

Supplementary Materials for
**Single-cell mapping of gene expression landscapes and lineage
in the zebrafish embryo**

Daniel E. Wagner, Caleb Weinreb, Zach M. Collins, James A. Briggs,
Sean G. Megason,* Allon M. Klein*

*Corresponding author. Email: sean_megason@hms.harvard.edu (S.G.M.); allon_klein@hms.harvard.edu (A.M.K.)

Published 26 April 2018 on *Science* First Release
DOI: 10.1126/science.aar4362

This PDF file includes:

Materials and Methods
Figs. S1 to S17
Captions for Tables S1 to S3
References

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/science.aar4362/DC1)

Tables S1 to S3 (Excel)

Materials and Methods

Zebrafish

Both AB and TU wild-type strains were used. Embryos were generated by natural spawning and the time of fertilization was used to stage each clutch. Later stages were confirmed using morphological criteria (41). Embryos were incubated at 28.5C for all wild-type time course experiments and processed for inDrops at the indicated times. All zebrafish were housed in a facility overseen by the Harvard Medical Area Standing Committee on Animals (our IACUC) which performs regular inspections and under which we have an approved protocol for all animal procedures.

Cell Preparation

Zebrafish embryos were grown to the indicated times and chorions were removed by incubating in 1mg/mL Pronase (Sigma P5147-1G) for 3-4 min followed by washing in 0.3X Danieau Buffer. [10X Danieau Buffer = 174 mM NaCl, 2.1 mM KCl, 1.2 mM MgSO₄, 1.8 mM Ca(NO₃)₂, 15 mM HEPES, pH 7.6]. Dissociation of embryonic tissues was performed similarly as previously described (42) with the following modifications. For wild-type time course experiments, 50-100 embryos were used to generate each sample. Embryo tissues were triturated to homogeneity in 1-5mL FACSmax cell dissociation solution (Genlantis T200100) and incubated for 4-5 minutes at room temperature. Cells were then filtered through a 40µm cell strainer mesh (Fisher 352340), and centrifuged in a swinging bucket rotor at 310g for 5 minutes. Cell pellets were resuspended in 1X DPBS (no Ca/Mg, Life Technologies 14190-144) containing 1% BSA (Sigma A3311-100G), and subjected to 2-3 additional rounds of centrifugation and resuspension. After washing, cells were resuspended in 0.05% BSA / DPBS containing 18% optiprep density medium (Sigma D1556-250ML). Cell density was quantified manually using INCYTO™ C-Chip™ Disposable Hemacytometers (Fisher 22-600-100), and adjusted to ~100,000 cells per mL. For single-embryo dissociations, all FACSmax and wash volumes were reduced to a volume 0.5 mL and were carried out in 0.5mL LoBind microcentrifuge tubes (Eppendorf 022431005) that had been pre-coated with 10% BSA/DPBS for 15 minutes at room temperature.

Single-cell Microfluidic Droplet Barcoding

Single-cell transcriptomes were barcoded using inDrops (4), as previously reported (7). Following the within-droplet reverse transcription step, emulsions were split into batches of approximately 1,000-2,000 cells, frozen at -80C, and subsequently processed as individual RNA-seq libraries (see Table S1).

Preparation of RNA-Seq Libraries

Standard transcriptome RNA-seq libraries were processed as previously reported (7). For TracerSeq experiments, TracerSeq-targeted RNA-Seq libraries were also prepared. These targeted libraries were reverse-transcribed from the product of the linear amplification *in vitro* transcription (IVT) reaction (step 143 of the Zilionis et al 2017 protocol, ref 7) without prior RNA fragmentation. Non-fragmented IVT product (5 uL) was mixed with 4 uL water, 1 uL 10mM dNTPs, and 1 uL of 10uM primer TracerRT. This reaction was incubated at 70C for 3 min, then moved to ice. To this reaction was added 4 uL 5X PrimeScript Buffer, 3.5 uL water, 1 uL RNASE-OUT (Thermo-Fisher 10777-019), and 0.5 uL PrimeScript Enzyme (Clontec 2680A). Reverse transcription (RT) was performed by incubating the reaction first at 30C for 10 min, followed by 42C for 1 hour, and inactivated at 70C for 15 min. RT products were purified using 1.2X AMPureXP beads (Beckman Coulter A63881), and eluted in 15 uL of RE Buffer (10 mM Tris pH 7.5, 0.1mM EDTA). TracerSeq cDNA was then PCR-amplified as follows: 5 uL RT

product was mixed with 3 uL of water, 1 uL each of inDropsTracerF and inDropsTracerR primers (10uM), and 10 uL of 2X Phusion Master Mix (Thermo-Fisher F548L). This reaction was thermal cycled: 98C for 30 sec; 10 cycles of [98C 10 sec; 63C 20 sec; 72C 30sec]; 72C for 3 min, and then cleaned up using 1.2X AMPureXP beads and eluting in 12 uL RE buffer. Illumina sequencing adapters and sample indices were then incorporated by resuming the standard inDrops library preparation protocol (step 157 of Zilionis et al 2017). Final libraries for each multiplexed sample index were quantified using the KAPA library quantification kit (Kapa KK4844) and pooled at equimolar ratios (2mM final concentration) prior to sequencing.

Sequencing and Read Mapping

All inDrops transcriptome and TracerSeq libraries were sequenced on an Illumina NextSeq 500 using the NextSeq 75 High Output Kits according to the following sequencing specifications. V2 libraries used custom sequencing primers and 35 cycles for Read1, 51 cycles for read2, and 6 cycles for IndexRead1 and included 15-25% PhiX spike-in. V3 libraries used standard Illumina sequencing primers and 61 cycles for Read1, 14 cycles for Read2, 8 cycles each for IndexRead1 and IndexRead2. Raw sequencing data (i.e. FASTQ files) were processed using the inDrops.py bioinformatics pipeline available at github.com/indrops/indrops. Transcriptome libraries were mapped to a zebrafish reference transcriptome built from the zebrafish GRCz10 genome assembly (Assembly Accession: GCF_000002035.5). Bowtie version 1.1.1 was used with parameter –e 200; inDrops.py UMI quantification was performed with parameter –u 2 (counts were ignored from UMIs split between more than 2 genes). For TracerSeq libraries, sequencing reads were filtered and sorted by inDrops.py and then processed by custom barcode filtering pipeline (see below).

Cell Filtering and Data Normalization

inDrops data were filtered to only include UMIs originating from abundant cell barcodes. This determination was made by manually inspecting a weighted histogram of UMI counts for each cell barcode, and thresholding only the top ~95% of the largest (and often the only) mode of the distribution. Transcript UMI counts from multiplexed libraries originating from each biological sample then concatenated into a single genes x cells table and adjusted by a total-count normalization.

Identification and Filtering of Variable Genes

For each normalized UMI counts table corresponding to a single biological sample, highly variable genes were identified by first computing gene Fano factors, and ranking all genes by an above-Poisson noise statistic, as previously described (4). The top 2000 variable genes according to this statistic were then filtered to include only genes whose single-cell transcript counts were minimally correlated (correlation coefficient > 0.2) to at least one other variable gene. A set of cell cycle and housekeeping-associated genes were then excluded from downstream analyses. This set was generated by “growing” a list including any gene that was similarly expressed (single-cell correlation coefficient > 0.4) to any of the following genes. Cell Cycle: *cdk1*, *mcm2*, *mcm7*, *rrm2*, *cenpa*, *cdc6*, *ccnf*, *cdca4*, *ccnd1*, *kif4*; Housekeeping: *hmgb1b*, *hmgb3a*, *hspd1*, *hspa9*, *rplp0*, *hnrnppab1*, *rps2*, *rps12*, *rpl12*, *rps13*, *rps14*, *rps15a*, *rpl10*, *rps3a*, *rpl31*, *rpl37*, *rps6*, *rpl9*, *rpl11*, *rpl34*, *rpl13*, *rpl36a*, *rpl26*, *rps8a*, *rpl21*, *rps27.1*, *rpl27a*, *cirbp*. The resulting list was then subjected to a second “growing” round, and all associated genes were discarded.

Low Dimensional Embedding and Clustering

Normalized gene x cell counts data were projected into low dimensional subspace by first standardizing counts for each gene (by z-score) and performing principal component analysis (PCA). The number of significant PCA dimensions was then estimated by comparing the

eigenvalue distribution of cell principal components to the eigenvalue distribution of randomized data, as previously described (4). Non-significant principal components were removed from subsequent steps of the analysis. Two-dimensional t-distributed stochastic neighbor embeddings (tSNE) (43) were then generated from cell PC scores using a perplexity setting of 30, with 1000-5000 iterations. Datasets containing >10,000 cells implemented a Barnes-Hut approximation (44). Groups of related cells in the resulting tSNE maps were then identified by density clustering (45). Cell cluster annotations were assigned according to known cell type and tissue expression patterns (<https://zfin.org>) and/or specific marker genes. For timepoints 6hpf, 14hpf, 18hpf, and 24hpf, a single small cluster was initially identified whose defining genes consisted solely of housekeeping factors (see “Identification of Variable Genes”). Cells contributing to these clusters were inferred to represent dead/unhealthy cells and were removed from subsequent analyses. In some cases, examination of cell or tissue-specific marker expression on the tSNE map revealed sub-groups of transcriptionally distinct cells that were not captured by the first round of density clustering. In these cases, clusters were isolated and sub-clustered. Cases in which two adjacent clusters failed to display any differentially expressed genes were merged into a single cluster.

Identification of Differentially Expressed Genes

Cluster-defining transcripts were identified by “Model-based Analysis of Single-cell Transcriptomics” / “MAST” (46) or a Wilcoxon rank-sum test, as indicated. Cluster-specific marker genes (Table S2) were identified by comparing cells of each cluster to cells from all other clusters in the same timepoint. Genes were considered differentially expressed based on fold-change, minimum expression, and adjusted p-value cutoffs, as indicated. Tests were implemented in Matlab (rank-sum) or R / Seurat 2.2.0 (rank-sum and MAST). Unless otherwise noted, p-values were adjusted for multiple hypotheses by either a Bonferroni correction (Seurat 2.2.0), or Storey (47) correction (Matlab/mafdr).

Automated Annotation of Cell States

Cells collected from *chordin* and *tyrosinase*-targeted embryos were classified using the fitcknn and predict functions in Matlab. A kNN classifier object was first trained using the PCA-projected wild-type dataset (14hpf) and corresponding cell-state assignments using the “exhaustive” search algorithm. The following parameters were chosen automatically via the “optimize hyperparameters” subroutine: Euclidean distance, and 5 nearest neighbors. Ties were broken by selecting the class with the nearest neighbor. Cell state assignments were then predicted for *chordin* and *tyrosinase* datasets by first standardizing counts for each gene (z-score) and projecting the data into PCA space defined by the 14hpf wild-type dataset. The predict function was implemented with default settings.

Construction of Single-Cell Graphs

A nearest-neighbor graph of cells represents the manifold of cell states observed in scRNA-Seq data by a set of nodes (cells) connected to their nearest neighbors by edges. Graph representations of single cell data have been used before, for example using k-nearest-neighbor (knn) graphs (48). The knn graph construction has the property that it allows regions of the graph to have variable neighborhood sizes. For scRNA-Seq embryo data, this means that each tissue can have a different natural scale for similarity between cells. However, such simple graph constructions are not suitable for the complexity of whole embryo time series data, because of two challenges: (1) different time points can reside in different sub-spaces of gene expression; (2) the absolute number of cells sampled from different tissues can vary significantly, such that the optimal connectivity of the graph required to identify features is not uniform across tissues. In addition to these two constraints, scRNA-Seq data still presents differing neighborhood sizes. To address constraint (1), we construct a single cell graph in a

step-wise manner using a subspace defined by cells from consecutive timepoints. In this case, we find that projecting cells into a future time point is sensible because each progressive time point is of higher complexity, but sufficiently close to justify a unidirectional choice. Other situations may warrant projecting cells back in time point or defining a consensus subspace shared by two or even multiple time points. To address constraint (2), one can invoke the advantages of more than one graph construction approach in order to account both for differences in neighborhood size and in the abundance of cells sampled (49, 50). We made sequential use of four neighbor-selection approaches: initially non-mutual k-nearest neighborhoods, then locally self-tuning neighborhoods, a globally-tuning neighborhood, and mutual k-nearest neighborhoods. This sequential approach is heuristic but succeeds in resolving the major aforementioned constraints. The specific steps used are as follows: (1) each cell in time point t_i forms an outgoing edge to its 200 nearest neighbors from all cells in time points (t_i, t_{i-1}) , where all cells are projected into the non-trivial PC subspace defined by the cells in t_i alone (see **Low Dimensional Embedding** section above for definition of the subspace). This strategy forces any edges between timepoints to directly compete with edges within a timepoint. Correlation was used as the distance metric. (2) Edges are then subjected to local neighborhood restriction (a self-tuning graph construction (50)): an outgoing edge from cell i to j was kept if the distance d_{ij} was less than a local threshold $e_{ij}=3*\min_{k\neq i}(d_{ik})$, i.e. retained neighbors were at most 3-fold as far as the cell's closest neighbor. (3) To avoid very sparse neighborhoods becoming connected to distant regions of the graph, edges were then subject to a global neighborhood restriction: edges were kept if they were below the average edge distance across all cells between time points (t_i, t_{i-1}) , or if they were within 1 standard deviation of the average edge distance within the same time point t_i . (4) The graph was further reduced by retaining at most 20 mutual nearest neighbor edges.

Construction of Coarse-Grained Graphs

A coarse-graining procedure to abstract the major features of the single-cell graph was performed as follows. First, single-cell nodes belonging to the same annotated tSNE cluster ID were collapsed into a single state node. Edges between each pair of state nodes were then weighted by calculating the Jaccard index of original shared single-cell edges (i.e. the ratio of shared single-cell edges to the total number of outgoing edges for that node pair). State edges were then discarded if they received a Jaccard index weight < 0.01 . Finally, a spanning tree was traced through the weighted edges as follows. Beginning with the final timepoint, edges for all nodes were ranked according to weight. Edges then were then removed recursively, starting with the weakest edges, unless doing so would increase the total number of graph connected components. This process was then repeated for each timepoint. The resulting spanning tree connects all nodes to a single 4hpf “root” node defined by all cells of the first timepoint.

Graph Visualization

Single-cell graphs were visualized using a force-directed layout (51), implemented in Gephi (0.9.1) (52). Coarse-grained graph layouts were also rendered in Gephi, using the Yifan Hu Proportional algorithm (53) and were based only on tree edges. The tailbud single-cell subgraph was visualized using the Matlab “subspace” layout subroutine implementing the visualization algorithm described in (54).

Generation of TracerSeq Embryos

TracerSeq experiments were performed by co-injecting 2-4 nL of TracerSeq library (~20ng/uL) together with Tol2 transposase mRNA (~50 ng/uL) in water and containing a 1:20 dilution of Phenol Red solution (Sigma P0290). Embryos were screened for mosaic GFP fluorescence (see Fig. S7A) the following day, and the brightest embryos were dissociated into single-cell suspensions and processed by inDrops.

Preparation of TracerSeq inDrops Libraries

TracerSeq barcode libraries were generated by isothermal assembly (18) (ISO) of two PCR-amplified dsDNA fragments via a single-stranded “bridge” oligonucleotide containing a stretch of 20 randomized bases (Fig. S6A). The initial TracerSeq libraries were based on a pMTB vector containing a superfolder-GFP gene driven by the *actb2* promoter and flanked by Tol2 sites. The two ISO fragments were PCR-amplified from the pMTP-sfGFP plasmid using Phusion Polymerase (Thermo-Fisher F548L) as follows: 2 uL pMTB-sfGFP plasmid (~ng/uL), 18uL water, 2.5 uL each of forward and reverse primers (see Table S3), and of 2X Phusion Master Mix (Thermo-Fisher F548L). Reactions were cycled: 98C for 30 sec; 25 cycles of [98C 10 sec; 66C 20 sec; 72C 30sec]; 72C for 3 min. PCR products were then cleaned up and concentrated with 1.0X AMPureXP beads and eluted in DS buffer (10 mM Tris pH 8.0, 0.1mM EDTA). Isothermal assembly was performed by combining 0.4 pmoles of each ISO fragment (1 and 2) with 1.2 pmoles of the TracerBridge oligo and water to a final volume of 5 uL. This mixture was then combined with ISO master mix and incubated at 50C for 30 min. The desired ~6.7kb final product was gel purified and then cleaned up using 1.0X AMPureXP beads and eluted in water. Aliquots of this final library were stored at -80C. To generate sufficient amounts of library for zebrafish embryo injections without the need for PCR amplification, the entire ISO procedure was scaled to 10-12X the recipe volumes listed above. All ISO reactions included a “no-Bridge” control in which the TracerBridge oligo was omitted from the reaction (see Fig. S6B). Some TracerSeq libraries were also generated by circularizing a single 6.7kb ISO fragment (amplified using pMTBF1_Reverse and pMTBF2_Forward primers) using the same TracerBridge oligo, and using the same reaction conditions as described above (Fig. S6A). In these cases, gel purification was based on differential migration of the open circular vs. linear forms of similarly sized DNA molecules. Homemade ISO master mix was prepared as previously described (18). Diversity of TracerSeq barcode libraries was initially assessed by Sanger sequencing. Briefly, the TracerSeq barcode junction was PCR-amplified from a fully assembled library, ligated into the pCR4-Blunt-TOPO cloning vector, and transformed into One-Shot Top10 competent cells. Sanger sequences from each of 19 individual bacterial clones are shown in Fig. S6C. TracerSeq library diversity was subsequently and more thoroughly assessed using RNA-seq data (Fig. S7B-D).

Processing of TracerSeq Sequencing Reads

TracerSeq sequencing data were processed by inDrops.py to perform read filtering and inDrops cell barcode correction. Sorted FASTQ files for TracerSeq cDNA reads (with inDrops cell barcodes and UMI sequences in the header) were then processed as follows. First, sequences flanking both sides of the N20 barcode were identified and trimmed; reads that did not contain at least 12bp of flanking sequence were discarded. Second, all reads corresponding to unique UMI-cell barcode pairs were combined and used to generate a multiple-sequence alignment consensus for each uniquely detected TracerSeq barcode in each cell. To determine which TracerSeq barcodes were derived from the same clonal insertion event, pairwise sequence comparisons were performed between all unique transcript barcodes detected across all single-cell libraries associated with a single TracerSeq embryo. Barcode diversity for a typical experiment is illustrated in Fig. S7B, in which 1,000 unique transcript barcodes were selected at random from the TracerSeq Fish1 embryo and subject to pairwise edit distance comparisons. The “edit” distance between two sequences was based on the Levenshtein distance, except that each single-base substitution was weighted as 2, while each indel was weighted as 1. The resulting distance matrix was clustered and plotted as a heatmap, revealing subsets of barcodes with identical or near-identical sequences. This signature was not detected when performing comparisons amongst random 20mers. A histogram plot of all pairwise edit distances (excluding self-pairs) confirmed a bimodal distribution in which the majority of pairwise

distances were consistent with those of the random distribution (Fig. S7C). The smaller peak in the distribution, centered at an edit distance of zero, was only observed when comparing barcodes within the same embryo (Fig. S7D). Based on these observations, TracerSeq barcodes were grouped into clones first by identifying and collapsing any barcodes with exact sequence matches. Remaining unique barcodes were then combined into clones if they were within an edit distance of “6” (e.g. up to 3 base substitutions) to a previously identified clone barcode. A UMI counts table of clones vs. cells was then generated for each TracerSeq embryo and subjected to downstream analysis. In order to directly compare UMI counts for both clones and transcripts for each individual cell, error-corrected inDrops cell barcodes (identified by inDrops.py) were retained for all cells. Analyses appearing in Figs. 4-5 and Figs. S9-11 were restricted to cell barcodes for which both TracerSeq and transcriptome data were recovered.

Calculation of TracerSeq Lineage Coupling Scores

Cells from all 5 TracerSeq embryos that were associated with both the clone and transcript datasets were used. First, the total number of shared TracerSeq clones was tabulated for each pair of 24hpf states. A TracerSeq clone was defined as “shared” when it contained at least 2 individual cells assigned to each state. Only clones with at least 5 total cells were considered. The number of shared clones was then compared to randomized data in which cell state assignments were permuted. A total of 20,000 random permutations were performed and used to calculate a z-score for each original “shared” clone count with respect to the random distribution. Lineage coupling z-scores were clustered and plotted as a hierarchically clustered heatmap in Fig. S9E. Positive z-scores indicate pairs of 24hpf states that shared significantly more TracerSeq clone barcode hits than expected by chance; a negative score indicates state pairs that were significantly less coupled than expected by chance. We then computed a correlation coefficient between z-scores for each pair of states. Lineage coupling correlations were plotted as a clustered heatmap in Fig 5A.

CRISPR

CRISPR experiments were performed by co-injecting 1-cell stage zebrafish embryos with *in vitro*-transcribed purified sgRNAs together with purified recombinant Cas9 protein. 2-4 nL injections were performed with sgRNAs at a combined concentration of ~100ng/uL (~3uM) and with Cas9 protein at ~7uM final concentration diluted in water and containing a 1:20 dilution of Phenol Red solution (Sigma P0290). Embryos that were damaged during the injection process (typically < 5%) were discarded. Embryos were incubated at 28C for 4-6 hours before moving to 23.5C so that they would develop to approximately the 10-14 somite stage (normally corresponding to ~14hpf) by the following morning. Endpoint stages for each clutch of embryos were confirmed as described above. For each CRISPR experiment, 20-30 individual *chordin* or *tyrosinase*-targeted embryos were dissociated into single cell suspensions and barcoded sequentially by inDrops as approximate stage-matched sample pairs. A total of 3 such pairs were sequenced and analyzed. The *tyrosinase* gene, when disrupted, generates a mild pigmentation defect that can serve as both a positive control for Cas9 activity and a negative control for injection-induced toxicity that can accompany CRISPR experiments (55). Embryos with disrupted *tyrosinase* develop otherwise normally and can be effectively treated as wild-type for the purposes of this experiment. *chordin* sgRNAs were designed using ChopChop (<http://chopchop.cbu.uib.no/>) to target exons 1-2 of the *chordin* locus. *tyrosinase* was targeted using a previously reported sgRNA sequence (55). sgRNAs were synthesized *in vitro* as previously described (34) using primer sequences listed in Table S3. Recombinant His-tagged Cas9 protein was expressed in *e.coli* and purified as previously reported (34).

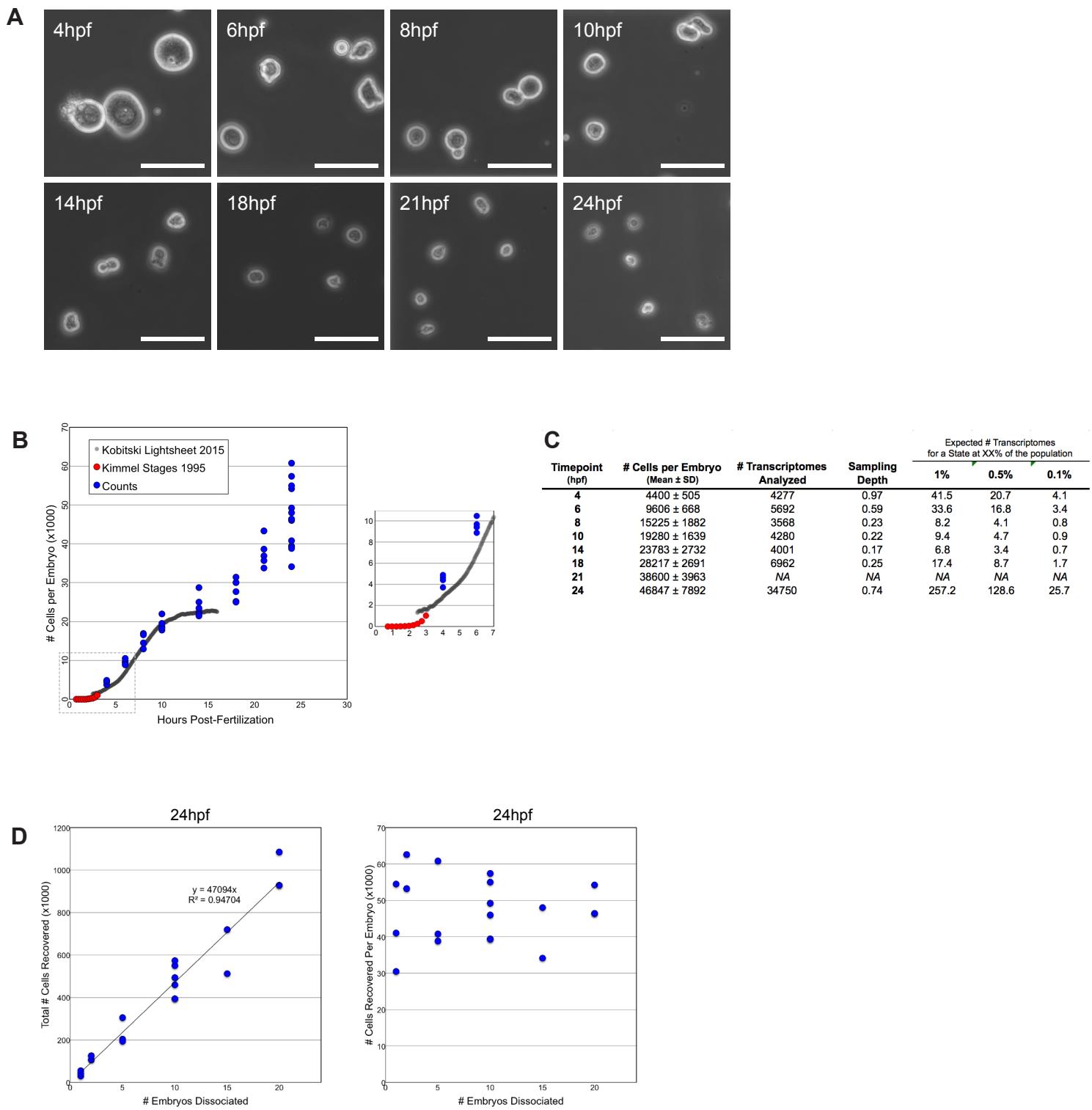
Projecting CRISPR inDrops datasets onto the Single-Cell Graph

A total of 6 inDrops datasets consisting of 3 biological replicates for both *chordin* and *tyrosinase* CRISPR-targeted samples were analyzed. To compare cells of CRISPR-targeted samples to those of the wild-type timecourse, each CRISPR dataset was indexed to include z-scored variable genes and projected into PCA subspaces that were defined by each of the following wild-type timepoints: 10hpf, 14hpf, and 18hpf. These timepoints were chosen to limit the analysis to wild-type states that were collected within a similar time frame as the CRISPR samples. A set of k-nearest neighbor edges from each CRISPR cell to one of the wild-type cells were then identified. The nearest neighbor number k was scaled to correct for different numbers of cells in each sample such that the total number of edges sought for each timepoint projection was set to 100 * (the number of wild-type cells in the target dataset). This process resulted in a vector of 6 projecting edge counts for each wild-type cell, 3 from each of the two CRISPR target genes. To identify cell state neighborhoods that were significantly over- or under-represented in *chordin*-targeted embryos (relative to *tyrosinase*-targeted embryos), a t-test was performed between the 3 embryo edge counts from each CRISPR target, for each cell node. Cell nodes participating in significantly different numbers of projecting edges were identified (FDR < 0.25), and their associated log2 mean edge count ratios were color-coded on the single-cell graphs in Fig. 6D and Fig. S16D.

Pseudo-Spatiotemporal Ordering and Identification of Dynamically Varying Genes

To assess the continuum of cell states comprised within the tailbud region of the single-cell graph, a subgraph was first extracted. The subgraph consisted of cells from the 8hpf-18hpf timepoints and included all cells assigned to posterior neural (hindbrain or spinal cord), tailbud, or pre-somitic mesoderm tSNE clusters. A pseudo-spatiotemporal ordering of cells along the resulting continuum was determined in a variation of Wanderlust (10), as follows. First, approximately 200 cells at each “end point” of the continuum were manually selected. Next, a series of shortest paths were calculated between these two sets of cells through a version of the subgraph in which 50% of all edges were randomly deleted. This process was repeated for a total of 100 iterations in which different sets of edges were randomly deleted. All cells/nodes discovered during this process were ordered based on their average position over all shortest paths in which they appeared, resulting in a continuous ordering of cells. A “start point” zone was then inferred based on expression of the *ta* / *brachyury* transcript. Genes that varied dynamically along this trajectory were then identified similarly to as previously described (5). Sliding windows of 100 cells were first scanned to identify two windows with maximum and minimum average expression levels for all genes, respectively. A t-test was then performed between these two sets of 100 expression measurements (FDR < 0.05). Gaussian-smoothened expression z-scores for significantly variable genes along the trajectory were then calculated. A subset of the significant genes identified are shown in Fig. S16B.

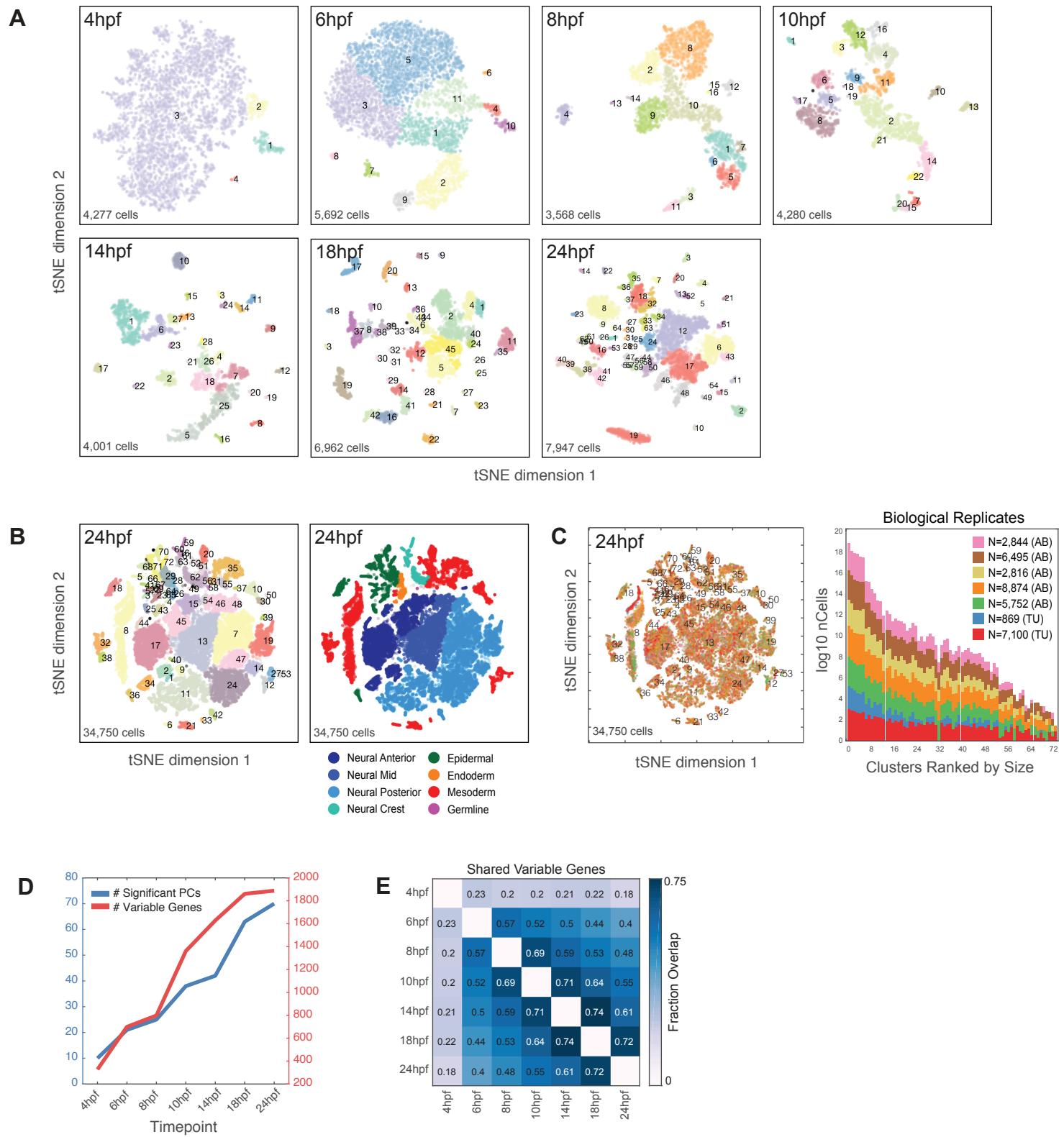
Figure S1



Supplemental Figure 1

- (A) Representative phase contrast images of cells dissociated from zebrafish embryos at the indicated timepoints. Scale bars: 50 μ m.
- (B) Scatterplot of estimated total numbers of cells per embryo vs. time. Each blue dot represents a separate biological sample (cells dissociated from a set of 5-10 embryos, see Methods) manually quantified with a hemocytometer. Red and grey dots denote previously reported estimates from (41) and (56), respectively.
- (C) Estimation of scRNA-seq cell sampling depth. Table lists total cells per embryo inferred from cell counts data in (B) alongside the total number of wild-type cells profiled by scRNA-seq, and associated sampling depths.
- (D) Demonstration of quantitative cell recovery. Estimated total cell yields resulting from the dissociation of sets of 1-20 embryos harvested at 24hpf.

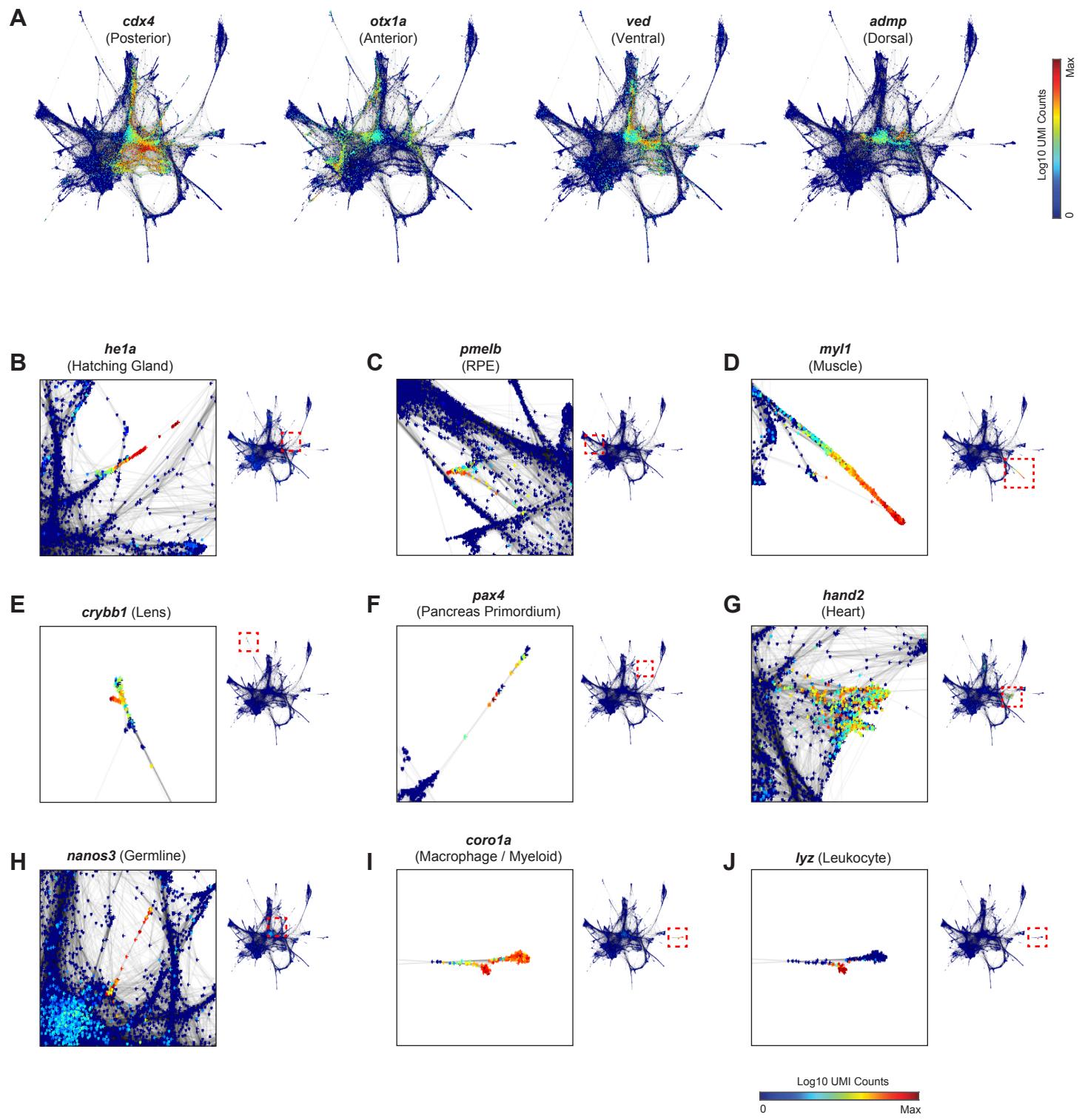
Figure S2



Supplemental Figure 2

- (A) tSNE embeddings of cells from 7 developmental timepoints, colored by local density cluster IDs.
- (B) tSNE embedding for the 24hpf timepoint, including additional biological replicates. Left: cells colored by local density cluster ID. Right: cells colored by inferred germ layer identity.
- (C) Left: tSNE embedding of cells for the 24hpf timepoint, colored by biological sample of origin. Right: stacked heatmap depicting the contribution of each biological replicate to each cluster. Clusters are ranked in order of decreasing cell number.
- (D) Numbers of significant principal component dimensions and highly variable genes identified for each timepoint (see Methods).
- (E) Fraction of overlap in sets of identified variable genes between all timepoint pairs.

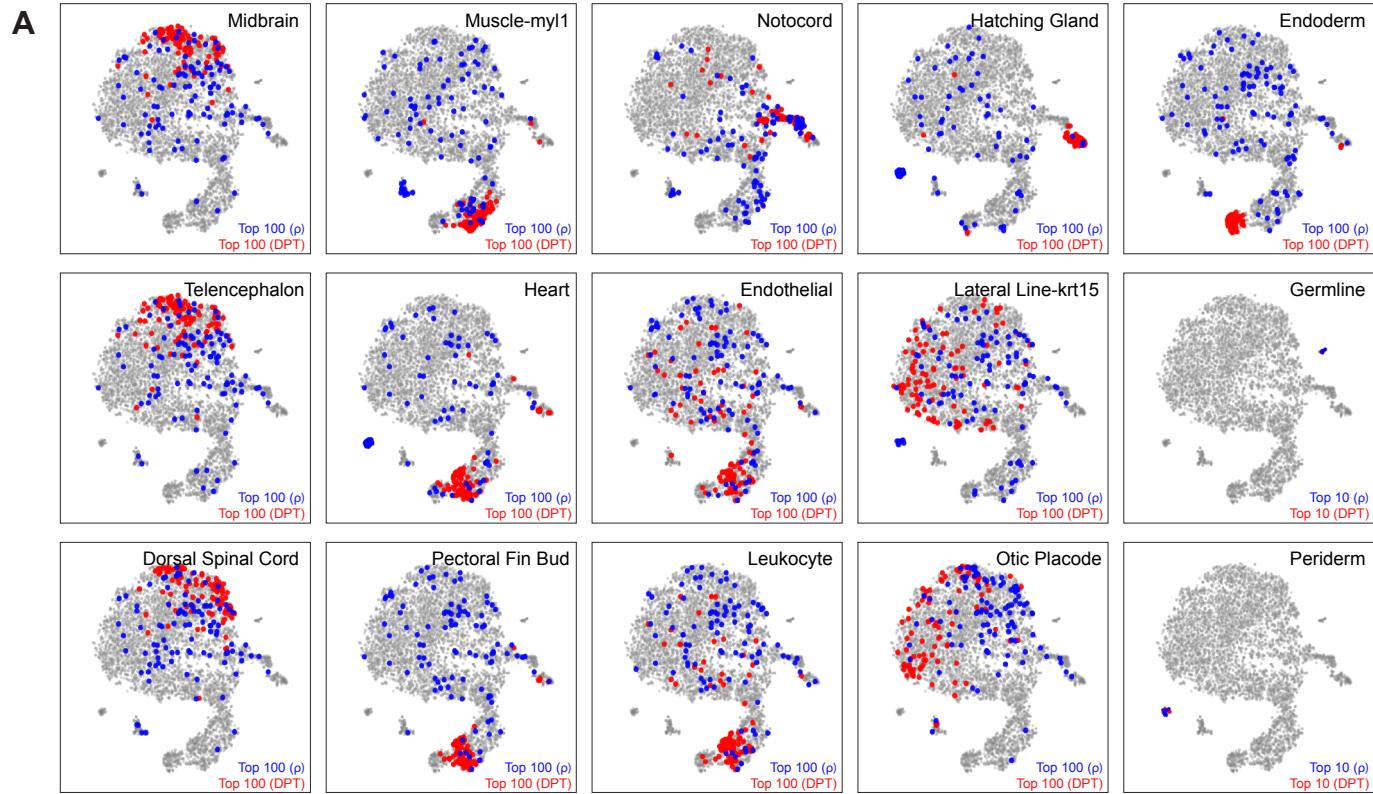
Figure S3



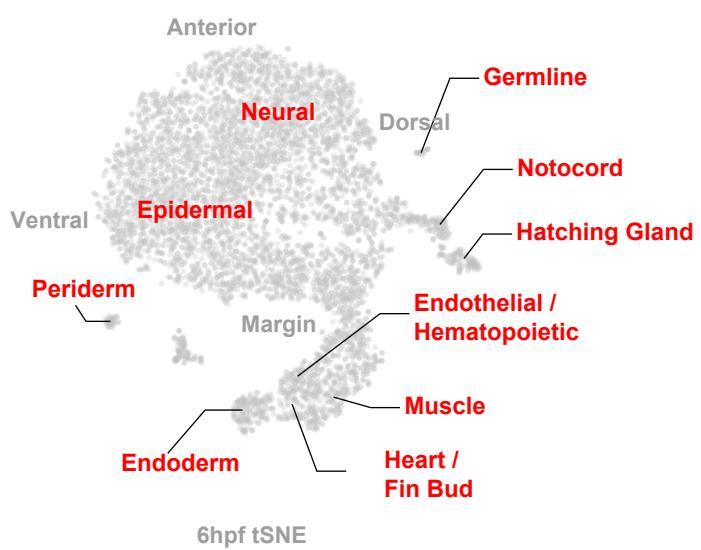
Supplemental Figure 3

Single-cell graphs (see also Fig 2). Nodes are colored by log10 expression counts for positional marker genes (A), and marker genes for particular cell/tissue types (B-J). Insets show zoomed regions marked by red boxes.

Figure S4



B



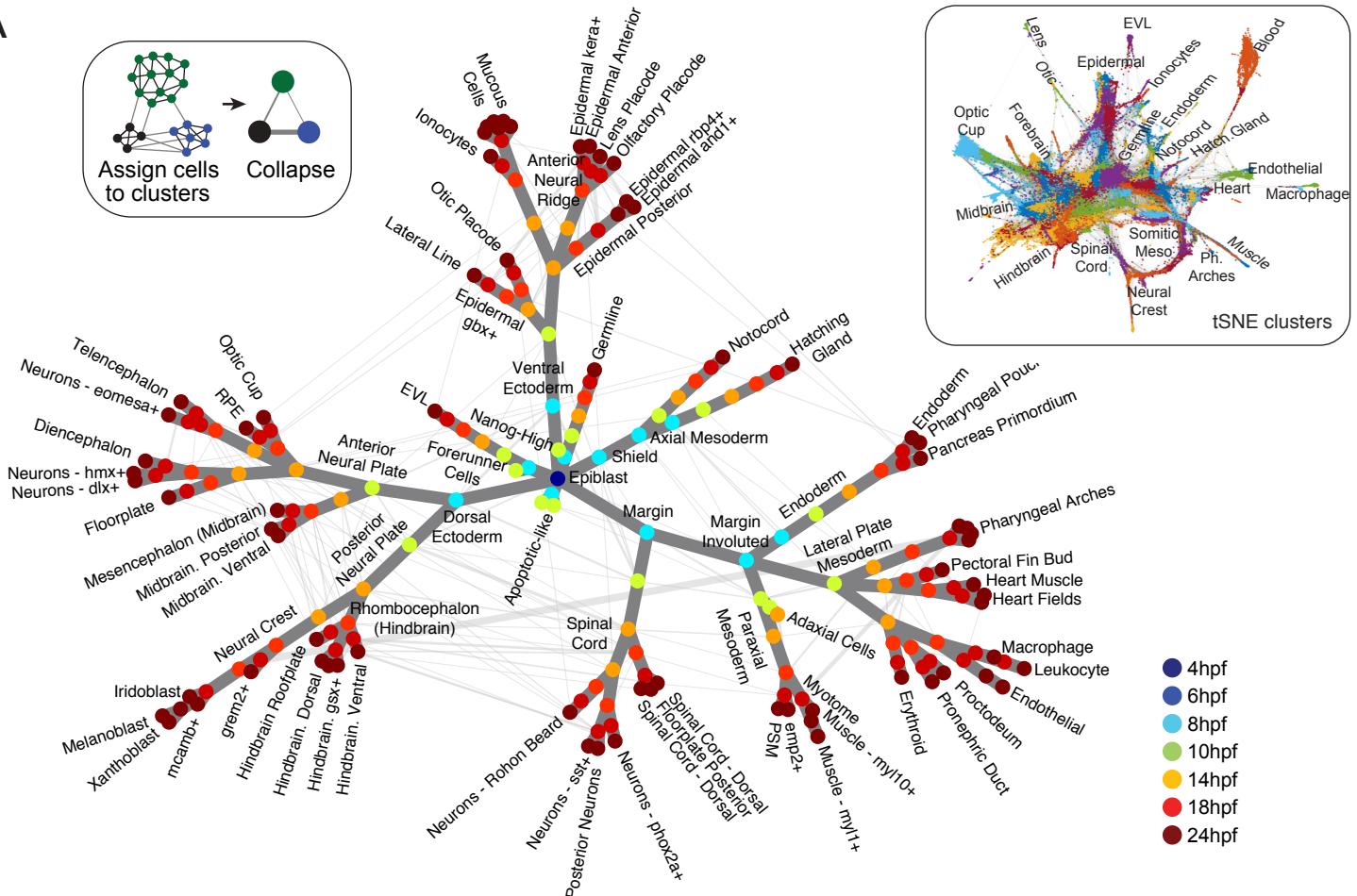
Supplemental Figure 4

(A) Additional tSNE maps of the 6hpf timepoint (see also Fig 3B), with the top 100 most proximal cells to specified 24hpf states highlighted in red (proximity calculated by single-cell graph diffusion distance) or blue (proximity calculated by correlation distance between gene expression profiles for each 6hpf cell and each 24hpf state). All regions of the original tSNE map are shown, including non-epiblast/non-hypoblast clusters. ρ , Pearson correlation. DPT, Diffusion Pseudotime.

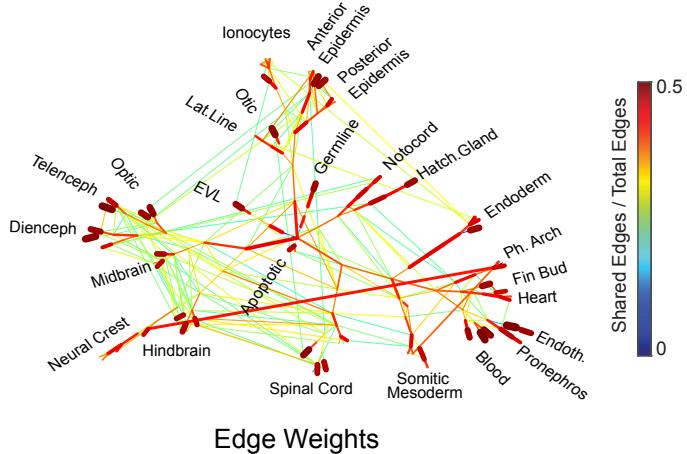
(B) Overview of the graph-predicted fate outcomes for each region of the 6hpf tSNE map.

Figure S5

A



B

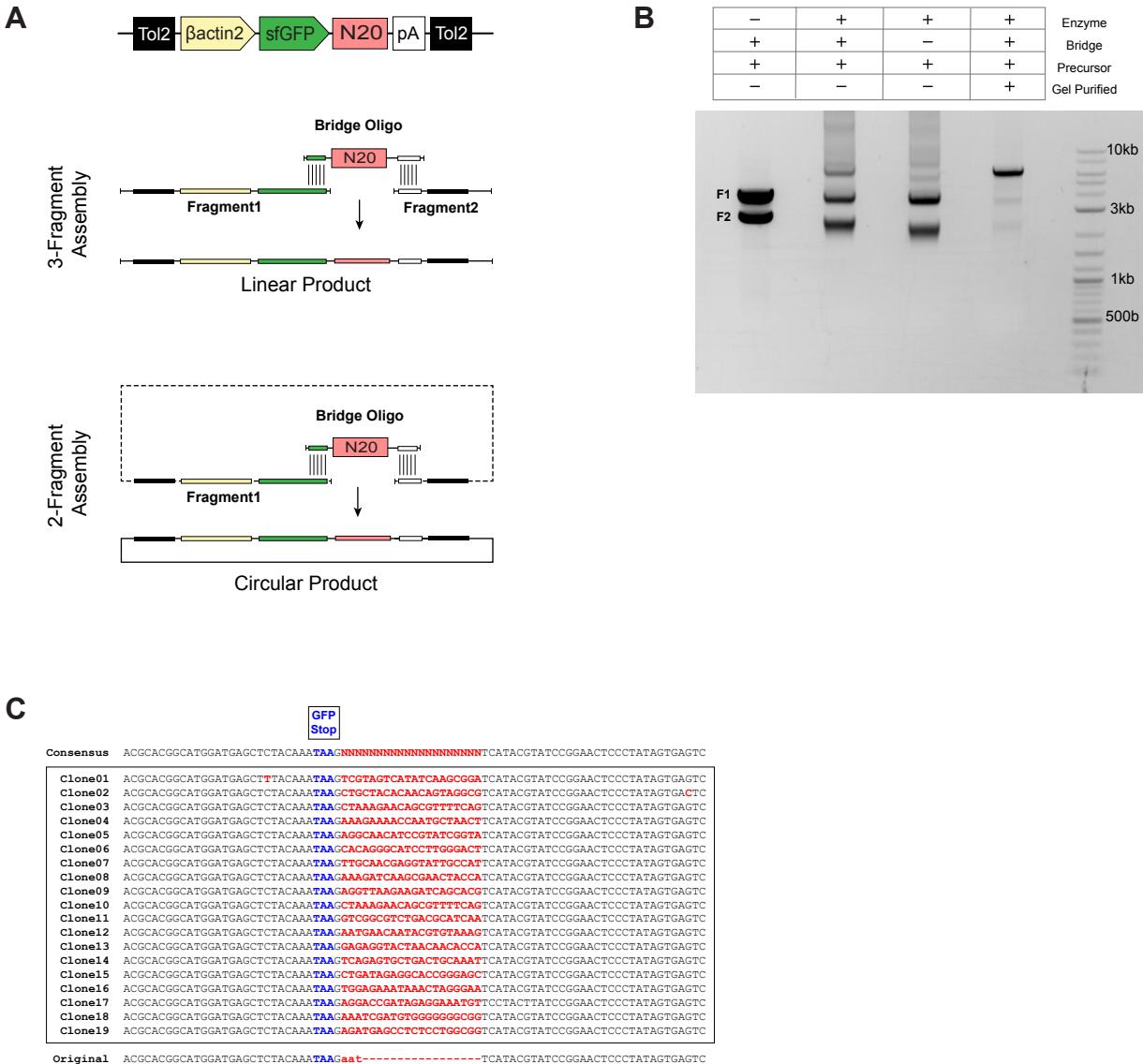


Supplemental Figure 5

(A) Detailed overview of the coarse-grained graph depicted in Fig 3C. Inset: single-cell graph (same layout as Fig 2A) with nodes colored by local density tSNE clusters that were used for collapsing nodes.

(B) Edge weights in the coarse-grained graph highlighted by color and edge thickness. Nodes are hidden from view.

Figure S6



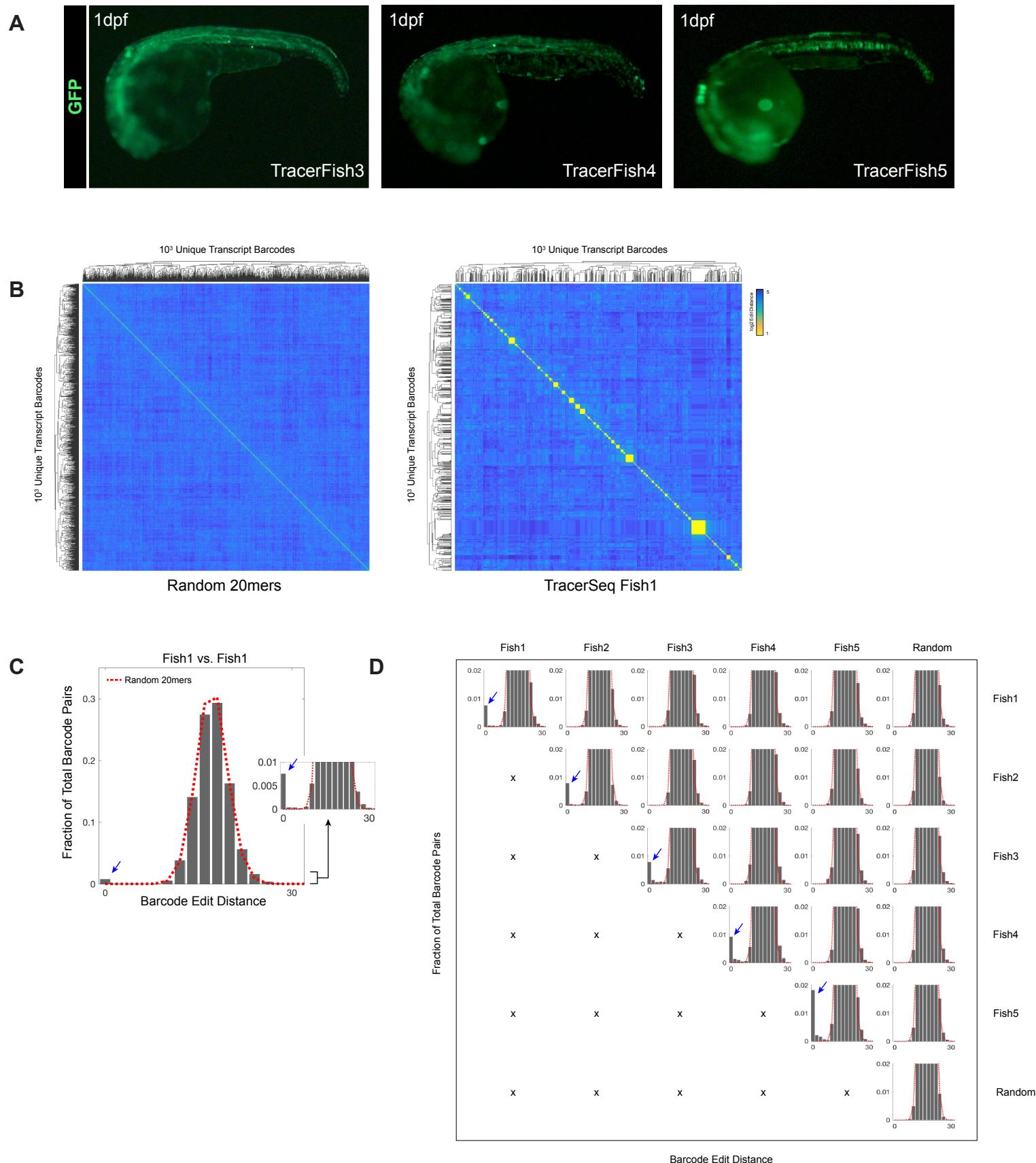
Supplemental Figure 6

(A) Overview of isothermal / Gibson assembly strategies for constructing TracerSeq libraries. In a 3-fragment assembly, two dsDNA fragments are joined via a single-stranded “bridge oligo” that contains a random 20mer, flanked by two 20bp homology arms to fragments 1 and 2. In a 2-fragment assembly, a single fragment is circularized via assembly at an identical junction.

(B) 1% TAE gel illustrating 2-fragment isothermal assembly of the ~6.7kb final product from the 4kb and ~2.7kb fragments.

(C) Sanger sequencing of the fully assembled TracerSeq junction. Sequencing reads from individual bacterial clones and their consensus sequence are indicated (note, the library is not passed through bacteria prior to use in fish). The N20 barcode region and GFP stop codon are marked in red and blue, respectively. Bottom, comparison to the original pMTB vector sequence.

Figure S7

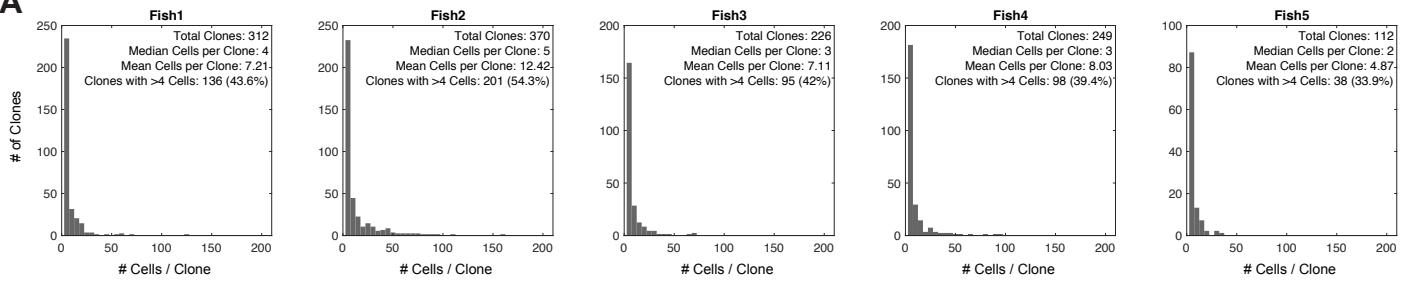


Supplemental Figure 7

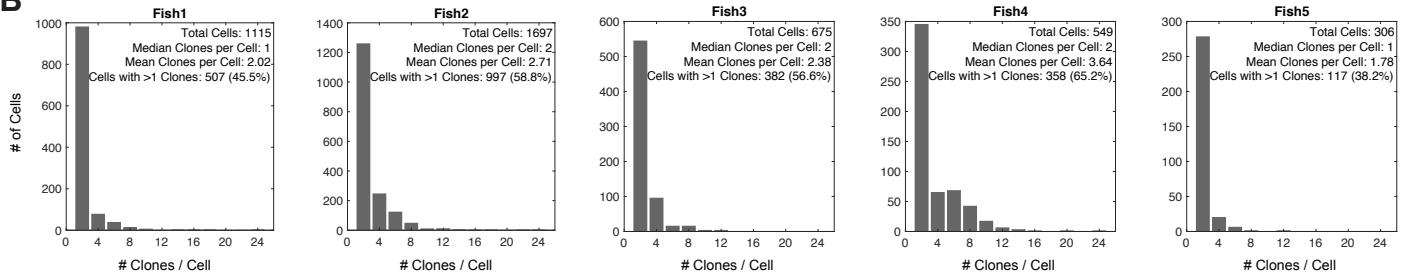
- (A) Live fluorescent microscopy of representative TracerSeq embryos illustrating mosaic GFP expression. Anterior, left.
- (B) Heatmap of pairwise edit distances between 1,000 random 20mer sequences (left) and 1,000 randomly selected unique TracerSeq transcript barcodes from Fish1 (right). Heatmap rows and columns were hierarchically clustered using correlation distance and average linkage. Clusters of sequences with small or zero edit distances (interpreted to derive from the same clonal insertion event) are evident for Fish1, but not for the random 20mers.
- (C) Histograms of all barcode edit distances plotted in (B), excluding self-barcode pairs. Grey bars denote pairwise distances between barcodes from Fish1, compared to distances between random 20mers (dotted red line). For Fish1, two histogram peaks are evident, with the first resembling the random distribution, and the second centered at an edit distance of zero (blue arrow).
- (D) Histograms comparing pairwise barcode edit distances within and between all 5 TracerSeq fish embryos vs. the random distribution. Peaks corresponding to identical or near identical barcodes (blue arrows) were only evident when comparing barcodes within a fish.

Figure S8

A



B



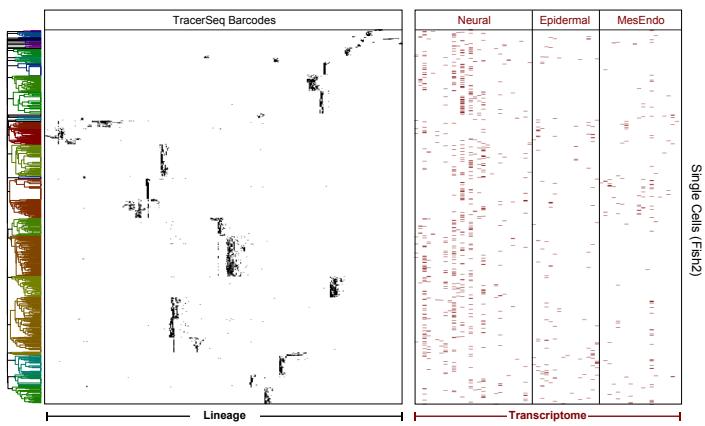
Supplemental Figure 8

(A) Histograms depicting the distribution in clone sizes (numbers of cells) over all 5 TracerSeq embryos.

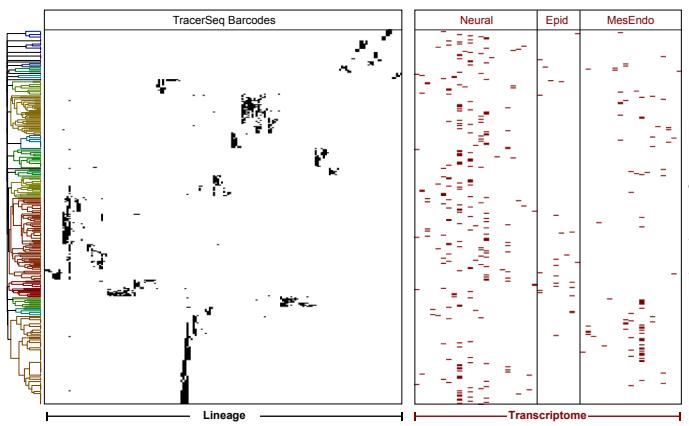
(B) Histograms depicting the distribution in numbers of clones detected per cell over all 5 TracerSeq embryos.

Figure S9

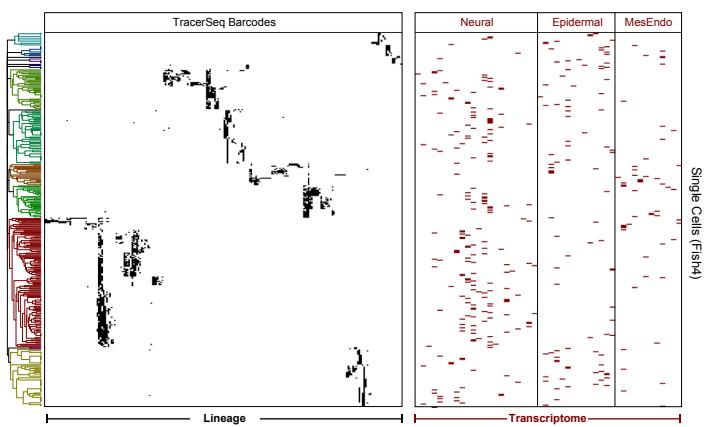
A



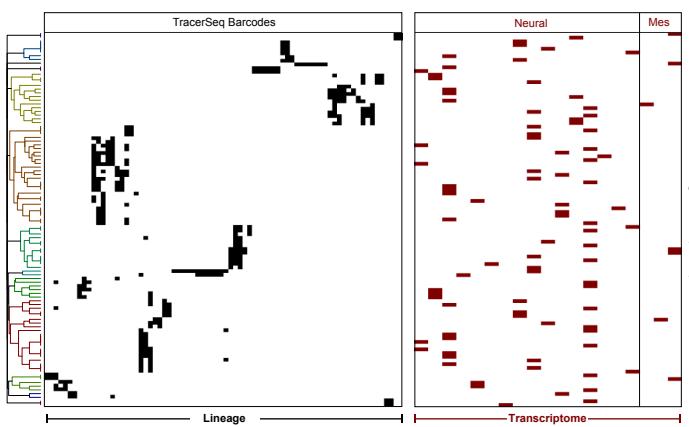
B



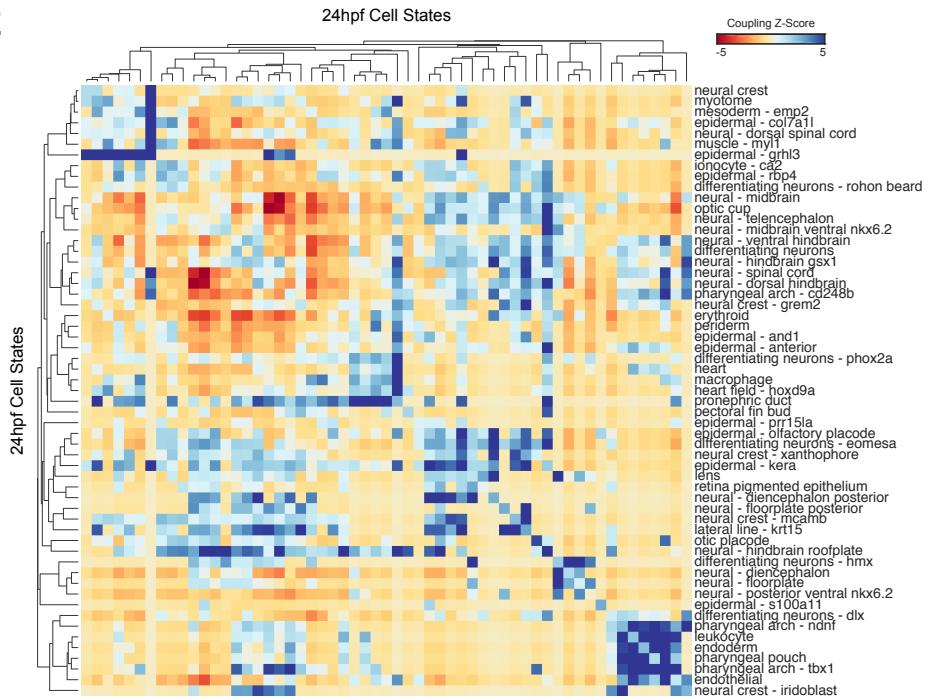
C



D



E

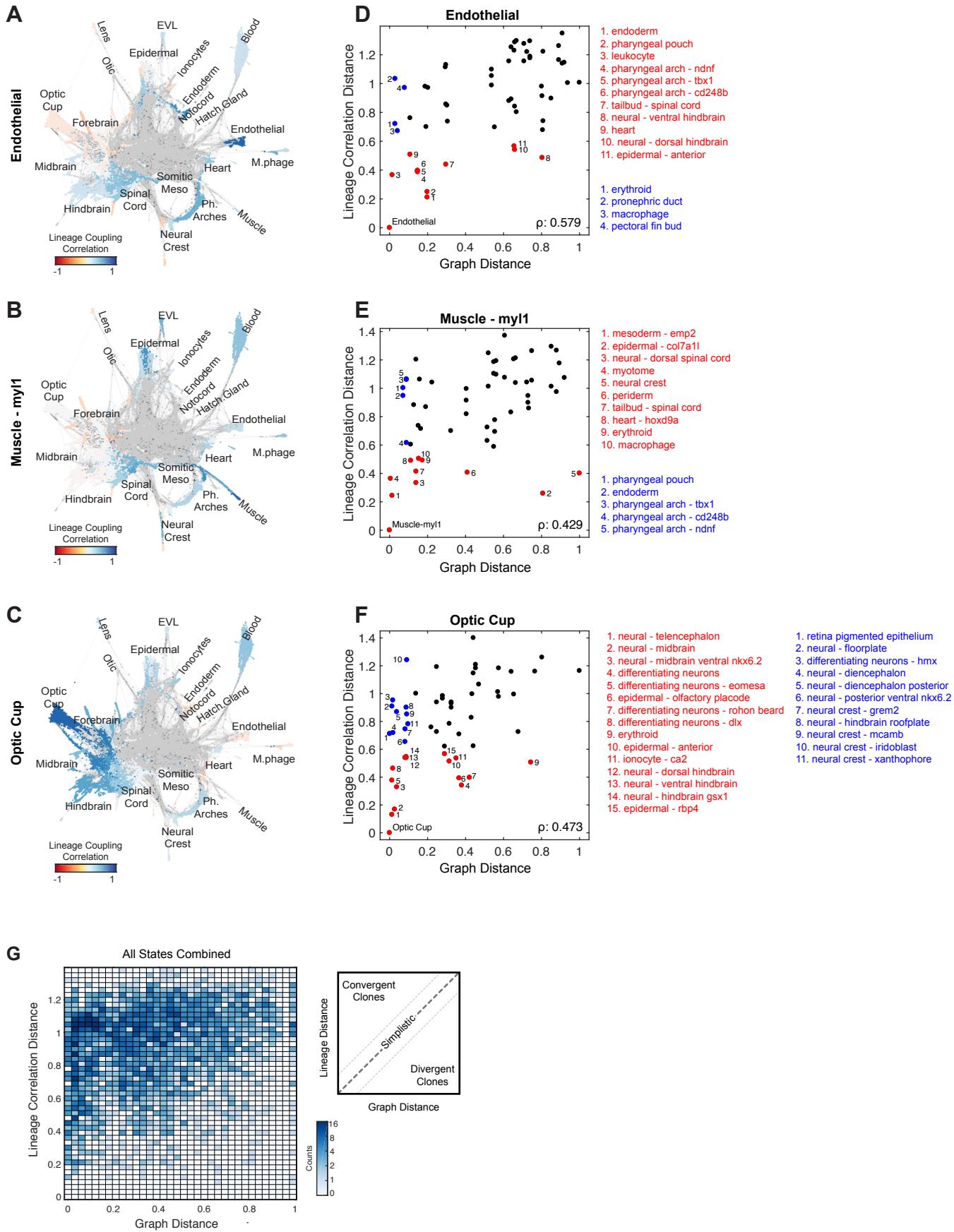


Supplemental Figure 9

(A-D) See also Fig 4B. Shown are additional heatmaps depicting dual lineage/transcriptome information for each cell in TracerSeq embryos 2-5. Heatmaps are clustered using Jaccard similarity and average linkage.

(E) Heatmap of TracerSeq lineage coupling z-scores between each pair of 24hpf states. Z-scores are hierarchically clustered by correlation distance and average linkage.

Figure S10



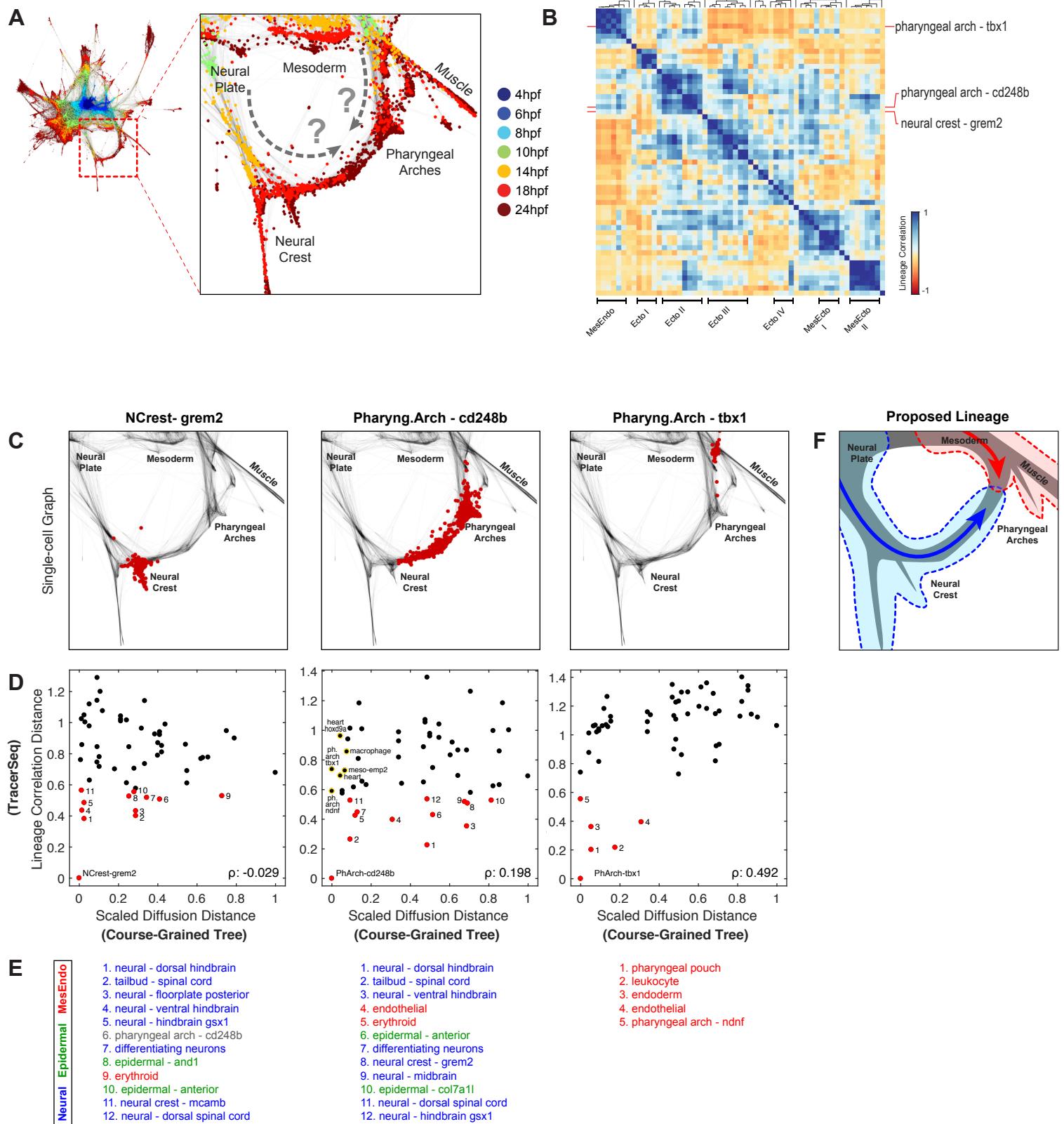
Supplemental Figure 10

(A-C) Lineage coupling correlation scores mapped onto the single-cell graph. Color values correspond to a single row of the matrix in Fig. 5A

(D-F) Scatterplots comparing lineage coupling correlation distances vs. scaled diffusion distances predicted from the state tree. Plots match those appearing in Fig. 5B, with additional points labeled. Red labels indicate states with significant non-zero TracerSeq lineage correlations (adj. p-value < 0.005), ranked by p-value. Blue labels indicate states in the upper left quadrant of the plot with short normalized graph diffusion distances (<0.1) and weak lineage correlations (correlation distance > 0.6), ranked by increasing diffusion distance.

(G) Left, two-dimensional histogram of lineage correlation distance vs. diffusion distance scores over all state pairs. Number of scatter points appearing in each 2D bin is indicated by a heatmap. Right, proposed interpretation of various state-lineage relationships. A simplistic model, assuming tree-like relationships, predicts that lineage and state distances will be highly correlated. Cell state pairs that are related by lineage but separated by large distances in state space suggest the presence of “Divergent Clones”. Cell state pairs that are highly similar in state but unrelated by lineage suggest the presence of “Convergent Clones”. ρ , Pearson correlation.

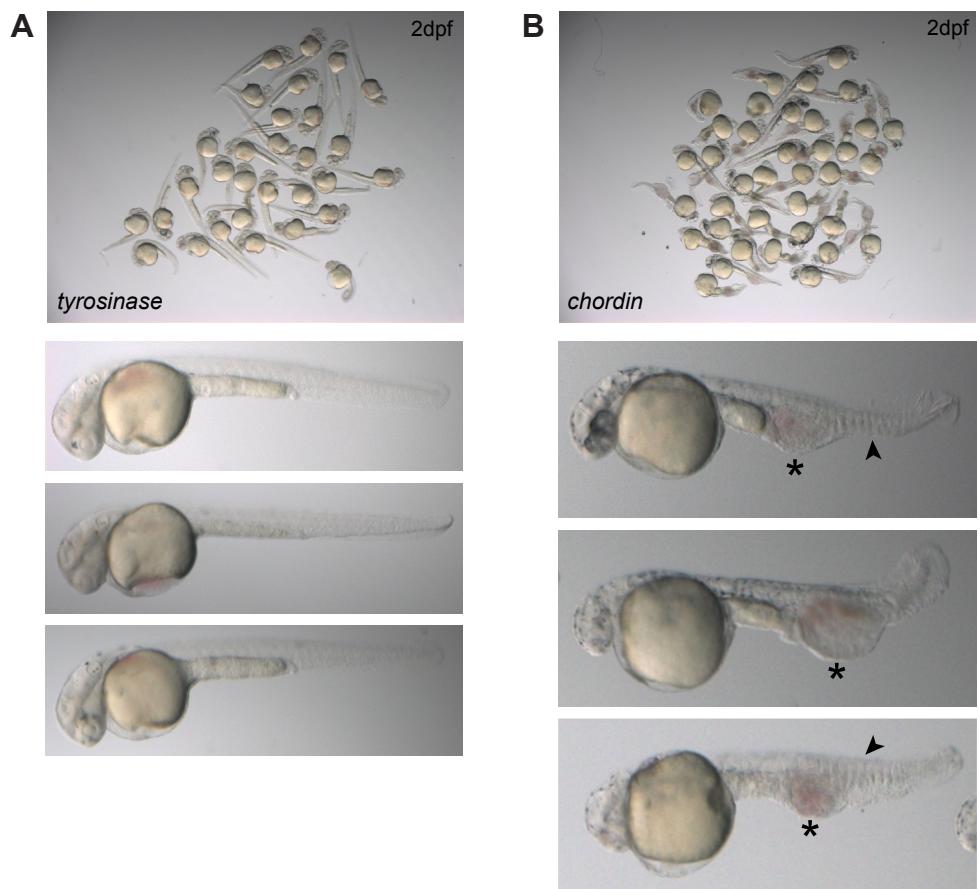
Figure S11



Supplemental Figure 11

- (A) Single-cell graph with close-up view of the pharyngeal arch / neural crest loop. Nodes are colored by collection timepoint. Arrows indicate two distinct hypothetical state trajectories into the loop with either neural plate or mesodermal origins.
- (B) Clustered heatmap of TracerSeq lineage correlation scores (from Fig. 5A). State names are hidden except those for three select states participating in the pharyngeal arch / neural crest loop. Together these states span two independent TracerSeq groups (MesEndo and Ectoll).
- (C) Close-up views of the single-cell graph. Graph edges are shown in dark grey. Red dots denote locations of cell nodes assigned to each state.
- (D) Scatterplots comparing lineage coupling correlation distances to scaled diffusion distances predicted from the state tree. Each plot is anchored on the states depicted in (B-C). Lineage correlation scores (y-axis) correspond to rows of the heatmap in (B). Middle panel: Additional relationships are highlighted in yellow between pharyngeal arch-cd248b and selected lateral plate mesodermal states. These states (plotted in the upper left quadrant) display weak lineage correlations to pharyngeal arch-cd248b, despite being in close transcriptional proximity on the state tree. ρ , Pearson correlation.
- (E) States with significant TracerSeq lineage correlations (adj. p-value < 0.005), ranked by p-value. Numbers correspond to labels in (D). State names are colored by germ layer of origin.
- (F) Proposed lineage model: A majority of cells in the pharyngeal arch / neural crest loop are neural-derived.

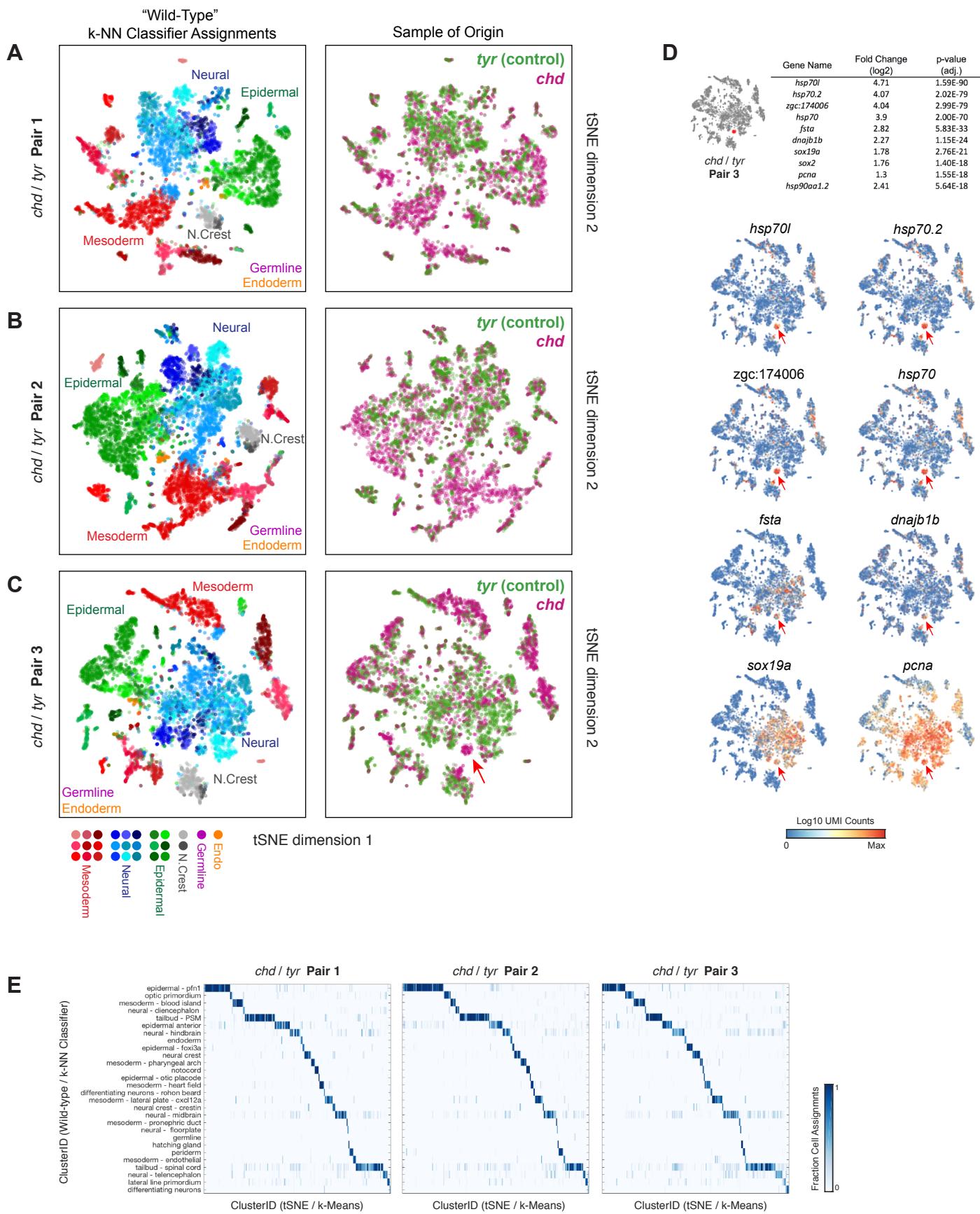
Figure S12



Supplemental Figure 12

Brightfield microscopy of live 2dpf embryos co-injected with CRISPR/Cas9 at the 1-cell stage with sgRNAs targeting *tyrosinase* (left) or *chordin* (right) genes. Following injection, damaged embryos were immediately identified and removed; all other embryos were retained. Embryos were chosen at random for inDrops sample preparation at stages corresponding to ~14-16hpf; remaining embryos were used to assess effectiveness of the CRISPR targeting. Top images: overview of the entire clutch including all embryos. Bottom images, representative examples of 3 different embryos from each condition. *tyrosinase*-targeted embryos displayed substantially lower levels of visible pigment, but are otherwise normal. *chordin*-targeted embryos displayed the classic “ventralization” phenotype with small heads, enlarged somites (arrowheads), and overabundance of ventral mesodermal tissue (asterisks).

Figure S13



Supplemental Figure 13

(A-C) Single-cell tSNE maps for three stage-matched pairs of *chordin* and *tyrosinase* CRISPR-targeted samples (~14-16hpf). Left: cells colored by 14hpf (wildtype) state identities assigned by a kNN-classifier (see Methods). Close clustering of cells with similar color codes demonstrates the comparability of the *chordin* cell state landscape to the original wildtype landscape. Right: cells colored by genotype. Data points were down-sampled to an equal number of *chordin* and *tyrosinase* cells in each plot. Local changes in the proportions of the two genotypes demonstrate shifts in state abundances in the *chordin* and *tyrosinase* samples. Red arrow in (C) indicates a cluster in which no *tyrosinase* cells appeared.

(D) Differential gene expression analysis of the “*chordin*-only” cell cluster identified in (C). Listed are the top 10 positively enriched genes for this cell cluster, determined by Wilcoxon rank-sum test (\log_2 fold change >1) ranked by adjusted p-value. Analysis was limited to transcripts with average expression level >25 transcripts per million (TPM). Counts overlays for selected differentially expressed genes confirm enrichment of transcripts associated with heat-shock/stress response and a neural progenitor / dividing cell state (*sox19a*, *pcna*).

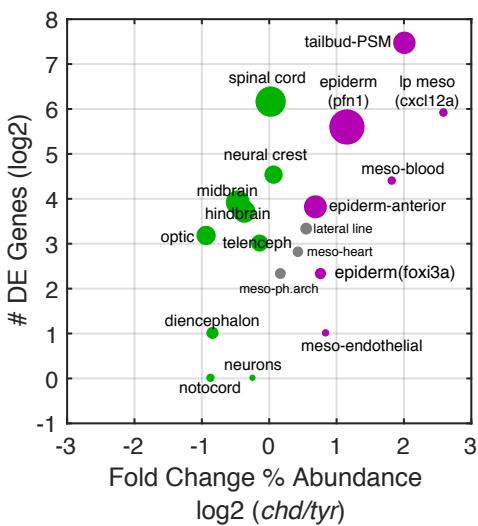
(E) Structural similarities between *chordin/tyrosinase* and wild-type datasets, assessed by a confusion matrix. Rows: wild-type kNN-classifier assignments (n=28). Columns: k-Means cell cluster assignments derived from the tSNE map (k = 200, cityblock distance). Data are column-normalized and columns are sorted by row maxima. Cells within each k-Means cluster predominantly map to a single wild-type state.

Figure S14

A

Cell State Name	Relative Abundance (<i>chd/tyr</i>)		# DE Genes		Top10 Novel Upregulated Genes (<i>chd/tyr</i>)	
	Log2 FC	Adj. P-Value	State-Defining	<i>chd/tyr</i>	<i>chd/tyr</i> (Novel)	
mesoderm - lateral plate - cxcl12a	2.6	0.0052	569	60	22 (36.7%)	<i>apoeb</i> , <i>ppp1r15a</i> , <i>id1</i> , <i>ddx21</i> , <i>mgst1.2</i> , <i>ta</i> , <i>gata3</i> , <i>histh1l</i> , <i>si:dkey-103d23.5</i>
tailbud - PSM	2.02	0.0001	712	176	80 (45.5%)	<i>apoeb</i> , <i>zgc:136826</i> , <i>zbtb16a</i> , <i>txnipa</i> , <i>ppp1r15a</i> , <i>id3</i> , <i>ier2</i> , <i>wu:fb55g09</i> , <i>id1</i> , <i>sesn3</i>
mesoderm - blood island	1.83	0.0187	1134	21	15 (71.4%)	<i>apoc1l</i> , <i>txnipa</i> , <i>apoeb</i> , <i>hsp70</i> , <i>ppp1r15a</i> , <i>h1f0</i> , <i>her6</i> , <i>fam46ba</i> , <i>tp53inp1</i> , <i>si:ch211-208m1.2</i>
epidermal - pfm1	1.16	0.0002	985	48	32 (66.7%)	<i>txnipa</i> , <i>h1f0</i> , <i>wu:fb55g09</i> , <i>fam46ba</i> , <i>hsp70</i> , <i>hsp70.2</i> , <i>ved</i> , <i>zgc:174710</i> , <i>pdkfb4l</i> , <i>LOC100330598</i>
hatching gland	1.13	0.0266	1407	0	0	
periderm	0.85	0.0184	854	0	0	
mesoderm - endothelial	0.85	0.0088	612	2	1 (50%)	<i>id1</i>
epidermal - foxi3a	0.77	0.0242	1032	5	2 (40%)	<i>h1f0</i> , <i>hsp70.2</i>
epidermal anterior	0.7	0.0077	340	14	13 (92.9%)	<i>txnipa</i> , <i>ppp1r15a</i> , <i>hsp70</i> , <i>hsp70.2</i> , <i>LOC103908986</i> , <i>h1f0</i> , <i>fam46ba</i> , <i>setd8b</i> , <i>tbx3a</i> , <i>szl</i>
lateral line primordium	0.56	0.0102	134	10	9 (90%)	<i>hsp70</i> , <i>hsp70.2</i> , <i>id3</i> , <i>ppp1r15a</i> , <i>h1f0</i> , <i>wu:fb55g09</i> , <i>hoxa9b</i> , <i>apoc1l</i>
differentiating neurons - rohon beard	0.5	0.0176	766	0	0	
epidermal - otic placode	0.48	0.0183	193	0	0	
mesoderm - heart field	0.44	0.033	476	7	6 (85.7%)	<i>hsp70.2</i> , <i>hsp70</i> , <i>ppp1r15a</i> , <i>fam46ba</i> , <i>LOC103911698</i> , <i>LOC103908986</i>
neural - floorplate	0.35	0.0314	276	0	0	
mesoderm - pharyngeal arch	0.17	0.0211	431	5	4 (80%)	<i>id1</i> , <i>hsp70.2</i> , <i>hsp70</i> , <i>LOC103908986</i>
germline	0.12	0.0557	430	0	0	
neural crest	0.07	0.0523	460	23	20 (87%)	<i>hsp70.2</i> , <i>hsp70</i> , <i>zgc:136826</i> , <i>txnipa</i> , <i>ppp1r15a</i> , <i>LOC103908986</i> , <i>drl</i> , <i>si:dkey-261j4.4</i> , <i>cepbp</i> , <i>mpp1</i>
neural crest - crestin	0.03	0.0549	204	0	0	
tailbud - spinal cord	0.03	0.05	491	71	43 (60.6%)	<i>zgc:136826</i> , <i>hsp70</i> , <i>hsp70.2</i> , <i>ppp1r15a</i> , <i>snai1a</i> , <i>ved</i> , <i>cdx1a</i> , <i>znf703</i> , <i>prdx1</i> , <i>vox</i>
endoderm	-0.09	0.049	264	0	0	
neural - telencephalon	-0.13	0.0223	501	8	4 (50%)	<i>hsp70.2</i> , <i>hsp70</i> , <i>ppp1r15a</i> , <i>dusp6</i>
mesoderm - pronephric duct	-0.24	0.0541	646	0	0	
differentiating neurons	-0.24	0.0504	1273	1	1 (100%)	<i>hsp70</i>
neural - hindbrain	-0.36	0.0091	695	13	9 (69.2%)	<i>hsp70</i> , <i>hsp70.2</i> , <i>ppp1r15a</i> , <i>fam46ba</i> , <i>pgam1a</i> , <i>trub2</i>
neural - midbrain	-0.46	0.0067	678	15	9 (60%)	<i>hsp70.2</i> , <i>hsp70</i> , <i>LOC101885394</i> , <i>pgam1a</i>
neural - diencephalon	-0.83	0.0188	429	2	2 (100%)	<i>hsp70</i> , <i>hsp70.2</i>
notocord	-0.86	0.0328	1771	1	1 (100%)	<i>ppp1r15a</i>
optic primordium	-0.93	0.015	653	9	7 (77.8%)	<i>yap1</i> , <i>ftr82</i> , <i>slit2</i> , <i>hsp70</i> , <i>ntn1a</i>

B



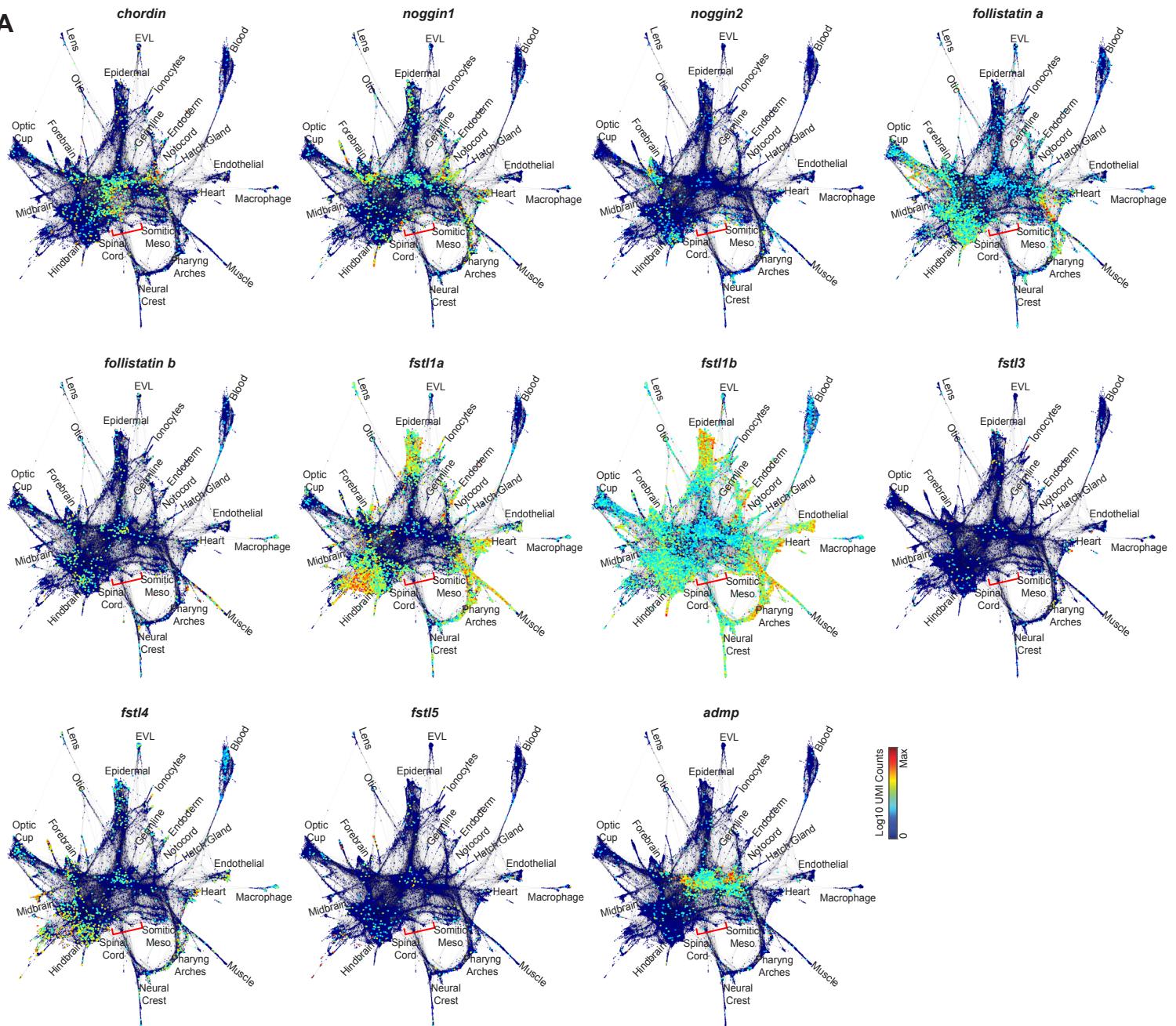
Supplemental Figure 14

(A) State-by-state analysis of the *chordin* phenotype. Cells from 3 *chordin*-targeted and 3 control (*tyrosinase*-targeted) samples were assigned to one of the 28 wild-type annotated states by a kNN-classifier (see Fig. S13 and Methods) and assessed for cell abundance changes and differential gene expression. After normalizing to total cell counts, the relative contribution of cells from the two CRISPR samples to each state was assessed by log2-fold change. Adjusted p-values report significant changes (2-tailed t-test). The extent to which *chordin* and control (*tyrosinase*) cells that were assigned to the same state differed in their transcriptional signatures was assessed by identifying differentially expressed genes (Wilcoxon rank-sum test, adj. pvalue < 0.01; absolute log2 fold change >1). This analysis was limited to transcripts with average expression level >25 transcripts per million (TPM). For each state, the number of differentially expressed genes (DEG) between *chordin* and *tyrosinase* cells was compared to the number of “state-defining” DEG. State-defining DEG were identified by comparing cells of each state to cells of all other states, using the same rank-sum test criteria specified above. Differences between states were systematically associated with much larger DEG numbers than were *chd/tyr* differences. On average, >70% of all *chd/tyr* DEG were “novel” (i.e. not present in the state-defining DEG list). For each state, up to 10 “novel” up-regulated DE genes, ranked by adj. p-value, are listed. These genes include most predominantly a set of multiple heat-shock / stress associated transcripts (e.g. *hsp70*, *hsp70.2*, *tp53inp1*).

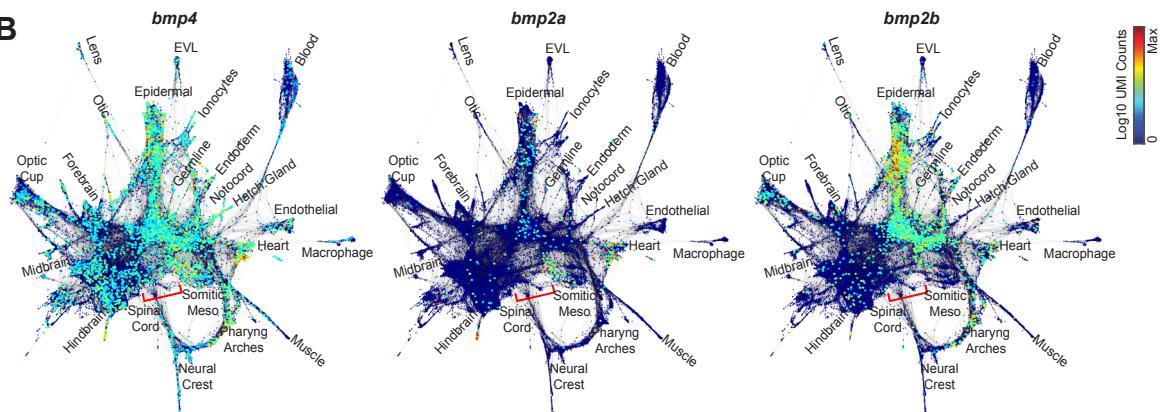
(B) Quantitative trends of the *chordin* phenotype, depicted in a scatterplot. Each cell state is colored by known dorsal / ventral locations in the embryo (magenta: ventral tissues; green: dorsal tissues; grey: “intermediate” tissues). The size of each point reflects the % abundance of each tissue in the wild-type embryo. X-values report the relative changes in state abundance (log2) in *chordin* vs control (*tyrosinase*) embryos. Y-values report the number of differentially expressed genes identified between *chordin* vs control cells in each cluster, same rank-sum test criteria as in (A).

Figure S15

A



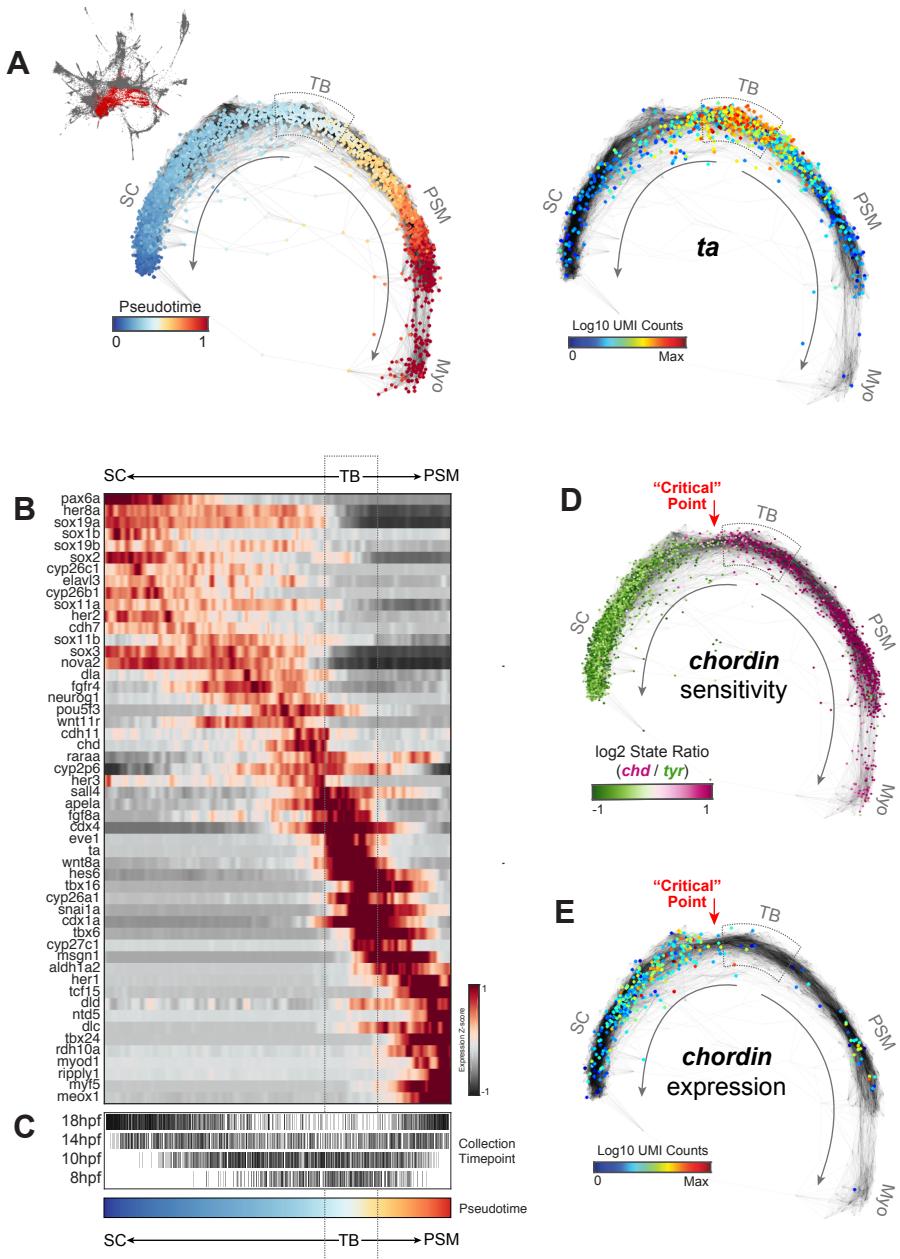
B



Supplemental Figure 15

Single-cell graphs, colored by normalized transcript counts for genes encoding select BMP inhibitors (A), and BMP proteins (B).

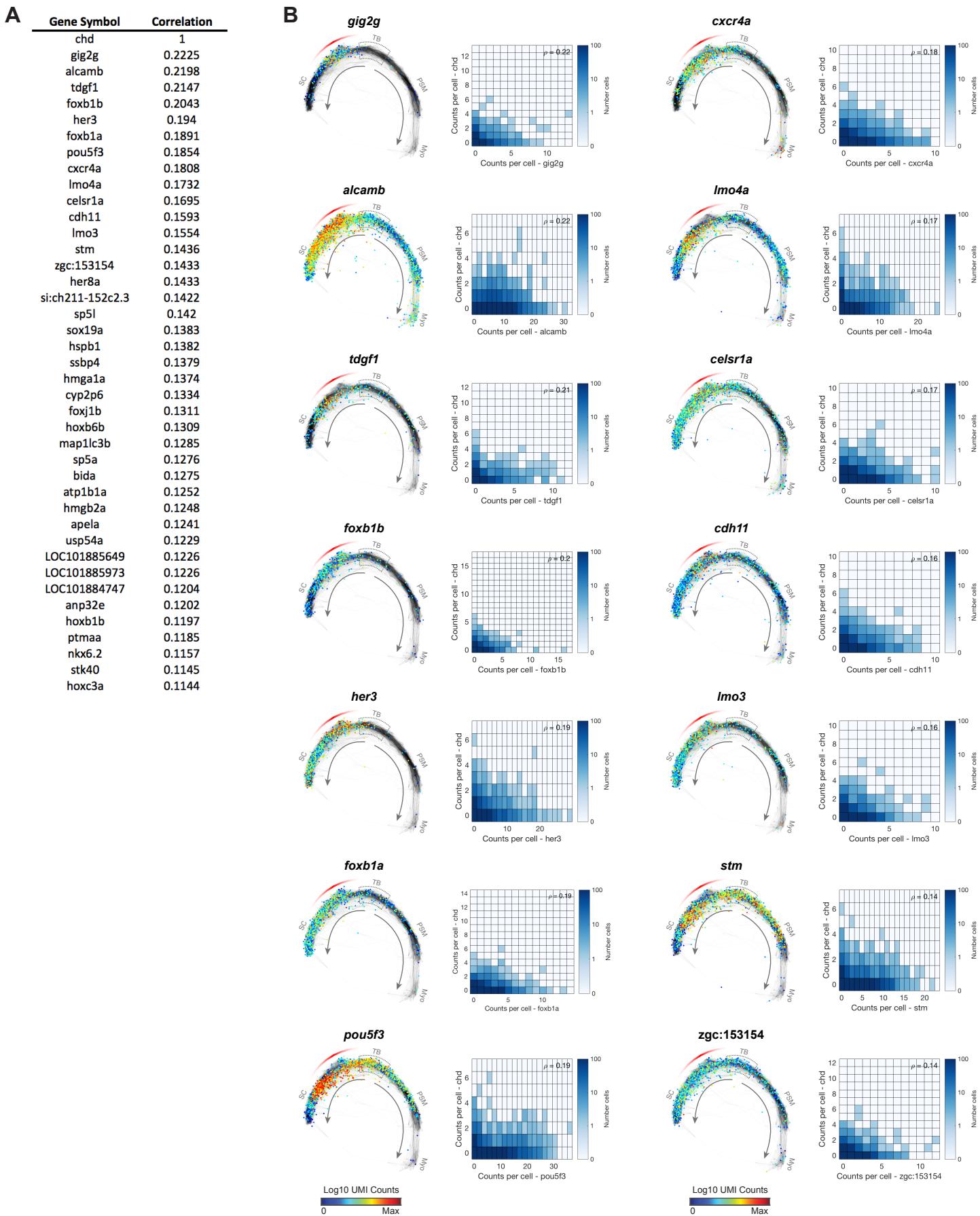
Figure S16



Supplemental Figure 16

- (A) Left: Subspace-projection layout of a subgraph ($n=5,634$ cells) corresponding to the tailbud region of the zebrafish embryo, colored by pseudo-spatiotemporal ordering (see Methods). Colormap was centered on peak *brachyury/ta* expression. TB: tailbud; SC: spinal cord; PSM: presomitic mesoderm. Right: Expression counts for *ta* / *brachyury*.
- (B) Heatmap of gaussian-smoothened expression z-scores for select dynamically expressed genes (see Methods). Distinct cascades of pro-neural or pro-mesodermal genes define two bifurcating trajectories emerging from the central tailbud region.
- (C) Trajectory cells originated from multiple collection timepoints. Black bars indicate the relative contributions of each timepoint to different regions of the tailbud trajectory. Early timepoint cells dominate the center of the trajectory; later timepoints dominate the tips.
- (D) Log2 ratios of differential cell state abundances in *chordin* vs. *tyrosinase* samples (same data as in Fig. 6D). Red arrow hypothesizes a “critical point” in the branching trajectory.
- (E) Overlay of normalized expression counts for *chordin*.

Figure S17



Supplemental Figure 17

Panel of genes co-expressed with *chordin* in the tailbud.

(A) Pearson correlation coefficients were calculated comparing the normalized transcript counts for *chordin* to all other genes, across all cells in the tailbud subgraph (Fig. 16A). Listed are the top 40 most correlated genes.

(B) Left: Gene expression overlays for the top 15 most correlated genes plotted for the tailbud subgraph. Red stripe indicates the graph region containing *chordin*+ cells. Right: Single-cell coexpression scatterplots of normalized transcript counts (vs. *chordin* counts) for the same 15 genes as in (A). p , Pearson correlation.

Captions for Supplemental Tables

Supplemental Table 1

Summary of sequencing statistics for all inDrops RNA-seq libraries.

Supplemental Table 2

Table of significantly enriched marker genes and corresponding annotations for all 195 cell state clusters identified in the study. The top 20 positive differentially expressed genes (ranked by fold enrichment) determined by MAST (46) and Wilcoxon Rank Sum Test were determined by comparing cells of each cluster to all other cells from the same collection timepoint. Differentially expressed genes were identified using the FindAllMarkers routine in Seurat 2.2.0 according to the following criteria: (1) a log₂-fold change >0.5, (2) Adjusted p-value <0.05. (3) >10% of cells in either test group must express at least one UMI.

Supplemental Table 3

Sequences of primers used in this study.

References

1. A. F. Schier, W. S. Talbot, Molecular genetics of axis formation in zebrafish. *Annu. Rev. Genet.* **39**, 561–613 (2005). [doi:10.1146/annurev.genet.37.110801.143752](https://doi.org/10.1146/annurev.genet.37.110801.143752) [Medline](#)
2. T. Hashimshony, N. Senderovich, G. Avital, A. Klochendler, Y. de Leeuw, L. Anavy, D. Gennert, S. Li, K. J. Livak, O. Rozenblatt-Rosen, Y. Dor, A. Regev, I. Yanai, CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77 (2016). [doi:10.1186/s13059-016-0938-8](https://doi.org/10.1186/s13059-016-0938-8) [Medline](#)
3. S. Islam, A. Zeisel, S. Joost, G. La Manno, P. Zajac, M. Kasper, P. Lönnerberg, S. Linnarsson, Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014). [doi:10.1038/nmeth.2772](https://doi.org/10.1038/nmeth.2772) [Medline](#)
4. A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, M. W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015). [doi:10.1016/j.cell.2015.04.044](https://doi.org/10.1016/j.cell.2015.04.044) [Medline](#)
5. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015). [doi:10.1016/j.cell.2015.05.002](https://doi.org/10.1016/j.cell.2015.05.002) [Medline](#)
6. L. Mazutis, J. Gilbert, W. L. Ung, D. A. Weitz, A. D. Griffiths, J. A. Heyman, Single-cell analysis and sorting using droplet-based microfluidics. *Nat. Protoc.* **8**, 870–891 (2013). [doi:10.1038/nprot.2013.046](https://doi.org/10.1038/nprot.2013.046) [Medline](#)
7. R. Zilionis, J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein, L. Mazutis, Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017). [doi:10.1038/nprot.2016.154](https://doi.org/10.1038/nprot.2016.154) [Medline](#)
8. B. Thisse, S. Pflumio, M. Fürthauer, B. Loppin, V. Heyer, A. Degrave, R. Woehl, A. Lux, T. Steffan, X. Q. Charbonnier, C. Thisse, Expression of the zebrafish genome during embryogenesis, ZFIN Direct Data Submission, (2001); <http://zfin.org>.
9. C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, J. L. Rinn, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014). [doi:10.1038/nbt.2859](https://doi.org/10.1038/nbt.2859) [Medline](#)
10. S. C. Bendall, K. L. Davis, A. D. Amir, M. D. Tadmor, E. F. Simonds, T. J. Chen, D. K. Shenfeld, G. P. Nolan, D. Pe'er, Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714–725 (2014). [doi:10.1016/j.cell.2014.04.005](https://doi.org/10.1016/j.cell.2014.04.005) [Medline](#)
11. J. Shin, D. A. Berg, Y. Zhu, J. Y. Shin, J. Song, M. A. Bonaguidi, G. Enikolopov, D. W. Nauen, K. M. Christian, G. L. Ming, H. Song, Single-cell RNA-seq with Waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* **17**, 360–372 (2015). [doi:10.1016/j.stem.2015.07.013](https://doi.org/10.1016/j.stem.2015.07.013) [Medline](#)
12. L. Haghverdi, M. Büttner, F. A. Wolf, F. Buettner, F. J. Theis, Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016). [doi:10.1038/nmeth.3971](https://doi.org/10.1038/nmeth.3971) [Medline](#)
13. C. B. Kimmel, R. M. Warga, T. F. Schilling, Origin and organization of the zebrafish fate map. *Development* **108**, 581–594 (1990). [Medline](#)
14. A. E. Melby, R. M. Warga, C. B. Kimmel, Specification of cell fates at the dorsal margin of the zebrafish gastrula. *Development* **122**, 2225–2237 (1996). [Medline](#)
15. R. M. Warga, C. Nüsslein-Volhard, Origin and development of the zebrafish endoderm. *Development* **126**, 827–838 (1999). [Medline](#)

16. K. Woo, S. E. Fraser, Order and coherence in the fate map of the zebrafish nervous system. *Development* **121**, 2595–2609 (1995). [Medline](#)
17. K. Kawakami, Tol2: A versatile gene transfer vector in vertebrates. *Genome Biol.* **8** (suppl. 1), S7 (2007). [doi:10.1186/gb-2007-8-s1-s7](https://doi.org/10.1186/gb-2007-8-s1-s7) [Medline](#)
18. D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison 3rd, H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009). [doi:10.1038/nmeth.1318](https://doi.org/10.1038/nmeth.1318) [Medline](#)
19. A. McKenna, G. M. Findlay, J. A. Gagnon, M. S. Horwitz, A. F. Schier, J. Shendure, Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016). [doi:10.1126/science.aaf7907](https://doi.org/10.1126/science.aaf7907) [Medline](#)
20. J. P. Junker, B. Spanjaard, J. Peterson-Maduro, A. Alemany, B. Hu, M. Florescu, A. van Oudenaarden, Massively parallel clonal analysis using CRISPR/Cas9 induced genetic scars. *BioRxiv* 056499 [Preprint]. 4 January 2017. <https://doi.org/10.1101/056499>.
21. R. M. Warga, D. A. Kane, R. K. Ho, Fate mapping embryonic blood in zebrafish: Multi- and unipotential lineages are segregated at gastrulation. *Dev. Cell* **16**, 744–755 (2009). [doi:10.1016/j.devcel.2009.04.007](https://doi.org/10.1016/j.devcel.2009.04.007) [Medline](#)
22. K. E. Whitlock, M. Westerfield, The olfactory placodes of the zebrafish form by convergence of cellular fields at the edge of the neural plate. *Development* **127**, 3645–3653 (2000). [Medline](#)
23. E. Tzouanacou, A. Wegener, F. J. Wyneersch, V. Wilson, J.-F. Nicolas, Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev. Cell* **17**, 365–376 (2009). [doi:10.1016/j.devcel.2009.08.002](https://doi.org/10.1016/j.devcel.2009.08.002) [Medline](#)
24. R. L. Davis, M. W. Kirschner, The fate of cells in the tailbud of *Xenopus laevis*. *Development* **127**, 255–267 (2000). [Medline](#)
25. J. P. Kanki, R. K. Ho, The development of the posterior body in zebrafish. *Development* **124**, 881–893 (1997). [Medline](#)
26. N. M. Le Douarin, E. Dupin, Multipotentiality of the neural crest. *Curr. Opin. Genet. Dev.* **13**, 529–536 (2003). [doi:10.1016/j.gde.2003.08.002](https://doi.org/10.1016/j.gde.2003.08.002) [Medline](#)
27. N. M. Le Douarin, S. Creuzet, G. Couly, E. Dupin, Neural crest cell plasticity and its limits. *Development* **131**, 4637–4650 (2004). [doi:10.1242/dev.01350](https://doi.org/10.1242/dev.01350) [Medline](#)
28. C. S. Le Lièvre, N. M. Le Douarin, Mesenchymal derivatives of the neural crest: Analysis of chimaeric quail and chick embryos. *J. Embryol. Exp. Morphol.* **34**, 125–154 (1975). [Medline](#)
29. Y. Sasai, B. Lu, H. Steinbeisser, D. Geissert, L. K. Gont, E. M. De Robertis, *Xenopus chordin*: A novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell* **79**, 779–790 (1994). [doi:10.1016/0092-8674\(94\)90068-X](https://doi.org/10.1016/0092-8674(94)90068-X) [Medline](#)
30. M. Hammerschmidt, F. Pelegri, M. C. Mullins, D. A. Kane, F. J. van Eeden, M. Granato, M. Brand, M. Furutani-Seiki, P. Haffter, C. P. Heisenberg, Y. J. Jiang, R. N. Kelsh, J. Odenthal, R. M. Warga, C. Nüsslein-Volhard, dino and mercedes, two genes regulating dorsal development in the zebrafish embryo. *Development* **123**, 95–102 (1996). [Medline](#)
31. S. Schulte-Merker, K. J. Lee, A. P. McMahon, M. Hammerschmidt, The zebrafish organizer requires *chordin*. *Nature* **387**, 862–863 (1997). [doi:10.1038/43092](https://doi.org/10.1038/43092) [Medline](#)
32. Y. Sasai, B. Lu, H. Steinbeisser, E. M. De Robertis, Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in *Xenopus*. *Nature* **378**, 419 (1995). [doi:10.1038/378419d0](https://doi.org/10.1038/378419d0) [Medline](#)
33. S. Piccolo, Y. Sasai, B. Lu, E. M. De Robertis, Dorsoventral patterning in *Xenopus*: Inhibition of ventral signals by direct binding of chordin to BMP-4. *Cell* **86**, 589–598 (1996). [doi:10.1016/S0092-8674\(00\)80132-4](https://doi.org/10.1016/S0092-8674(00)80132-4) [Medline](#)

34. J. A. Gagnon, E. Valen, S. B. Thyme, P. Huang, L. Akhmetova, A. Pauli, T. G. Montague, S. Zimmerman, C. Richter, A. F. Schier, Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLOS ONE* **9**, e98186 (2014). [doi:10.1371/journal.pone.0098186](https://doi.org/10.1371/journal.pone.0098186) Medline
35. A. Y. Leung, E. M. Mendenhall, T. T. F. Kwan, R. Liang, C. Eckfeldt, E. Chen, M. Hammerschmidt, S. Grindley, S. C. Ekker, C. M. Verfaillie, Characterization of expanded intermediate cell mass in zebrafish chordin morphant embryos. *Dev. Biol.* **277**, 235–254 (2005).
[doi:10.1016/j.ydbio.2004.09.032](https://doi.org/10.1016/j.ydbio.2004.09.032) Medline
36. V. E. Miller-Bertoglio, S. Fisher, A. Sánchez, M. C. Mullins, M. E. Halpern, Differential regulation of *chordin* expression domains in mutant zebrafish. *Dev. Biol.* **192**, 537–550 (1997).
[doi:10.1006/dbio.1997.8788](https://doi.org/10.1006/dbio.1997.8788) Medline
37. R. H. Row, D. Kimelman, Bmp inhibition is necessary for post-gastrulation patterning and morphogenesis of the zebrafish tailbud. *Dev. Biol.* **329**, 55–63 (2009).
[doi:10.1016/j.ydbio.2009.02.016](https://doi.org/10.1016/j.ydbio.2009.02.016) Medline
38. R. H. Row, S. R. Tsotras, H. Goto, B. L. Martin, The zebrafish tailbud contains two independent populations of midline progenitor cells that maintain long-term germ layer plasticity and differentiate in response to local signaling cues. *Development* **143**, 244–254 (2016).
[doi:10.1242/dev.129015](https://doi.org/10.1242/dev.129015) Medline
39. M. Gouti, J. Delile, D. Stamatakis, F. J. Wymeersch, Y. Huang, J. Kleinjung, V. Wilson, J. Briscoe, A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. *Dev. Cell* **41**, 243–261.e7 (2017). [doi:10.1016/j.devcel.2017.04.002](https://doi.org/10.1016/j.devcel.2017.04.002) Medline
40. P. Gönczy, Mechanisms of asymmetric cell division: Flies and worms pave the way. *Nat. Rev. Mol. Cell Biol.* **9**, 355–366 (2008). [doi:10.1038/nrm2388](https://doi.org/10.1038/nrm2388) Medline
41. C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, T. F. Schilling, Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995). [doi:10.1002/aja.1002030302](https://doi.org/10.1002/aja.1002030302) Medline
42. M. Manoli, W. Driever, Fluorescence-activated cell sorting (FACS) of fluorescently tagged cells from zebrafish larvae for RNA isolation. *Cold Spring Harb. Protoc.* **2012**, pdb.prot069633 (2012).
[doi:10.1101/pdb.prot069633](https://doi.org/10.1101/pdb.prot069633) Medline
43. L. Van der Maaten, G. Hinton, Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
44. L. Van der Maaten, Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).
45. A. Rodriguez, A. Laio, Clustering by fast search and find of density peaks. *Science* **344**, 1492–1496 (2014). [doi:10.1126/science.1242072](https://doi.org/10.1126/science.1242072) Medline
46. 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). [doi:10.1186/s13059-015-0844-5](https://doi.org/10.1186/s13059-015-0844-5) Medline
47. J. D. Storey, A direct approach to false discovery rates. *J. R. Stat. Soc. Series B Stat. Methodol.* **64**, 479–498 (2002). [doi:10.1111/1467-9868.00346](https://doi.org/10.1111/1467-9868.00346)
48. C. Weinreb, S. Wolock, A. M. Klein, SPRING: A kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248 (2018).
[doi:10.1093/bioinformatics/btx792](https://doi.org/10.1093/bioinformatics/btx792) Medline
49. Q. Zhu, J. Feng, J. Huang, Weighted natural neighborhood graph: An adaptive structure for clustering and outlier detection with no neighborhood parameter. *Cluster Comput.* **19**, 1385–1397 (2016).
[doi:10.1007/s10586-016-0598-1](https://doi.org/10.1007/s10586-016-0598-1)

50. D. Ting, L. Huang, M. Jordan, An analysis of the convergence of graph Laplacians. [arXiv:1101.5435](https://arxiv.org/abs/1101.5435) [stat.ML] (28 January 2011).
51. M. Jacomy, T. Venturini, S. Heymann, M. Bastian, ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE* **9**, e98679 (2014). [doi:10.1371/journal.pone.0098679](https://doi.org/10.1371/journal.pone.0098679) [Medline](#)
52. M. Bastian, S. Heymann, M. Jacomy, “Gephi: an open source software for exploring and manipulating networks,” in *Third International AAAI Conference on Weblogs and Social Media* (Association for the Advancement of Artificial Intelligence, AAAI, 2009).
53. Y. Hu, Efficient, high-quality force-directed graph drawing. *Mathematica J.* **10**, 37–71 (2006).
54. Y. Koren, Drawing graphs by eigenvectors: Theory and practice. *Comput. Math. Appl.* **49**, 1867–1888 (2005). [doi:10.1016/j.camwa.2004.08.015](https://doi.org/10.1016/j.camwa.2004.08.015)
55. L. E. Jao, S. R. Wente, W. Chen, Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13904–13909 (2013). [doi:10.1073/pnas.1308335110](https://doi.org/10.1073/pnas.1308335110) [Medline](#)
56. A. Y. Kobitski, J. C. Otte, M. Takamiya, B. Schäfer, J. Mertes, J. Stegmaier, S. Rastegar, F. Rindone, V. Hartmann, R. Stotzka, A. García, J. van Wezel, R. Mikut, U. Strähle, G. U. Nienhaus, An ensemble-averaged, cell density-based digital model of zebrafish embryo development derived from light-sheet microscopy data with single-cell resolution. *Sci. Rep.* **5**, 8601 (2015). [doi:10.1038/srep08601](https://doi.org/10.1038/srep08601) [Medline](#)