defining structures derived from pharyngeal sub-clusters
| Gene name  | Gene id | Closest region | open DAR (GRCz11)     |
|------------|---------|----------------|----------------------|
| ENSDART00000193255 | *ebf2*  | gap            | chr5-67413046-67413546 |
| ENSDART00000043595  | *twist1a* | utr            | chr19-2201369-2201869  |
| ENSDART00000043857  | *irx5a*  | exon           | chr7-35731311-35731811  |
| ENSDART00000172171  | *ebf3a*  | gap            | chr14-35083548-35084048  |
| ENSDART00000144915  | *tbx1*   | exon           | chr5-15173726-15174226  |
| ENSDART00000193255  | *ebf2*   | gap            | chr5-67422402-67422902  |
| ENSDART00000138176  | *foxd2*  | utr            | chr8-19691367-19691867  |
| ENSDART00000138176  | *foxd2*  | utr            | chr8-19698479-19698979  |

Table S6 Open DARs near conserved pharyngeal genes

Data S1 Single-cell RNA-seq and ATAC-seq experiment metrics

Data S2 Full result of the gene set enrichment analysis of the mouse orthologous genes in the SHF-like sub-cluster

Data S3 ATAC-seq peaks coordinates (GRCz11)

Data S4 ATAC-seq functional (GREAT) and motif (Homer, CentriMO) enrichment analysis results

Data S5 Motif scan results of the four DARs at the *tbx1* locus.
REFERENCES AND NOTES

1. B. G. Bruneau, Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harb. Perspect. Biol.* **5**, a008292 (2013).

2. A. Holtzinger, T. Evans, Gata5 and Gata6 are functionally redundant in zebrafish for specification of cardiomyocytes. *Dev. Biol.* **312**, 613–622 (2007).

3. J. Sam, E. J. Mercer, I. Torregroza, K. M. Banks, T. Evans, Specificity, redundancy and dosage thresholds among gata4/5/6 genes during zebrafish cardiogenesis. *Biol. Open* **9**, 053611 (2020).

4. R. Zhao, A. J. Watt, M. A. Battle, J. Li, B. J. Bondow, S. A. Duncan, Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. *Dev. Biol.* **317**, 614–619 (2008).

5. K. Ragkousi, J. Beh, S. Sweeney, E. Starobinska, B. Davidson, A single GATA factor plays discrete, lineage specific roles in ascidian heart development. *Dev. Biol.* **352**, 154–163 (2011).

6. K. Gajewski, N. Fossett, J. D. Molkentin, R. A. Schulz, The zinc finger proteins Pannier and GATA4 function as cardiogenic factors in Drosophila. *Development* **126**, 5679–5688 (1999).

7. S. U. Morton, D. Quiat, J. G. Seidman, C. E. Seidman, Genomic frontiers in congenital heart disease. *Nat. Rev. Cardiol.* **19**, 26–42 (2021).

8. J. D. Molkentin, Q. Lin, S. A. Duncan, E. N. Olson, Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev.* **11**, 1061–1072 (1997).

9. J. F. Reiter, J. Alexander, A. Rodaway, D. Yelon, R. Patient, N. Holder, D. Y. Stainier, Gata5 is required for the development of the heart and endoderm in zebrafish. *Genes Dev.* **13**, 2983–2995 (1999).

10. T. Peterkin, A. Gibson, R. Patient, Redundancy and evolution of GATA factor requirements in development of the myocardium. *Dev. Biol.* **311**, 623–635 (2007).
11. T. Peterkin, A. Gibson, R. Patient, Common genetic control of haemangioblast and cardiac development in zebrafish. *Development* **136**, 1465–1474 (2009).

12. M. Heikinheimo, J. M. Scandrett, D. B. Wilson, Localization of transcription factor GATA-4 to regions of the mouse embryo involved in cardiac development. *Dev. Biol.* **164**, 361–373 (1994).

13. W. P. Devine, J. D. Wythe, M. George, K. Koshiba-Takeuchi, B. G. Bruneau, Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife* **3**, e03848 (2014).

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

15. F. Lescroart, X. Wang, X. Lin, B. Swedlund, S. Gargouri, A. Sànchez-Dànes, V. Moignard, C. Dubois, C. Paulissen, S. Kinston, B. Göttgens, C. Blanpain, Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science* **359**, 1177–1181 (2018).

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

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

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

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

20. I. Harel, Y. Maezawa, R. Avraham, A. Rinon, H.-Y. Ma, J. W. Cross, N. Leviatan, J. Hegesh, A. Roy, J. Jacob-Hirsch, G. Rechavi, J. Carvajal, S. Tole, C. Kioussi, S. Quaggin, E. Tzahor,
Pharyngeal mesoderm regulatory network controls cardiac and head muscle morphogenesis. Proc. Natl. Acad. Sci. U.S.A. 109, 18839–18844 (2012).

21. B. Guner-Ataman, J. M. González-Rosa, H. N. Shah, V. L. Butty, S. Jeffrey, M. Abrial, L. A. Boyer, C. G. Burns, C. E. Burns, Failed progenitor specification underlies the cardiopharyngeal phenotypes in a zebrafish model of 22q11.2 deletion syndrome. Cell Rep. 24, 1342–1354.e5 (2018).

22. L. Tirosh-Finkel, H. Elhanany, A. Rinon, E. Tzahor, Mesoderm progenitor cells of common origin contribute to the head musculature and the cardiac outflow tract. Development 133, 1943–1953 (2006).

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

24. J. Liu, H. Cheng, M. Xiang, L. Zhou, B. Wu, I. P. Moskowitz, K. Zhang, L. Xie, Gata4 regulates hedgehog signaling and Gata6 expression for outflow tract development. PLOS Genet. 15, e1007711 (2019).

25. W.-F. Tseng, T.-H. Jang, C.-B. Huang, C.-H. Yuh, An evolutionarily conserved kernel of gata5, gata6, otx2 and prdm1a operates in the formation of endoderm in zebrafish. Dev. Biol. 357, 541–557 (2011).

26. D. E. Wagner, C. Weinreb, Z. M. Collins, J. A. Briggs, S. G. Megason, A. M. Klein, Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science 360, 981–987 (2018).

27. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of community hierarchies in large networks. J. Stat. Mech. Theory Exp. JSTAT, 1–6 (2008).

28. J. A. Farrell, Y. Wang, S. J. Riesenfeld, K. Shekhar, A. Regev, A. F. Schier, Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science 360, eaar3131 (2018).
29. K. D. Prummel, H. L. Crowell, S. Nieuwenhuize, E. C. Brombacher, S. Daetwyler, C. Soneson, J. Kresoja-Rakic, M. Ronner, A. Kocere, A. Ernst, Z. Labbaf, D. E. Clouthier, A. B. Firulli, H. Sanchez-Iranzo, R. O’Rourke, E. Raz, N. Mercader, A. Burger, E. Felley-Bosco, J. Huiskken, M. D. Robinson, C. Mosimann, Hand2 delineates mesothelium progenitors and is reactivated in mesothelioma. bioRxiv, 2020.11.11.355693 (2020).

30. M. Tremblay, O. Sanchez-Ferras, M. Bouchard, GATA transcription factors in development and disease. Development 145, dev164384 (2018).

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

32. K. E. Martin, J. S. Waxman, Atrial and sinoatrial node development in the zebrafish heart. J. Cardiovasc. Dev. Dis. 8, 15 (2021).

33. R. Ilagan, R. Abu-Issa, D. Brown, Y.-P. Yang, K. Jiao, R. J. Schwartz, J. Klingensmith, E. N. Meyers, Fgf8 is required for anterior heart field development. Development 133, 2435–2445 (2006).

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

35. T. Y. de Soysa, S. S. Ranade, S. Okawa, S. Ravichandran, Y. Huang, H. T. Salunga, A. Schricker, A. Del Sol, C. A. Gifford, D. Srivastava, Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. Nature 572, 120–124 (2019).

36. L. A. Dyer, M. L. Kirby, The role of secondary heart field in cardiac development. Dev. Biol. 336, 137–144 (2009).

37. A. B. Rydeen, J. S. Waxman, Cyp26 enzymes are required to balance the cardiac and vascular lineages within the anterior lateral plate mesoderm. Development 141, 1638–1648 (2014).
38. N. Paffett-Lugassy, R. Singh, K. R. Nevis, B. Guner-Ataman, E. O’Loughlin, L. Jahangiri, R. P. Harvey, C. G. Burns, C. E. Burns, Heart field origin of great vessel precursors relies on nkx2.5-mediated vasculogenesis. *Nat. Cell Biol.* **15**, 1362–1369 (2013).

39. N. Paffett-Lugassy, N. Novikov, S. Jeffrey, M. Abrial, B. Guner-Ataman, S. Sakthivel, C. E. Burns, C. G. Burns, Unique developmental trajectories and genetic regulation of ventricular and outflow tract progenitors in the zebrafish second heart field. *Development* **144**, 4616–4624 (2017).

40. T. F. Schilling, C. B. Kimmel, Musculoskeletal patterning in the pharyngeal segments of the zebrafish embryo. *Development* **124**, 2945–2960 (1997).

41. B. R. Keegan, D. Meyer, D. Yelon, Organization of cardiac chamber progenitors in the zebrafish blastula. *Development* **131**, 3081–3091 (2004).

42. G. Begemann, P. W. Ingham, Developmental regulation of Tbx5 in zebrafish embryogenesis. *Mech. Dev.* **90**, 299–304 (2000).

43. T. B. Duong, P. Ravisankar, Y. C. Song, J. T. Gafranek, A. B. Rydeen, T. E. Dohn, L. A. Barske, J. G. Crump, J. S. Waxman, Nr2f1a balances atrial chamber and atrioventricular canal size via BMP signaling-independent and -dependent mechanisms. *Dev. Biol.* **434**, 7–14 (2018).

44. D. Nagelberg, J. Wang, R. Su, J. Torres-Vázquez, K. L. Targoff, K. D. Poss, H. Knaut, Origin, specification, and plasticity of the great vessels of the heart. *Curr. Biol.* **25**, 2099–2110 (2015).

45. F. O. Kok, M. Shin, C.-W. Ni, A. Gupta, A. S. Grosse, A. van Impel, B. C. Kirchmaier, J. Peterson-Maduro, G. Kourkoulis, I. Male, D. F. DeSantis, S. Sheppard-Tindell, L. Ebarasi, C. Betsholtz, S. Schulte-Merker, S. A. Wolfe, N. D. Lawson, Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev. Cell* **32**, 97–108 (2015).

46. L. A. Cirillo, F. R. Lin, I. Cuesta, D. Friedman, M. Jarnik, K. S. Zaret, Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
47. X. Yuan, M. Song, P. Devine, B. G. Bruneau, I. C. Scott, M. D. Wilson, Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. *Nat. Commun.* **9**, 4977 (2018).

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

49. W. Wang, X. Niu, T. Stuart, E. Jullian, W. M. Mauck, R. G. Kelly, R. Satija, L. Christiaen, A single-cell transcriptional roadmap for cardiopharyngeal fate diversification. *Nat. Cell Biol.* **21**, 674–686 (2019).

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

51. F. Razy-Krajka, K. Lam, W. Wang, A. Stolfi, M. Joly, R. Bonneau, L. Christiaen, Collier/OLF/EBF-dependent transcriptional dynamics control pharyngeal muscle specification from primed cardiopharyngeal progenitors. *Dev. Cell* **29**, 263–276 (2014).

52. S. Zhang, W. Cui, Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J. Stem Cells* **6**, 305–311 (2014).

53. C. Vicente, A. Conchillo, M. A. García-Sánchez, M. D. Odero, The role of the GATA2 transcription factor in normal and malignant hematopoiesis. *Crit. Rev. Oncol. Hematol.* **82**, 1–17 (2012).

54. M. Iwafuchi-Doi, K. S. Zaret, Cell fate control by pioneer transcription factors. *Development* **143**, 1833–1837 (2016).

55. M. Koutsourakis, A. Langeveld, R. Patient, R. Beddington, F. Grosveld, The transcription factor GATA6 is essential for early extraembryonic development. *Development* **126**, 723–732 (1999).

56. C. S. Simon, L. Zhang, T. Wu, W. Cai, N. Saiz, S. Nowotschin, C.-L. Cai, A.-K. Hadjantonakis, A *Gata4* nuclear GFP transcriptional reporter to study endoderm and cardiac development in the mouse. *Biol. Open* **7**, bio036517 (2018).
57. T. B. Duong, A. Holowiecki, J. S. Waxman, Retinoic acid signaling restricts the size of the first heart field within the anterior lateral plate mesoderm. *Dev. Biol.* **473**, 119–129 (2021).

58. A. He, F. Gu, Y. Hu, Q. Ma, L. Yi Ye, J. A. Akiyama, A. Visel, L. A. Pennacchio, W. T. Pu, Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat. Commun.* **5**, 4907 (2014).

59. C. Racioppi, K. A. Wiechecki, L. Christiaen, Combinatorial chromatin dynamics foster accurate cardiopharyngeal fate choices. *eLife* **8**, e49921 (2019).

60. N. Stutt, M. Song, M. D. Wilson, I. C. Scott, Cardiac specification during gastrulation—The yellow brick road leading to Tinman. *Semin. Cell Dev. Biol.* 10.1016/j.semcdb.2021.11.011 (2021).

61. Y.-S. Ang, R. N. Rivas, A. J. S. Ribeiro, R. Srivas, J. Rivera, N. R. Stone, K. Pratt, T. M. A. Mohamed, J.-D. Fu, C. I. Spencer, N. D. Tippens, M. Li, A. Narasimha, E. Radzinsky, A. J. Moon-Grady, H. Yu, B. L. Pruitt, M. P. Snyder, D. Srivastava, Disease model of GATA4 mutation reveals transcription factor cooperativity in human cardiogenesis. *Cell* **167**, 1734–1749.e22 (2016).

62. M. Ieda, J.-D. Fu, P. Delgado-Olguin, V. Vedantham, Y. Hayashi, B. G. Bruneau, D. Srivastava, Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).

63. J. K. Takeuchi, B. G. Bruneau, Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* **459**, 708–711 (2009).

64. L. Zhou, Y. Liu, L. Lu, X. Lu, R. A. F. Dixon, Cardiac gene activation analysis in mammalian non-myoblastic cells by Nkx2-5, Tbx5, Gata4 and Myocd. *PLOS ONE* **7**, e48028 (2012).

65. E. Beuling, T. Bosse, D. J. aan de Kerk, C. M. Piaseckyj, Y. Fujiwara, S. G. Katz, S. H. Orkin, R. J. Grand, S. D. Krasinski, GATA4 mediates gene repression in the mature mouse small intestine through interactions with friend of GATA (FOG) cofactors. *Dev. Biol.* **322**, 179–189 (2008).

66. N. Doni Jayavelu, A. Jajodia, A. Mishra, R. D. Hawkins, Candidate silencer elements for the human and mouse genomes. *Nat. Commun.* **11**, 1061 (2020).
67. M. Westerfield, *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Danio rerio)* (University of Oregon Press, Eugene, Oregon, 2007).

68. G. K. Varshney, W. Pei, M. C. LaFave, J. Idol, L. Xu, V. Gallardo, B. Carrington, K. Bishop, M. Jones, M. Li, U. Harper, S. C. Huang, A. Prakash, W. Chen, R. Sood, J. Ledin, S. M. Burgess, High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res.* **25**, 1030–1042 (2015).

69. C. Thisse, B. Thisse, High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* **3**, 59–69 (2008).

70. T. Dobrzycki, M. Krecsmarik, R. Monteiro, Genotyping and quantification of in situ hybridization staining in zebrafish. *J. Vis. Exp.*, e59956 (2020).

71. I. C. Scott, B. Masri, L. A. D’Amico, S.-W. Jin, B. Jungblut, A. M. Wehman, H. Baier, Y. Audigier, D. Y. R. Stainier, The g protein-coupled receptor agtrl1b regulates early development of myocardial progenitors. *Dev. Cell* **12**, 403–413 (2007).

72. Y. C. Song, T. E. Dohn, A. B. Rydeen, A. V. Nechiporuk, J. S. Waxman, HDAC1-mediated repression of the retinoic acid-responsive gene ripply3 promotes second heart field development. *PLOS Genet.* **15**, e1008165 (2019).

73. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, 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).

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

75. L. Christiaen, E. Wagner, W. Shi, M. Levine, Isolation of sea squirt (*Ciona*) gametes, fertilization, dechorionation, and development. *Cold Spring Harb. Protoc.* **2009**, pdb.prot5344 (2009).
76. S. Gandhi, M. Haeussler, F. Razy-Krajka, L. Christiaen, A. Stolfi, Evaluation and rational design of guide RNAs for efficient CRISPR/Cas9-mediated mutagenesis in Ciona. *Dev. Biol.* **425**, 8–20 (2017).

77. A. Woznica, M. Haeussler, E. Starobinska, J. Jemmett, Y. Li, D. Mount, B. Davidson, Initial deployment of the cardiogenic gene regulatory network in the basal chordate, *Ciona intestinalis*. *Dev. Biol.* **368**, 127–139 (2012).

78. X. Gu, F. Xu, X. Wang, X. Gao, Q. Zhao, Molecular cloning and expression of a novel CYP26 gene (cyp26d1) during zebrafish early development. *Gene Expr. Patterns* **5**, 733–739 (2005).

79. M. Takeuchi, S. Yamaguchi, Y. Sakakibara, T. Hayashi, K. Matsuda, Y. Hara, C. Tanegashima, T. Shimizu, S. Kuraku, M. Hibi, Gene expression profiling of granule cells and Purkinje cells in the zebrafish cerebellum. *J. Comp. Neurol.* **525**, 1558–1585 (2017).

80. M. Tambalo, R. Mitter, D. G. Wilkinson, A single cell transcriptome atlas of the developing zebrafish hindbrain. *Development* **147**, dev184143 (2020).

81. C. Inomata, T. Yuikawa, Y. Nakayama-Sadakiyo, K. Kobayashi, M. Ikeda, M. Chiba, C. Konishi, A. Ishioka, S. Tsuda, K. Yamasu, Involvement of an Oct4-related PouV gene, pou5f3/pou2, in neurogenesis in the early neural plate of zebrafish embryos. *Dev. Biol.* **457**, 30–42 (2020).

82. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

83. G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A. K. Shalek, C. K. Slichter, H. W. Miller, M. J. McElrath, M. Prlic, P. S. Linsley, R. Gottardo, MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).

84. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
85. The Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).

86. W. Saelens, R. Cannoodt, H. Todorov, Y. Saeys, A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).

87. K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, S. Dudoit, Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

88. S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data (2010); www.bioinformatics.babraham.ac.uk/projects/fastqc/.

89. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

90. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

91. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

92. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

93. R. Stark, G. Brown, DiffBind: Differential binding analysis of ChIP-Seq peak data (2011).

94. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
95. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

96. T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, W. S. Noble, MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).

97. O. Fornes, J. A. Castro-Mondragon, A. Khan, R. van der Lee, X. Zhang, P. A. Richmond, B. P. Modi, S. Correard, M. Gheorghe, D. Baranašić, W. Santana-Garcia, G. Tan, J. Chêneby, B. Ballester, F. Parcy, A. Sandelin, B. Lenhard, W. W. Wasserman, A. Mathelier, JASPAR 2020: Update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).

98. M. Hiller, S. Agarwal, J. H. Notwell, R. Parikh, H. Guturu, A. M. Wenger, G. Bejerano, Computational methods to detect conserved non-genic elements in phylogenetically isolated genomes: Application to zebrafish. *Nucleic Acids Res.* **41**, e151 (2013).

99. R. M. Kuhn, D. Haussler, W. J. Kent, The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).

100. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

101. K. L. Howe, P. Achuthan, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, B. El Houdaigue, R. Fatima, A. Gall, C. Garcia Giron, T. Grego, C. Guijarro-Clarke, L. Haggerty, A. Hemrom, T. Hourlier, O. G. Izuogu, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. Gonzalez Martinez, J. C. Marugán, T. Maurel, A. C. McMahon, S. Mohanan, B. Moore, M. Muffato, D. N. Oheh, D. Paraschas, A. Parker, A. Parton, I. Prosovetskaia, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, E. Steed, M. Szpak, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, B. Walts, A. Winterbottom,
M. Chakiachvili, A. Chaubal, N. De Silva, B. Flint, A. Frankish, S. E. Hunt, G. R. IIsley, N. Langridge, J. E. Loveland, F. J. Martin, J. M. Mudge, J. Morales, E. Perry, M. Ruffier, J. Tate, D. Thybert, S. J. Trevanion, F. Cunningham, A. D. Yates, D. R. Zerbin, P. Flicek, Ensembl 2021. Nucleic Acids Res. 49, D884–D891 (2021).

102. K. Kikuchi, J. E. Holdway, R. J. Major, N. Blum, R. D. Dahn, G. Begemann, K. D. Poss, Retinoic acid production by endocardium and epicardium is an injury response essential for zebrafish heart regeneration. Dev. Cell 20, 397–404 (2011).

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

104. D. Traver, B. H. Paw, K. D. Poss, W. T. Penberthy, S. Lin, L. I. Zon, Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. Nat. Immunol. 4, 1238–1246 (2003).

105. S. Higashijima, H. Okamoto, N. Ueno, Y. Hotta, G. Eguchi, High-frequency generation of transgenic zebrafish which reliably express GFP in whole muscles or the whole body by using promoters of zebrafish origin. Dev. Biol. 192, 289–299 (1997).

106. C.-J. Huang, C.-T. Tu, C.-D. Hsiao, F.-J. Hsieh, H.-J. Tsai, Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. Dev. Dyn. 228, 30–40 (2003).

107. J. Wang, D. Panáková, K. Kikuchi, J. E. Holdway, M. Gemberling, J. S. Burris, S. P. Singh, A. L. Dickson, Y.-F. Lin, M. K. Sabeh, A. A. Werdich, D. Yelon, C. A. MacRae, K. D. Poss, The regenerative capacity of zebrafish reverses cardiac failure caused by genetic cardiomyocyte depletion. Development 138, 3421–3430 (2011).

108. S.-W. Jin, D. Beis, T. Mitchell, J.-N. Chen, D. Y. R. Stainier, Cellular and molecular analyses of vascular tube and lumen formation in zebrafish. Development 132, 5199–5209 (2005).
109. A. J. Sehnert, A. Huq, B. M. Weinstein, C. Walker, M. Fishman, D. Y. R. Stainier, Cardiac
troponin T is essential in sarcomere assembly and cardiac contractility. *Nat. Genet.* 31, 106–110
(2002).

110. D. Yelon, S. A. Horne, D. Y. R. Stainier, Restricted expression of cardiac myosin genes reveals
regulated aspects of heart tube assembly in zebrafish. *Dev. Biol.* 214, 23–37 (1999).

111. J. N. Chen, M. C. Fishman, Zebrafish tinman homolog demarcates the heart field and initiates
myocardial differentiation. *Development* 122, 3809–3816 (1996).

112. D. P. Szeto, K. J. P. Griffin, D. Kimelman, HrT is required for cardiovascular development in
zebrafish. *Development* 129, 5093–5101 (2002).

113. S. Lazic, I. C. Scott, Mef2cb regulates late myocardial cell addition from a second heart field-like
population of progenitors in zebrafish. *Dev. Biol.* 354, 123–133 (2011).

114. J. J. Schoenebeck, B. R. Keegan, D. Yelon, Vessel and blood specification override cardiac
potential in anterior mesoderm. *Dev. Cell* 13, 254–267 (2007).

115. F. Reifers, H. Böhlí, E. C. Walsh, P. H. Crossley, D. Y. Stainier, M. Brand, Fgf8 is mutated in
zebrafish acerebellar (ace) mutants and is required for maintenance of midbrain-hindbrain
boundary development and somitogenesis. *Development* 125, 2381–2395 (1998).

116. I. Ruvinsky, A. C. Oates, L. M. Silver, R. K. Ho, The evolution of paired appendages in
vertebrates: T-box genes in the zebrafish. *Dev. Genes Evol.* 210, 82–91 (2000).

117. H. Grandel, K. Lun, G.-J. Rauch, M. Rhinn, T. Piotrowski, C. Houart, P. Sordinò, A. M. Küchler,
S. Schulte-Merker, R. Geisler, N. Holder, S. W. Wilson, M. Brand, Retinoic acid signalling in the
zebrafish embryo is necessary during pre-segmentation stages to pattern the anterior-posterior axis
of the CNS and to induce a pectoral fin bud. *Development* 129, 2851–2865 (2002).

118. T. Piotrowski, C. Nüsslein-Volhard, The endoderm plays an important role in patterning the
segmented pharyngeal region in zebrafish (*Danio rerio*). *Dev. Biol.* 225, 339–356 (2000).
119. A. Cheung, K. E. Trevers, M. Reyes-Corral, P. Antinucci, R. Hindges, Expression and roles of teneurins in zebrafish. *Front. Neurosci.* **13**, 158 (2019).

120. B. Thisse, Fast Release Clones: A High Throughput Expression Analysis. ZFIN Direct Data Submission (2004); http://zfin.org.