RESEARCH ARTICLE

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*

High-throughput mapping of cellular differentiation hierarchies from single-cell data promises to empower systematic interrogations of vertebrate development and disease. Here we applied single-cell RNA sequencing to >92,000 cells from zebrafish embryos during the first day of development. Using a graph-based approach, we mapped a cell-state landscape that describes axis patterning, germ layer formation, and organogenesis. We tested how clonally related cells traverse this landscape by developing a transposon-based barcoding approach (TracerSeq) for reconstructing single-cell lineage histories. Clonally related cells were often restricted by the state landscape, including a case in which two independent lineages converge on similar fates. Cell fates remained restricted to this landscape in embryos lacking the *chordin* gene. We provide web-based resources for further analysis of the single-cell data.

A major goal of developmental biology is to understand the progression of embryonic cell lineages from pluripotency to adulthood (*1*). Fate mapping and analysis of mutant phenotypes have explained much of what we know of development, yet we still lack a systematic atlas of all cell states in a developing embryo. Owing to technical advances in single-cell RNA sequencing (scRNA-seq) (*2–6*), it is now possible to assemble comprehensive single-cell atlases describing complex and dynamic in vivo biological processes. Here we utilized inDrops scRNA-seq (*4, 7*) to collect more than 92,000 single-cell transcriptomes from dissociated wild-type and mutant zebrafish embryos during the first 24 hours of embryonic development (Fig. 1 and fig. S1). For different developmental stages, we sampled 17 to 97% of the total cells per embryo, sufficient to detect cell states as rare as 0.1 to 0.5% of all cells (fig. S1C), including germ cells, which were detected in all time points (Fig. 1B and table S2). From this dataset, clustering of the wild-type transcriptomes revealed an expanding set of epidermal, neural, mesodermal, and endodermal cell states over developmental time, many of which could be specifically annotated on the basis of expression of marker genes (Fig. 1B, fig. S2A, and table S2) (*8*). We collected seven biological replicates for the final time point [24 hours postfertilization (hpf)], which demonstrated consistency of both transcriptional signatures and cell-state proportions across independent specimens (fig. S2, B and C).

Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.
*Corresponding author. Email: sean_megason@hms.harvard.edu (S.G.M.); allon_klein@hms.harvard.edu (A.M.K.)

A single-cell graph of cell-state progression in the developing zebrafish embryo

We sought to map trajectories of cell state during development by linking cell states across time. Several computational approaches exist to infer orderings of asynchronous processes from scRNA-seq data (*9–11*), typically by projecting all cells into a single low-dimensional latent space. Such strategies may be ill-suited to map gene expression in developing embryos, which exhibit dramatically increasing cell-state dimensionality and continuous changes in the sets and numbers of cell state–defining genes (fig. S2, D and E). To overcome these obstacles, we developed a graph-based strategy for locally embedding consecutive time points on the basis of the biological variation that they share, rather than using a global coordinate system for all time points. This approach first constructs a single-cell *k*-nearest neighbor graph for each time point t_i, with nodes representing cells and edges that link neighbors in a low-dimensional subspace; it then joins the graphs by identifying neighboring cells in pairs of adjacent time points, by using a coordinate system learned from the future (t_{i+1}) time point (see methods). The resulting graph spans all time points and allows application of formal graph-based methods for data analysis. When applied to our zebrafish data, the full graph forms a branching network (Fig. 2A). Inspection of numerous domain and cell type–specific transcriptional markers shows that major initial branches represent neural, epidermal, and mesendodermal states undergoing progressive and spatially restricted differentiation (Fig. 2, B and C, and fig. S3). We also noted distinct and early branching events for germline, notochord, enveloping layer epidermis, and the prechordal plate.

To test whether this graph recapitulates known lineage relationships, we used a measure of graph distance (diffusion pseudotime, or DPT) (*12*) to
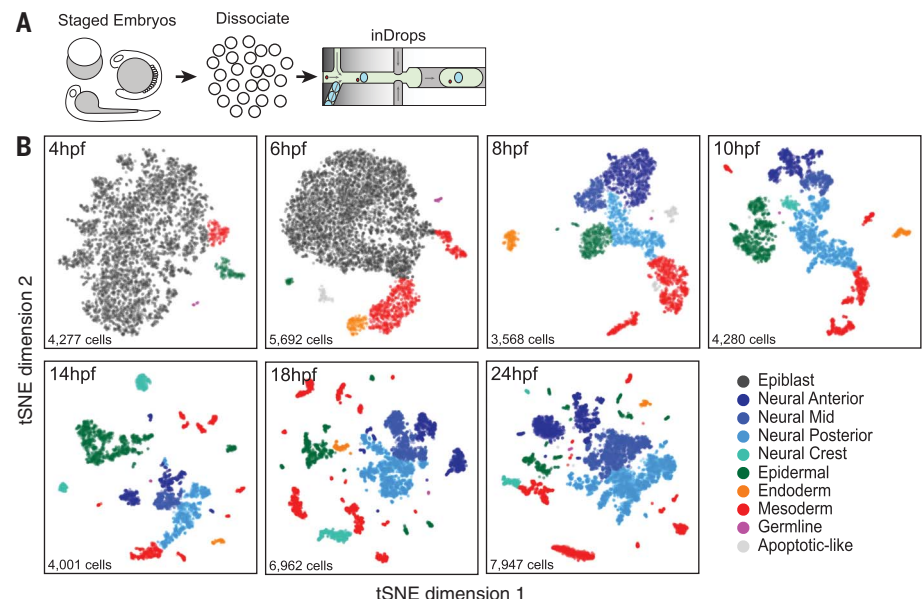

Fig. 1. A single-cell transcriptional atlas of the zebrafish embryo. (**A**) Experimental workflow. Single-cell suspensions were dissociated from staged zebrafish embryos and introduced into the inDrops microfluidic device. Single-cell transcriptome libraries were prepared and sequenced by RNA-seq. (**B**) tSNE maps for each time point, constructed in dimensionality-reduced principal component analysis subspace defined by highly covariable genes (see methods). Cells are colored by germ layer identities inferred from expressed marker genes (see also fig. S2A and table S2).

explore long-range temporal connections between cell states. Cell states of the early gastrula (shield stage, 6 hpf) are defined largely by positional marker genes (Fig. 3A), yet these cells are connected, through the single-cell graph, to tissue-specific states that emerge later (for example, pharyngula stage, 24 hpf). We found that the shield-stage cells with the shortest mean graph distance to each particular 24-hpf tissue were clustered and expressed spatial marker genes predicted from previous in vivo fate-mapping studies (13–16); for example, 24-hpf neural tissues mapped to the 6-hpf dorsal anterior epiblast (Fig. 3B and fig. S4). Conversely, direct comparison

of 6- and 24-hpf gene expression states failed to capture lineage relationships (Fig. 3B and fig. S4, blue points).

We next tested the extent to which the single-cell graph represents a simple treelike hierarchy of discrete states. For this, we "coarse-grained" the graph by collapsing groups of similar cells into state nodes; edges between state nodes were weighted by the number of original single-cell connecting edges. A spanning tree was then traced through the most densely weighted edges to a 4-hpf root state (Fig. 3C and fig. S5A). This spanning tree (the "state tree") reflects many specific aspects of early development. In the neural plate,

we observe notable branch points for the optic cup, diencephalon, telencephalon, mesencephalon, and rhombencephalon, with associated states for region-specific postmitotic neurons (for example, *eomes*[+] and *dlx1*[+] neurons in distinct forebrain branches). The neural plate also includes neural crest, which branches to include cell states for melanoblasts, iridoblasts, and xanthoblasts. In the lateral plate and ventral mesoderm, the state tree encodes extensive branching into hematopoietic cells, endothelial cells, heart, pharyngeal arches, the pronephritic duct, and fin buds. In the endoderm, two branch points give rise to cell states for pancreatic primordium (which includes *insulin*[+]
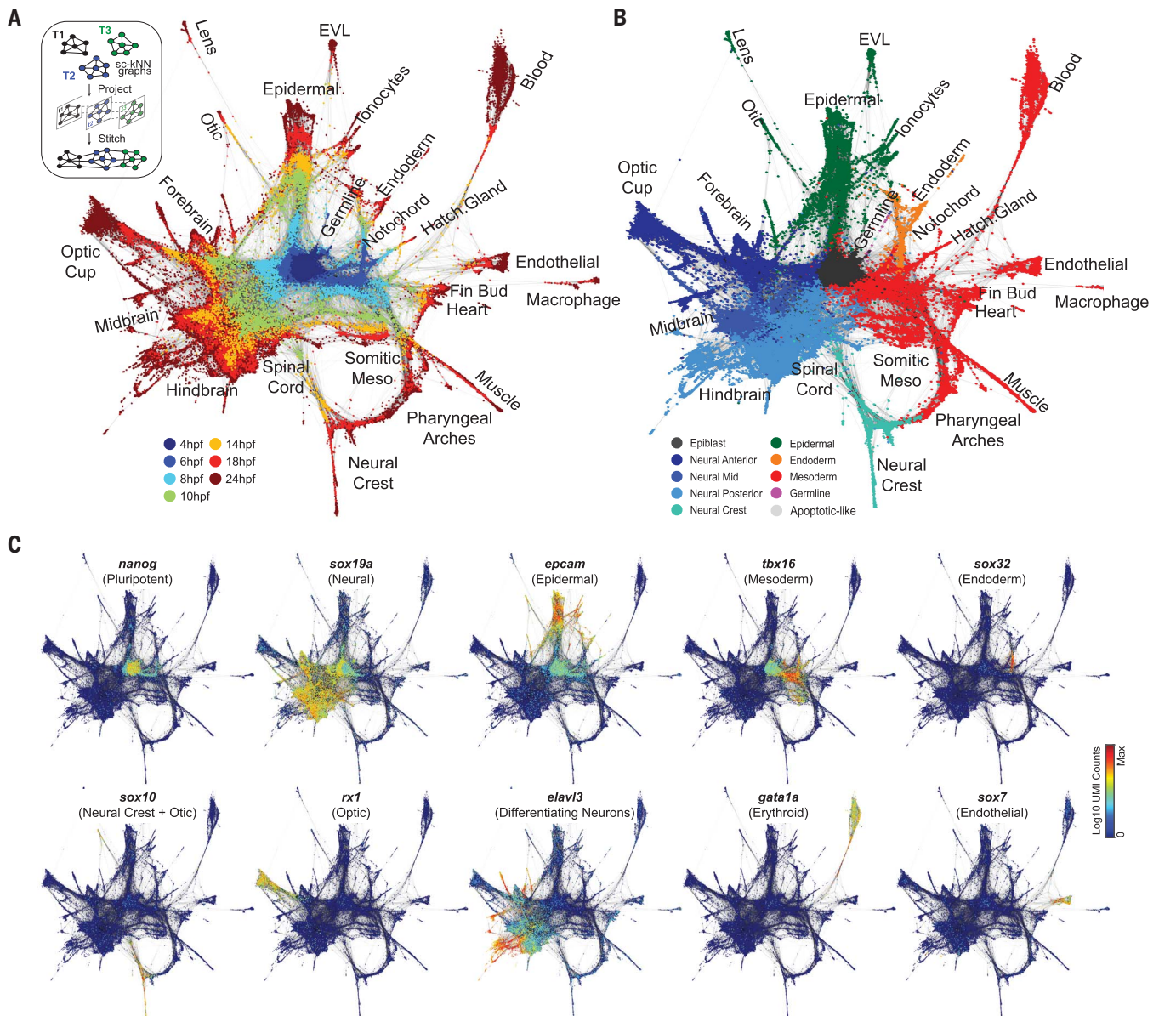

Fig. 2. Single-cell graph reveals a continuous developmental landscape of cell states. (**A**) Overview of graph construction strategy and a force-directed layout of the resulting single-cell graph (nodes colored by collection time point). For each cell, up to 20 within– or between–time point mutual nearest neighbor edges are retained. sc-kNN, single cell–k-nearest neighbor; EVL, enveloping layer epidermis. (**B**) Single-cell graph colored by germ layer identities inferred from differentially expressed marker genes (see table S2). (**C**) Single-cell graphs, colored by log$_{10}$ expression counts for indicated cell type–specific marker genes. UMI, unique molecular identifier.

cells) and the pharyngeal pouch. In the epidermal lineage, branch points differentiate the otic placode, lateral line, ionocytes, and several states expressing markers for annotated mucous-secreting cells (8). To facilitate data exploration, we developed web-based interfaces for the state tree and the full single-cell graph (www.tinyurl.com/scZfish2018). These tools permit interactive examination of the inferred state hierarchy, expression for any gene of interest, and differential expression analysis between states, state combinations, or single cells.

Although many major cell-state transitions are captured in the state tree, more complex features are evident in the coarse-grained and single-cell graphs. Off-tree interconnections between states, for example, were evident for the neural crest and pharyngeal arches, spinal cord and somitic mesoderm, the neural plate, and others (Fig. 3C and fig. S5A). To formalize the degree to which the developmental landscape can be approximated as a hierarchy with discrete, non-looping branches, we defined a "canalization score" (Fig. 3D, see legend for definition), which reflects the off-tree connectivity of each coarse-grained state node. This analysis revealed widespread regions of low canalization, particularly in the neural plate and somitic mesoderm. These observations suggest that, in contrast to the classic notion of a cell lineage, the zebrafish cell-state landscape cannot be fully represented as a tree.

Cell-lineage history does not invariantly reflect cell-state graph topology

Although the single-cell and coarse-grained graphs represent an inferred landscape of developmental cell states, they do not reveal how individual cells traverse these states. A simple prediction would be that individual cell histories mirror graph topology. We tested this prediction by developing an inDrops-compatible strategy for recording in vivo lineage histories at the single-cell level: sequencing of transcribed clonally encoded random barcodes (TracerSeq). TracerSeq utilizes the Tol2 transposase system (17) to randomly integrate green fluorescent protein (GFP) reporter cassettes driven by the β-actin promoter (actb2) into the zebrafish genome. To render each integration event unique and detectable by RNA-seq, we utilized Gibson assembly (18) without subsequent amplification to introduce a random 20-nucleotide oligomer sequence barcode into the GFP 3′ un-translated region (Fig. 4A and fig. S6). Because transgenic insertions can occur asynchronously over successive cell divisions, TracerSeq barcodes can facilitate the construction of lineage trees (Fig. 4A). TracerSeq offers an advantage over related Cas9-based approaches (19, 20), which can generate identical edits and/or large barcode deletions in independent lineages at nontrivial frequencies. By contrast, TracerSeq barcodes are uniformly distributed over a large sequence space (for example, 4^{20} nucleotide combinations = 10^{12} unique barcode sequences), facilitating straightforward calling of genetic clones (fig. S7). The small (20–base pair) locus size also greatly simplifies the construction, sequencing, and analysis of TracerSeq inDrops libraries.

The use of TracerSeq to analyze potentially small clones of cells (each restricted to a single embryo) requires high-efficiency tissue dissociation and transcriptomic barcoding methods. We therefore optimized a high-yield cell dissociation and recovery protocol for individual 24-hpf zebrafish embryos (fig. S1D and methods) and leveraged the high cell barcoding efficiency (>80%) of the inDrops platform (7). We then sequenced individual embryos (n = 5) at 24 hpf (fig. S7) that were injected at the one-cell stage with the TracerSeq library, generating combined lineage and transcriptome datasets for 1269 clonal barcodes distributed over 4342 single cells (fig. S8). Of these cells, 2361 (54%) were each marked by ≥2 distinct barcode integrations; 624 cells (14%) were marked by ≥5 integrations (fig. S8). Hierarchical clustering of TracerSeq barcodes organized these cells into more than a hundred distinct founder clones with internal nested clone structures (Fig. 4B and fig. S9, A to D). We then compared the lineage history and inferred transcriptional history of each founder clone by embedding its constituent cells onto the single-cell graph (Fig. 4C). We found that the largest clones often marked a wide diversity of cell states. In multiple cases, however, additional barcode integrations in the same founder clone marked cells that were state restricted. For example, one such clone (34F1) marked cells of the neural plate, epidermal tissues, and muscle but contained a subclone restricted to ectoderm. Similar lineage restriction events could be described for other founder clones (Fig. 4C). These observations suggest that the current timing of TracerSeq integrations encompasses the transition from unrestricted pluripotency to the first fate-restriction events appearing in the zebrafish embryo.

To investigate lineage relationships more systematically, we assessed the likelihood of recovering shared TracerSeq barcodes from all pairs of transcriptional states in the 24-hpf zebrafish embryo. We first calculated a lineage coupling score (fig. S9E and methods), defined as the number of shared barcodes relative to randomized data (z-score standardized), with values ranging from positive (coupled fates) to negative (anti-coupled fates). Hierarchical clustering of the pairwise correlation between coupling scores revealed structured groups of cell states (Fig. 5A), which comprised related tissues and/or inferred
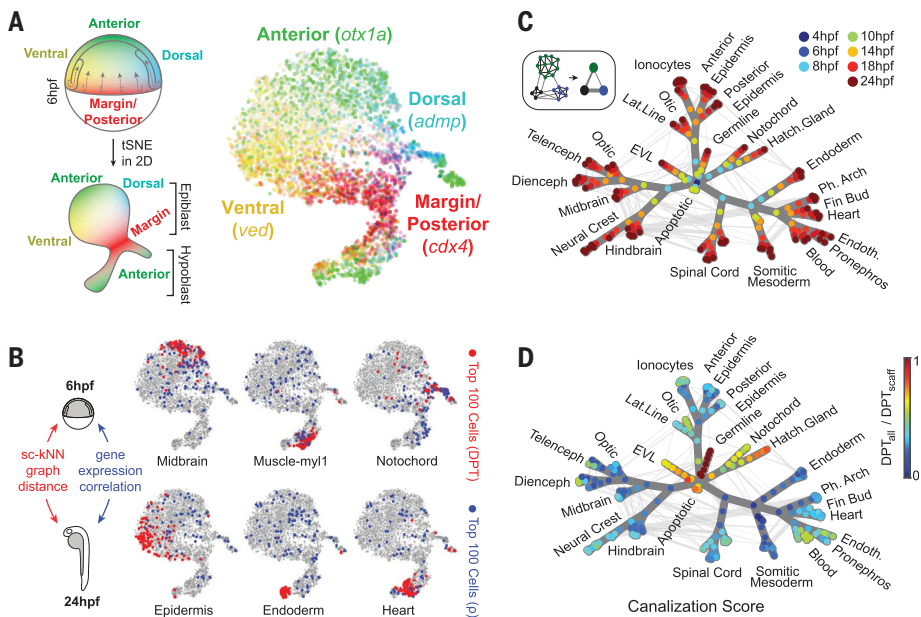

Fig. 3. Single-cell and coarse-grained graphs encode progenitor-fate relationships. (**A**) tSNE map of 6-hpf epiblast and hypoblast states, colored by normalized transcript counts for select positional marker genes. Overlapping color gradients demonstrate continuous expression domains defined by position. Diagram relates positions of cells in the tSNE map to theoretical positions in the embryo. 2D, two dimensions. (**B**) In silico fate predictions for 6-hpf embryo cells. The top 100 cells with predicted 24-hpf fate outcomes are indicated for shortest graph diffusion distances (red) or direct single-cell gene expression correlation distances (blue) between 6-hpf cells and 24-hpf cluster centroids. ρ, Pearson correlation. (**C**) Construction and overview of the coarse-grained graph (see also fig. S5). Nodes indicate states (groups of transcriptionally similar cells), colored by time point. Weighted edges connect similar states within or between time points. Spanning tree edges connecting each node to the 4-hpf root state through the top weighted edges are highlighted in dark gray. (**D**) Coarse-grained graph nodes are colored by a canalization score, defined as the ratio of diffusion distances between each node (DPT_{all}) and the 4-hpf root node through state tree edges only (DPT_{scaff}) versus through all graph edges. Highly canalized regions of the graph correspond to branches with the fewest off-tree edges.

germ layer derivatives. These included one distinct group that contained both mesodermal and endodermal derivatives, four groups containing ectodermal derivatives, and two groups containing mixtures of ectoderm and mesoderm. Several of these lineage groups are corroborated by prior fate-mapping studies. We discuss here three examples. The first major lineage group (MesEndo) includes derivatives of both lateral plate mesoderm and endoderm. These tissues originate from the marginal blastomeres of the early zebrafish gastrula, which involute first during gastrulation to form the hypoblast and then rapidly migrate toward the animal pole (13, 15, 21). The observed lineage isolation of these tissues is thus consistent with an early spatial partitioning of this region, further reflected in Fig. 5A by negative lineage correlations to most other states. A second group (Fig. 5A, Ecto III) captures strong lineage couplings both between anterior neural tissues—including the optic cup, midbrain, and telencephalon (16)—and to anterior epidermal derivatives such as the olfactory placode (22). These tissues are coupled to a lower degree with another group (Ecto II), which includes couplings between the hindbrain, spinal cord, and neural crest (grem2[+]). The third example we note is a group coupling ectoderm and mesoderm

(Fig. 5A, MesEcto II), including muscle (myl1[+]), myotome, spinal cord, posterior neural crest, and epidermal states. These correlations mirror the development of posterior body regions, which trace their origins to blastomeres proximal to the medial and ventral margin (13). These mesodermal–spinal cord couplings might also be explained by the presence of a later population of transient, multipotent neuromesodermal progenitor cells in the embryonic tailbud, which give rise to both of these populations (23–25). Interestingly, these lineage groups tend to be organized by position (for example, along the anterior-posterior axis) rather than strictly by germ layer or tissue origin (for example, neural, epidermal, and mesodermal).

We next questioned how clonal relationships compared with cell-state relationships. A simplistic model of development is that cells progressively diverge in state as they diverge in lineage. Developing embryos, however, could violate this prediction in at least two ways: First, clonally distinct embryonic fields can give rise to similar cell types ("convergent clones"); second, major transcriptional changes might drive related cells into qualitatively dissimilar states, possibly even late in development ("divergent clones"). Overlaying TracerSeq lineage correlation scores on

the cell-state graph and comparing these scores to graph-derived state distances (Fig. 5B and fig. S10) revealed that some nearby states on the state graph were indeed clonally correlated, as expected by the simplistic model. However, nearby cell states also frequently displayed weak clonal correlations, suggesting convergent differentiation. These patterns were evident among state relationships for endothelial, optic cup, and muscle tissues (Fig. 5B and fig. S10, A to F) and systematically when examining all states (fig. S10G).

We observed considerably fewer cases of divergent clonal behavior (fig. S10G). However, one notable example manifested as apparent looping of the neural crest into the pharyngeal arches, which originate in the graph from both neural plate and lateral plate mesoderm and merge at 18 to 24 hpf (Fig. 2, A and B, and fig. S11A). Although the contribution of neural crest to various mesenchymal tissues is well established (26–28), the transcriptional information reflected by the graph loop alone does not reveal which annotated pharyngeal arch states arise from neural crest. TracerSeq data, however, provide a clear signature of distinct clonal patterns between pharyngeal arch states: One pharyngeal arch state (ph.arch-tbx1) is a member of the MesEndo lineage group with mesodermal clonal associations, whereas

Fig. 4. Single-cell transcriptomic barcoding of cell lineages using TracerSeq. (**A**) Method overview. Tol2 transposase system integrates barcode-containing GFP reporter cassettes into zebrafish genome. Asterisks denote integration events. Colors (red, blue, black, and green) indicate unique barcode sequences. (**B**) Clustered heatmap for one of five TracerSeq embryos (see also fig. S9, A to D), displaying lineage and transcriptome information for each cell. Heatmap rows are single cells for which both transcriptome and >1 TracerSeq barcodes were recovered. Columns denote unique TracerSeq barcodes (left: black squares, ≥1 UMI) and tissue identities (right: red squares) inferred from cluster annotations (table S2). Heatmaps were clustered using Jaccard similarity and average linkage. (**C**) Examples of TracerSeq founder clones with positions of constituent cells (colored nodes) overlaid on the single-cell graph. Graph edges are shown in dark gray. Colors indicate the first lineage bifurcation within each founder clone. In the three cases shown, the founder clone included cells that differentiated into both ectodermal (E) and mesodermal (M) states, whereas one of the two first subclones was restricted to ectoderm.
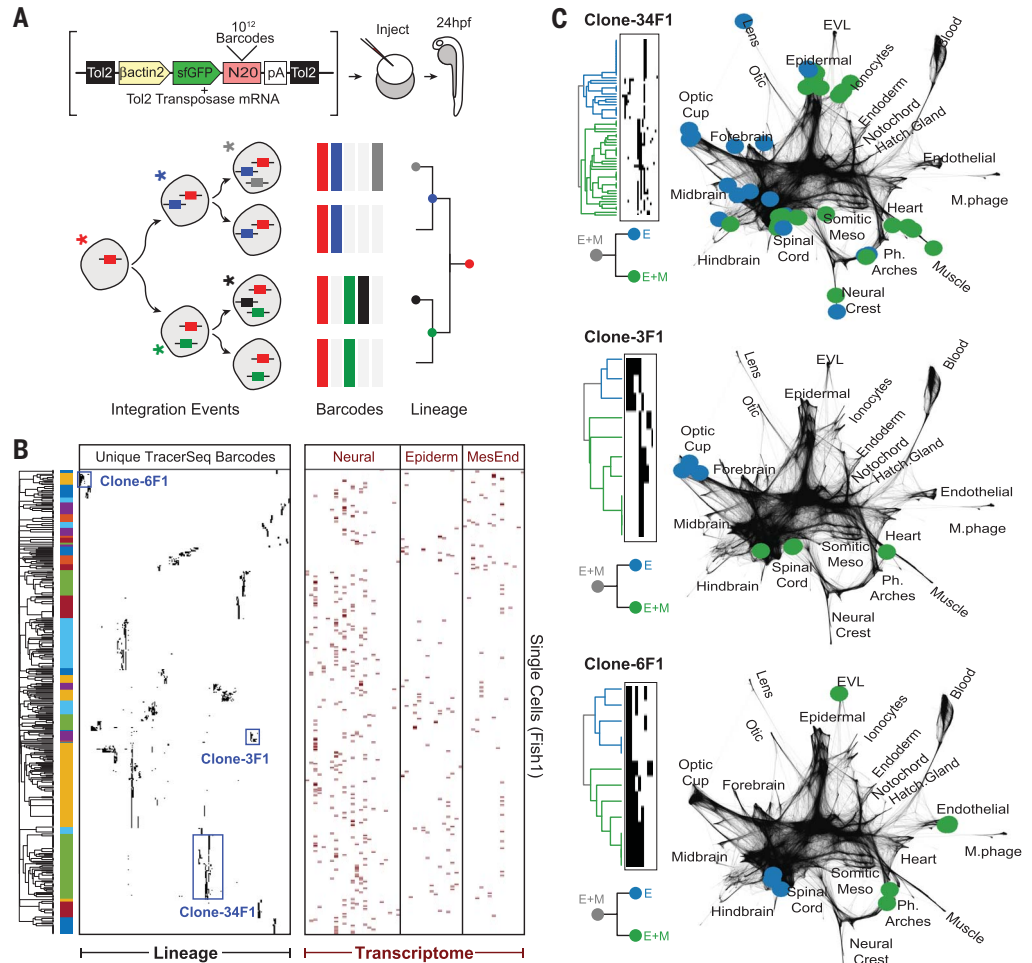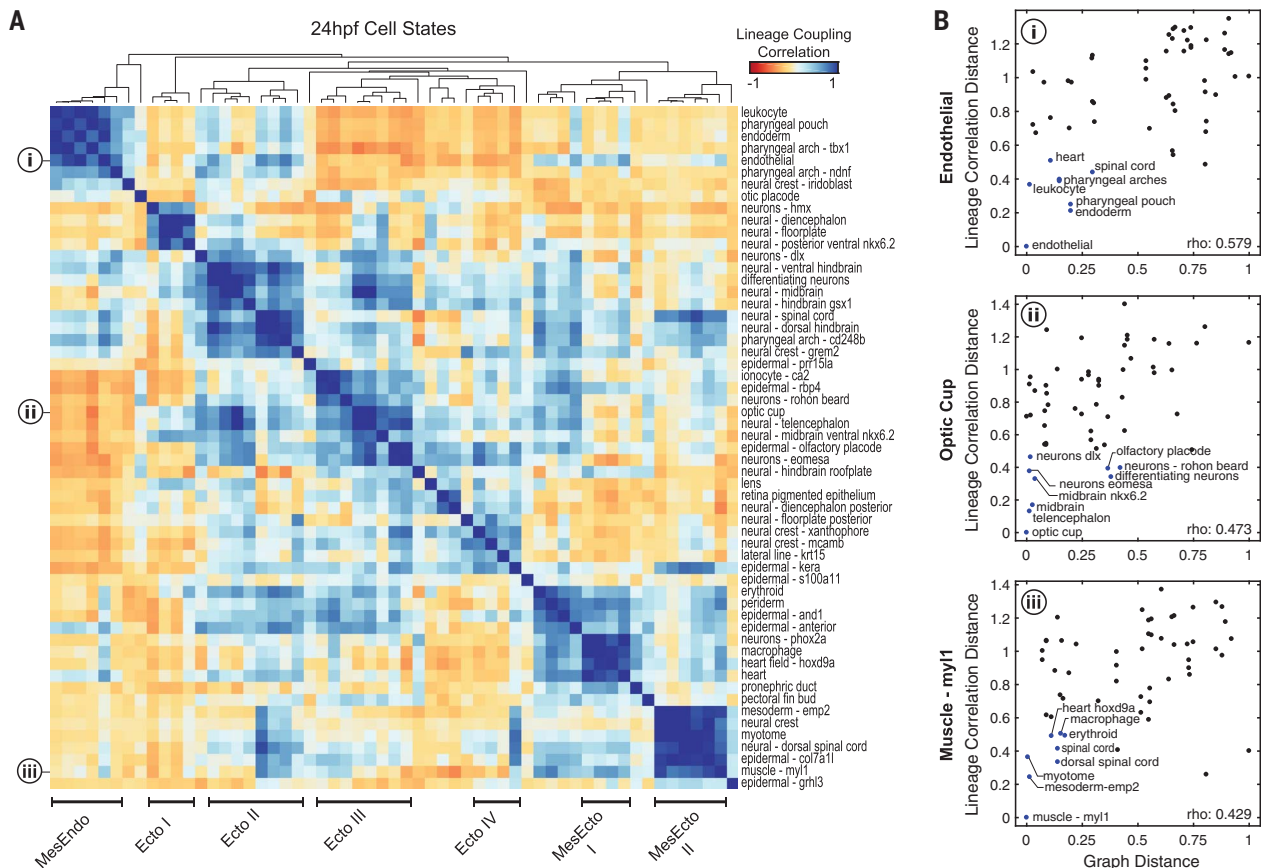
Fig. 5. TracerSeq reveals systematic relationships between cell lineage and cell state. (A) Heatmap of TracerSeq lineage coupling scores (see methods) between pairs of 24-hpf states, clustered by correlation distance and average linkage. Groups of states with similar lineage coupling signatures are annotated. **(B)** Quantitative relationships between lineage coupling correlation distances and scaled state tree diffusion distances for (i) endothelial, (ii) optic cup, and (iii) myl^+ muscle states (see also fig. S10, A to F).

the second pharyngeal arch state (ph.arch-cd248b) is clonally related to neural crest and posterior neural states (Fig. 5A and fig. S11, B to F). These data indicate that cells in the ph.arch-cd248b state diverged from a neural plate lineage and subsequently converged with other lateral plate–derived states. The ability of embryonic clones to undergo dramatic converging and diverging behaviors thus underscores a continued need for independent measurements of both cell state and lineage in the mapping of cell-fate hierarchies.

Robustness of cell-type transcriptional programs after a signaling perturbation

Single-cell maps of vertebrate development can, in principle, facilitate unbiased, systematic analyses of mutant phenotypes and disease states. We used scRNA-seq to analyze the mutant phenotype for *chordin*, a well-studied developmental gene that encodes a secreted bone morphogenetic protein (BMP) inhibitor expressed in the organizer and required for patterning the early dorsal-ventral axis (29–33). Disruption of *chordin* leads to changes in gross embryo morphology, with an expansion of ventral tissues and a reduction of dorsal tissues (30). The scRNA-seq method is well suited to address how all cell types in the embryo change in frequency, and

in gene expression, while also allowing detection of qualitatively new states, or combinations of states, if they occur.

We used CRISPR-Cas9 (34) to disrupt the *chordin* locus, resulting in highly penetrant clutches of mutant zebrafish embryos (fig. S12). We performed inDrops profiling on *chordin*-targeted and control embryos (*tyrosinase*-targeted, see methods) in a narrow time series corresponding to ~14 to 16 hpf (Fig. 6A). After sequencing, we classified each of the *chordin*- and control-targeted cells to reference cell clusters of the 14-hpf wild-type embryo (fig. S13 and methods) and tested for altered gene expression. We reasoned that a qualitatively new cell state, if formed as a result of the aberrant patterning, would manifest as widespread changes in gene expression after mutation, with a magnitude comparable to the differences between wild-type embryonic states. Applying this criterion, we found no evidence of a qualitatively new cell state after *chordin* depletion. Rather, the number of genes differentially expressed within states was modest compared to the differences defining the wild-type states of the 14-hpf embryo (Fig. 6B and fig. S14A). Moreover, a t-distributed stochastic neighbor embedding (tSNE) mapping of CRISPR-targeted cells (fig. S13, A to C) identified only a single clus-

ter solely occupied by *chordin*-mutant cells (fig. S13D), distinguished primarily by a heat shock–like transcriptional signature. This same stress signature was increased in multiple states in *chordin*-targeted embryos (fig. S14A).

We next tested whether *chordin* disruption led to changes in abundance of particular classified cell types. As expected, expansion of states corresponding to ventral tissues (for example, somitic mesoderm, epidermis, hatching gland, blood, and endothelial tissues) at the expense of dorsal tissues (for example, the neural plate and notochord) was observed (fig. S14, A and B) (30, 35). Additional features could be appreciated by projecting the CRISPR datasets directly onto the wild-type single-cell graph (Fig. 6, C and D). For example, a sharp boundary bisected the lateral plate mesoderm into two compartments of opposing *chordin* sensitivity, separating the heart and fin bud progenitor fields. Similar juxtaposed domains of opposing *chordin* sensitivity were evident in the axial mesoderm, partitioning notochord from hatching gland, and in the tailbud separating spinal cord from somitic mesoderm (Fig. 6D). Notably, each of these pairs of phenotypic domains appeared to be organized downstream of an inferred branch-point in the cell-state landscape. These domain

pairs, therefore, likely reflect binary fate choices that are tuned by BMP signaling in wild-type embryos.

In a final analysis, we searched for the putative identity of the cells responding to *chordin* in the tailbud, as this is the site showing the largest expansion (somitic mesoderm) and loss (spinal cord) after perturbation. In zebrafish, *chordin* is expressed in the embryonic shield, adaxial cells, posterior tailbud region, and also transiently in the neural plate (*36*). All of these expression patterns were confirmed in our single-cell graphs (fig. S15A). Furthermore, in contrast to its earlier expression in the shield, continued expression of *chordin* in the tailbud was distinct among a large panel of known BMP inhibitor genes (fig. S15A) and was tightly apposed by expression domains for multiple *bmp* transcripts (fig. S15B). These expression characteristics might explain the increased *chordin* sensitivity of posterior body regions. To examine this region in greater detail, we isolated a subgraph of tailbud and descendent cells. Consistent with previous studies, two cell-state trajectories branching from a common neuromesodermal-like *brachyury*+;*sox2*+ progenitor state were identified, each expressing markers of neural fates (*sox3*, *sox19a*, *pax6a*, and *neurog1*) or somitic fates (*tbx16*, *tbx6*, *tbx24*, *msgn1*, and *myod1*) (fig. S16, A to C) (*25*, *37–39*). Notably, the neural-mesodermal branchpoint coincided with the boundaries of both *chordin* expression and sensitivity (fig. S16, D and E). The *chordin*-expressing cells in this region of the single-cell graph exhibited a distinct expression profile (fig. S17), including a cadherin (*cdh11*), early neurogenic markers (*her3*, *her8a*, and *sox19a*), and several relatively uncharacterized genes (*gig2g*, *foxb1b*, and *foxb1a*). We hypothesize that these cells represent a key transition state, at which point tailbud cells initiate a posterior neurogenic program in a *chordin*-dependent manner.

Discussion

Our study demonstrates a graph-based approach for mapping whole-embryo developmental landscapes, over time, from scRNA-seq data. The graph was constructed with minimal assumptions about development and describes individual cell states transitioning from pluripotent blastomeres to a large array of cell types and tissues during the first day of zebrafish embryogenesis. This dataset can now be mined to identify temporal and tissue associations for any gene, cell type, or biological process of interest. As with genome annotation efforts over the years, we expect that the annotation of identified cell states may undergo refinement with community input.

As single-cell atlases and landscapes of embryo development become routinely available, one is challenged to reconsider the relationship between a cell lineage (by definition, a tree) and the considerably more topologically complex gene expression landscape through which these cells traverse. Using TracerSeq, we confirmed that differentiating cells of the zebrafish embryo do not invariantly follow treelike hierarchies. In-

stead, we observed both widespread convergence in cell states for clonally distant cells and instances in which clonally related cells diverged into distant states. Non-treelike convergence of cell states could be explained by the differentiation of well-separated spatial domains of the embryo into the same basic cell types (for example, along the anterior-posterior axis), whereas divergence could involve mechanisms such as asymmetric cell division or exposure to spatially varying signals (*40*). We anticipate that the synthesis of single-cell lineage and transcriptome information will continue to be crucial for deciphering how cells

traverse state trajectories with complex topologies (for example, loops or continua).

Single-cell mapping of genetic perturbation data presents a powerful framework for identifying regulatory features of a developmental landscape. After disruption of *chordin*, which encodes a BMP inhibitor, we showed that the defining transcriptional features of the landscape remained mostly unchanged, yet cell-state abundances could be dramatically and reciprocally altered, as if the landscape were "tilted" but cell fates remain canalized. Future systematic mapping of signaling perturbations could be
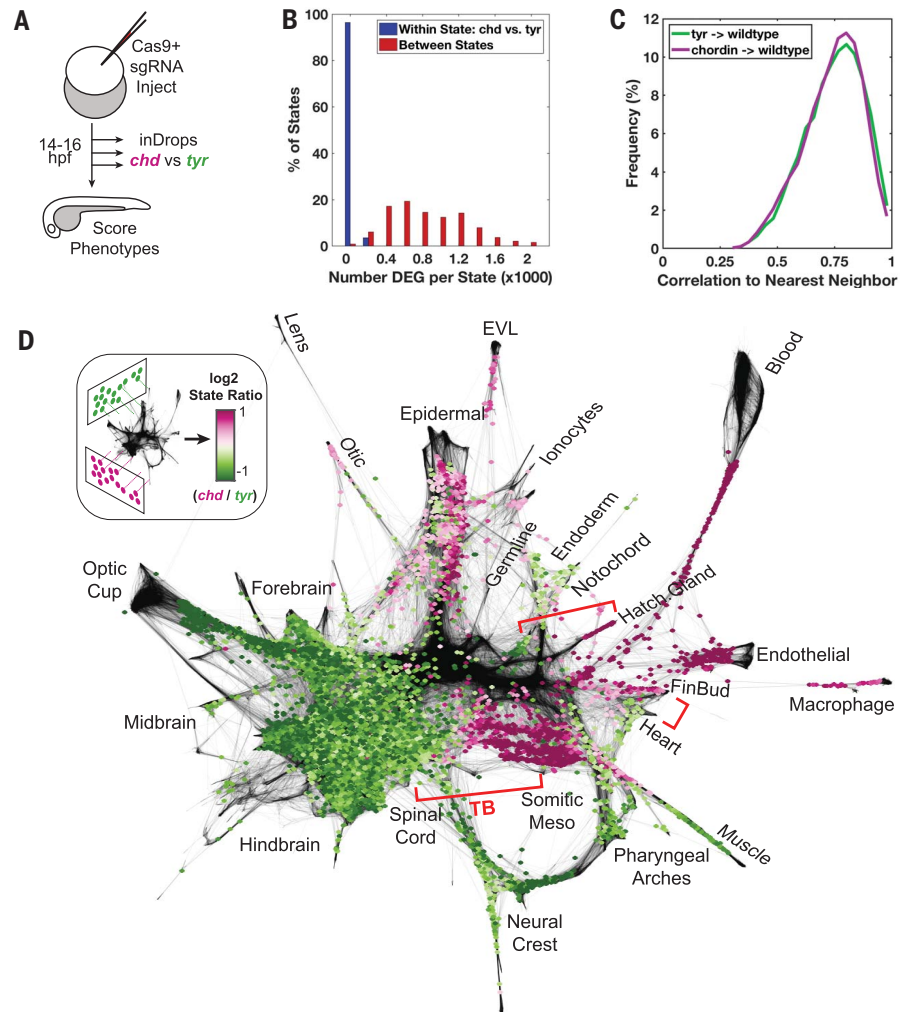

Fig. 6. Regulatory features of the developmental landscape identified by genetic perturbation.
(**A**) Overview of the CRISPR experiment. Three pairs of *chordin-* and *tyrosinase* (control)−targeted samples were prepared and processed by inDrops at ~14 to 16 hpf. (**B**) Histogram depicting numbers of differentially expressed genes (DEGs) identified in *chordin* versus control (*tyrosinase*) cells for each state (blue bars), compared to DEG numbers when comparing between all state pairs (red bars). DEGs were identified by Wilcoxon rank sum test (adjusted P value < 0.01, absolute $\log_2$ fold change > 1, average expression > 25 transcripts per million). (**C**) Histogram of Pearson correlation similarities (after PCA projection) between each *chordin* or *tyrosinase* cell and its nearest neighbor from 10-, 14-, and 18-hpf wild-type datasets (see methods). (**D**) $\log_2$ ratios of cell states with significant differential abundance (false discovery rate < 0.25) in the *chordin* versus *tyrosinase* samples. Purple and green regions correspond to wild-type cell states that are over- or underrepresented in the *chordin* mutant, respectively. Adjacent graph domains with opposing *chordin* sensitivity are highlighted by brackets. TB, tailbud region (see *cdx4* expression in fig. S3).

used to reveal the complete signaling logic of the embryo, as cells are specified toward their final fates. Together, these studies demonstrate the power, modularity, and quantitative benefits of unbiased scRNA-seq–based interrogations of embryonic development. We anticipate that similar large-scale datasets will facilitate explorations of additional developmental stages, tissues, and species.

REFERENCES AND NOTES

1. A. F. Schier, W. S. Talbot, *Annu. Rev. Genet.* **39**, 561–613 (2005).
2. T. Hashimshony *et al.*, *Genome Biol.* **17**, 77 (2016).
3. S. Islam *et al.*, *Nat. Methods* **11**, 163–166 (2014).
4. A. M. Klein *et al.*, *Cell* **161**, 1187–1201 (2015).
5. E. Z. Macosko *et al.*, *Cell* **161**, 1202–1214 (2015).
6. L. Mazutis *et al.*, *Nat. Protoc.* **8**, 870–891 (2013).
7. R. Zilionis *et al.*, *Nat. Protoc.* **12**, 44–73 (2017).
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 *et al.*, *Nat. Biotechnol.* **32**, 381–386 (2014).
10. S. C. Bendall *et al.*, *Cell* **157**, 714–725 (2014).
11. J. Shin *et al.*, *Cell Stem Cell* **17**, 360–372 (2015).
12. L. Haghverdi, M. Büttner, F. A. Wolf, F. Buettner, F. J. Theis, *Nat. Methods* **13**, 845–848 (2016).
13. C. B. Kimmel, R. M. Warga, T. F. Schilling, *Development* **108**, 581–594 (1990).
14. A. E. Melby, R. M. Warga, C. B. Kimmel, *Development* **122**, 2225–2237 (1996).
15. R. M. Warga, C. Nüsslein-Volhard, *Development* **126**, 827–838 (1999).
16. K. Woo, S. E. Fraser, *Development* **121**, 2595–2609 (1995).
17. K. Kawakami, *Genome Biol.* **8** (suppl. 1), S7 (2007).
18. D. G. Gibson *et al.*, *Nat. Methods* **6**, 343–345 (2009).
19. A. McKenna *et al.*, *Science* **353**, aaf7907 (2016).
20. J. P. Junker *et al.*, bioRxiv 056499 [Preprint]. 4 January 2017.
21. R. M. Warga, D. A. Kane, R. K. Ho, *Dev. Cell* **16**, 744–755 (2009).
22. K. E. Whitlock, M. Westerfield, *Development* **127**, 3645–3653 (2000).
23. E. Tzouanacou, A. Wegener, F. J. Wymeersch, V. Wilson, J.-F. Nicolas, *Dev. Cell* **17**, 365–376 (2009).
24. R. L. Davis, M. W. Kirschner, *Development* **127**, 255–267 (2000).
25. J. P. Kanki, R. K. Ho, *Development* **124**, 881–893 (1997).
26. N. M. Le Douarin, E. Dupin, *Curr. Opin. Genet. Dev.* **13**, 529–536 (2003).
27. N. M. Le Douarin, S. Creuzet, G. Couly, E. Dupin, *Development* **131**, 4637–4650 (2004).
28. C. S. Le Lièvre, N. M. Le Douarin, *J. Embryol. Exp. Morphol.* **34**, 125–154 (1975).
29. Y. Sasai *et al.*, *Cell* **79**, 779–790 (1994).
30. M. Hammerschmidt *et al.*, *Development* **123**, 95–102 (1996).
31. S. Schulte-Merker, K. J. Lee, A. P. McMahon, M. Hammerschmidt, *Nature* **387**, 862–863 (1997).
32. Y. Sasal, B. Lu, H. Steinbelsser, E. M. De Robertis, *Nature* **378**, 419 (1995).
33. S. Piccolo, Y. Sasai, B. Lu, E. M. De Robertis, *Cell* **86**, 589–598 (1996).
34. J. A. Gagnon *et al.*, *PLOS ONE* **9**, e98186 (2014).
35. A. Y. Leung *et al.*, *Dev. Biol.* **277**, 235–254 (2005).
36. V. E. Miller-Bertoglio, A. Fisher, A. Sánchez, M. C. Mullins, M. E. Halpern, *Dev. Biol.* **192**, 537–550 (1997).
37. R. H. Row, D. Kimelman, *Dev. Biol.* **329**, 55–63 (2009).
38. R. H. Row, S. R. Tsotras, H. Goto, B. L. Martin, *Development* **143**, 244–254 (2016).
39. M. Gouti *et al.*, *Dev. Cell* **41**, 243–261.e7 (2017).
40. P. Gönczy, *Nat. Rev. Mol. Cell Biol.* **9**, 355–366 (2008).

Science

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 and Allon M. Klein

Mapping the vertebrate developmental landscape

As embryos develop, numerous cell types with distinct functions and morphologies arise from pluripotent cells. Three research groups have used single-cell RNA sequencing to analyze the transcriptional changes accompanying development of vertebrate embryos (see the Perspective by Harland). Wagner *et al.* sequenced the transcriptomes of more than 90,000 cells throughout zebrafish development to reveal how cells differentiate during axis patterning, germ layer formation, and early organogenesis. Farrell *et al.* profiled the transcriptomes of tens of thousands of embryonic cells and applied a computational approach to construct a branching tree describing the transcriptional trajectories that lead to 25 distinct zebrafish cell types. The branching tree revealed how cells change their gene expression as they become more and more specialized. Briggs *et al.* examined whole frog embryos, spanning zygotic genome activation through early organogenesis, to map cell states and differentiation across all cell lineages over time. These data and approaches pave the way for the comprehensive reconstruction of transcriptional trajectories during development.
Science, this issue p. 981, p. eaar3131, p. eaar5780; see also p. 967

ARTICLE TOOLS	http://science.sciencemag.org/content/360/6392/981
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2018/04/25/science.aar4362.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/360/6387/367.full http://science.sciencemag.org/content/sci/360/6392/eaar3131.full http://science.sciencemag.org/content/sci/360/6392/eaar5780.full http://science.sciencemag.org/content/sci/360/6392/967.full
REFERENCES	This article cites 53 articles, 16 of which you can access for free http://science.sciencemag.org/content/360/6392/981#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service