



Supplementary Materials for

Characterization of a common progenitor pool of the epicardium and myocardium

Richard C.V. Tyser^{1, †}, Ximena Ibarra-Soria^{2, †}, Katie McDole³, Satish A. Jayaram¹, Jonathan Godwin¹, Teun A.H. van den Brand⁴, Antonio M.A. Miranda^{1,5}, Antonio Scialdone^{6,7,8}, Philipp J. Keller³, John C. Marioni^{2,4,9,#}, Shankar Srinivas^{1,#}

Correspondence to: john.marioni@cruk.cam.ac.uk, shankar.srinivas@dpag.ox.ac.uk

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S5
Captions for Movies S1 to S14
Captions for Data S1 to S4

Other Supplementary Materials for this manuscript include the following:

Movies S1 to S14
Data S1 to S4

Materials and Methods

Mouse strains, husbandry and embryo collection

All animal experiments complied with the UK Animals (Scientific Procedures) Act 1986, approved by the local Biological Services Ethical Review Process and were performed under UK Home Office project licenses PPL 30/3420 and PCB8EF1B4. To obtain wild-type embryos, C57BL/6 males (in house) were crossed with 8-16 week old CD1 females (Charles River, England). All mice were maintained in a 12-hr light-dark cycle. Noon of the day when a vaginal plug was found was designated E0.5. To dissect the embryos, the pregnant females were culled by cervical dislocation in accordance with schedule one of the Animals (Scientific Procedures) Act. Embryos of the appropriate stage were dissected in M2 medium (Sigma-Aldrich, Cat No. M7167).

Dissection and unbiased collection of cells from the mouse cardiac crescent

Following dissection, embryos were collected, keeping the yolk sac intact, in fresh M2 media before being grouped based on their cardiac crescent stage. Progressive crescent stages (-1 to 3) were defined based on morphological criteria including the shape of the cardiac crescent and the ratio between width (medio-lateral axis) and maximum height (rostral-caudal axis) of the crescent¹. Stage -1 embryos were defined based on the incomplete fusion of the mesodermal wings in the anterior portion of the embryo. Embryos were considered to be at the LHT stage once both sides of the cardiac crescent were completely folded and fused. Embryos prior to cardiac crescent formation were staged using the Lawson and Wilson staging system of mouse development².

Once embryos were staged, crescent regions were micro-dissected using tungsten needles, keeping the overlying endoderm intact (Figure 1B and Figure S1A). By maintaining the integrity of the overlying endoderm at cardiac crescent and LHT stages, regions of both yolk sac endoderm and embryonic endoderm were also collected as well as the forming pericardium. To isolate cardiac

crescent regions, the anterior half of the embryo was dissected from the posterior before a cut was made along the midline separating the embryo into left and right sides. The cardiac crescent could be then isolated using the sagittal profile to collect both dorsal and ventral tissue. Left and right sides of the sub-dissected regions were disaggregated and processed separately, but analyzed as one. At LHT stage the heart tube was isolated by dissections from the ventral surface, due to both the morphology and size of the tissue, therefore it is unlikely we collected regions of the dorsal pericardial wall. Following isolation of the cardiac crescent region the tissue was pooled from multiple embryos (Embryos pooled: Stage LHT: 5, Stage 3; 7, Stage 2; 27, Stage 1; 18, Stage 0; 7, Stage -1; 2). The pooled samples were dissociated into single cells using 200µl Accutase (ThermoFisher, Cat No. A1110501) for 12 minutes at 37°C, being agitated every 2 minutes, before adding 200µl heat-inactivated FBS (ThermoFisher, Cat No. 10500) to quench the reaction. Cells were then centrifuged at 1000rpm for 3 minutes at 4°C before being suspended in 100µl HBSS (ThermoFisher, Cat No. 14025) + 1% FBS, and stored on ice. Single cells were collected using a Sony SH800 FACS machine with a stringent single-cell collection protocol and sorted into 384 well plates containing SMART-seq2 lysis buffer³ plus ERCC spike-ins. To ensure we collected good quality cells, a live/dead dye (Abcam, Cat No. ab115347) was used; 100µl was added to the cell suspension at a 2x concentration in HBSS 10 minutes before collection, and live cells were collected based on their FITC intensity. Once cells were collected, plates were sealed, spun down, and frozen using dry ice before being stored at -80°C. This complete process, from dissection to single-cell collection, took approximately 2-3 hours. Cells were collected in multiple batches.

Single-cell RNA sequencing

mRNA from single cells was isolated and amplified (25 PCR cycles) using the SMART-seq2 protocol³. Multiplexed sequencing libraries were generated from cDNA using the Illumina Nextera

XT protocol and 125 bp paired-end sequencing was performed on an Illumina HiSeq 2500 instrument (V4 chemistry). Samples were processed in seven different batches, over a period of ~1.5 years. Detailed metadata is provided in Extended Data 3.

Data processing and quality control

Sequencing data were aligned to the mouse reference genome GRCm38 (plus the sequences of ERCC spike-ins) using GSNAP (version 2016-11-07)⁴ with parameters `-A sam -B 5 -n 1 -N 1`. Additional processing of the alignment files was performed using samtools v1.2⁵. The number of reads aligned to each gene was quantified using the htseq-count program from HTSeq-0.6.1p1⁶ with options `-m union -s no`, against the mouse transcriptome annotation from the Ensembl database⁷, version 87 (<http://dec2016.archive.ensembl.org/index.html>), supplemented with the ERCC spike-ins. Raw counts are provided in Data S1.

The dataset consisted of 4,180 single cells. As a measure of the quality of each sample, we examined the total number of reads mapped to exons, the total number of genes detected per cell and the proportion of reads mapped to ERCC spike-ins or mitochondrial genes. We excluded any sample with fewer than 50,000 total reads mapped to exons; less than 6,000 detected genes; more than 30% of reads mapped to ERCC spike-ins; or more than 15% of reads mapped to mitochondrial genes. This filtering resulted in the retention of 3,105 single cells for downstream analyses (Figure S2A).

Data normalization

The data were normalized for cell-specific biases using the deconvolution method proposed in Lun et al. (2016)⁸ and implemented in the Bioconductor package *scran*⁹. To calculate size factors we first used the *quickCluster* function to obtain an initial clustering of the cells (with *min.size=100*

and *method=igraph*). We then estimated size factors with the *computeSumFactors* function using only endogenous genes, the aforementioned clusters and a minimum mean of 1. These were used to normalize the count data for all endogenous genes expressed in at least one cell.

Identification of highly variable genes

To identify highly variable genes (HVGs), we used the function *modelGeneCV2* from the *scran*⁹ package to model the trend between the squared coefficient of variation (CV^2) and mean expression, and fit a trend to account for the mean-variance relationship across genes¹⁰. We provided the date of sample collection as a blocking factor to account for batch effects. We then used the *getTopHVGs* function to select the 2,000 most variable genes, defined as those with the highest deviations from the trend. We excluded from this list mitochondrial and sexually dimorphic genes (*Xist* and genes from the Y chromosome) and used the remaining 1,995 genes as our set of HVGs.

Batch correction

To correct technical batch effects originating from collecting the data at different dates, we used the function *fastMNN* from the *batchelor* Bioconductor package, which is a fast implementation of the method developed by Haghverdi et al. (2018)¹¹. We performed the correction using the \log_2 -transformed normalized counts of HVGs and obtained corrected expression estimates for all genes.

Dimensionality reduction

To visualize the dataset in low dimensions we used uniform manifold approximation and projection (UMAP)¹² as implemented in the *scater* package¹³, using the *runUMAP* function with both normalized and batch-corrected data (Supplementary Figure 2B and C).

Clustering into distinct subpopulations and identification of marker genes

To classify cells into different clusters we first used hierarchical clustering (*hclust* function, *method=average*) on the distance matrix computed with the function *dist*, from the batch-corrected normalized counts of HVGs. Clusters were defined with the dynamic hybrid cut algorithm from the *dynamicTreeCut* package¹⁴ (*cutreeDynamic* function with *method=hybrid* and *minClusterSize=40*). One cell could not be assigned to any cluster by the algorithm and was therefore removed from downstream analyses. This procedure resulted in the definition of 12 clusters (Figure 1C).

To find genes specifically expressed in certain clusters we applied the *findMarkers* function from *scran*⁹ to the log₂-transformed normalized counts, restricted to genes with a minimum mean expression of 0.1, and blocking for the date of sample collection to account for batch effects. We restricted the analysis to finding genes with higher expression in a given cluster (*direction=up*) when compared to all other clusters, and performed the test with two settings of the argument *pval.type*: a stringent analysis that returns the most specific set of markers by setting it to *all*, thus requiring the gene to be differentially expressed compared to all other clusters; and a more lenient strategy that only requires the gene to be differential against three quarters of the clusters (*pval.type=some*, *min.prop=0.75*) and thus recovers genes that are markers shared by closely related subpopulations. The results from these analyses can be browsed at <https://marionilab.cruk.cam.ac.uk/heartAtlas/>. We annotated each cluster based on the expression of these marker genes combined with previous knowledge from the literature. The heatmaps in Figure 1 were generated with the *ComplexHeatmap* package¹⁵.

Cell cycle phase inference

To infer the cell cycle stage of each cell we used the classification method developed by Scialdone et al. (2015)¹⁶, as implemented in *scran*⁹. We ran the *cyclone* function, using the provided training data of mouse cell cycle markers, on the normalized count matrix of endogenous genes.

Whole mount immunostaining

Dissected intact embryos were fixed for 1 hr at room temperature with 4% paraformaldehyde (SantaCruz, Cat No. sc281692) in PBS (Sigma-Aldrich, Cat No. P4417). The embryos were then rinsed twice in PBST-0.1% (PBS with 0.1% Triton X-100 (Sigma-Aldrich, Cat No. T8787)) before being permeabilized in PBST-0.25% for 40 mins and rinsed again 3x in PBT-0.1%. The embryos were next transferred to blocking solution (5% donkey serum (Sigma-Aldrich, Cat No. D9663), 1%BSA (Sigma-Aldrich, Cat No. A7906) in PBST-0.1%) overnight (o/n) at 4°C. Primary antibodies (Supplementary Table 5) were then added to the solution and were incubated o/n at 4°C. The embryos were washed 3x 15mins in PBST-0.1% and incubated o/n at 4°C in PBST-0.1% with the secondary antibodies, and DAPI (Supplementary Table 5), then subsequently washed 3x PBT-0.1% for 15 min and placed in Vectashield mounting medium with DAPI (Vector Labs, Cat No. H-1200) for at least 24 hr at 4°C. Each staining combination was repeated on at least 3 embryos. Samples were imaged using a Zeiss 880 confocal microscope with a 40x oil (1.36 NA) objective. Images were captured at 512 × 512 pixel dimension using multiple tiles with a Z-step of between 1.5 and 0.5 μm. Based on the working distance of the Zeiss 880 microscope and tissue penetration, embryos were imaged to a depth of between 120 – 150μm. Sagittal sections were generated using optical reconstruction of multiple Z-images. Movies of 3D volume renderings were created using Arivis 4D software.

Anatomically defined micro-dissection.

Anatomically defined reference regions were collected from wild type embryos using micro-dissection. Once embryos were isolated and staged (using the criteria described above), the rostral region of the embryo was removed and placed into 2.5% pancreatin in PBS (Sigma-Aldrich, Cat No. P3292) for 3mins at room temperature before being placed back into M2 media; germ layers were separated using tungsten needles. The isolated mesoderm was then further dissected using tungsten needles into 4 sub regions including cardiac, cranial, dorsal and presomitic mesoderm (Figure 2A). Following isolation, samples were dissociated into single cells and collected using the methods previously described. Each of the four regions collected came from 3 individual embryos, all collected in the same experiment. Detailed metadata is provided in Extended Data 4.

Spatially localized scRNA-seq data processing

Single cell RNA-seq data from the micro-dissected anatomical regions was processed in the same way as the unbiased dataset. We excluded any sample with fewer than 50,000 total reads mapped to exons; less than 4,000 detected genes; more than 50% of reads mapped to ERCC spike-ins; or more than 10% of reads mapped to mitochondrial genes. The dataset consisted of 748 cells across the four anatomical regions. From these, 326 were retained for downstream analyses. Raw data were normalized as described above.

Random forest classifier on spatially localized data

To assess the most likely anatomical origin of each cell from the cardiac mesoderm clusters in the unbiased data (Me3-8), we used the cells with known anatomical localization information as a reference dataset. First, we computed a PCA of the unbiased data on the log-scale batch corrected counts for the highly variable genes (*prcomp* function), and projected the reference dataset onto the first two PCs. Since the collection of cells obtained from any given anatomical region in the embryo is a heterogeneous mixture, with some degree of contamination from the adjacent

endoderm layer as well as endothelial and blood cells, we removed any cells that mapped to the endoderm (En1-3), endothelial (Me2) or blood (Me1) clusters. When clustering the reference data on its own, these cells formed independent, well-separated clusters that did not correlate with the anatomical region of origin, supporting the notion that they are distinct cell types and thus not helpful for determining the anatomical origin of cardiac mesoderm cells. Thus, the final reference set we used comprised 278 cells, with around 60 cells per anatomical region except for the dorsal mesoderm group, which contained 94 cells.

Reference cells were clustered using the same methods described above (with HVGs recomputed from this dataset), resulting in the definition of five clusters. These clusters showed good correlation with the anatomical labels. However, all mesoderm classes but the cranial mesoderm were split between several clusters, suggesting the anatomically defined regions contain heterogeneous cell populations.

To classify the unbiased cells we used a random forest classifier. We combined the dissection labels with the clustering results to define classes (Figure S5A). One of these classes, containing caudal PSM cells, comprised only 8 cells, and was excluded from the analysis. The remaining cells were split 70:30 as training and validation sets. The input for the random forest was the log-normalized counts of HVGs in the training set. We used 10-fold cross-validation to determine the optimal parameters for the number of trees and number of randomly selected variables at each split. The random forest was trained using the *randomForest* function from the *randomForest* package (with `importance=TRUE`;¹⁷), and achieved an overall accuracy of 0.9091 on the validation set (95% confidence interval: 0.82-0.96). The 50 genes deemed most important by the random forest are shown in Figure S5B.

To classify the cells from the unbiased dataset we used the *predict* function to obtain the classification labels and their associated probabilities. We only classified cells where the difference between the probabilities of the two top classes was greater than 0.15 (Figure S5C).

Diffusion map analysis and pseudotime inference

To study the dynamics between the different cardiac progenitor populations and the mature cardiomyocytes (clusters Me3-7) we used diffusion maps¹⁸ to visualize the data and infer possible trajectories. We excluded from this analysis cells from the Me8 cluster since the majority were assigned to non-cardiac phenotypes by the random forest classifier (Figure 2D and E); we further removed any other cells classified with a cranial mesoderm label. To compute the diffusion map, we first performed PCA (*prcomp* function) on the batch-corrected normalized counts of HVGs and used the first 50 principal components as input for the diffusion map analysis (*DiffusionMap* function from the Bioconductor package *destiny*¹⁹, with *sigma=local* and *distance=euclidean*).

To order cells along the differentiation trajectory, we calculated the diffusion pseudotime with the *DPT* function. Branching analysis of the ordered cells revealed a branching point that separated cells into three branches: one composed by cells from Me3, Me4 and Me6; another containing the majority of Me5 cells; and a third corresponding to the Me7 cluster (Figure S9A). To better define the two trajectories connecting the progenitor clusters to the Me3 cells, we repeated the diffusion pseudotime analysis excluding Me5 or Me7, respectively. When Me7 cells were excluded, the diffusion analysis showed a branching point separating Me6, Me4+Me5 and Me3 into different branches, suggesting that the Me5 and Me4 progenitors lie in a trajectory leading to Me3 cells, that is separate from Me6 cells (Figure S9B). On the other hand, the analysis excluding Me5 cells showed a branching point that separated a fraction of the Me4 cluster containing the most distinct cells. The rest of the cluster was part of the other two branches, that contained Me3 and Me6+Me7

cells respectively. These results thus suggest that only the most distinct Me4 cells are excluded from the trajectory linking Me6-7 cells to Me3, with some Me4 cells compatible with these transcriptional transitions (Figure S9C).

We observe that cells from all stages except stage -1 are distributed evenly across the whole diffusion space (Figure S9D and E), and there is no correlation between the pseudotime ordering of cells and their stage, except for stage -1 cells that are restricted to the earliest part of the trajectory including the Me7 progenitors. This suggests that there are no major transcriptional differences across stages but, instead, a continuous differentiation process that is maintained across these developmental stages.

Identifying genes with dynamic expression along pseudotime

To identify genes that change their expression levels along pseudotime we regressed the normalized counts (on the log scale) along pseudotime for cells consistent with each trajectory (only genes with mean expression of at least 0.1 were considered). More precisely, we fitted a constant or a degree 2 polynomial using local regression with the *locfit* package (<https://CRAN.R-project.org/package=locfit>) (*locfit* function, with $nn=1$) and calculated Akaike's information criterion (AIC) for each. We selected genes that were better fitted by the degree 2 model by computing the difference of the AIC (ΔAIC) of the degree 2 model minus the AIC of the null model (similar to²⁰). We retained all genes with $\Delta AIC < -200$. The results from this analysis can be browsed and downloaded at <https://marionilab.cruk.cam.ac.uk/heartAtlas/>. For each gene, the expression levels are plotted against the corresponding pseudotime values; if the gene was identified as dynamically expressed (see above) the local fit is also plotted.

In Situ Hybridization Chain reaction (HCR)

In situ HCR kit (ver.3) containing amplifier set, hybridization, amplification, wash buffers, and DNA probe sets (Supplementary Table 5), were purchased from Molecular Instruments (molecularinstruments.org) and the protocol described in Choi et al. (2018) was followed with slight modifications²¹. Embryos were collected as previously described and fixed overnight in 4% paraformaldehyde at 4⁰C with shaking. Following fixation, the embryos were kept on ice and washed twice in cold nuclease free 0.1% PBS Tween (nucPBST) before being dehydrated in methanol with a series of graded MeOH/0.1% nucPBST washes (25% MeOH/75% nucPBST; 50% MeOH/50% nucPBST; 75% MeOH/25% nucPBST; 2x 100% MeOH) for 10 minutes on ice before being stored at -20⁰c in 100% MeOH. For HCR staining, embryos were rehydrated with a series of graded MeOH/PBST washes for 10 minutes each on ice (75% MeOH/25% nucPBST; 50% MeOH/50% nucPBST; 25% MeOH/75% nucfree PBST; 100% nucPBST; 100% nucPBST at room temperature). After rehydration, embryos were treated with 10 µg/ml proteinase K solution for 5 minutes at room temperature, washed twice for 5 minutes with nucPBST, post-fixed with 4% paraformaldehyde for 20 minutes at room temperature, then washed three times with nucPBST for 5 minutes. Embryos were next pre-hybridized with 50% probe hybridization buffer/ 50% nucPBST at room temperature for two hours before being incubated with probe hybridization buffer for 30 minutes at 37°C prior to being incubated in probe solution (2 pmol of each probe in 500µl of 30% probe hybridization buffer) overnight (12-16 hours) at 37°C. Probe libraries were designed and manufactured by Molecular Instruments using *Mus musculus* sequences from NCBI database. Excess probe was removed by washing with pre-warmed probe wash buffer at 37°C four times for 15 minutes each and then two times 5 minutes with 5X SSCT buffer (5x sodium chloride sodium citrate, 0.1% Tween 20 in ultrapure H₂O). For amplification, embryos were incubated in amplification buffer for 30 minutes at room temperature before being placed into hairpin solution

and incubated overnight in the dark at room temperature. 30 pmol hairpin solution was prepared by snap cooling (95°C for 90 seconds, then cooled to room temperature) stock hairpins in storage buffer and then diluted in amplification buffer at room temperature. Following overnight incubation, excess hairpins were removed by washing with 5X SSCT buffer at room temperature (2x 5 mins, 2x 30 mins with DAPI and 1x 5 min). Embryos were then placed into Vectashield antifade mounting medium (Vector Laboratories, Cat no. H-1000) or 87% glycerol solution, then imaged on a Zeiss 880 confocal microscope as previously described in the immunohistochemistry section. Each HCR combination was repeated on at least 3 embryos. Movies of 3D volume renderings were created using Arivis 4D software.

Live cell tracking

Embryos were prepared for light-sheet live imaging as previously described in²². Briefly, preparation includes dissection at E7.0, selection for light-sheet imaging, mounting in glass capillary tubes (2.5 mm inner diameter) filled with Matrigel by embedding the ectoplacental cone in the Matrigel column, insertion of custom Teflon FEP tubing into the capillary, withdrawal of dissection media with a fine-tip pipette and replacing the dissection media with imaging media containing 40-50% rat serum in Fluorobrite DMEM. Mounted embryos were then placed directly into the environmentally controlled sample chamber of an adaptive multi-view light-sheet microscope²². A mixture of 5% CO₂ and 5% O₂ was supplied by a customized Okolab Bold line system which also maintained the sample chamber temperature at a constant 37°C for the duration of the experiment. Embryos were imaged every 5 minutes with a z-step size of 2.031 μm, a laser power of ~150-250 μW, and a laser sweep time of 20 ms per image. Multi-view light-sheet microscopy data sets were then registered, fused and visualized using the image processing pipeline described in²². Three such embryos were examined and between 3 to 6 Me5 cells were

located based on their anatomical position and the absence of Nkx2.5 at stage -1. These cells were then tracked both forward and back in time in order to assess their movement and expression of Nkx2-5.

Generation of Mab2112 iCreERT2 transgenic mouse line

The Mab2112:CreERT2 Knock-in mouse line was generated on a C57BL/6J background homozygous for the R26R-YFP reporter²³ using 2C-HR-CRISPR²⁴. Briefly, we injected the two blastomeres of 2-cell embryos with 75 ng/μl of Cas9-mSA mRNA, 50 ng/μl of sgRNA targeting the Mab2112 translation initiation site (TTCTAATACGACTCACTATAGCGGCGATCATGTTGGGGCACGTTTTAGAGCTAGA) and 20 ng/μl of the biotinylated repair construct²⁴. The repair construct consisted of 1Kb homology arms on either side of the Mab2112 start codon and had inserted in frame immediately downstream of the start codon sequence encoding TdTomato-T2A-iCreERT2-E2A. Injected embryos were transferred into pseudopregnant recipient females. Founder mice were identified by PCR against Cre sequence, and then the integrity of the modified *Mab2112* locus verified by Sanger sequencing. Founders with correct insertions were outbred for at least three generations to segregate off-target lesions before being used for lineage labelling experiments.

Tamoxifen labeling

Tamoxifen (Sigma) was dissolved in 37°C ethanol (Sigma) to give a concentration of 100mg/ml before being diluted in 37°C peanut oil to give a final stock concentration 10mg/ml. 150ul (1.5mg) of tamoxifen was administered via intraperitoneal injection in the afternoon of embryonic day(E) 6.5 (~17:00). In order to control for incorrect tamoxifen labelling control experiments were conducted by administering 1.5mg tamoxifen at E5.5. No YFP cells could be detected in embryos

at E10.5 (0/11) after injection at E5.5 confirming specificity of our labelling approach and ruling out persistent tamoxifen leading to late recombination.

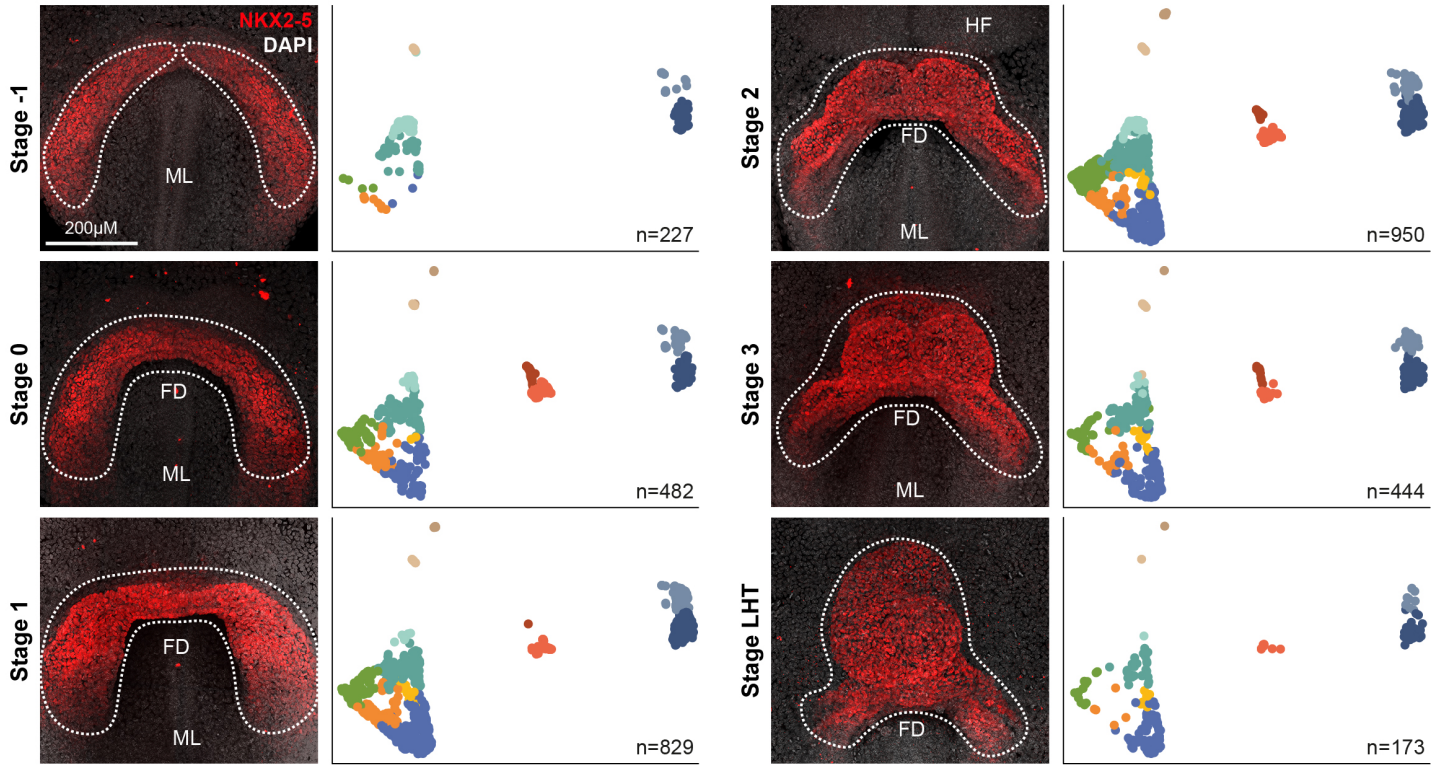
Methods and Materials References

1. Tyser, R. C. V. *et al.* Calcium handling precedes cardiac differentiation to initiate the first heartbeat. *Elife* **5**, (2016).
2. Lawson, K. A. & Wilson, V. A Revised Staging of Mouse Development Before Organogenesis. in *Kaufman's Atlas of Mouse Development Supplement* (2016). doi:10.1016/b978-0-12-800043-4.00003-8
3. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* (2014). doi:10.1038/nprot.2014.006
4. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq057
5. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp352
6. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* (2015). doi:10.1093/bioinformatics/btu638
7. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkv1157
8. Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* (2016). doi:10.1186/s13059-016-0947-7
9. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level

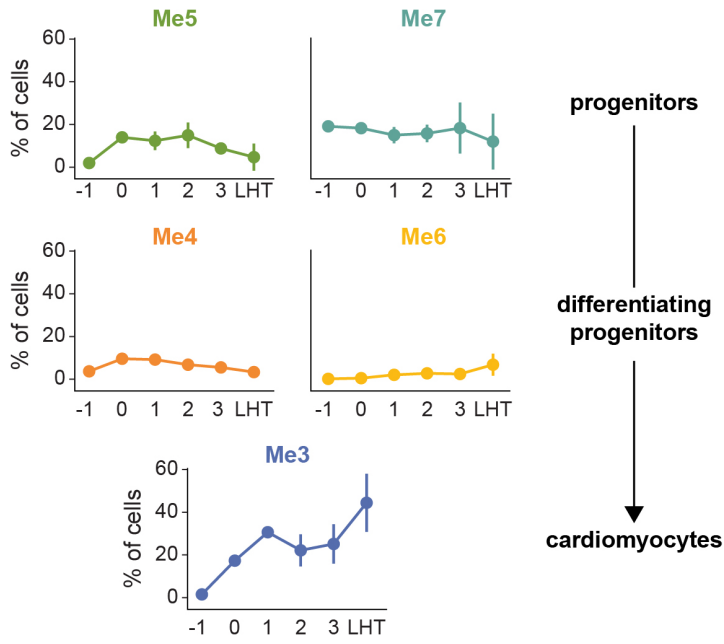
- analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* (2016).
doi:10.12688/f1000research.9501.2
10. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* (2013). doi:10.1038/nmeth.2645
 11. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4091
 12. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* (2018). doi:10.21105/joss.00861
 13. McCarthy, D. J., Campbell, K. R., Lun, A. T. L. & Wills, Q. F. Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* (2017). doi:10.1093/bioinformatics/btw777
 14. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* (2008).
doi:10.1093/bioinformatics/btm563
 15. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* (2016).
doi:10.1093/bioinformatics/btw313
 16. Scialdone, A. *et al.* Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* (2015). doi:10.1016/j.ymeth.2015.06.021
 17. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* (2002).
 18. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell

- analysis of differentiation data. *Bioinformatics* (2015). doi:10.1093/bioinformatics/btv325
19. Angerer, P. *et al.* Destiny: Diffusion maps for large-scale single-cell data in R. *Bioinformatics* (2016). doi:10.1093/bioinformatics/btv715
 20. Ibarra-Soria, X. *et al.* Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat. Cell Biol.* (2018). doi:10.1038/s41556-017-0013-z
 21. Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Dev.* (2018). doi:10.1242/dev.165753
 22. McDole, K. *et al.* In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level. *Cell* **175**, 859–876 (2018).
 23. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* (2001). doi:10.1186/1471-213X-1-4
 24. Gu, B., Posfai, E. & Rossant, J. Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4166

A



B



C

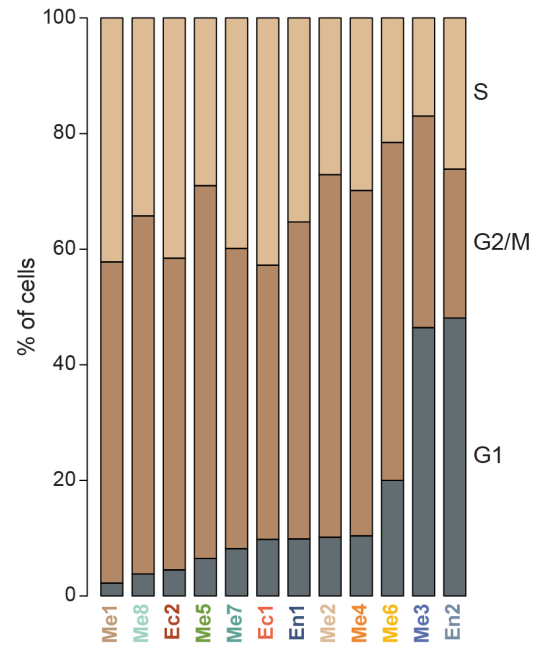


Fig. S1. Staged unbiased scRNA-seq data collection during cardiac crescent development.

A, Maximum intensity projections of whole mount embryo immunostaining showing the expression domain of NKX2-5 over the 6 stages of cardiac crescent development sampled. The dotted region highlights the area micro-dissected for the unbiased scRNA-seq experiments. Alongside, stagewise UMAP plots showing the clusters present at the different stages of cardiac crescent development. Cluster definition and UMAP coordinates are the same as in Fig. 1C, but only cells from the given stage are shown. FD, foregut diverticulum; ML, midline; HF, headfold. B, Plots indicating the proportion of cells in each of the cardiac mesoderm clusters (Me3-Me7) at each of the stages profiled. The fraction of cells assigned to the cardiomyocyte transcriptional profile (Me3) increases as development proceeds, whereas the progenitor pools (Me5 and Me7) remain roughly constant. The filled circles correspond to the mean fraction for each stage from different replicates (batches), and the line indicates ± 1 standard deviation, except for stage -1 where no replicates are available and thus no standard deviation is shown. C, Barplots of the proportion of cells from each cluster assigned to the different cell cycle phases, as shown in Figure 1G. The most immature clusters Me5, Me7 and Me8 have consistently low proportions of cells in G1, whereas the maturing progenitor clusters Me4 and Me6 contain higher proportions of non-proliferative cells as development proceeds. The mature cardiomyocytes in Me3 show the highest fraction of non-proliferative cells.

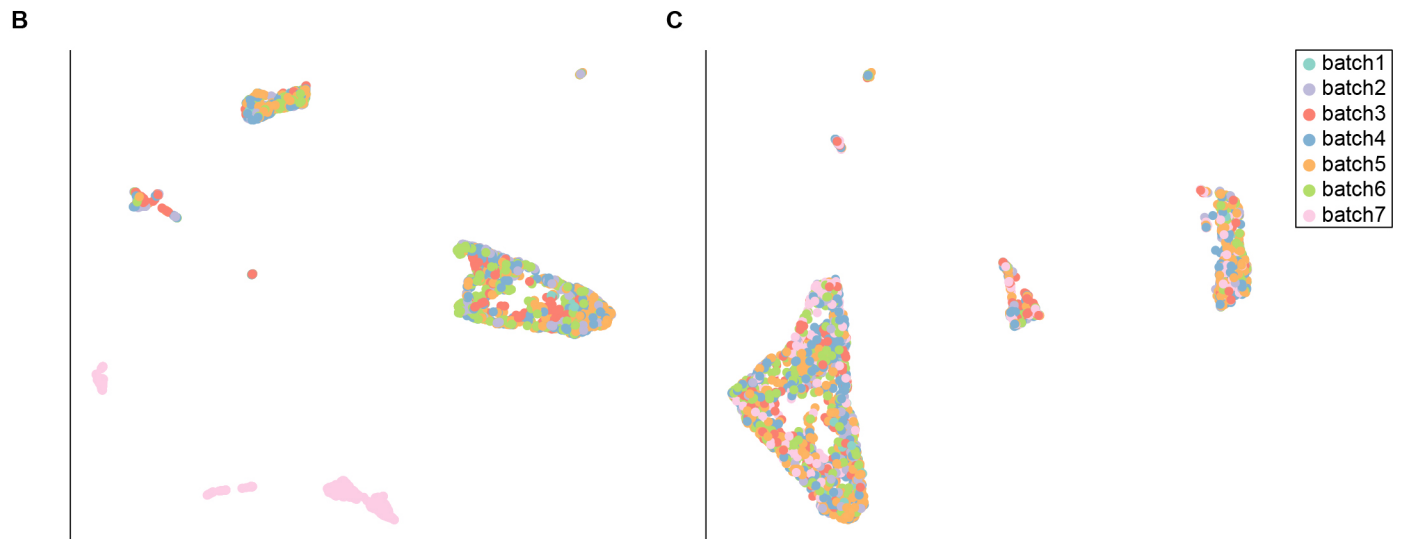
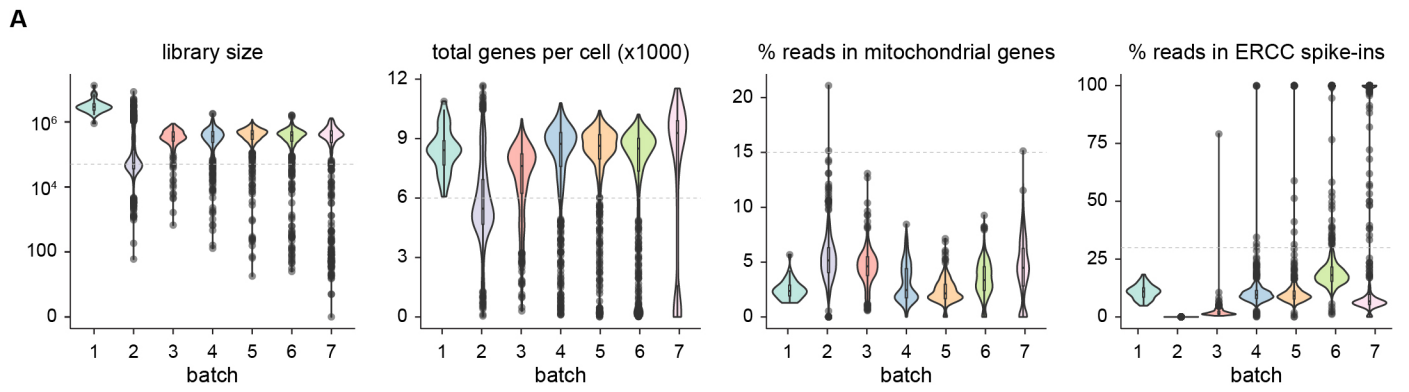


Fig. S2. Quality control of unbiased scRNA-seq dataset.

A, Metrics used to assess the quality of the scRNA-seq libraries. Cells passed quality control if they had library sizes greater than 50,000 reads, more than 6,000 genes detected, less than 15% of the reads mapped to mitochondrial genes and less than 30% of the reads mapped to ERCC spike-ins. B, UMAP plot of all the cells that passed quality control computed from highly variable genes, using normalized expression levels. Cells are colored according to the batch they were processed in. There are clear batch effects, observed from the cells of certain batches clustering separately. C, Same as B, but using batch corrected expression estimates (Methods). Cells no longer cluster by their batch of origin, indicating successful removal of the batch effects.

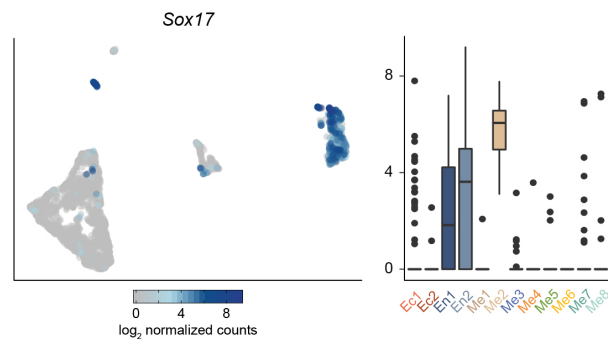
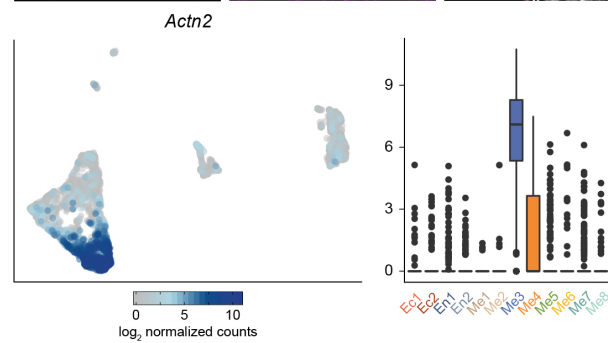
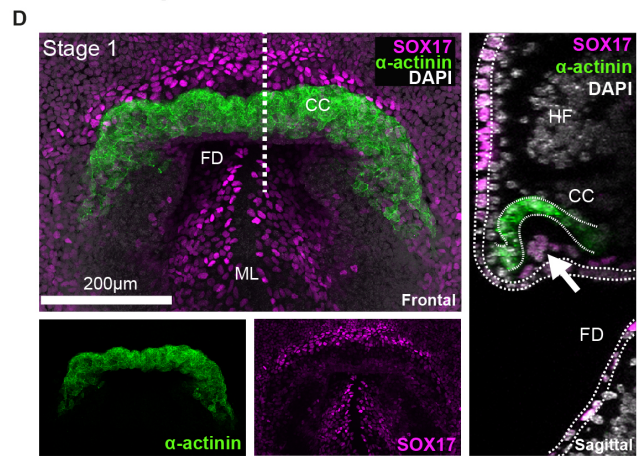
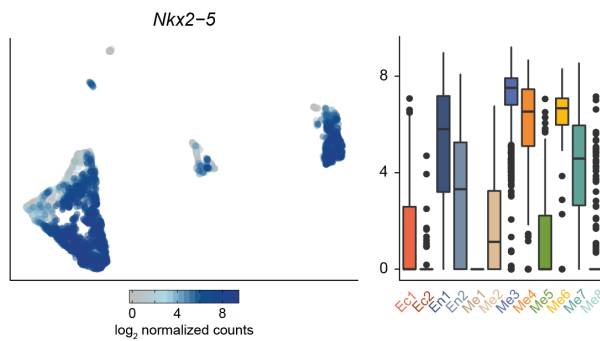
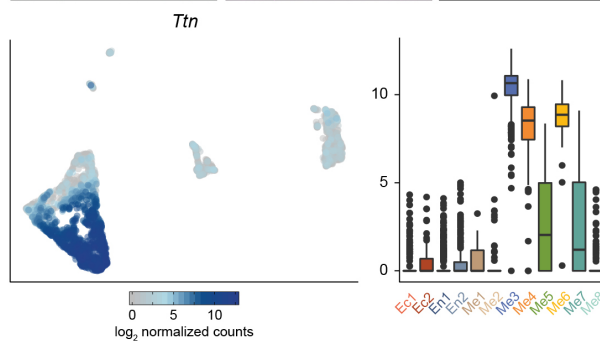
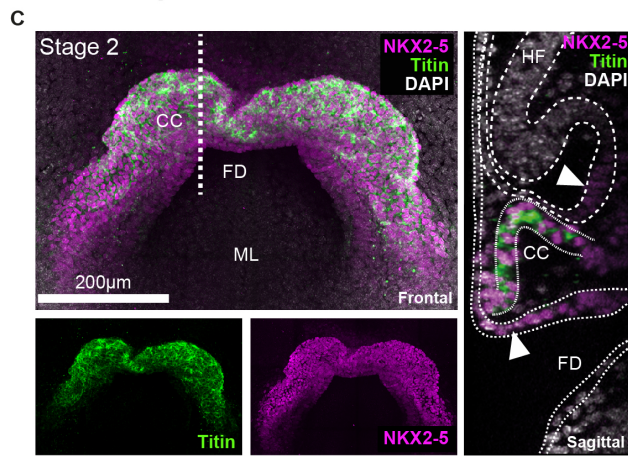
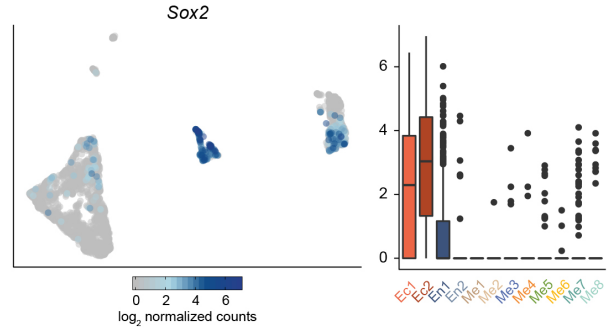
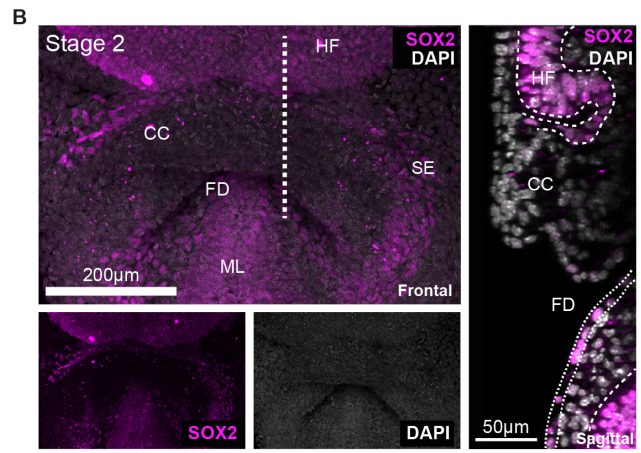
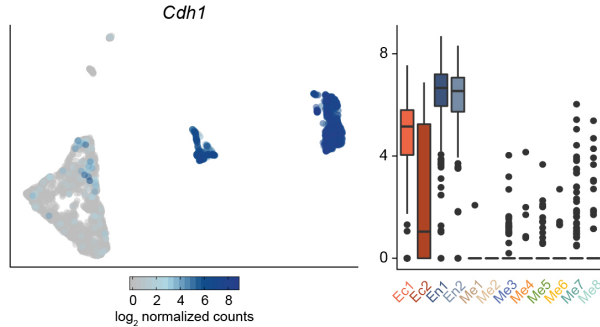
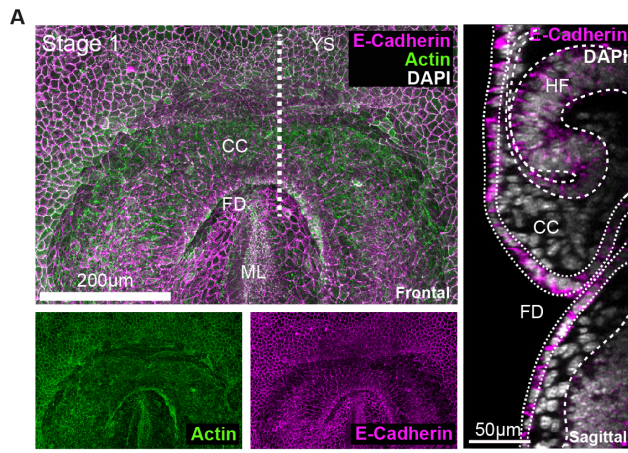


Fig. S3. Correlation of gene expression and anatomical identification of clusters in Fig. 1C. Maximum intensity projections (MIP) and sagittal sections of whole mount immunostaining showing staining of marker genes. Below, UMAP and boxplots show the corresponding gene expression per cluster. A, E-cadherin staining was localized to both endoderm (En1-2) and ectoderm (Ec1-2) and was not expressed in mesoderm clusters. Dotted lines represent the localization of sagittal sections. B, SOX2 staining marked the headfolds (neuroectoderm), surface ectoderm/amnion and definitive endoderm as represented by clusters Ec1, Ec2 and En1. C, Immunostaining for TITIN revealed the localization of cardiomyocytes within the cardiac crescent. *Ttn*, was most strongly expressed in mesoderm cluster Me3 (cardiomyocytes) but also in other mesoderm clusters. NKX2-5 staining colocalized with TITIN expression in the cardiac crescent but also marked a number of different cell types including cells in the endoderm and surface ectoderm (arrowheads). This was reflected in the expression of *Nkx2-5* in En1, En2, Ec1 and varying levels within the mesoderm clusters (Me2-8). D, Representative MIP image shown in Fig. 1E showing individual channels for SOX17 and sarcomeric α -actinin. Arrow highlights expression of SOX17 in endothelial cells of the forming endocardium. CC, cardiac crescent; FD, foregut diverticulum; ML, midline; HF, headfold; YS, yolk sac; SE, surface ectoderm/amnion.

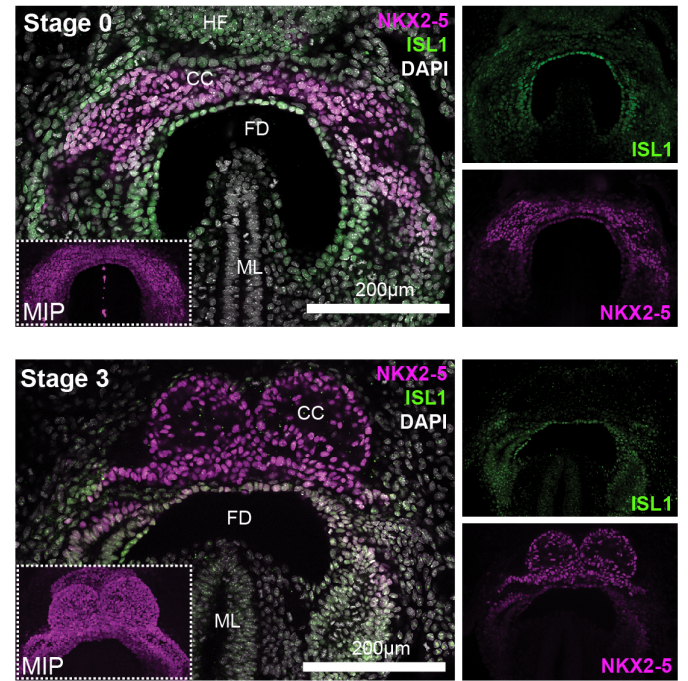
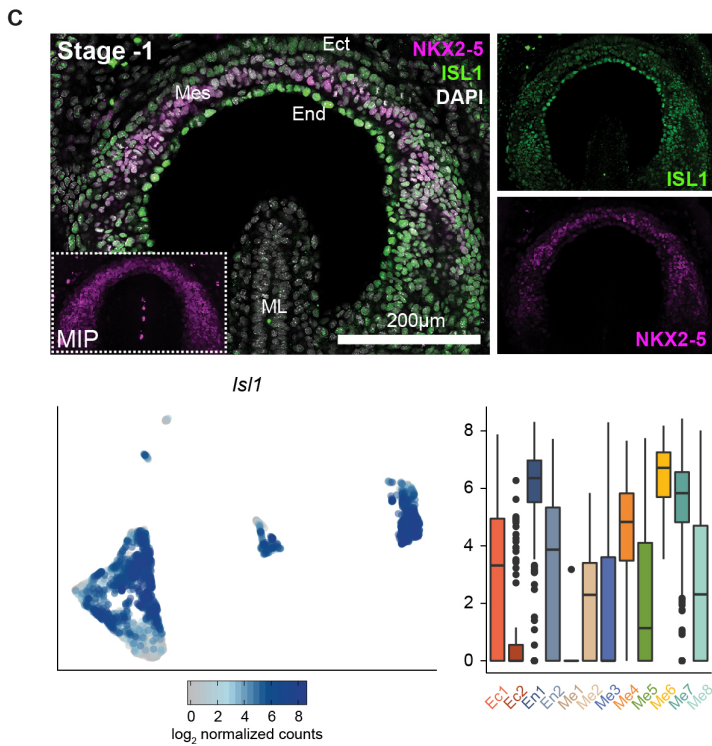
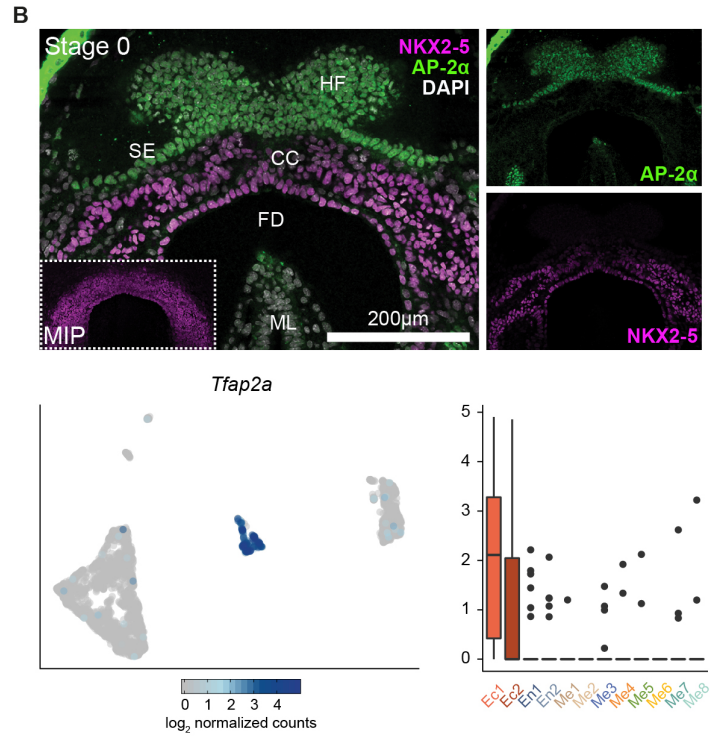
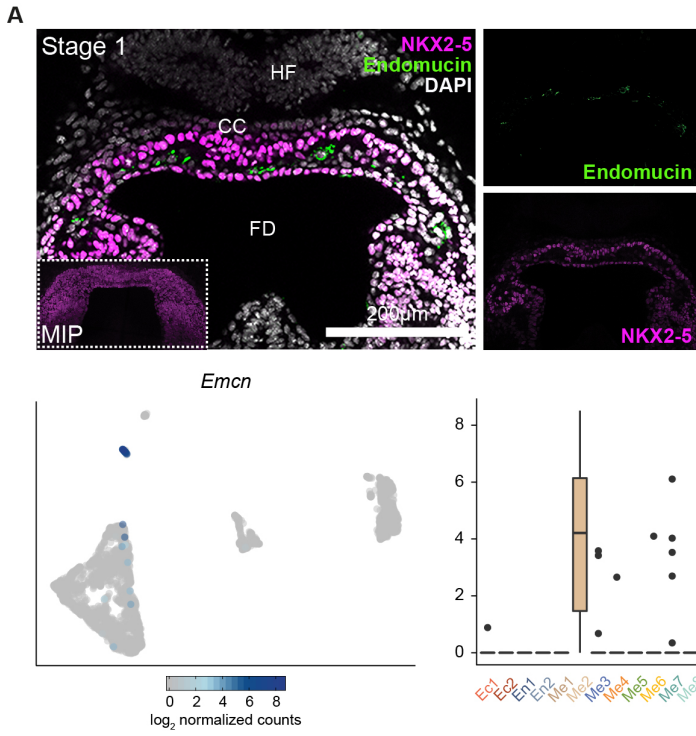


Fig. S4. Correlation of gene expression and anatomical localization of cell clusters.

Maximum intensity projections (MIP) and sagittal sections of whole mount immunostaining showing staining of marker genes. Below, UMAP and boxplots showing the corresponding gene expression per cluster. A, *Emcn* expression was localized specifically to mesoderm cluster Me2. Localization of the corresponding protein, Endomucin, could be detected within the developing cardiac crescent, marking endothelial cells of the forming endocardium. Main figures represent single Z-section from volume datasets. Inset dotted boxes show maximum intensity projections of NKX2-5 staining highlighting overall cardiac crescent morphology. B, *Tfap2a*, was expressed specifically in the ectoderm populations Ec1 and Ec2. This corresponded with the protein (AP-2 α) being localized in the headfolds and surface ectoderm and absent from the cardiac crescent and endoderm. C, Whole mount immunostaining revealed that at early stages of cardiac crescent development ISLET1 was expressed throughout cardiac progenitors in the cardiac crescent (stage -1 and 0) as well as in adjoining ectoderm and endoderm. At later stages of cardiac crescent development, expression of ISLET1 was not detected in the NKX2-5 positive cardiac crescent region or ectoderm but was still detected in the more dorsal and lateral mesoderm as well as overlying endoderm. UMAP and boxplots showing the broad expression of *Isl1* expression during cardiac crescent development. *Isl1* was widely expressed throughout the majority of clusters including endoderm, ectoderm and mesoderm. HF, headfold; CC, cardiac crescent; FD, foregut diverticulum; ML, midline; SE, surface ectoderm/amnion; Ect, ectoderm; Mes, mesoderm; End, endoderm.

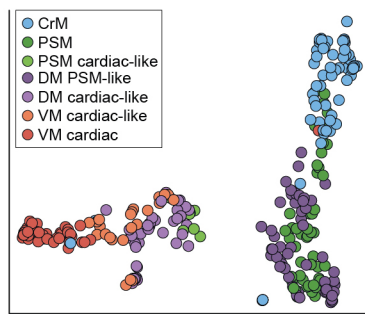
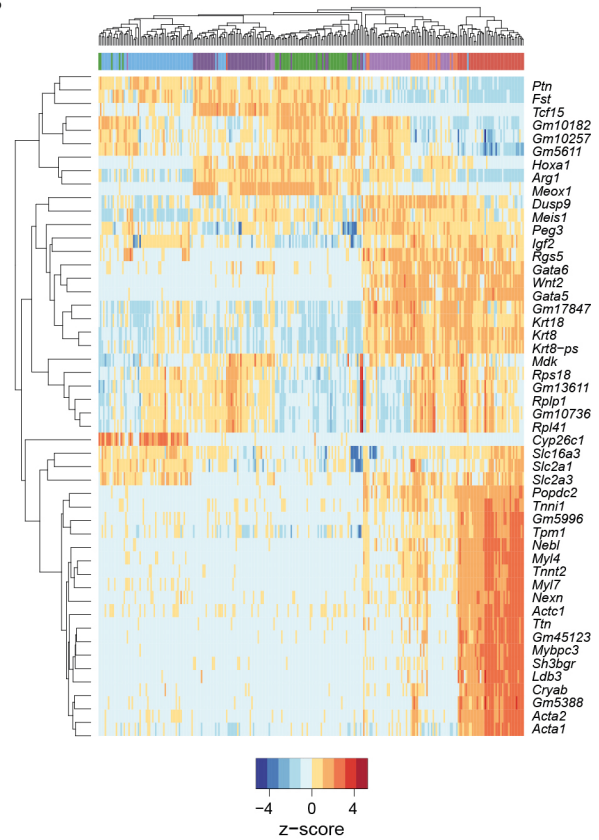
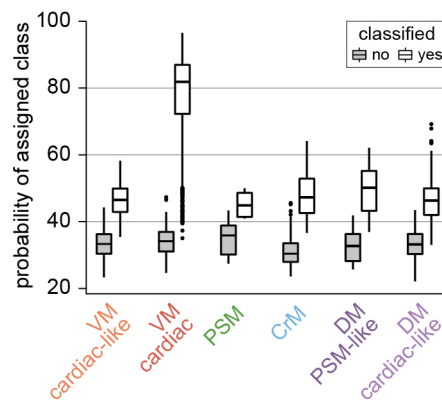
A**B****C**

Fig. S5. Construction of a random forest to classify the cardiac mesoderm cells.

A, UMAP plot as in Figure 2B, but colored based on the groups used as classes to train the random forest, which combine the anatomical information from the dissections together with the clustering of the transcriptomes. B, Heatmap of the 50 most important genes for the random forest. Cells are clustered by hierarchical clustering. The class of each cell is indicated at the top, following the same color scheme as in A. C, Boxplots indicating the probability with which each cell from the unbiased dataset is classified. Cells are split by the class they were assigned to. We required at least a 15% difference between the top and second assigned classes to classify a cell.

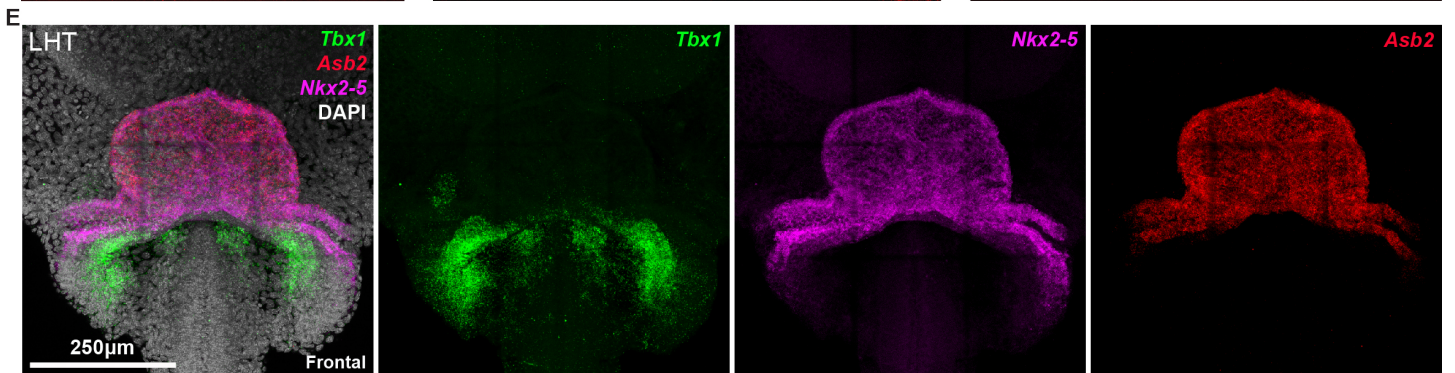
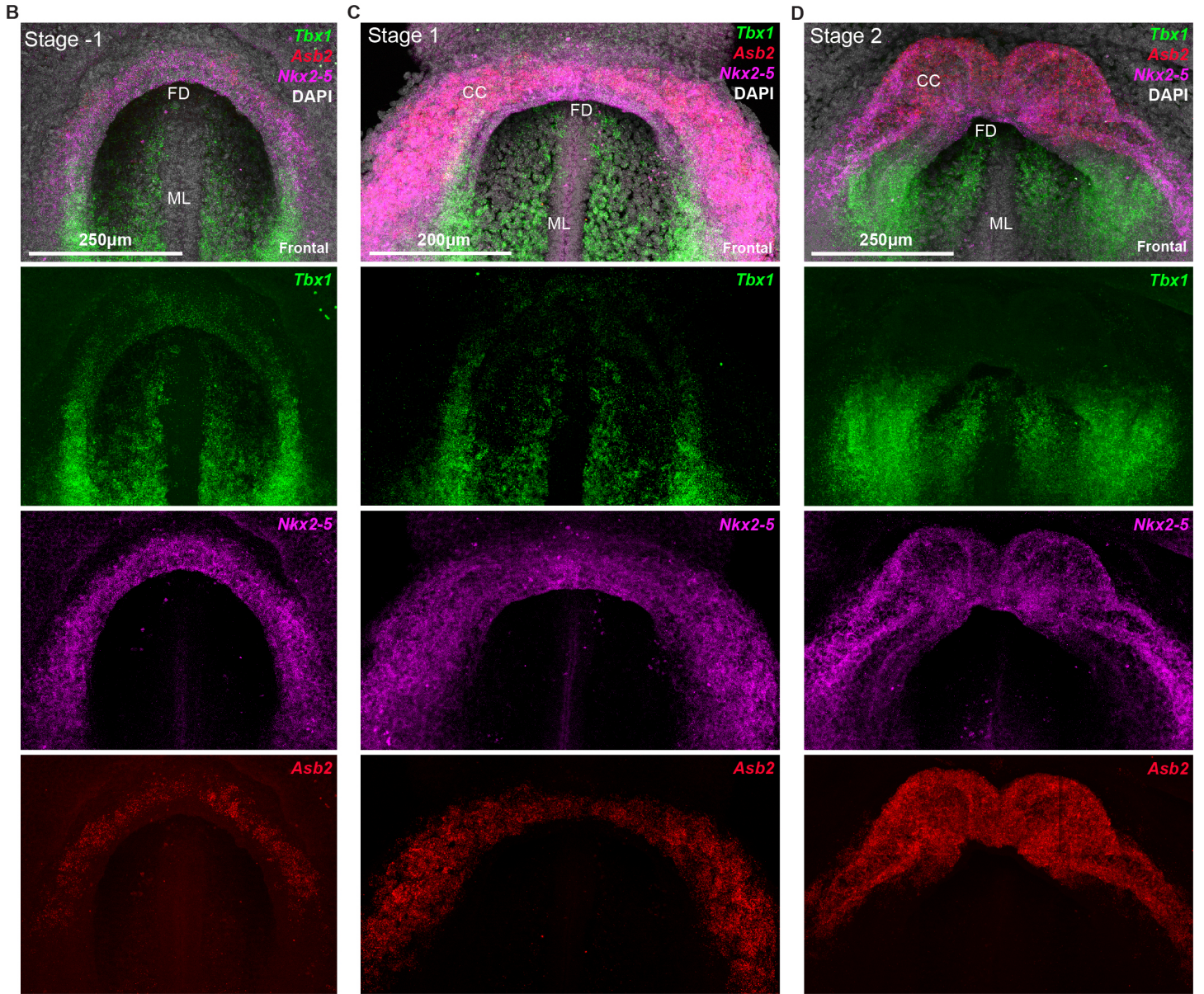
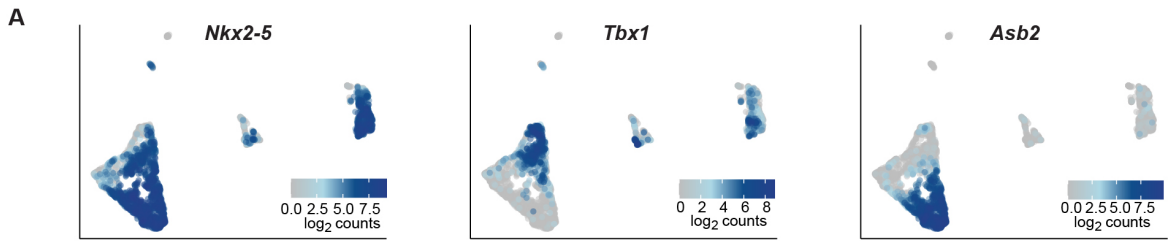


Fig. S6. Localization of progenitor clusters Me7 and Me8 during crescent formation.

A, UMAP plots showing the expression of *Nkx2-5*, *Tbx1* and *Asb2*. B-E, Maximum intensity projection (MIP) showing merged and individual expression of *Tbx1*, *Asb2* and *Nkx2-5* using multiplexed in situ Hybridization Chain Reaction (HCR) to identify clusters Me6, Me7 and Me8 at different stages of cardiac crescent development, as indicated in each panel. Merged image in D is the same as in Figure 2G. CC, cardiac crescent; FD, foregut diverticulum; ML, midline. See also Supplementary Figure 7.

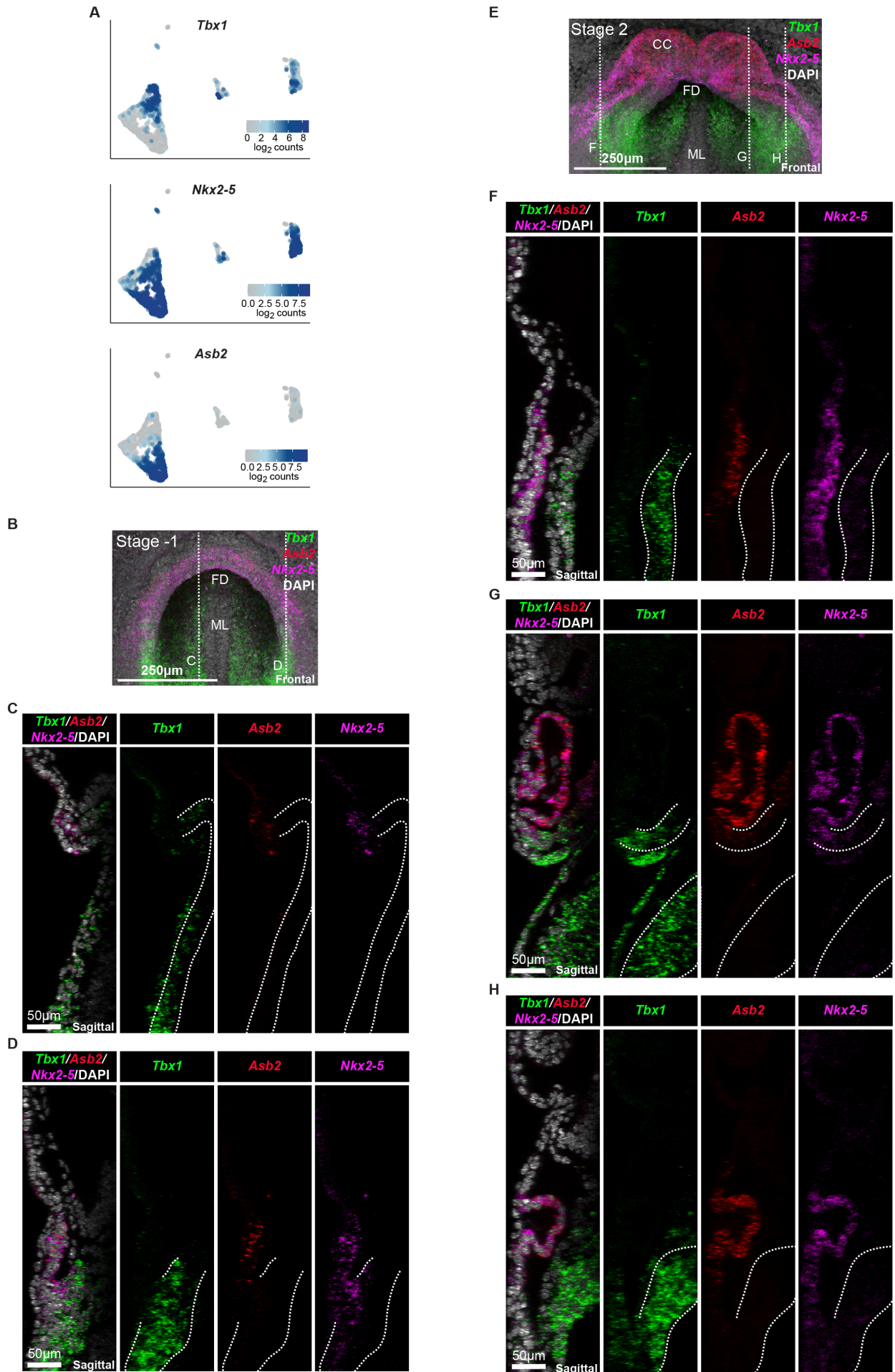
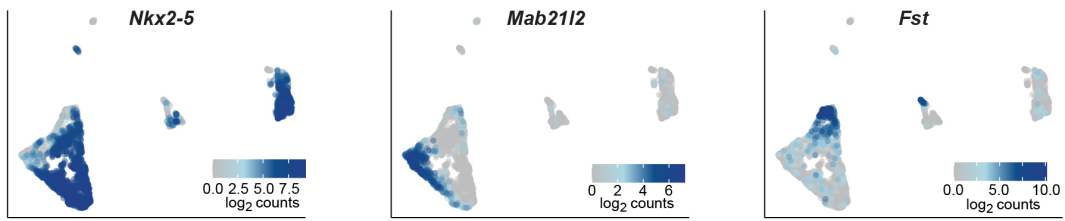


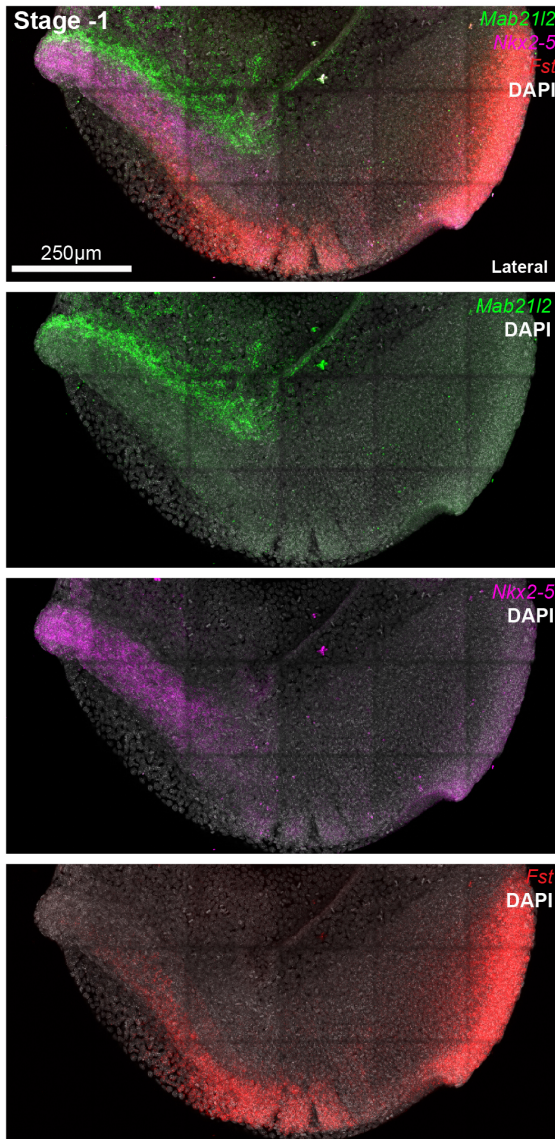
Fig. S7. Sagittal sections showing location of clusters Me6, Me7 and Me8.

A, UMAP plots showing the expression of *Tbx1*, *Nkx2-5* and *Asb2*. B, Maximum intensity projection (MIP) of stage -1 embryo from Supplementary Figure 6B, showing the location of sagittal sections shown in C and D (dotted lines). C-D, Sagittal sections showing merged and individual expression of *Tbx1*, *Asb2* and *Nkx2-5* in a medial (C) and lateral (D) sections. Dotted lines highlight *Tbx1* expressing mesoderm. E, MIP of stage 2 embryo from supplementary figure 6C, showing the location of sagittal sections shown in F-H (dotted lines). CC, cardiac crescent; FD, foregut diverticulum; ML, midline. F-H, Sagittal sections showing merged and individual expression of *Tbx1*, *Asb2* and *Nkx2-5* across a stage 2 cardiac crescent. Dotted lines highlight *Tbx1* expressing mesoderm. Panels represent different regions throughout the cardiac crescent highlighting differences between medial and lateral sections.

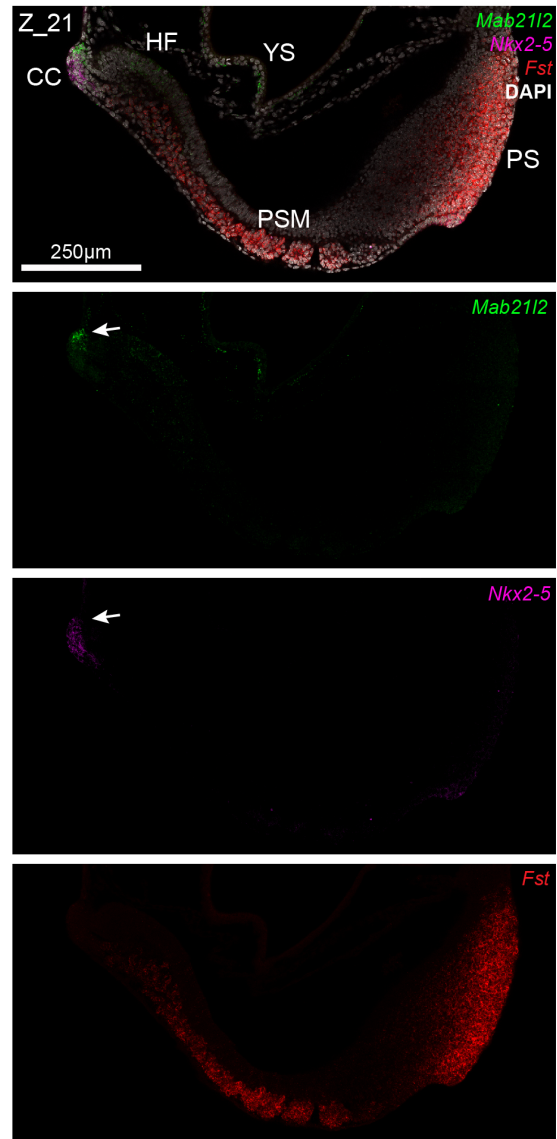
A



B



C



D

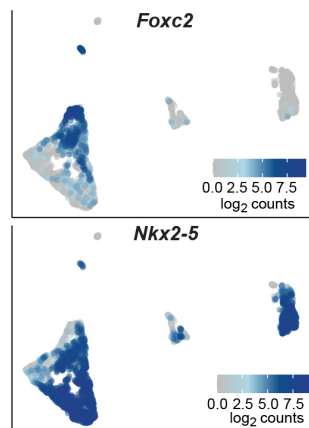
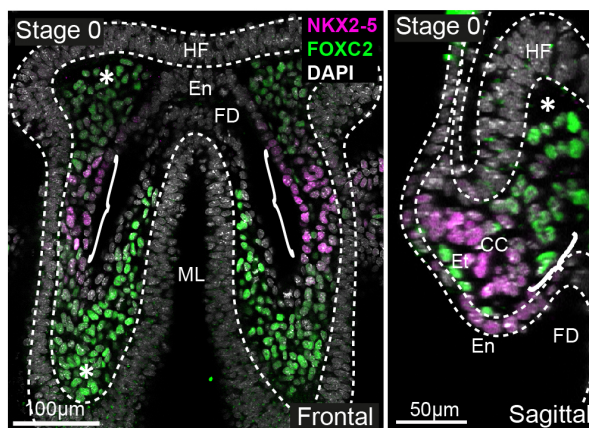


Fig. S8. Localization of progenitor clusters Me7, Me8 and Me5.

A, UMAP plots showing the expression of *Vsnl1*, *Mab21l2* and *Nkx2-5*. B, Maximum intensity projection (MIP) showing individual channels of the in-situ hybrid chain reaction (HCR) staining for *Mab21l2*, *Nkx2-5* and *Fst* in a stage -1 cardiac crescent formation, as shown in Fig. 5E. C, Individual Z sections taken from image in B highlighting *Fst* expression in both the presomitic mesoderm, primitive streak and extending dorsally behind the cardiac crescent towards the forming cranial mesoderm. D, transverse section (frontal view) of a stage 0 cardiac crescent showing expression of NKX2-5 and FOXC2 using whole mount immunostaining to identify Me7 & Me8. HF, headfold; En, Endoderm; FD, foregut diverticulum; ML, Midline. The brackets represent regions of mesoderm co-expressing FOXC2 and NKX2-5 whilst asterisks represents regions of mesoderm expression FOXC2 but lacking NKX2-5. Right panel, sagittal section showing expression of NKX2-5 and FOXC2 in a dorsal region behind the cardiac crescent extending into the cranial mesoderm. Endothelial (Et) expression of FOXC2 could also be detected.

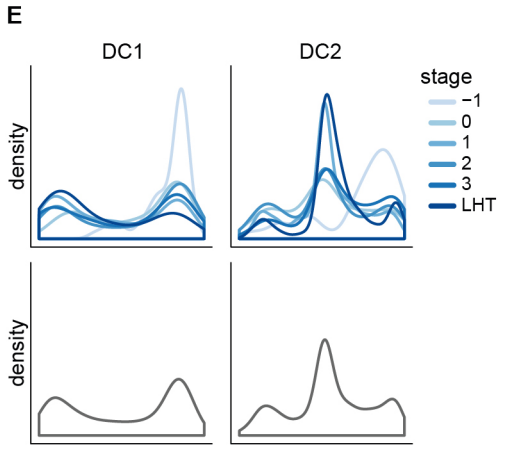
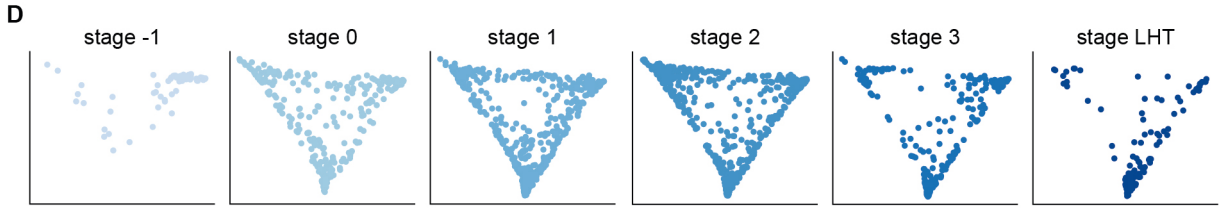
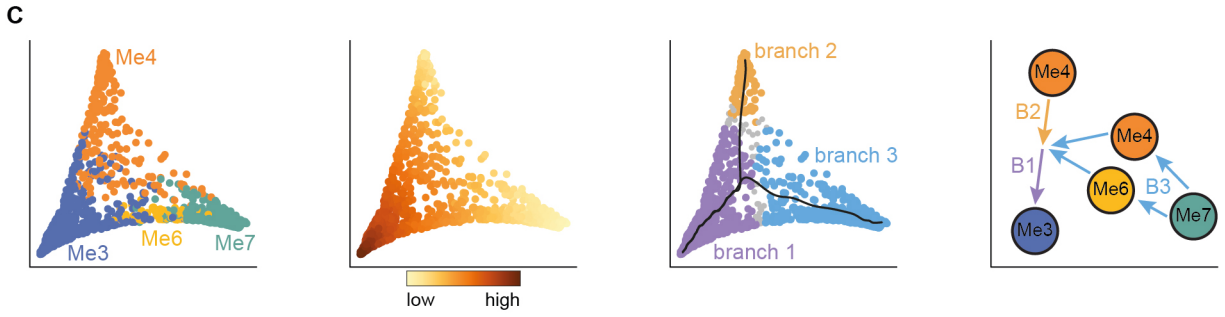
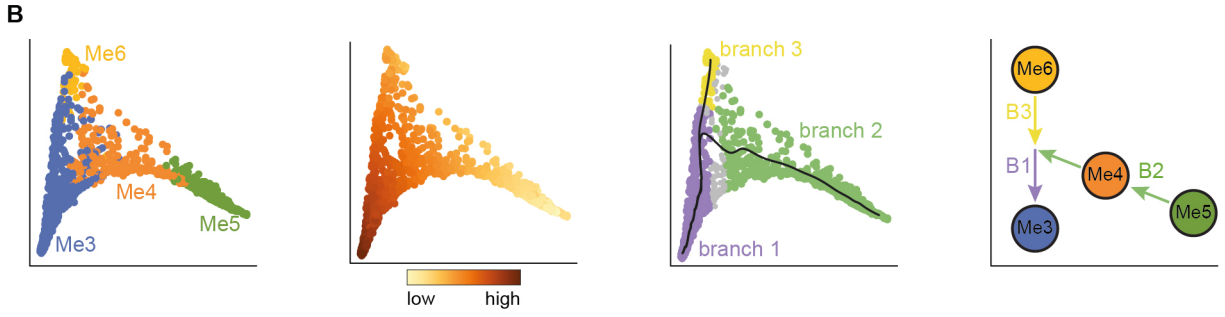
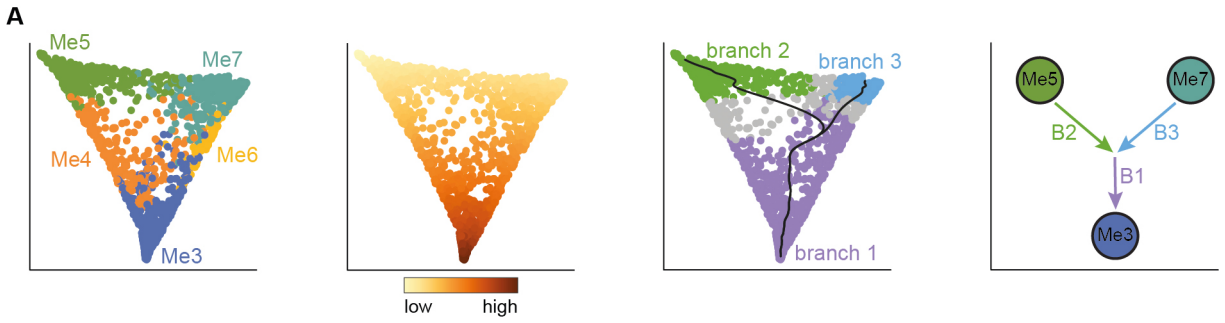


Fig. S9. Trajectory inference suggests two alternative pathways to cardiomyocytes.

Diffusion map analysis of the cardiac mesoderm cells suggested a branching point separating the Me5, Me7 and M3-Me6 clusters into independent branches (Figure 2H). For each panel, the first three plots are diffusion maps colored by cluster identity, diffusion pseudotime and branch assignment; for the latter, the average paths connecting the cells at the tip of each branch are indicated by lines. The fourth plot is a diagram summarizing the trajectories inferred based on the branching points. A, Same diffusion map as shown in Figure 2H, including all Me3-Me7 clusters. B, Diffusion map excluding cells from the Me7 cluster, to explore the trajectory involving the Me5 cluster. The branching point suggests that cells from Me6 are in a different trajectory to that followed by cells from Me5-Me4-Me3. C, Diffusion map excluding the cells from the Me5 cluster. In this case, the most distinct cells from the Me4 cluster are in a different trajectory to the one followed by the Me7-Me6-Me3 cells. D, Diffusion map from A split by developmental stage. All stages but stage -1 cover the whole diffusion space. E, Density plots of cells from each stage along the first and second diffusion components. At the bottom, the density plot for all cells. All stages but stage -1 follow a similar distribution.

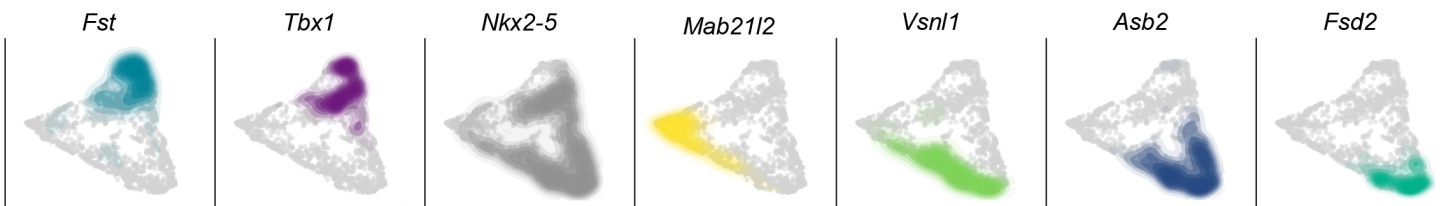
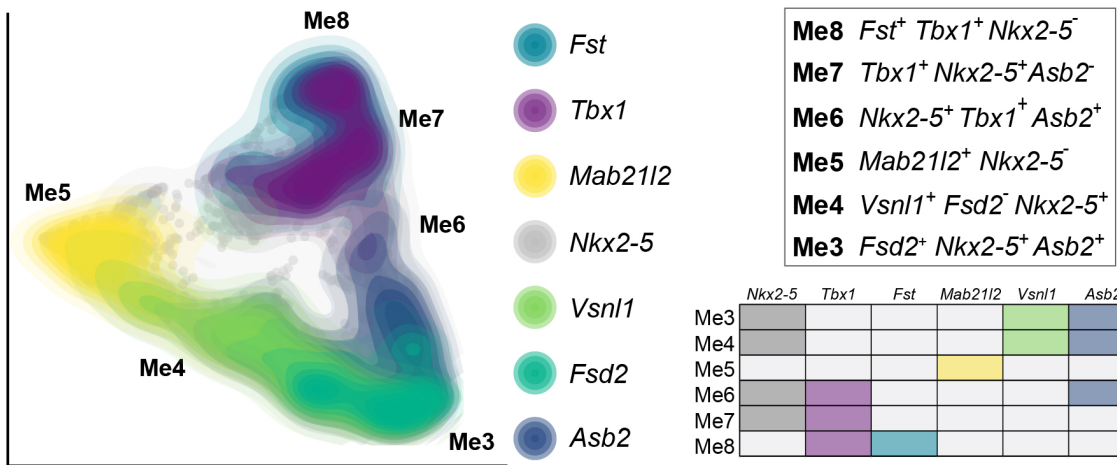


Fig. S10. Combinatorial marker genes used to anatomically localize cardiac mesoderm clusters.

The combinatorial expression of seven marker genes allows identification of the cardiac mesoderm clusters Me3-Me8. Marker genes were selected based on their ability to allow the specific identification of the 6 cardiac mesoderm related clusters. Each marker gene either uniquely marked a single cluster or could be used in combination to identify a single cluster. These were manual selected based on the log fold expression per cluster. The UMAP plots are overlaid with contour plots that represent the density of cells expressing the gene above log₂ normalized counts of 1. Both merged and individual plots are shown.

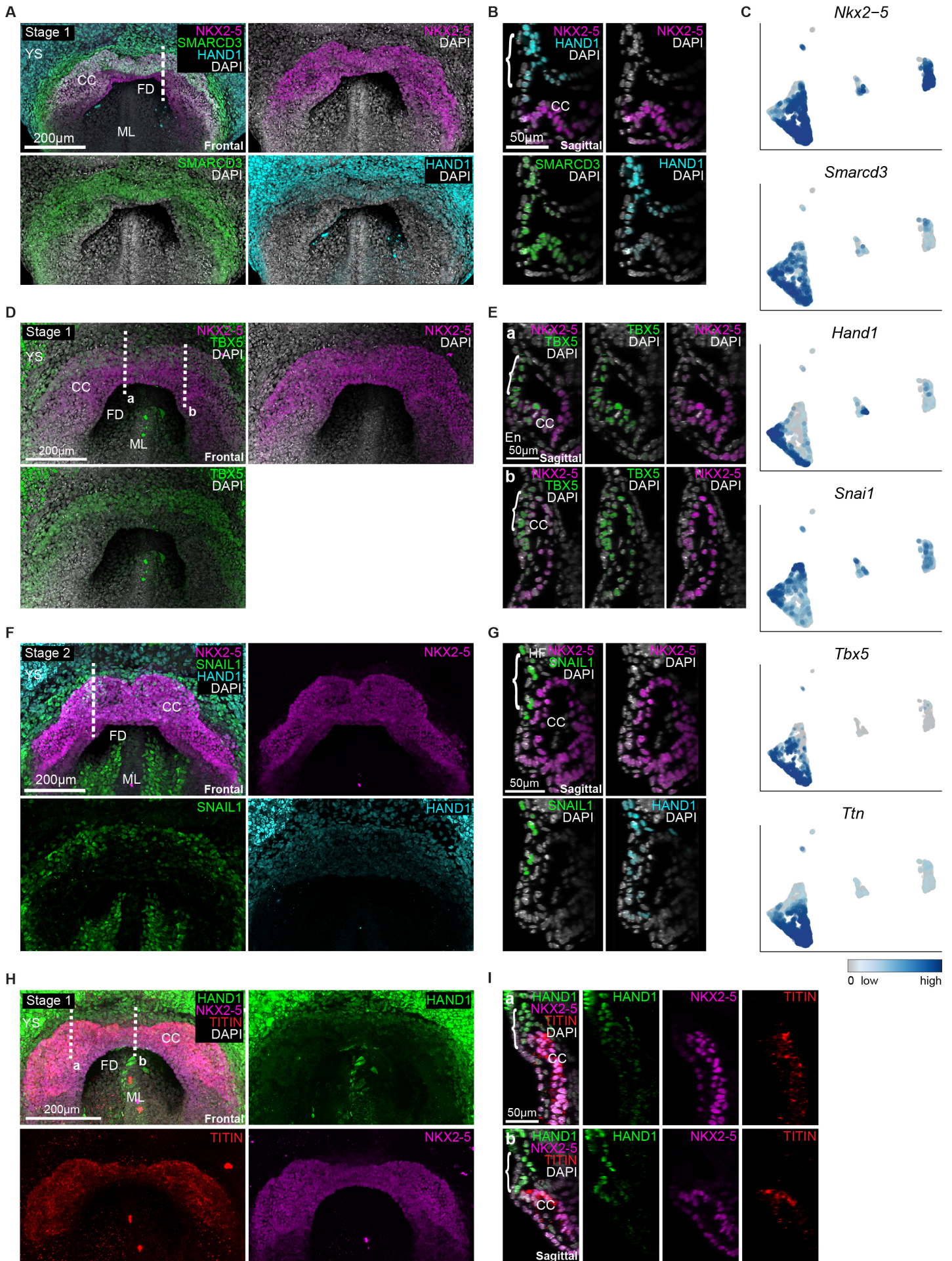


Fig. S11. Identification of Me5 as an anatomical distinct cell type.

A, Maximum intensity projections (MIP) for the embryo used in the whole mount immunostaining shown in Fig. 3A. Dotted line indicates the location of sagittal section. B, Sagittal sections showing individual channels used in the whole mount immunostaining in Fig. 3A. NKX2-5 immunostaining reveals cardiac crescent morphology, whilst SMARCD3 and HAND1 (bracket) can be seen extending outside of this NKX2-5+ domain, to a more rostral position. CC, cardiac crescent; FD, foregut diverticulum; ML, midline; YS, yolk sac. C, UMAP plots showing the expression of genes used to characterize the juxta-cardiac field. D, MIP of a stage 1 cardiac crescent stained for TBX5 and NKX2-5 using whole mount immunofluorescence, showing both merged and individual channels. Dotted lines indicate the location of sagittal sections shown in panel E. E, Sagittal sections at a medial (a) and lateral (b) position highlighting the relative position of TBX5 positive, NKX2-5 negative juxta cardiac field cells (bracket). F, MIP for the embryo used in whole mount immunostaining in Fig. 3C showing both merged and individual channels. Dotted line indicates the location of sagittal section shown in G. G, Sagittal sections showing individual channels used in Fig. 3C. SNAIL1 could be detected in the juxta cardiac field (bracket) confirming the sc-RNAseq data which showed its expression in Me5 (Fig. 1D). H, MIP of a stage 1 cardiac crescent stained for HAND1, NKX2-5 and TITIN using whole mount immunofluorescence showing both merged and individual channels. Dotted lines indicate the location of sagittal sections shown in panel I. I, Sagittal sections at a medial (a) and lateral (b) position highlighting the relative position of HAND1 positive, NKX2-5 negative juxta cardiac field cells (bracket) in relation to cardiomyocytes of the cardiac crescent as marked by TITIN expression.

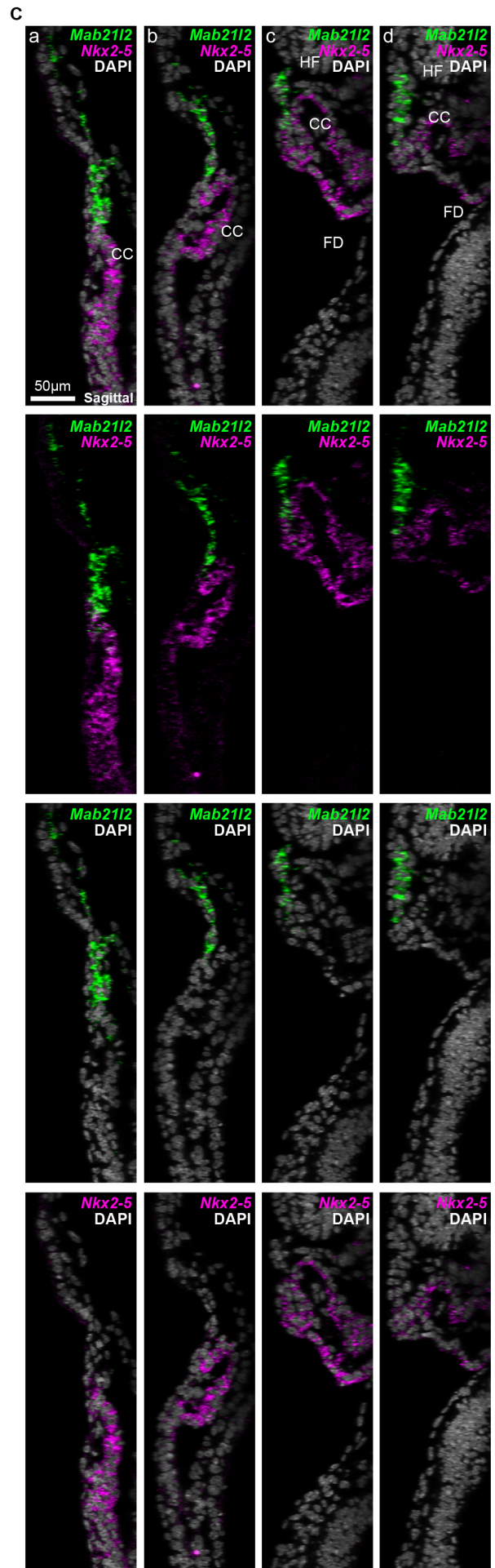
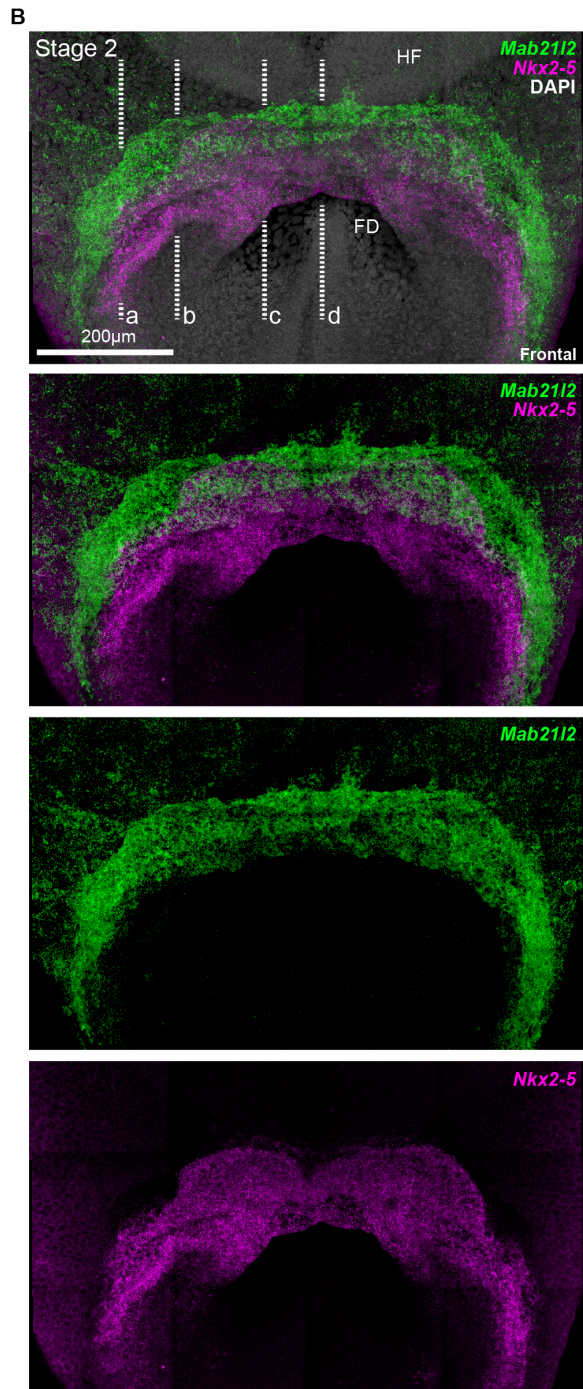
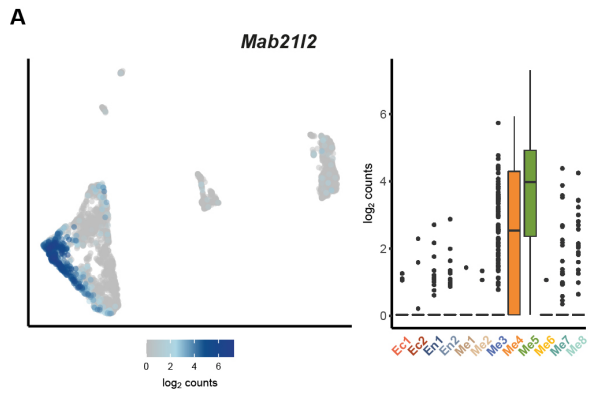
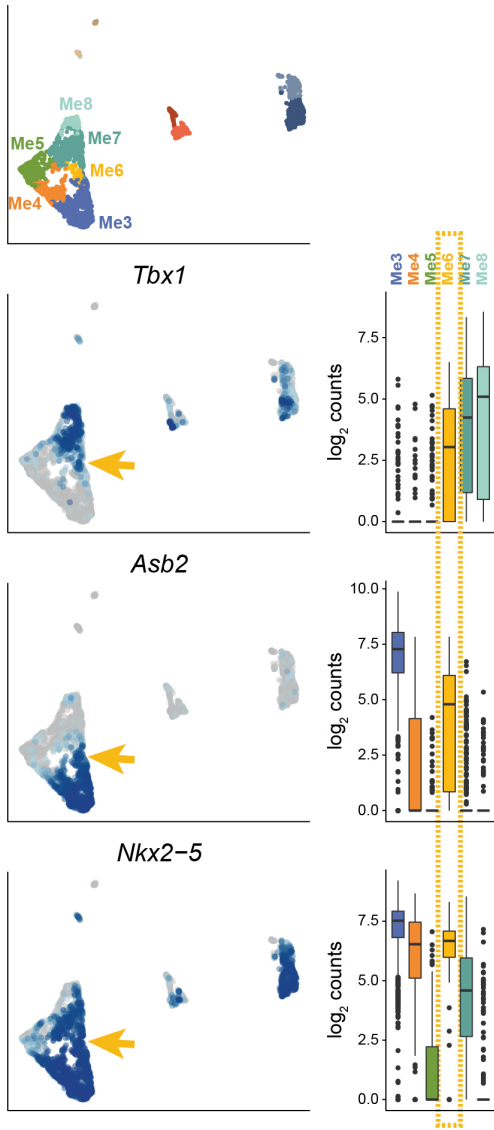


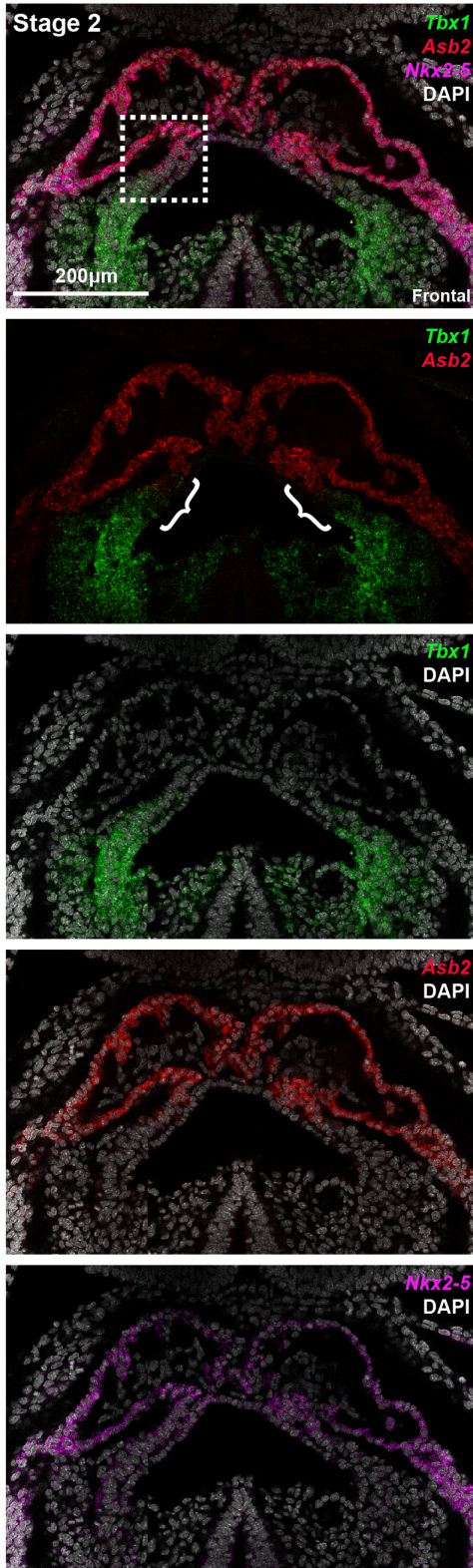
Fig. S12. *Mab21l2* marks an anatomically distinct region corresponding with Me5.

A, UMAP and boxplots showing expression of *Mab21l2*. B, Maximum intensity projection (MIP) of in-situ hybrid chain reaction (HCR) staining showing merged and individual expression of *Mab21l2* and *Nkx2-5* from figure 3E. Dotted lines indicate the location of sagittal sections in C. HF, headfold; FD, foregut diverticulum. C, Sagittal sections taken from the stage 2 embryo shown in figure 3E, at different locations throughout the cardiac crescent, highlighting the specific location of Me5. CC, cardiac crescent; FD, foregut diverticulum.

A



B



C

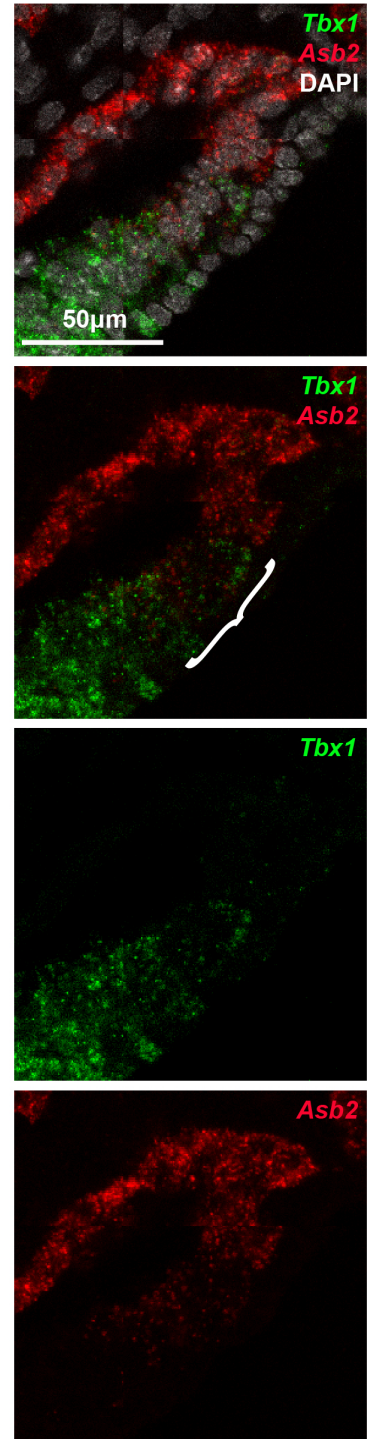


Fig. S13. Localization of the Me6 transition state.

A, UMAP and boxplots showing expression of SHF marker *Tbx1*, cardiomyocyte marker *Asb2* and cardiac progenitor marker *Nkx2-5*. Cluster Me6 was the only population of cells expressing all three markers (arrows), allowing us to specifically identify the anatomical localization of this cell state. B, Single transverse Z-section shown in Fig. S6C showing merged and individual expression of *Tbx1*, *Asb2* and *Nkx2-5*, using multiplexed in situ Hybridization Chain Reaction (HCR). Brackets show the anatomical location of Me6 as determined by the co-expression of both *Tbx1* and *Asb2*. Dotted box represents region shown at higher magnification in panel C.

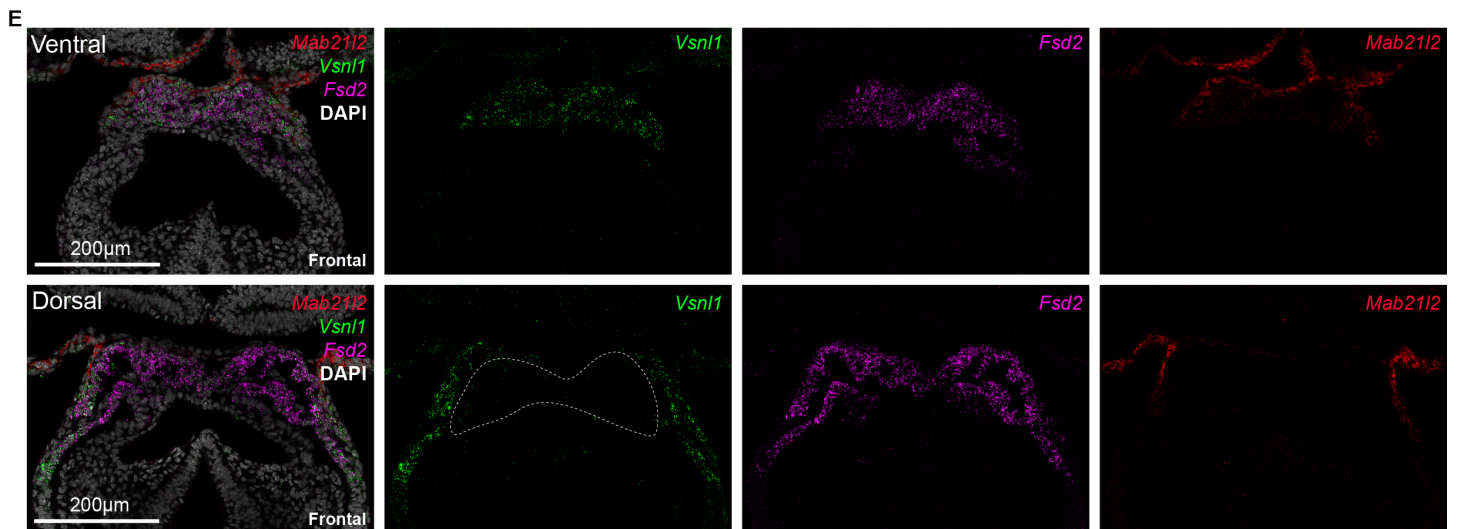
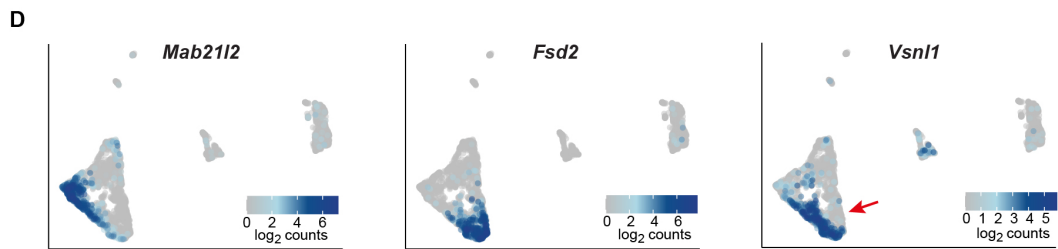
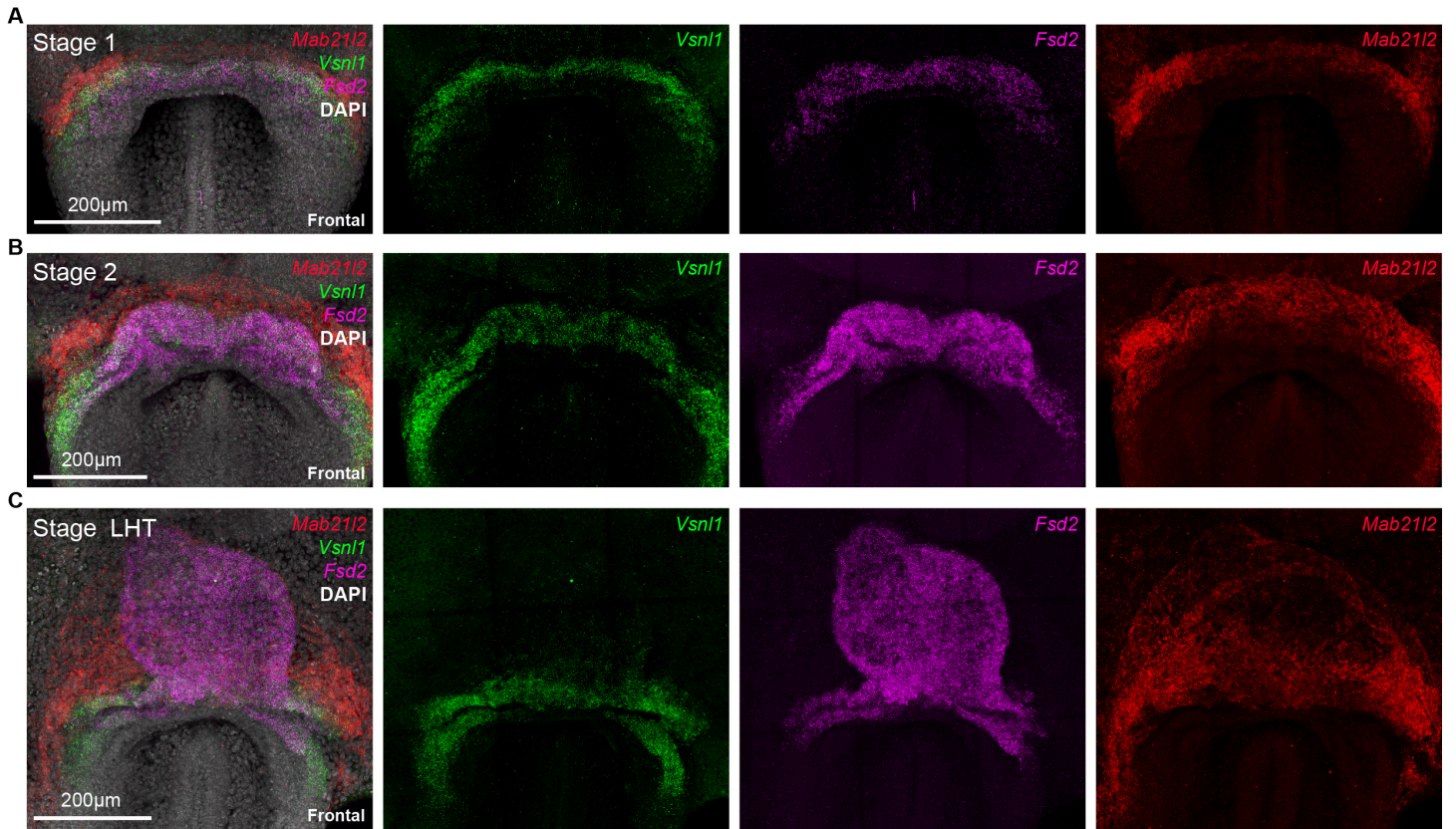


Fig. S14. Localization of the Me4 transition state during cardiac crescent development.

A-C, Maximum intensity projection (MIP) showing merged and individual expression of *Fsd2*, *Vsn11* and *Mab2112* using multiplexed in situ Hybridization Chain Reaction (HCR) to identify cluster Me4 at different stages of cardiac crescent development, as indicated in each panel. (merged image in A is the same as in Figure 4B). Me4 could be identified as the only population of cells which expressed *Vsn11* but did not express *Fsd2* whilst Me3 expressed both *Vsn11* and *Fsd2*. D, UMAP plots showing the expression of *Mab2112*, *Fsd2* and *Vsn11*. Note *Vsn11* expression is restricted to only a fraction of Me3 cells (arrow). E, Single ventral and dorsal transverse sections of the stage 2 embryo in B, showing *Vsn11* expression in ventral cardiomyocytes, in contrast to dorsal *Fsd2*-expressing cardiomyocytes that do not express *Vsn11* (dotted region).

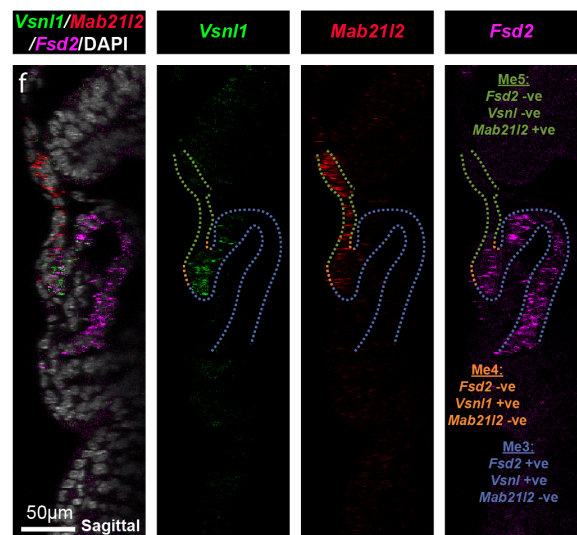
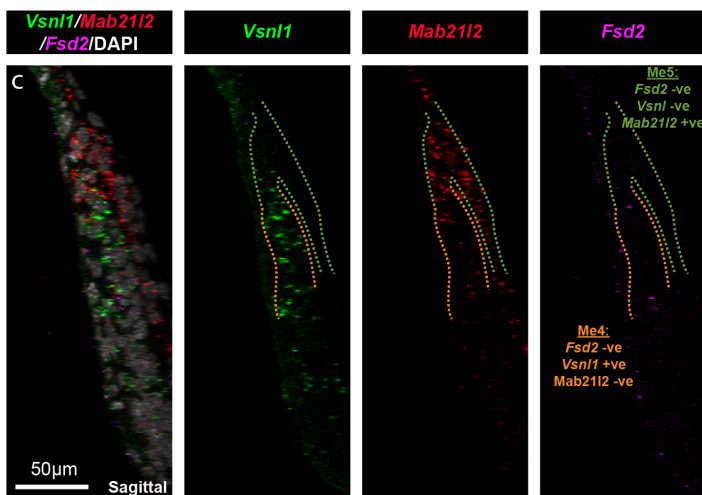
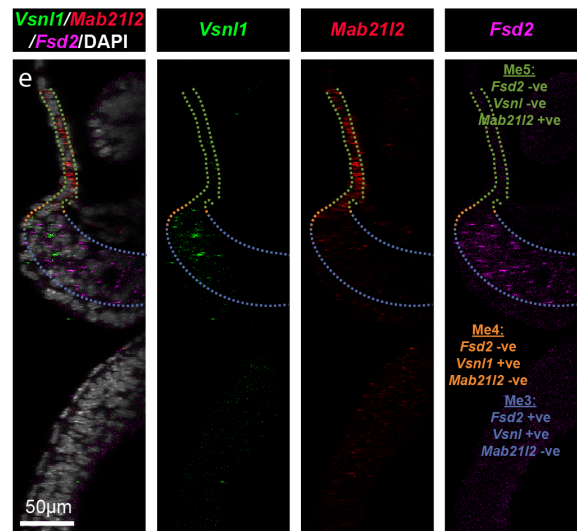
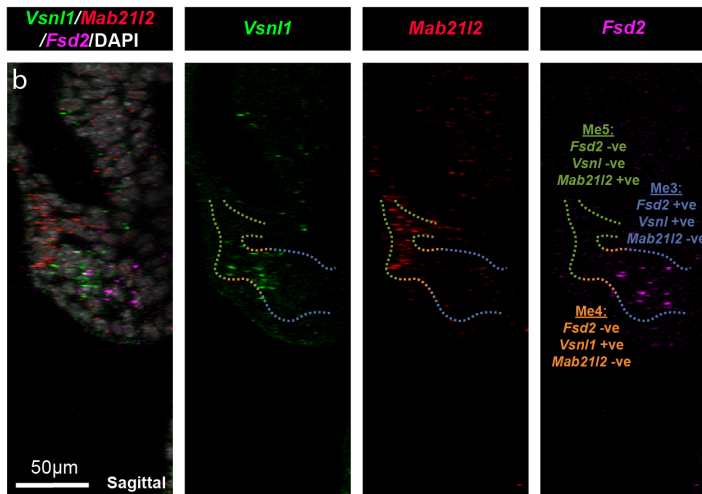
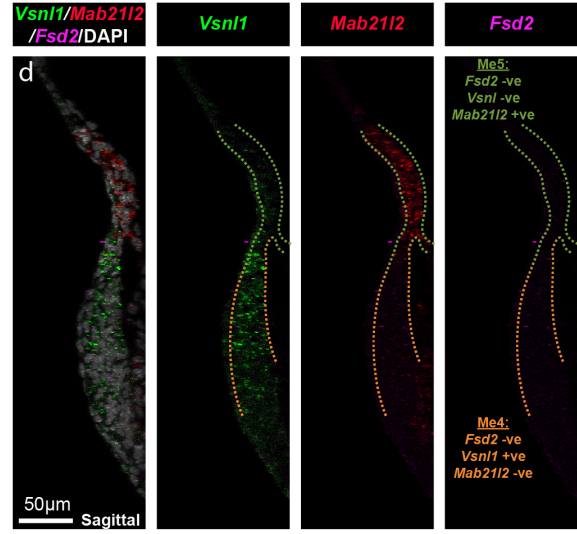
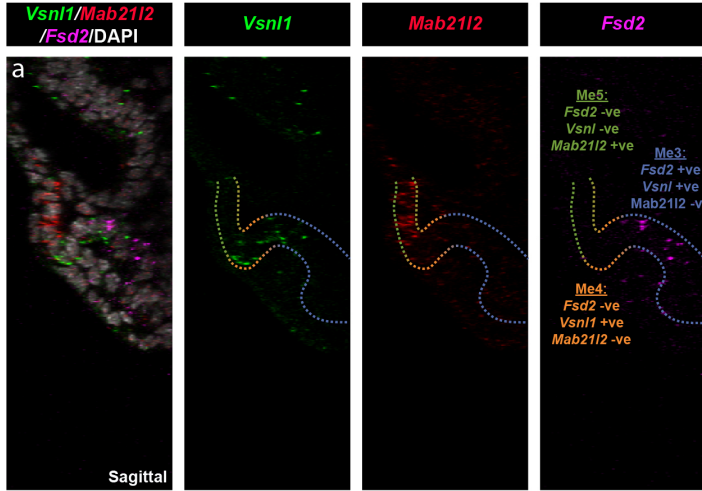
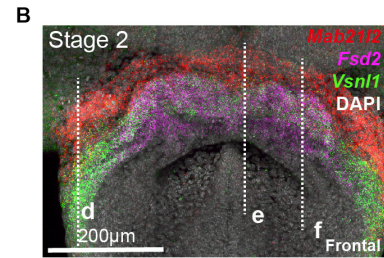
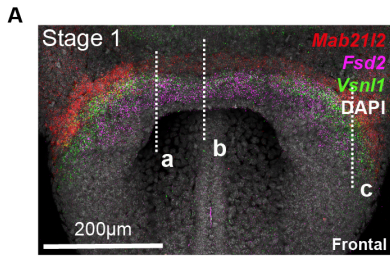


Fig. S15. Sagittal sections reveal the transition from Me5-Me4-Me3 cardiomyocyte trajectory.

A, Maximum intensity projection (MIP) of a stage 1 embryo from Figure 4B, showing the location of sagittal sections in panels a-c. a-c, Merged and individual expression of *Vsn11*, *Mab2112* and *Fsd2* using multiplexed in situ Hybridization Chain Reaction (HCR) to identify cluster Me4. Me4 could be located between cluster Me5 and Me3. B, Same as A, but for the stage 2 embryo from Supplementary Figure 14B. Me4 could be located between clusters Me5 and Me3.

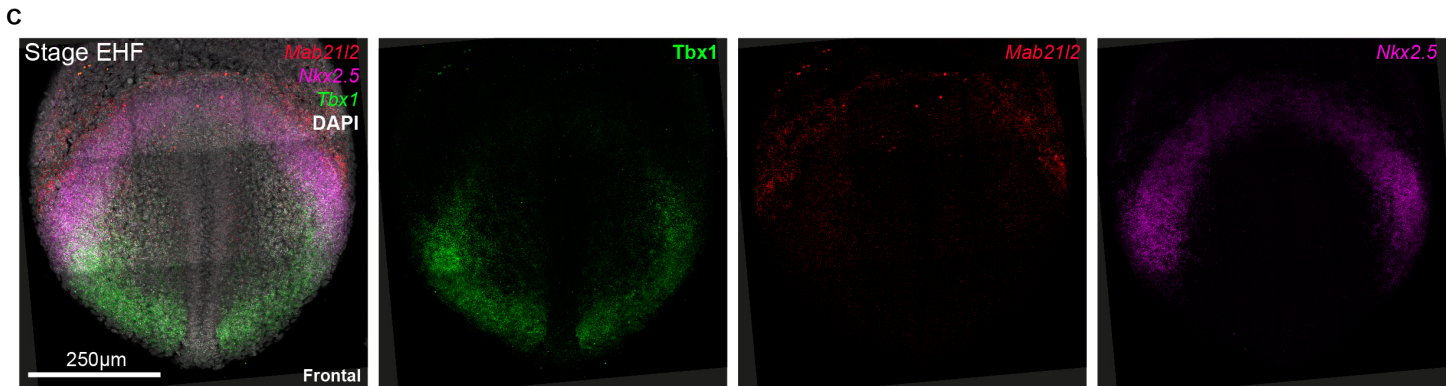
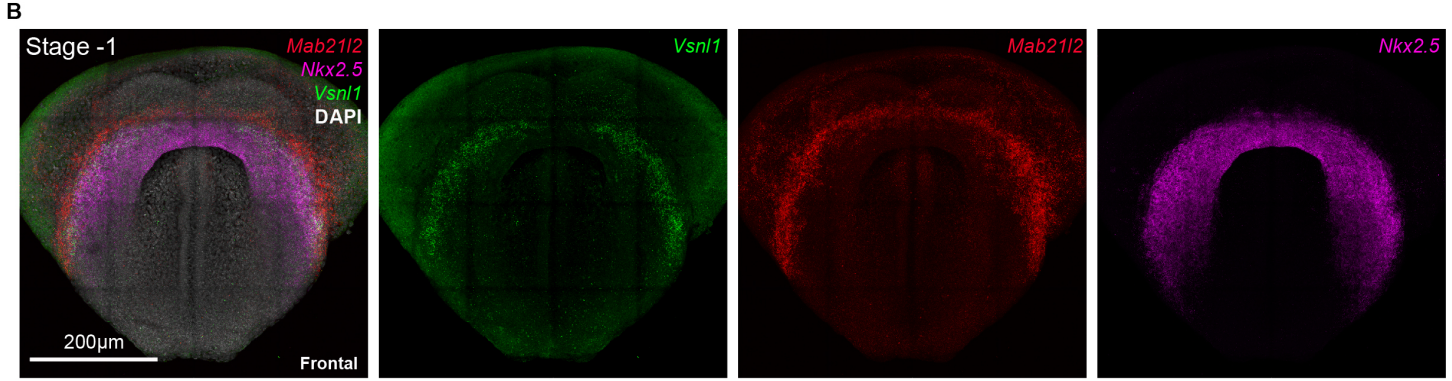
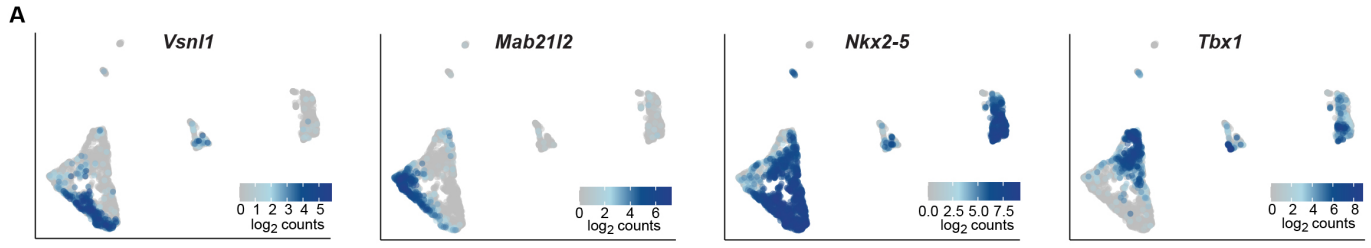


Fig. S16. Me4 could be first detected just prior to cardiac crescent formation.

A, UMAP plots showing the expression of *Vsnl1*, *Mab21l2*, *Nkx2-5*, *Tbx1*. B, Maximum intensity projection (MIP) from Figure 4D showing merged and individual expression of *Vsnl1*, *Mab21l2* and *Nkx2-5*, using multiplexed in situ Hybridization Chain Reaction (HCR) to identify cluster Me4 at stage -1. C, Maximum intensity projection (MIP) from Figure 4E showing merged and individual expression of *Tbx1*, *Mab21l2* and *Nkx2-5*, using multiplexed in situ Hybridization Chain Reaction (HCR) to identify cluster Me7 in relation to Me5 and Me4 at Early headfold Stage (EHF).

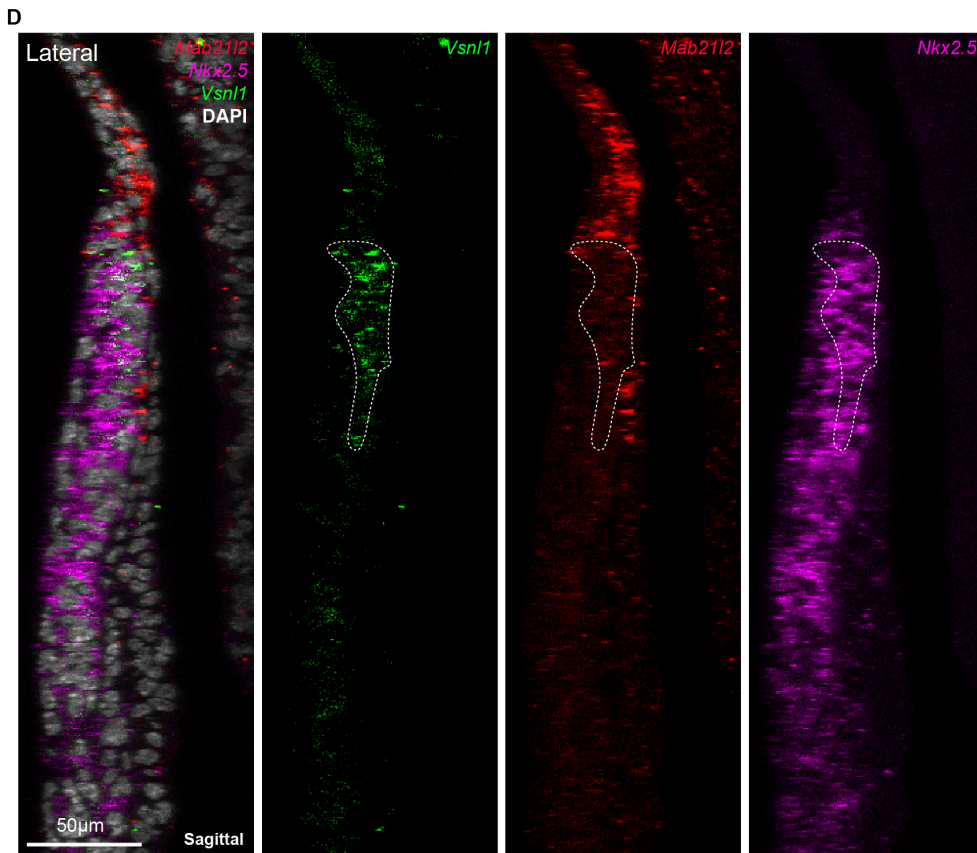
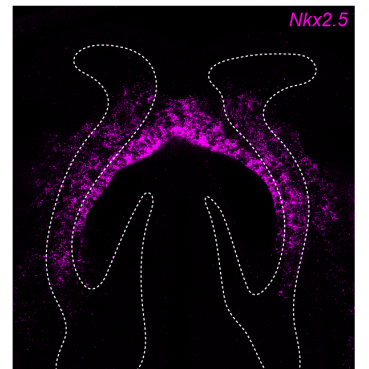
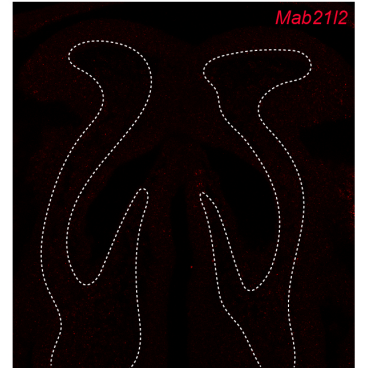
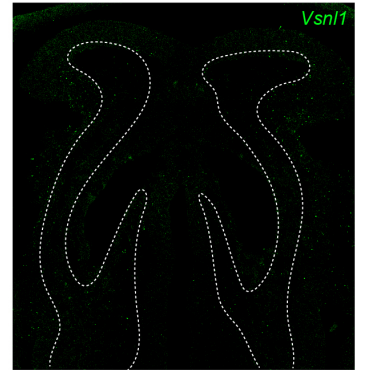
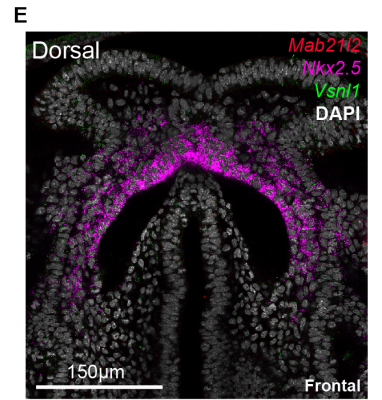
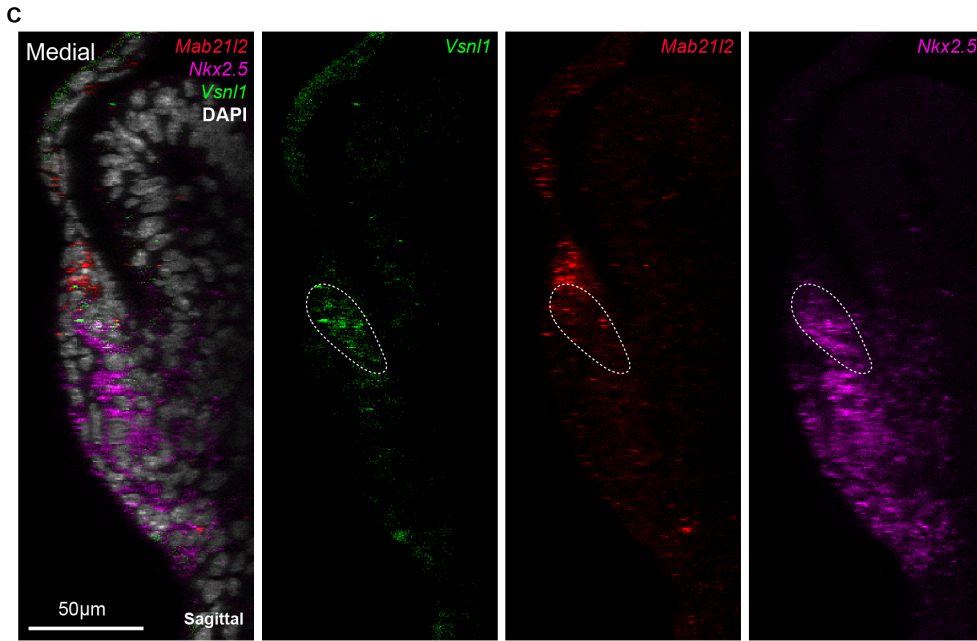
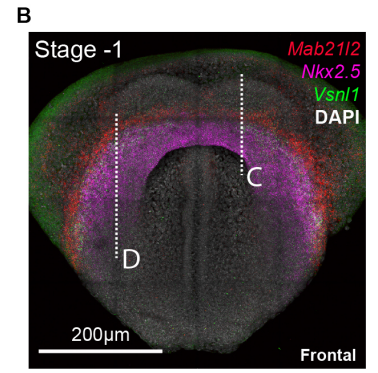
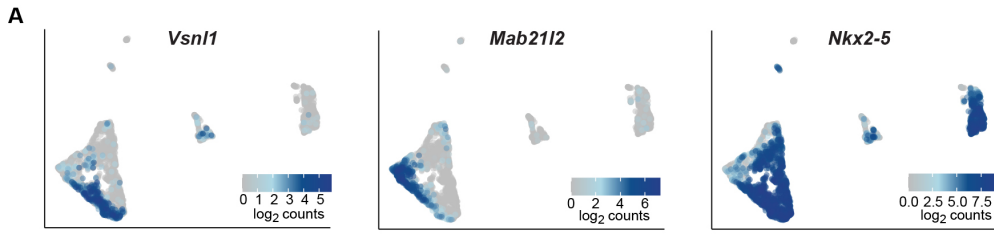


Fig. S17. Sagittal sections reveal the anterior location of Me4 prior to cardiomyocyte formation.

A, UMAP plots showing the expression of *Vsn11*, *Mab2112* and *Nkx2-5*. B, Maximum intensity projection (MIP) showing individual channels of the in-situ hybrid chain reaction (HCR) staining for *Mab2112*, *Nkx2-5* and *Vsn11* at stage -1, as shown in Figure 4D. Dotted lines indicate the location of sagittal sections shown in C and D. C, Sagittal section from a medial region of embryo shown in B showing individual channels of the in-situ HCR for *Vsn11*, *Mab2112* and *Nkx2-5*. Dotted regions highlight cluster Me4 as marker by the expression of *Vsn11*. D, Same as C but in a lateral region of the forming cardiac crescent. E, Transverse section through a dorsal region of the embryo shown in B, highlighting the absence of *Vsn11* and *Mab2112* expression in the dorsal mesoderm. Dotted region highlights mesoderm.

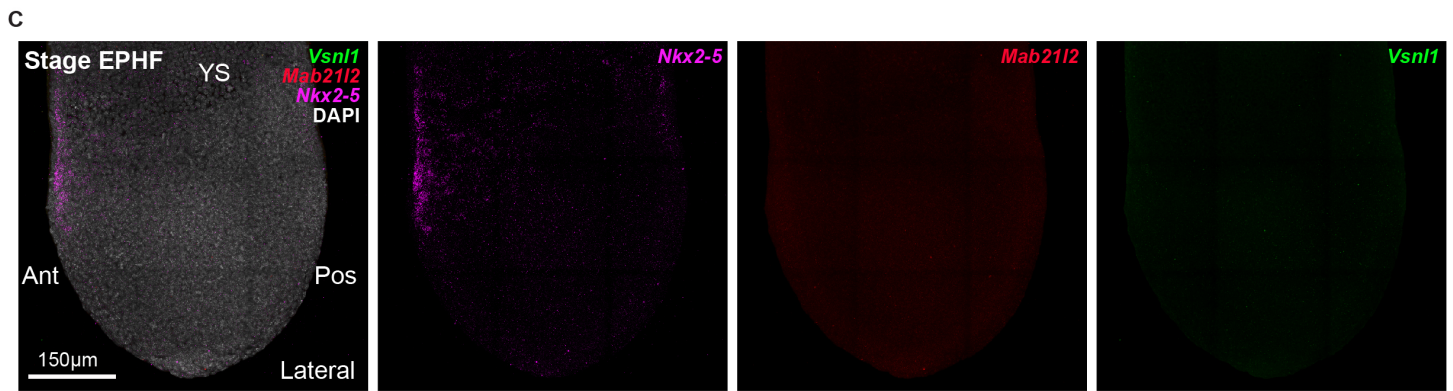
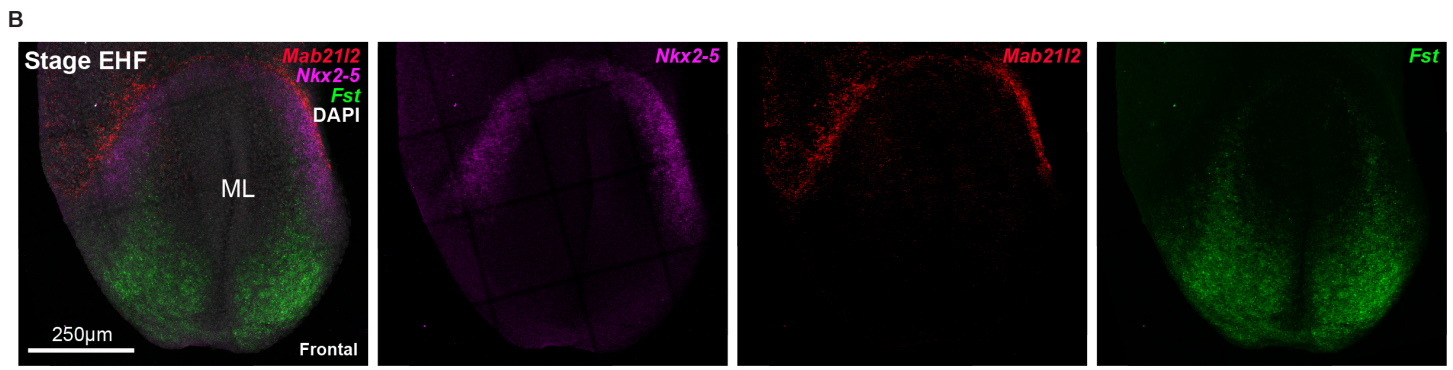
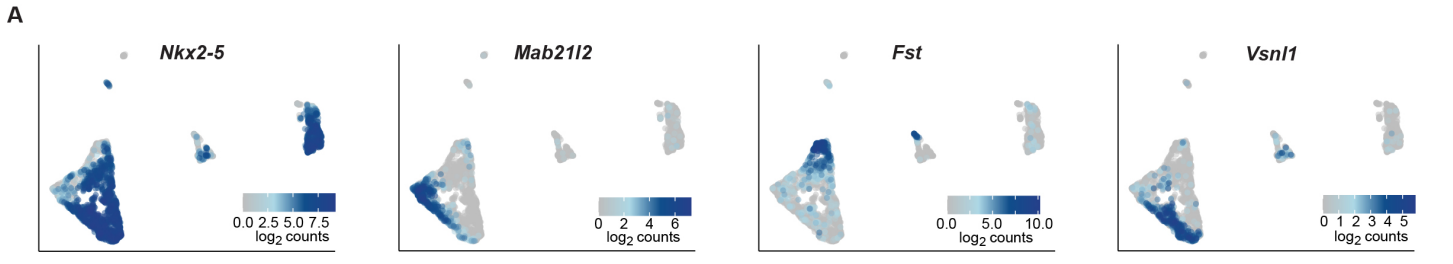


Fig. S18. *Mab21l2* is expressed prior to cardiac crescent formation in a region distinct from *Nkx2-5* expressing progenitors.

A, UMAP plots showing the expression of *Nkx2-5*, *Mab21l2*, *Fst* and *Vsn1l*. B, Maximum intensity projection (MIP) showing individual channels of the in-situ hybrid chain reaction (HCR) staining for *Mab21l2*, *Nkx2-5* and *Fst* at Early headfold (EHF) stage, prior to cardiac crescent formation, as shown in Figure 5E. C, MIP showing merged and individual channels of the in-situ hybrid chain reaction (HCR) staining for *Mab21l2*, *Nkx2-5* and *Vsn1l* at Early pre-headfold (EPHF) stage. These highlight that the expression of *Nkx2-5* can be detected prior to both the Me4 marker *Vsn1l* and Me5 marker *Mab21l2*. ML, midline; YS, yolk sac; Ant, anterior; Pos, posterior.

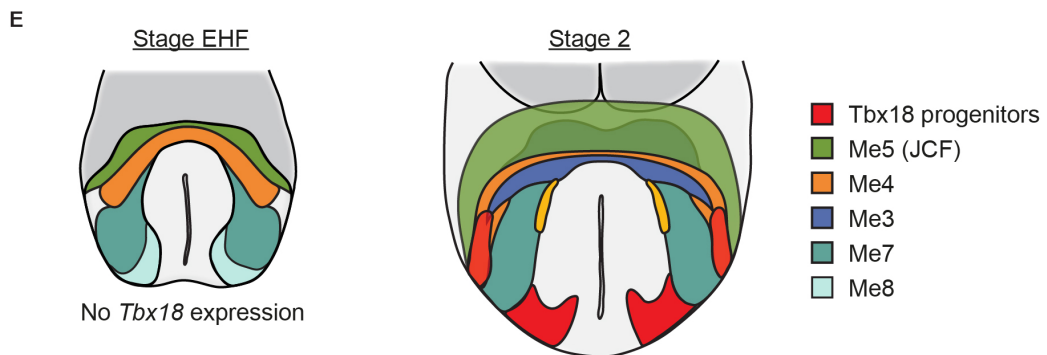
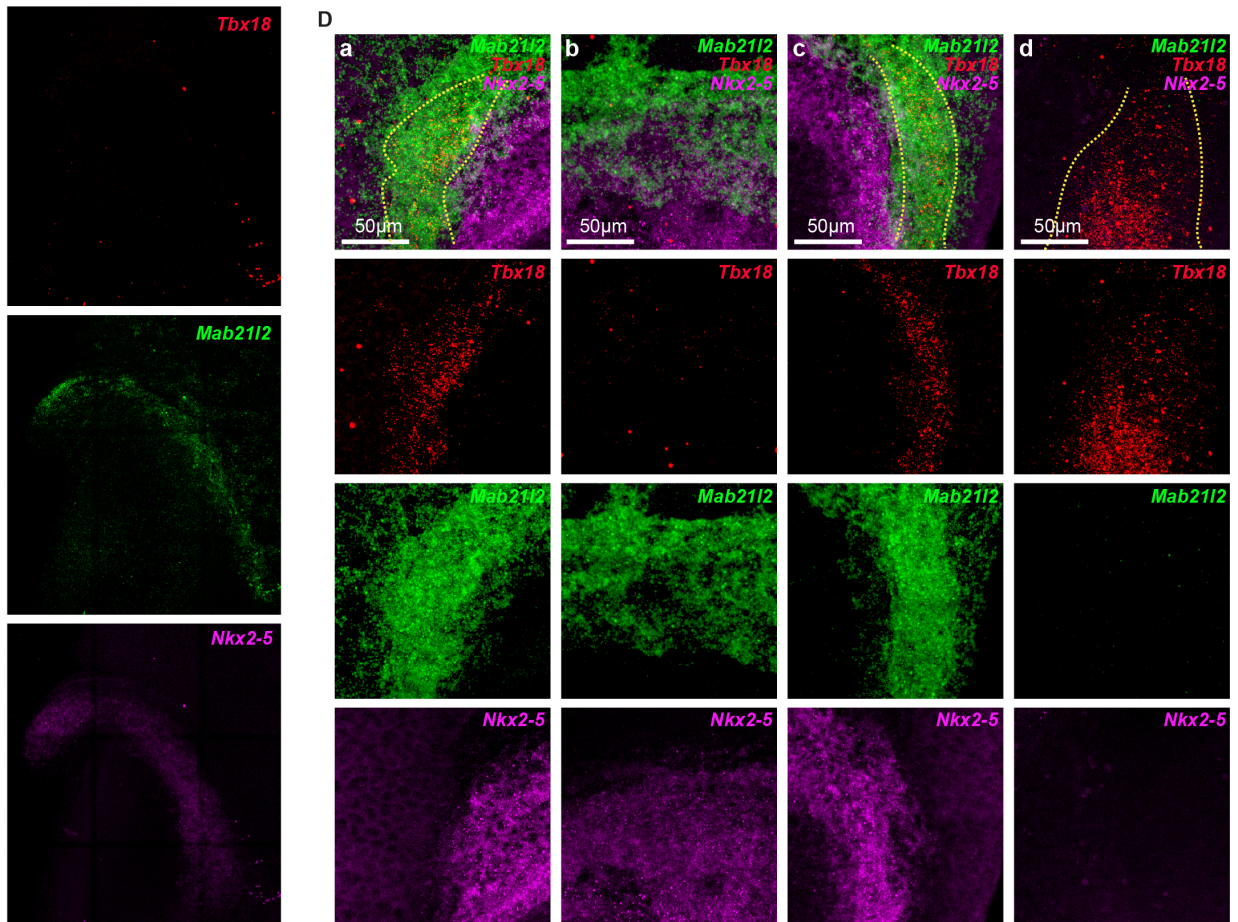
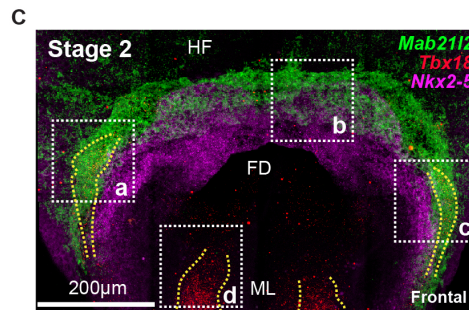
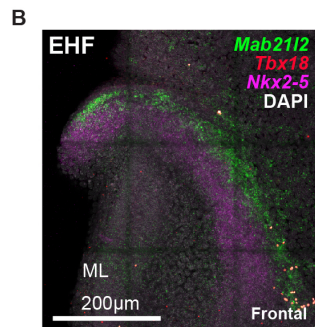
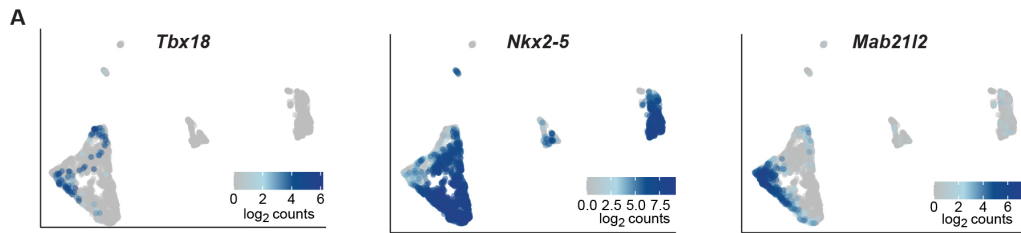


Fig. S19. *Mab2112* marks Juxta-cardiac field (JCF) progenitors in a region distinct from *Tbx18* positive progenitors.

A, UMAP plots showing expression of *Tbx18*, *Nkx2-5* and *Mab2112*. B, Maximum intensity projection (MIP) of an embryo at early headfold (EHF) stage (pre-cardiac crescent) stained using in-situ HCR for *Mab2112*, *Tbx18* and *Nkx2-5*. At this stage, *Tbx18* could not be detected in the embryo; however, *Mab2112* was localized in a region between the extra-embryonic mesoderm and *Nkx2-5* progenitors. C, MIP of the in-situ hybrid chain reaction (HCR) staining in Fig. 5F. Dotted boxes (a-c) represent regions shown in D. Yellow dotted lines highlight *Tbx18* expression. HF; headfold; FD, foregut diverticulum; ML, midline. D, Higher magnification of regions highlighted in C. *Tbx18* could be weakly detected in the lateral regions of the cardiac crescent starting between stage 1 and 2 (yellow dotted regions in box a and c; but not in the more medial *Mab2112* expressing regions (box b). *Tbx18* could also be detected in a region corresponding with Me8 (box d). E, Schematic diagram showing the location of the JCF (Me5) in relation to *Nkx2-5* progenitors as well as *Tbx18* expressing progenitors (located in clusters Me8, Me4, Me5 and Me7) during formation of the cardiac crescent.

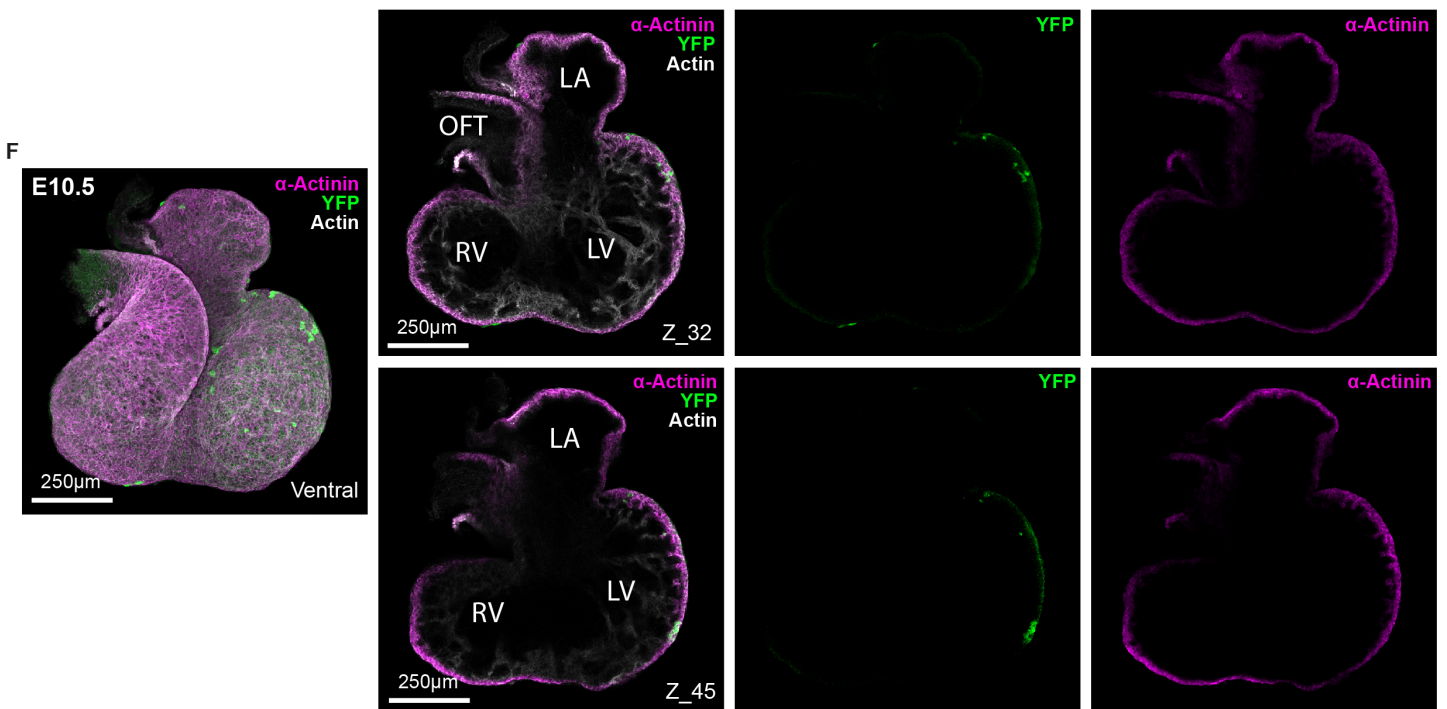
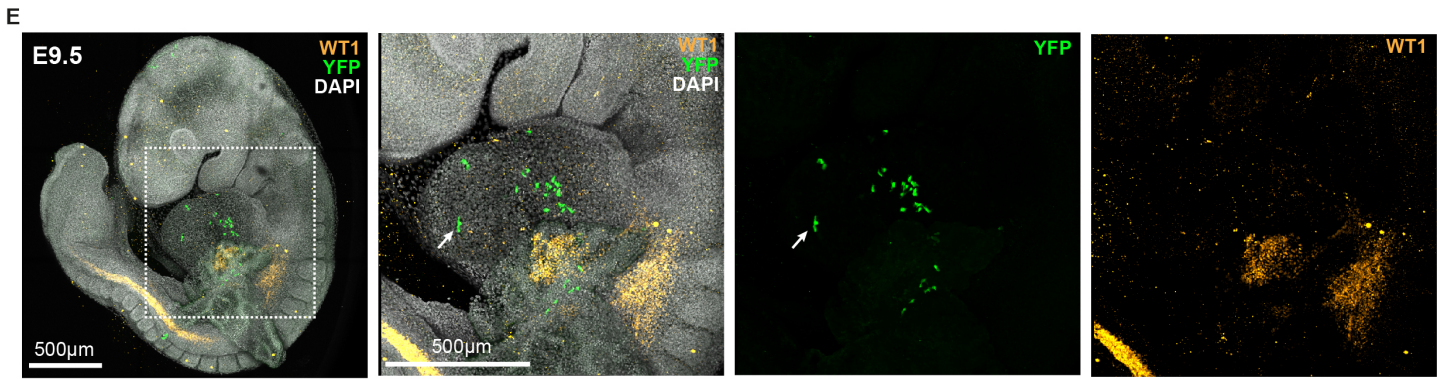
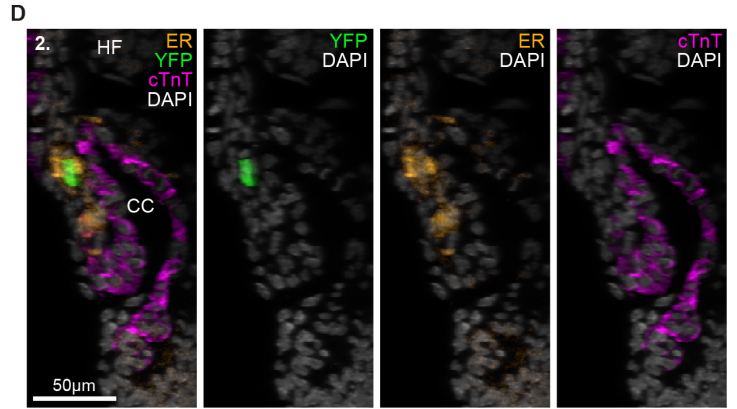
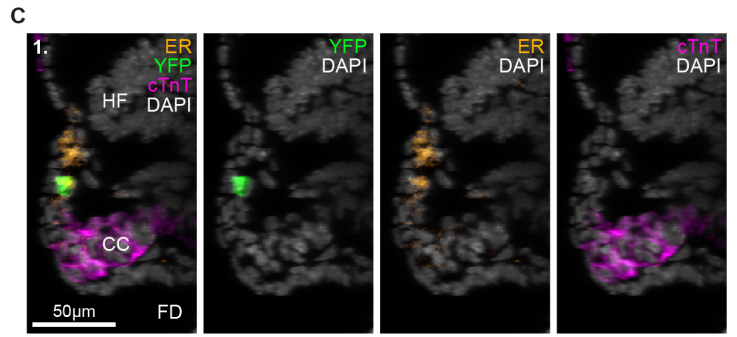
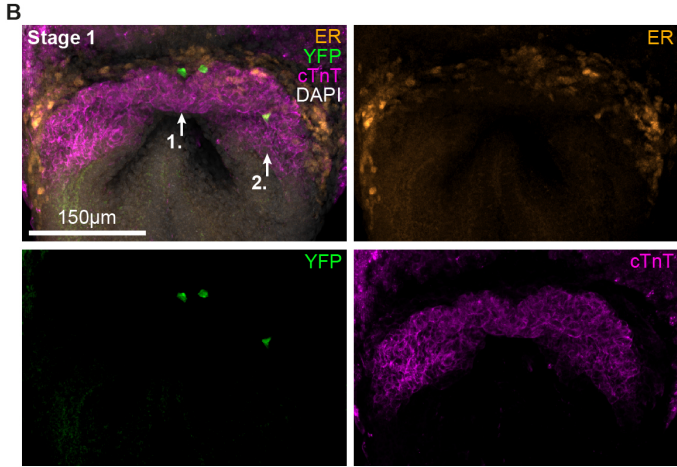
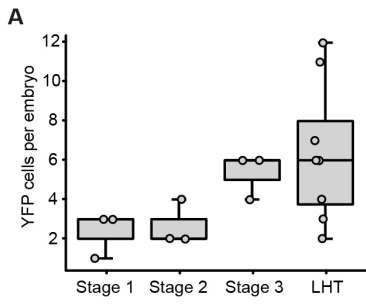


Fig. S20. Validation and analysis of a novel *Mab2112*-iCreERT2 transgenic mouse line.

A, Quantification of YFP expressing cells labeled at cardiac crescent to linear heart tube (LHT) stages following tamoxifen injection at E6.5 (n=17). The location of YFP cells within LHT stage embryos is characterised in Supplementary Table 1, all YFP cells at earlier cardiac crescent stages were located within the JCF. B, Maximum intensity projection (MIP) of a stage 1 cardiac crescent stained for YFP, cardiac Troponin T (cTnT) and the Estrogen receptor (ER) using whole mount immunofluorescence. Both merged and individual channels are shown. YFP staining shows the location of recombined cells after tamoxifen induction at E6.5. Staining for ER highlighted the regions in which iCreERT2 was expressed and confirmed that our novel *Mab2112*-iCreERT2 transgenic mouse line was correctly targeting the juxta-cardiac field. Arrows represent location of sagittal sections in C and D. Sagittal sections confirm that recombined YFP positive cells and ER expression is outside of the cTnT expressing cardiac crescent. C, Sagittal sections showing individual and merged channels used in the whole mount immunostaining in B, at medial location 1. D, Same as C but at lateral location 2. HF; headfold; FD, foregut diverticulum; CC, cardiac crescent. E, MIP of an E9.5 embryo whole mount immunofluorescent staining for WT1 and YFP. The dotted box represents the location of zoomed in panels shown to right, and highlights (arrows) WT1 negative YFP positive cells in the left ventricle. F, Left hand panel, Ventral MIP of whole mount immunohistochemistry for α -actinin and YFP in hearts at E10.5, highlighting the dual lineage potential of the JCF. The right panels show single Z sections through the E10.5 heart shown on the left, highlighting the co-expression of α -actinin and YFP in the left ventricle. LV; Left ventricle, RV; Right ventricle, LA; Left Atria, OFT, Outflow tract.

Table S1.

Embryo ID	JCF	IFT	Primitive Ventricle	Outflow
1	2	0	0	0
2	3	3	0	0
3	8	2	2	0
4	4	0	0	0
5	3	0	0	0
6	5	1	0	0
7	7	0	0	0
8	6	0	0	0
9	8	3	0	0
Total	46	9	2	0
Percentage	81%	16%	4%	0%

Location of YFP cells detected at E8.5 (Linear heart tube stage)

Table showing the number of YFP cells detected in different cardiac related regions of linear heart tube stage embryos after tamoxifen at 6.5dpc. Analysis was based on whole-mount ventral images of intact E8.5 embryos. Location of labelled cells within the cardiac region was based on morphology as well as immunofluorescence for sarcomeric α -Actinin (myocardium). JCF; Juxtacardiac Field, IFT; Inflow tract.

Table S2.

Embryo ID	Inflow/Atria	AVC	Left Ventricle	SC	ProEpi
1	4	6	0	10	17
2	14	4	5	1	8
3	12	12	5	4	22
4	10	0	6	10	0
5	2	3	0	0	9
6	12	10	0	0	7
Total	54	35	16	25	63
Percentage	28%	18%	8%	13%	33%

Location of YFP cells detected at E9.5

Table showing the number of YFP cells detected in different cardiac related regions at E9.5 after tamoxifen at 6.5dpc. Analysis was based on whole-mount image volumes of intact E9.5 embryos. All embryos were imaged from the left side due to their size, image volumes did not penetrate through the entire looped heart. In two of these embryos that were remounted on their right side and imaged, no YFP cells could be detected in the right ventricle or outflow. Also see tables S3 and S4 for information regarding contribution to the right ventricle and OFT at E10.5. Cell location was based on morphology as well as immunofluorescence for Sarcomeric α -Actinin (myocardium) and WT1 (proepicardium/epicardium). AVC; Atrioventricular canal, SC; Cells located on the surface of the heart, ProEpi; Proepicardium.

Table S3.

Embryo ID	Atria	AVC	Left Ventricle	Right Ventricle	Outflow Tract	Epicardium
1	0	0	7	0	0	6
2	0	0	2	0	0	8
3	0	6	3	0	0	41
4	0	0	9	0	0	23
5	0	1	34	0	0	22
6	0	1	11	0	0	14
Total	0	8	66	0	0	114
Percentage	0%	4%	35%	0%	0%	61%

Location of YFP cells detected at E10.5 from the ventral side.

Table showing the number of YFP cells detected in different heart regions at E10.5 after tamoxifen at 6.5dpc. Analysis was based on whole-mount images of isolated hearts at E10.5 imaged from the ventral side. Cell location was based on morphology as well as immunofluorescence for Sarcomeric α -Actinin (myocardium) and WT1 (Epicardium). AVC; Atrioventricular canal.

Table S4.

Embryo ID	Atria	AVC	Left Ventricle	Right Ventricle	Outflow Tract	Epicardium
1	0	9	1	0	0	12
2	3	0	0	0	0	49
3	0	16	6	0	0	23
4	0	21	22	0	0	56
5	0	23	24	0	0	48
6	0	10	16	0	0	49
7	0	3	0	0	0	0
Total	3	82	69	0	0	237
Percentage	0.8%	21.0%	17.6%	0.0%	0.0%	60.6%

Location of YFP cells detected at E10.5 from the dorsal side.

Table showing the number of YFP cells detected in different heart regions at E10.5 after tamoxifen at 6.5dpc. Analysis was based on whole-mount images of isolated hearts at E10.5 imaged from the dorsal side. Cell location was based on morphology as well as immunofluorescence for Sarcomeric α -Actinin (myocardium) and WT1 (Epicardium). AVC; Atrioventricular canal.

Table S5.

Primary Antibodies

<u>Target</u>	<u>Species</u>	<u>Dilution</u>	<u>Manufacturer</u>	<u>Cat No.</u>
NKX2-5	Goat	1:100	Santa Cruz	SC-8697
E-Cadherin	Rat	1:100	Sigma	U3254
SOX2	Goat	1:100	Santa Cruz	SC-17320
SOX17	Goat	1:100	R and D Systems	AF1924
Sarcomeric α -actinin	Rabbit	1:500	Abcam	ab68167
Titin	Rabbit	1:50	Gift from Dr E. Ehler	25376
FOXC2	Mouse	1:100	Santa Cruz	sc-515234
SMARCD3	Rabbit	1:100	Abcam	ab204745
HAND1	Mouse	1:100	Santa Cruz	SC-390376
Snail1	Rabbit	1:100	Cell Signalling	C1503
Endomucin	Rat	1:100	Santa Cruz	SC-53941
Islet1	Rabbit	1:100	Abcam	ab20670
TFAP2A	Mouse	1:100	Santa Cruz	SC-12726
TBX5	Mouse	1:100	Santa Cruz	SC-515536
GFP/YFP	Chicken	1:1000	Abcam	ab13970
Cardiac Troponin T	Mouse	1:500	Abcam	ab8295
WT1	Rabbit	1:50	Abcam	ab52933
Estrogen Receptor	Rabbit	1:100	Abcam	ab16660

Secondary Antibodies

<u>Target</u>	<u>Fluorophore</u>	<u>Origin</u>	<u>Manufacturer</u>	<u>Cat No.</u>
Goat	488	Donkey	Invitrogen	A11055
Goat	633	Donkey	Invitrogen	A21082
Mouse	555	Donkey	Invitrogen	A31570
Mouse	647	Donkey	Invitrogen	A31571
Rabbit	488	Donkey	Invitrogen	A21206
Rabbit	555	Donkey	Invitrogen	A31572
Rat	647	Goat	Invitrogen	A21247
Rat	488	Donkey	Invitrogen	A21208
Chicken	488	Goat	Invitrogen	A11039

HCR Probes

<u>Probe Name</u>	<u>Amplifier</u>
<i>Mab21l2</i>	B2
<i>Nkx2-5</i>	B1
<i>Fst</i>	B3
<i>Tbx1</i>	B3
<i>Asb2</i>	B4
<i>Tbx18</i>	B3
<i>Fsd2</i>	B1
<i>Vsnl1</i>	B3

HCR Amplifiers

<u>Amplifier</u>	<u>Label</u>
B1	Alexa 647
B2	Alexa 546
B3	Alexa 488
B4	Alexa 546

Antibody and HCR probe details

List of Primary and Secondary antibodies used for whole mount immunofluorescence including manufacturer and catalogue number. List of HCR probes including amplifier details and fluorophores used.

Movie S1.

3D volume rendering movie of a stage 1 cardiac crescent using whole mount immunofluorescence to label the cardiac precursor marker NKX2-5 (red). DAPI labelling of nuclei shown in grey. Related to Figure 1B.

Movie S2.

3D volume rendering movie of a stage 1 cardiac crescent using whole mount immunofluorescence to label the contractile protein Sarcomeric Alpha-Actinin (green) and the transcription factor SOX17 (magenta). Sarcomeric Alpha-Actinin marks cardiomyocytes and highlights the cardiac crescent. SOX17 marks the endoderm overlying the heart, and endothelial cells. DAPI labeling of nuclei is shown in grey. Related to Figure 1D.

Movie S3.

3D volume rendering movie of a stage 2 cardiac crescent using multiplexed in situ Hybridization Chain Reaction (HCR) to label *Tbx1* (green), *Asb2* (red) and *Nkx2-5* (magenta). Related to Figure 2G.

Movie S4.

3D volume rendering movie of a stage -1 cardiac crescent using multiplexed in situ Hybridization Chain Reaction (HCR) to label *Tbx1* (green), *Asb2* (red) and *Nkx2-5* (magenta). Related to Supplementary Figure 6B.

Movie S5.

3D volume rendering movie of a LHT using multiplexed in situ Hybridization Chain Reaction (HCR) to label *Tbx1* (green), *Asb2* (red) and *Nkx2-5* (magenta). Related to Supplementary Figure 6D.

Movie S6.

3D volume rendering movie of a stage 2 cardiac crescent using multiplexed in situ Hybridization Chain Reaction (HCR) to label *Mab2112* (green), *Tbx18* (red) and *Nkx2-5* (magenta). Related to Figure 3E.

Movie S7.

3D volume rendering movie of a stage 0 cardiac crescent using multiplexed in situ Hybridization Chain Reaction (HCR) to label *Fst* (green), *Mab2112* (red) and *Nkx2-5* (magenta). Related to Figure 3F.

Movie S8.

3D volume rendering movie of a stage 1 cardiac crescent using multiplexed in situ Hybridization Chain Reaction (HCR) to label *Vsn11* (green), *Mab2112* (red) and *Fsd2* (magenta). Related to Figure 4B.

Movie S9.

3D volume rendering movie of a stage 2 cardiac crescent using multiplexed in situ Hybridization Chain Reaction (HCR) to label *Vsn11* (green), *Mab2112* (red) and *Fsd2* (magenta). Related to Supplementary Figure 14B.

Movie S10.

3D volume rendering movie of a LHT using multiplexed in situ Hybridization Chain Reaction (HCR) to label *Vsn11* (green), *Mab2112* (red) and *Fsd2* (magenta). Related to Supplementary Figure 14C.

Movie S11.

3D volume rendering movie of a stage -1 cardiac crescent using whole mount immunofluorescence to label *Vsn11* (green), *Mab2112* (red) and *Nkx2-5* (magenta). Related to Figure 4D.

Movie S12.

Time-lapse video of a *Nkx2-5*-Cre; R26-nTnG embryo imaged over a 16-hour period with adaptive light-sheet microscopy showing an Me5 cell (circled) migrating towards the heart from a region rostral to the developing cardiac crescent. As cells migrated towards the cardiac crescent, the cardiac progenitor marker *Nkx2-5* was upregulated, as observed by an increase in the expression of nuclear GFP. Circles indicate the same cell tracked across the time-series shown in different channels.

Movie S13.

Time-lapse video of a *Nkx2-5*-Cre; R26-nTnG embryo imaged over a 16-hour period with adaptive light-sheet microscopy showing an Me5 cell (circled) migrating towards the heart from a region rostral to the developing cardiac crescent. As cells migrated towards the cardiac crescent, the cardiac progenitor marker *Nkx2-5* was upregulated, as observed by an increase in the expression of nuclear GFP. Circles indicate the same cell tracked across the time-series shown in different channels.

Movie S14.

3D volume rendering movie of a early headfold (EHF) embryo using whole mount immunofluorescence to label *Fst* (green), *Mab2112* (red) and *Nkx2-5* (magenta). Related to Figure 5E.

Data S1. (separate file)

Table of raw counts for all cells in the unbiased scRNA-seq dataset. Associated metadata is provided in Table S5.

Data S2. (separate file)

Table of raw counts for all cells in the reference scRNA-seq dataset. Associated metadata is provided in Table S7.

Data S3. (separate file)

Metadata corresponding to all the cells sequenced, including individual sample identifiers in the ENA database for the unbiased scRNA-seq dataset.

Data S4. (separate file)

Metadata corresponding to all the cells sequenced, including individual sample identifiers in the ENA database for the reference scRNA-seq dataset.