

RESEARCH ARTICLE SUMMARY

DEVELOPMENT

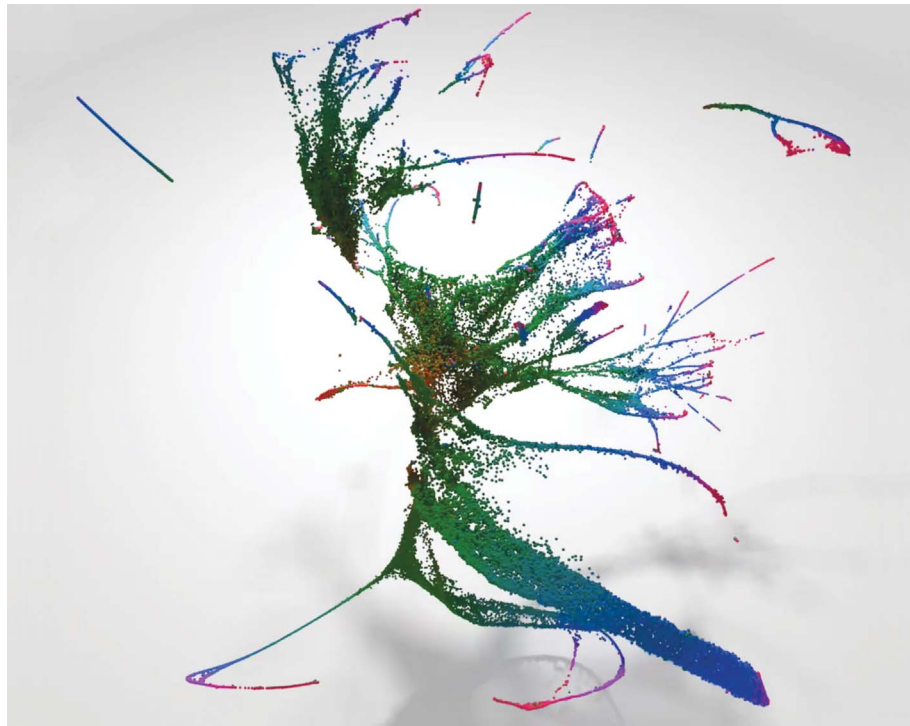
A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution

Jonathan S. Packer*, Qin Zhu*, Chau Huynh, Priya Sivaramakrishnan, Elicia Preston, Hannah Dueck, Derek Stefanik, Kai Tan, Cole Trapnell, Junhyong Kim†, Robert H. Waterston‡, John I. Murray†

INTRODUCTION: During development, a single-cell zygote undergoes repeated cell divisions to produce an embryo that contains many distinct cell types. This sequence of cell divisions is called the organism's cell lineage. Each cell in the lineage expresses a different set of genes in various quantities (the cell's transcriptome), thus directing cells to differentiate into specific cell types over time. Identifying these transcriptome changes and understanding how they regulate cell type specification are fundamental challenges in biology. Advances in methods that assay the mRNA content of single cells [e.g., single-cell RNA sequencing (sc-RNA-seq)] have made it possible to

directly measure these gene expression patterns for each cell at a whole-organism scale.

RATIONALE: The nematode worm *Caenorhabditis elegans* has only 558 cells at hatching. Despite its limited cell number, however, *C. elegans* contains a wide diversity of cell types and a complex anatomy. Its cell lineage is invariant—every individual *C. elegans* animal is produced by the exact same sequence of cell divisions—and this lineage has been fully mapped. Together, these features make it an ideal animal in which to comprehensively describe the gene expression patterns of each cell and to define the transcriptome dynamics during cell differentiation and development.



UMAP visualization of 86,024 single-cell transcriptomes from *C. elegans* embryos.

Single-cell transcriptomes were projected into a 3D space by using the uniform manifold approximation and projection (UMAP) algorithm. Each point represents a cell. The projection reconstructs developmental trajectories of *C. elegans* cell types, spanning from mid-gastrulation (orange and green) to terminal differentiation (blue, purple, and pink).

RESULTS: We profiled the transcriptomes of 86,024 single cells from *C. elegans* embryos at developmental stages ranging from gastrulation to terminal cell differentiation. Using computational methods, gene expression patterns from the literature, and gene expression data obtained from three-dimensional (3D) movies of fluorescent reporter lines, we mapped each single-cell transcriptome to its corresponding position in the known *C. elegans* cell

ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aax1971>

lineage tree. In total, we identified 502 distinct terminal and preterminal cell types, which correspond to 1068 individual branches of the lineage tree. We computed a transcriptional profile for each detected cell type and determined the gene expression differences between mother and daughter cells, and between sister cells, for >200 cell division events in the lineage.

Analyzing these data, we find that:

1) A cell's lineage history and its transcriptome are transiently correlated. This correlation increases from middle to late gastrulation, then falls substantially as cells adopt their terminal fates.

2) Genes that distinguish sister cells are often first coexpressed in the parent and then selectively retained in one daughter but not the other. This phenomenon, known as “multilineage priming,” is notably prevalent throughout the *C. elegans* lineage.

3) Most distinct lineages that produce the same anatomical cell type converge to a homogeneous transcriptomic state, with little or no residual signature of their lineage identity.

4) In many cases, purely computational reconstruction of developmental trajectories from the single-cell transcriptomic data does not accurately reproduce the known cell lineage. Marker genes known to be expressed in specific lineages were critical for correct annotation. This is particularly evident for lineages in which gene expression changes rapidly.

CONCLUSION: Our dataset defines the succession of gene expression changes associated with almost every cell division in an animal's embryonic cell lineage. It provides an extensive resource that will guide future investigations of gene regulation and cell fate decisions in *C. elegans*. It can also serve as a benchmark dataset that will facilitate rigorous evaluation of computational methods for reconstructing cell lineages from sc-RNA-seq data. ■

The list of author affiliations is available in the full article online.

*These authors contributed equally to this work.

†Corresponding author. Email: junhyong@sas.upenn.edu (J.K.); watersto@uw.edu (R.H.W.); jnmurr@pennmedicine.upenn.edu (J.I.M.)

Cite this article as J. S. Packer et al., *Science* 365, eaax1971 (2019). DOI: 10.1126/science.aax1971

RESEARCH ARTICLE

DEVELOPMENT

A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution

Jonathan S. Packer^{1*}, Qin Zhu^{2*}, Chau Huynh¹, Priya Sivaramakrishnan³, Elicia Preston³, Hannah Dueck^{3†}, Derek Stefanik⁴, Kai Tan^{3,5,6,7}, Cole Trapnell¹, Junhyong Kim^{4‡}, Robert H. Waterston^{1‡}, John I. Murray^{3‡}

Caenorhabditis elegans is an animal with few cells but a wide diversity of cell types. In this study, we characterize the molecular basis for their specification by profiling the transcriptomes of 86,024 single embryonic cells. We identify 502 terminal and preterminal cell types, mapping most single-cell transcriptomes to their exact position in *C. elegans*' invariant lineage. Using these annotations, we find that (i) the correlation between a cell's lineage and its transcriptome increases from middle to late gastrulation, then falls substantially as cells in the nervous system and pharynx adopt their terminal fates; (ii) multilineage priming contributes to the differentiation of sister cells at dozens of lineage branches; and (iii) most distinct lineages that produce the same anatomical cell type converge to a homogenous transcriptomic state.

To understand how cell fates are specified during development, it is essential to know the temporal sequence of gene expression in cells during their trajectories from early uncommitted precursors to differentiated terminal cell types. Gene expression patterns near branch points in these developmental trajectories can help identify candidate regulators of cell fate decisions (1). Single-cell RNA sequencing (sc-RNA-seq) has made it possible to obtain comprehensive measurements of gene expression in whole animals (2–7) and embryos (8–14). sc-RNA-seq profiling of multiple developmental stages in a time series can be particularly informative, as algorithms can use the data to reconstruct the developmental trajectories followed by specific cell types. However, confounding factors can generate misleading trajectories. For example, progenitor cell populations with distinct lineage origins may be conflated if their transcriptomes are too similar; and abrupt changes in gene expression can result in discontinuous

trajectories. Thus, information from independent assays is necessary to conclusively validate an inferred trajectory as an accurate model of development.

In this work, we comprehensively reconstruct and validate developmental trajectories for the embryo of the nematode worm *Caenorhabditis elegans*. *C. elegans* develops through a known and invariant cell lineage from the fertilized egg to an adult hermaphrodite with 959 somatic cells (15, 16), which creates the potential for a truly comprehensive understanding of its development. Using sc-RNA-seq, the known *C. elegans* lineage, and imaging of fluorescent reporter genes (17, 18), we produce a lineage-resolved single-cell atlas of embryonic development that includes trajectories for most individual cells in the organism. Our atlas expands on previous studies of the earliest embryonic blastomeres (8, 19), covering 87% of embryonic lineage branches.

We use this dataset to quantitatively model the relationship between the cell lineage and the temporal dynamics of gene expression. We find that during gastrulation, lineage distance between cells is a strong predictor of transcriptome dissimilarity. The strength of this correlation increases from the middle to the end of gastrulation. After gastrulation, expression patterns of closely related cells diverge as the cells adopt their terminal fates. Body wall muscle (BWM), hypodermis, and the intestine are exceptions to this trend, as they are produced by semiclinal lineage clades that maintain within-clade transcriptomic similarity. In the ectoderm, the final two rounds of cell division produce distinct neuron and glia cell types, which rapidly differentiate, often resulting in discontinuities in computa-

tional reconstructions of their developmental trajectories. In several cases, the transcriptomes of distant lineages converge as the cells adopt the same terminal fate, and they simultaneously diverge from their close relatives in the lineage.

Our ability to reconstruct these complex gene expression dynamics highlights both the utility of the known *C. elegans* lineage and the challenges presented when sc-RNA-seq is used in an attempt to reconstruct the lineages of other organisms.

Single-cell RNA-seq of *C. elegans* embryos

We sequenced the transcriptomes of single cells from *C. elegans* embryos with the 10X Genomics platform. We assayed loosely synchronized embryos enriched for preterminal cells as well as embryos that had been allowed to develop for ~300, ~400, and ~500 min after the first cleavage of the fertilized egg. We processed the datasets with the Monocle software package (20). After quality control, the final integrated dataset contained 86,024 single cells, representing a >60-fold oversampling of the 1341 branches in the *C. elegans* embryonic lineage.

We estimated the embryo stage of each cell by comparing its expression profile with a high-resolution whole-embryo RNA-seq time series (21) (fig. S1). We then visualized the data with the uniform manifold approximation and projection (UMAP) (22, 23) algorithm, which projects the data into a low-dimensional space and is well suited for data with complex branching structures (23). We found that trajectories (24) in the UMAP projection reflect a smooth progression of embryo time (Fig. 1A), with cells collected from later time points usually occupying more-peripheral positions (Fig. 1B). Unique transcripts per cell, as estimated with unique molecular identifiers (UMIs), decreased with increasing embryo time throughout the period of embryonic cell division, consistent with decreasing physical cell size (fig. S2). These observations suggest that UMAP trajectories corresponded to developmental progression and that embryo time estimates are a reasonable proxy for developmental stage for most cells. Approximately 75% of the cells recovered (64,384 cells) were from embryos spanning 210 to 510 min after first cleavage, corresponding to mid-gastrulation (~190-cell stage) to terminal differentiation (threefold stage of development) (Fig. 1C); however, cells were also recovered from earlier embryos (<210 min, 9886 cells) and later embryos (>510 min, 11,754 cells).

We clustered cells in the UMAP using the Louvain algorithm (25) and annotated clusters with cell type identities using marker genes from the literature on *C. elegans* gene expression (26). Markers used for each annotation are listed in table S1. The global UMAP arranges cells into a central group of progenitor cells and branches corresponding to eight major tissues (Fig. 1A and fig. S3): muscle and mesoderm, epidermis, pharynx, ciliated neurons, nonciliated neurons, glia and excretory cells, intestine, and germ line. Although some individual cell types were identifiable in this global UMAP, many were not, especially

¹Department of Genome Sciences, University of Washington, Seattle, WA, USA. ²Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA, USA. ³Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁴Department of Biology, University of Pennsylvania, Philadelphia, PA, USA. ⁵Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA, USA. ⁶Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁷Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

*These authors contributed equally to this work. †Present address: National Cancer Institute, Bethesda, MD, USA.

‡Corresponding author. Email: junhyong@sas.upenn.edu (J.K.); watersto@uw.edu (R.H.W.); jmurr@perelman.upenn.edu (J.I.M.)

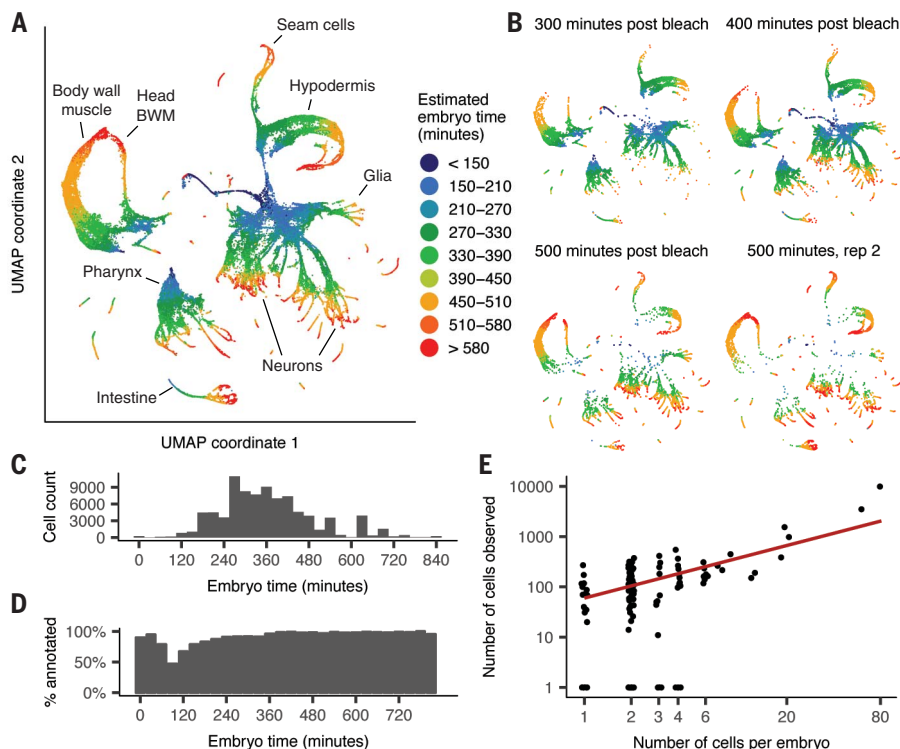


Fig. 1. UMAP projection shows tissues and developmental trajectories in *C. elegans* embryogenesis.

(A) UMAP projection of the 81,286 cells from our sc-RNA-seq dataset that passed our initial quality control step. This UMAP does not include 4738 additional cells that were initially filtered but were later whitelisted and included in downstream analyses. Color indicates the age of the embryo from which a cell was produced, estimated from correlation to a whole-embryo RNA-seq time series (21) and measured in minutes after an embryo's first cell cleavage. (B) Positions of cells from four samples of synchronized embryos on the UMAP plot. (C) Histogram of estimated embryo time for all cells in the dataset. (D) Bar plot showing the percentage of cells that we were able to assign to a terminal cell type or preterminal lineage for each embryo time bin. (E) Scatter plot showing correlation of the number of cells of a given anatomical cell class in a single embryo (x axis, log scale) with the number of cells recovered in our data (y axis, log scale). Each point corresponds to a cell class. Only cells with estimated embryo time ≥ 390 min are included in the counts (many earlier cells are still dividing). The red line is a linear fit, excluding points with $y = 0$.

progenitor lineages. To gain resolution, we hierarchically created separate UMAPs of each tissue (figs. S4 to S13). These “sub-UMAPs” enabled better resolution of specific cell types, allowing us to make extensive, fine-grained annotations.

A combination of marker genes, lineage assignments, and developmental time allowed us to locate 112 specific terminal anatomical cell types, including every lineage input to BWM, every distinct subtype of pharyngeal muscle (pm1-2, pm3-5, pm6, pm7, and pm8) and hypodermis (hyp1-2, hyp3, hyp4-6, hyp7, hyp8-11, seam, and P cells), and every non-neuronal cell type in the mesoderm. We identified 69 of 82 nonpharyngeal neuron types and 9 of 12 glial cell types present in the embryo (table S2). We could not identify 12 of 14 pharyngeal neuron types. A cluster corresponding to the most differentiated pm3-5 pharyngeal muscle cells had a low level of expression of neuron-specific genes, suggesting that we failed to dissociate the neurons that innervate these muscles in late embryos.

We successfully annotated 93% of cells in our dataset with a cell type (for terminal cells) or a cell lineage (for progenitor cells, discussed below) (Fig. 1D). The number of cells annotated for each cell type was variable but roughly fit the expectation on the basis of the number of cells of that type present in a single embryo (Fig. 1E; $r = 0.64$, $P \leq 2.4 \times 10^{-13}$, t test).

Annotation of progenitor lineages

The structure of the global and single-tissue UMAPs was dominated by trajectories of terminal cell differentiation. We hypothesized that closely related lineages could be better resolved by separately analyzing progenitor cells before terminal differentiation. Thus, we ran UMAP on subsets of cells with ≤ 150 -, 250-, or 300-min embryo times and found branching patterns that reflect lineage identities (Fig. 2 and figs. S14 to S16). Intestine and germline cells commit to their terminal fates very early and have very divergent expression that distorts the projections,

so they were removed and analyzed separately (figs. S7 and S12). The 300-min UMAP contained several large quasi-connected groups corresponding approximately to major founding lineages, roughly organized by the major fates produced by each founder cell lineage (MS muscle, MS pharynx, C and D muscle, and AB-derived lineages that produce pharynx, neurons and glia, or hypodermis). We were able to resolve additional details by recursively making sub-UMAP projections of these cell subsets.

To annotate progenitor lineages, we exploited lineage marker genes from the literature and the Expression Patterns in *Caenorhabditis* (EpiC) database, which contains single-cell-resolution expression profiles extracted by cell-tracking software from confocal movies of *C. elegans* embryos expressing fluorescent reporters (17). In addition to the 180 previously described patterns (17, 27), we have collected movies for 71 additional genes, increasing the total number of patterns in EpiC to 251 genes (table S3). We annotated branches with lineage identities between the 28-cell and 350-cell stages by finding genes that were differentially expressed both between sister lineages in the EpiC data and between branches of the sub-UMAP trajectories in a concordant manner (Fig. 2, figs. S14 to S16, and tables S4 and S5). For example, expression of *celh-51* is restricted to the MS (mesoderm-producing) lineage (28), allowing us to label the single group of *celh-51*(+) cells in 150-min UMAP as part of the MS lineage (Fig. 2, A and B). Within this lineage, we used expression of *pha-4* to annotate the anterior granddaughters of MS (MSaa and MSpa) and *hnd-1* to annotate the posterior granddaughters (MSap and MSpP) (Fig. 2C). We applied this same logic iteratively across the different UMAPs and lineage marker genes to annotate each branch with its lineage identity (table S4).

In most cases, branches in the progenitor lineage UMAPs corresponded directly to sister cells in the lineage (Fig. 2, D and E), but some branches were unclear or misleading, and marker gene expression was critical to annotate lineages correctly. For example, ABpxpaaaa and ABpxpaapa are cousin lineages but appear to branch as sisters in the UMAP trajectory, and the same is true for their sisters (ABpxpaap and ABpxpaapp) (Fig. 2D). In other cases, such as the ABpxppap lineage (Fig. 2D), marker gene combinations were required to annotate lineages that were not contiguous with their parent or sister lineages in the UMAP. These misleading branches demonstrate the importance of having independent expression or lineage data to correctly interpret trajectories visualized in low-dimensional embeddings of sc-RNA-seq data.

To complete our annotations, we used UMAPs of selected subsets of cells with embryo time ≤ 350 or 400 min to reconstruct trajectories leading from the grandparents and parents of terminal cells to their terminal descendants (fig. S17). Most terminal cell types were thus identified by two methods: first using marker genes for the differentiated cell type, and second by following UMAP trajectories from the cell's progenitors. Notably,

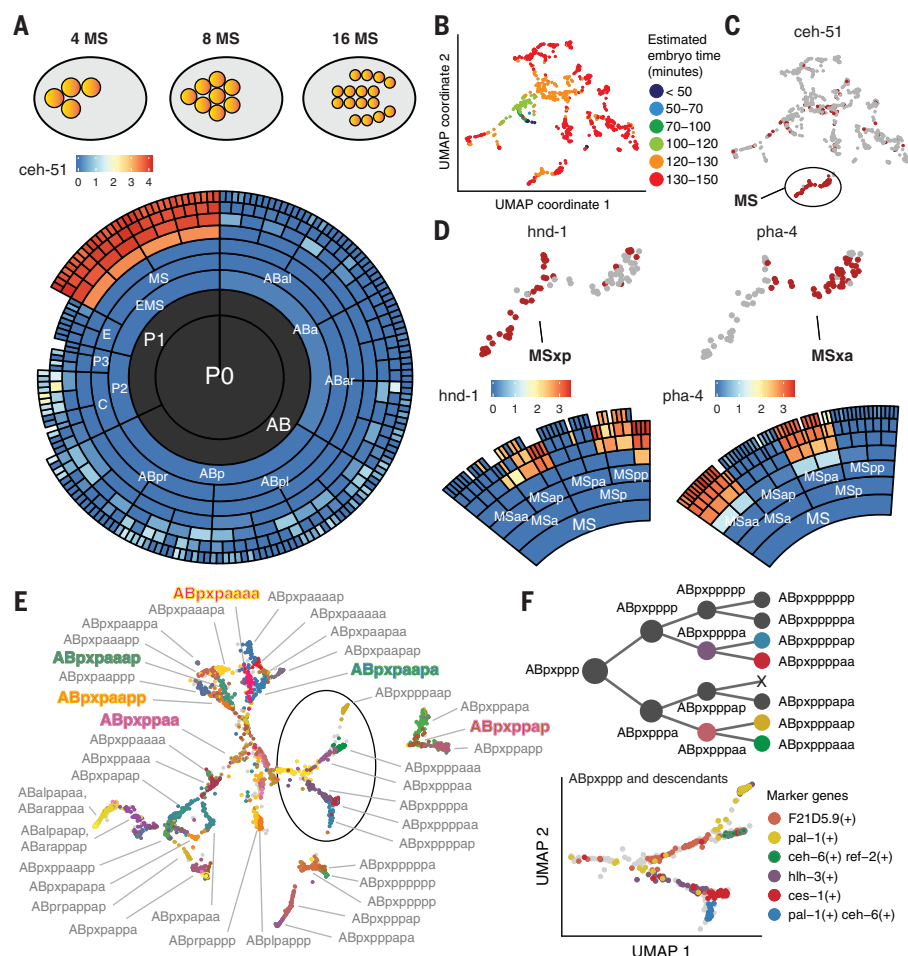


Fig. 2. Annotation of the early lineage. (A) (Top) Diagram showing the position of early mesoderm (MS lineage) cells marked by expression of *ceH-51*. (Bottom) Lineage radiograph showing the average fluorescent intensity (log₁₀ scaled) of a CEH-51::GFP protein fusion measured by live imaging. The inner rings show the generation of the founder cells, AB (which produces almost exclusively ectoderm and pharynx), MS (mesoderm and pharynx), C (muscle and ectoderm) and P3, which gives rise to P4 (germ line) and D (muscle). Daughter cells are named by their relative positions at mitosis (e.g., ABa is the anterior daughter of AB, ABal is left daughter of ABa). (B) UMAP projection of 926 early-stage cells (estimated embryo time ≤150 min), colored by embryo time. E lineage and germline cells are excluded and shown separately in figs. S7 and S12, as they differentiate early relative to other lineages. (C) Same UMAP as in (B), colored by *ceH-51* expression (red indicates cells with nonzero UMIs for *ceH-51*). (D) Expression of *hnd-1* and *pha-4* measured by sc-RNA-seq (UMAP) and live imaging of GFP protein fusions (radiograph). (E) Cropped section of a UMAP of 8083 neuron, glia, and rectal progenitor cells with embryo time ≤250 min (fig. S15). This plot shows the section of that UMAP that corresponds to the 3233 cells from the ABxpp ectodermal lineage (“ABxpp” is shorthand for two symmetric lineages, ABplp and ABprp). Colored bold annotations highlight specific lineages that are discussed in the text. (F) Lineage tree for the ABxpppp sublineage, highlighting cells that are present in the circled section of (E). The (co)expression pattern of marker genes identifies branches in the UMAP that correspond to specific ABxpppp descendants. Additional ABxpppp descendants not shown in this panel are annotated in (E), below the circled section.

in all cases, the cell type predictions of these two mostly independent methods were concordant.

A near-complete atlas of the embryonic transcriptome

In total, we annotated 502 distinct cell lineages. Most lineage annotations correspond to a symmetric pair of cells, with the exception of some

terminal cell types in which 3 to 18 cells converge to a homogenous transcriptomic state and could not be further resolved. Our annotations account for 1068 of 1228 individual branches in the *C. elegans* embryonic lineage (fig. S18), excluding the 113 branches that lead to programmed cell death. Combined with the dataset of Tintori *et al.* (8), which profiles the 1- to 16-cell

stages, we now have a near-complete molecular atlas of *C. elegans* embryogenesis.

The lineages included in our atlas partially overlap with the Tintori *et al.* dataset (8) at the 16-cell stage. Gene expression profiles for lineages annotated in both datasets were concordant (fig. S19). Additionally, gene expression profiles for terminal cells in our data were concordant with previously published microarray data (29) (fig. S20).

In table S6, we provide a map of anatomical cell names to annotations defined in this study. In tables S7 and S8, we provide aggregate gene expression profiles for each terminal cell type (binned by embryo time) and each cell lineage annotation. We use bootstrap resampling to estimate a confidence interval for the expression level of each gene in each cell type. In tables S9 to S11, we provide lists of differentially expressed genes between all pairs of sister lineages and all pairs of parent versus daughter lineages within our annotations. Lastly, we systematically re-annotated our previous data from the L2 stage (2), identifying 118 cell types (more than twice as many as reported in the initial publication). Tables S12 to S14 list marker genes, annotation statistics, and expression profiles for the L2 data.

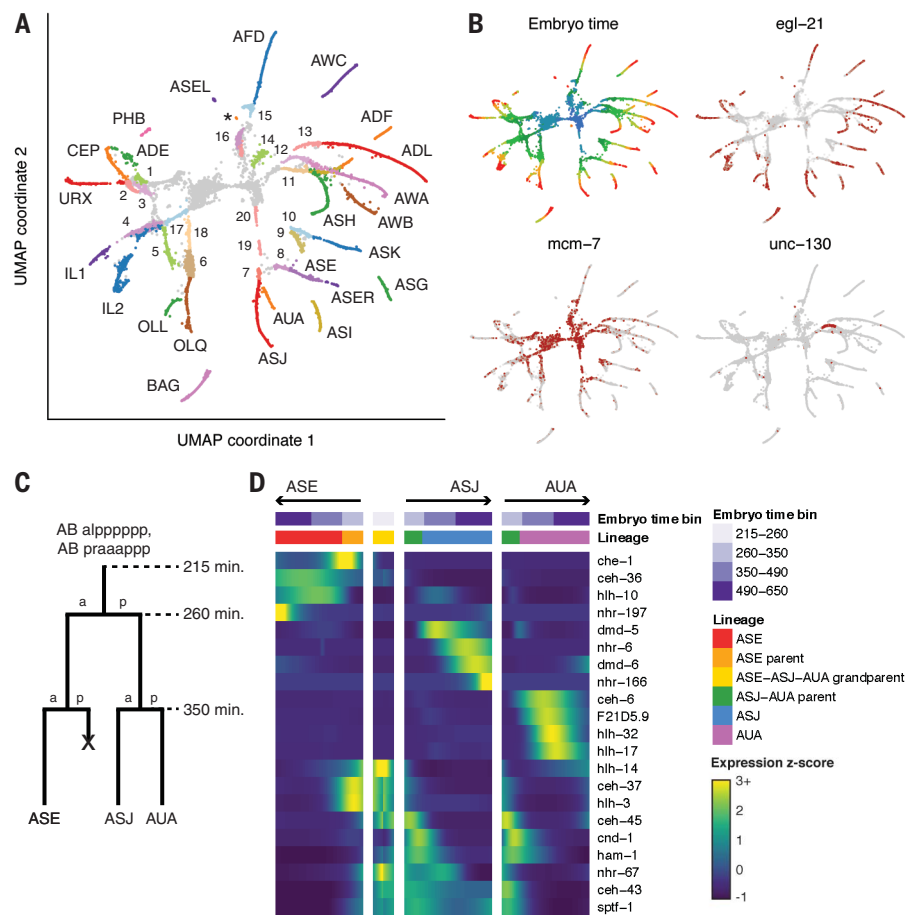
Bifurcating cell fates and multilineage priming

Developmental trajectories in which a parent cell divides to produce two terminal daughter cells of different cell types are a basic type of cell fate decision. Bifurcations like these are common in neuronal lineages in *C. elegans*, such as those that produce ciliated neurons. To examine the molecular basis for such developmental decisions, we used recursive UMAP projections of ciliated neurons (Fig. 3A) to identify developmental trajectories for all but one of the 22 ciliated neuron types and their parents, missing only the PHA phasmid neurons. The distinction between neuroblasts and terminal neurons was supported by embryo time estimates consistent with terminal cell division times (30), by the expression patterns of cell cycle-associated genes and transcription factors (TFs) (Fig. 3B), and by the structure of the UMAP projection. A 3D version of the UMAP featured better continuity for several trajectories, including those connecting the ASG-AWA, ADF-AWB, and ASJ-AUA neuroblasts with their daughter cells, as well as the branching of the laterally asymmetric left and right ASE neurons (fig. S21).

To identify potential regulators of cell fate decisions, we identified genes that were differentially expressed between the branches of each bifurcating ciliated neuron lineage (table S9). The lineage of the ASE, ASJ, and AUA neurons (spanning embryo time ~215 to 650 min) serves as a representative example (Fig. 3C). Approximately three or four TFs are specific to each terminal neuron type in this lineage (Fig. 3D). Similar numbers of branch-specific TFs were observed for other lineage bifurcations (fig. S22). Beyond these simple cases, we also found several TFs that were expressed in a parent cell and had

Fig. 3. Developmental trajectories of ciliated neurons.

(A) UMAP of 10,740 ciliated neurons and precursors. Colors correspond to cell identity. Text labels indicate terminal cell types. Numbers 1 to 16 indicate parents of ADE-ADA (1), CEP-URX (2), PHB-HSN (3), IL1 (4), OLL (5), OLQ (6), ASJ-AUA (7), ASE (8), ASI (9), ASK (10), ADF-AWB (11), ASG-AWA (12), ADL (13), ASH-RIB (14), AFD-RMD (15), and AWC-SAA (purple; 16) and BAG-SMD (red; 16). Numbers 4 to 6, 8 to 10, and 13 are listed as parents of only one cell type, as the sister cells die. Numbers 17 to 20 indicate grandparents of IL1 (= IL2 parent; 17) OLQ-URY (18), and ASE-ASJ-AUA (19 and 20). Differentiated PHA was not conclusively identified but may cocluster with PHB. The parent of PHA is not present in this UMAP but was located separately within the area annotated as “rectal cells” in the UMAP in fig. S3. The tiny cluster labeled with an asterisk is putatively AWC-ON on the basis of *srt-28* expression. **(B)** UMAP plot colored by embryo time (colors matched to Fig. 1A) and gene expression (red indicates nonzero reads for the listed gene). *egl-21* codes for an enzyme that is essential for processing neuropeptides (49). Its expression is used as a proxy for the onset of neuron differentiation. *mcm-7* codes for a DNA replication licensing factor. Loss of *mcm-7* expression in each UMAP trajectory approximately marks the boundary between neuroblasts and terminal cells. *unc-130* is known to be expressed in the ASG-AWA neuroblast but neither terminal cell (50). **(C)** Cartoon illustrating the lineage of the ASE, ASJ, and AUA neurons. Lowercase letters indicate the relative position of the daughters after division (anterior, posterior, left, right). **(D)** Heatmap showing patterns of differential TF expression associated with branches in the ASE-ASJ-AUA lineage. Expression values are log-transformed, then centered and scaled by standard deviation for each row (gene).



expression selectively maintained in one daughter but not the other. For example, the TFs *ceh-36/37/43/45*, *ham-1*, and *hlh-3* are all coexpressed within single ASE-ASJ-AUA neuroblast cells. *ceh-36/37* and *hlh-3* expression was maintained in only one daughter of this neuroblast, the ASE parent, whereas *ceh-43/45* and *ham-1* expression was maintained in only the other daughter, the ASJ-AUA neuroblast (fig. S23).

This pattern, in which a progenitor cell co-expresses genes specific to each of its daughters, has been termed “multilineage priming” and has been observed in several organisms and developmental contexts (10, 31–35). Our transcriptomic atlas of the *C. elegans* cell lineage allows us to provide an unbiased quantification of the prevalence of multilineage priming throughout the organism’s ectoderm and mesoderm (we lack sufficient resolution in our annotations of the endoderm, which produces only one cell type, the intestine). There are 172 instances in which we have data for a parent cell and both of its distinct daughters. Of these, 52% exhibit multilineage priming. Multilineage priming events are distributed throughout several generations of both the ectoderm and mesoderm (fig. S24), demonstrating that it is a common and pervasive mechanism of gene regulation. The expression

patterns of many TFs involved in multilineage priming, e.g., *hlh-3* (fig. S23D), are confirmed by the movies in EPiC (17).

TFs that are both required for neuron type specification and have expression maintained throughout the lifetime of the neuron are referred to as “terminal selectors” (36). To identify potential terminal selectors, we looked for TFs that were (i) expressed in a neuron type but not its sister in the embryo and (ii) expressed in the same neuron type at the L2 stage. This analysis replicated 23 known neuron-TF associations (36) and identified 116 previously unknown associations (table S15). Other known associations may have been missed owing to the extreme sparsity of the L2 stage data, as well as the fact that many terminal selectors are expressed at low levels in fully differentiated neurons or are expressed in both daughters of a terminal division. In cases where a neuron’s sister undergoes programmed cell death, we looked for TFs that are both enriched in the terminal cell’s most recent ancestor that has a surviving sister cell (compared to that sister) and also have expression maintained throughout the life span of the terminal neuron. This revealed previously unknown associations, including *ceh-6* for AVH, *ceh-8* for RIA, *unc-62* for RIC, and *lin-11* for RIC and RIM, in which the

putative terminal selector TF is expressed in a neuroblast before the terminal cell is produced, suggesting that these lineages commit early to a cell fate.

Only two neurons (ASE and AWC) are known to have left-right asymmetric gene expression (37, 38). For both neuron types, the lineages of the left and right neurons diverge in the early embryo at the four-cell stage (<50 min). Asymmetric gene expression in our data, however, emerges only much later in embryogenesis. The transcriptomes of ASEL and ASER diverged in our UMAP at ~650 to 700 min, with *lim-6* expressed specifically in the ASEL branch, consistent with previous studies (39, 40). AWC left-right asymmetry occurs stochastically, with one neuron becoming “AWC-ON” and the other becoming “AWC-OFF” (38). We identified a small cluster in the UMAP with embryo time >700 min as AWC-ON based on *srt-28* expression (Fig. 3A) (41). AWC-OFF is putatively part of the main AWC trajectory. No evidence of left-right asymmetry was observed in neurons other than ASE and AWC.

Transcriptional convergence of cofated lineages

Although most bilaterally symmetric cells were not distinguishable by UMAP (as expected),

several cell types with >twofold symmetry are produced by multiple nonsymmetric lineage inputs. These lineage inputs tended to cluster separately in our progenitor cell UMAPs, whereas in our late-cell tissue UMAPs, we saw almost no evidence of heterogeneity within the terminal cell types that they produce. This difference suggested that the transcriptomes of these cofated lineages were converging during differentiation.

One example of apparent molecular convergence of cells from distinct lineages was the IL1/IL2 neuroblasts. The six IL1 and six IL2 neurons are produced by three symmetric pairs of neuroblast lineages. Each neuroblast pair produces a pair of bilaterally symmetric IL1 neurons and, likewise, a pair of IL2 neurons. A UMAP of IL1/IL2 neurons and progenitors revealed trajectories for these neuroblasts that converge gradually over their life span (Fig. 4A). The TF *ast-1* was transiently expressed at extremely high levels

[>10,000 transcripts per million (TPM)] during this process, suggesting that it might play a role in homogenizing the IL1/IL2 neuroblast transcriptomes (Fig. 4B). Correspondingly, expression of genes differentially expressed between the input lineages decreased over time, whereas expression of genes specific to terminal neurons increased (Fig. 4, C and D). We observed similar lineage convergence via continuous gene expression trajectories for other cell types, including hypodermis (fig. S8), head BWM (fig. S17), and GLR cells (fig. S17).

Like the IL1/IL2 neurons, IL socket glia (ILso) are produced by three symmetric pairs of lineages. In contrast to the examples discussed above, trajectories formed by the ILso progenitors and their terminal descendants were discontinuous in UMAP space (fig. S25). Discontinuous trajectories were also observed for several other cell types from multiple tissues, including other glia,

several neuron types, the excretory gland, colomocytes, and somatic gonad precursors (Z1/Z4) (fig. S25). Several lines of evidence suggest that these discontinuities reflect sudden changes in the transcriptome rather than technical artifacts of sc-RNA-seq or UMAP. Discontinuous trajectories had more genes differentially expressed between the parent and daughter cells than did continuous trajectories (fig. S26). Almost all discontinuous trajectories were observed in lineages where a parent cell gives rise to two daughters of different broadly defined cell types—e.g., a glia and a nonglial cell, or a ciliated neuron and a nonciliated neuron (fig. S26). These discontinuities were seen in both the global and the tissue-specific UMAPs, as well as with different UMAP parameters. Finally, for most discontinuous trajectories, cells had a continuous distribution of embryo times (fig. S27). However, a few trajectories, such as that of the BAG neuron, had gaps in the embryo time distribution indicative of potential sampling bias.

BWM was notable in that lineage-related heterogeneity persisted throughout differentiation. BWM is produced by multiple distinct lineages (C, D, MS) and occupies a wide range of positions along the anterior-posterior (A-P) axis of the animal. A UMAP of BWM cells identified distinct trajectories for the first row of head BWM versus all other BWM (Fig. 4E). The non-first-row trajectory was formed by input trajectories that corresponded to lineages and progressed in parallel along the temporal axis. Using marker genes expressed in domains along the A-P axis (17, 42–44), we divided BWM cells in the UMAP into six “bands” (Fig. 4E) and identified the specific anatomical cells present in each band (Fig. 4F and table S16). We found that the Jensen-Shannon (JS) distance, a measure of transcriptome difference, between the transcriptomes of posterior BWM (C lineage) versus both the first and second rows of BWM (D and MS lineage) did not decrease over time (Fig. 4G), indicating that BWM heterogeneity persists throughout differentiation.

Temporal dynamics of the lineage-transcriptome relationship

The presence of discontinuities between progenitor cells and terminal cells in the UMAP projections suggested that the terminal division could mark a shift from lineage-correlated to fate-correlated gene expression. We investigated how well the distance between two cells in the lineage predicts the difference between their transcriptomes (as defined with the JS distance). We focused on the AB lineage, which produces mostly ectoderm and accounts for ~70% of the terminal cells in the embryo. The AB lineage undergoes roughly synchronized cell divisions, allowing us to group cells by generation. For example, we refer to the 32 cells produced by five divisions of AB as “AB5,” and so on.

In AB5 (early to mid-gastrulation; 50-cell stage), the earliest stage at which our lineage annotations were near complete, sister cells were more similar than distant relatives but the

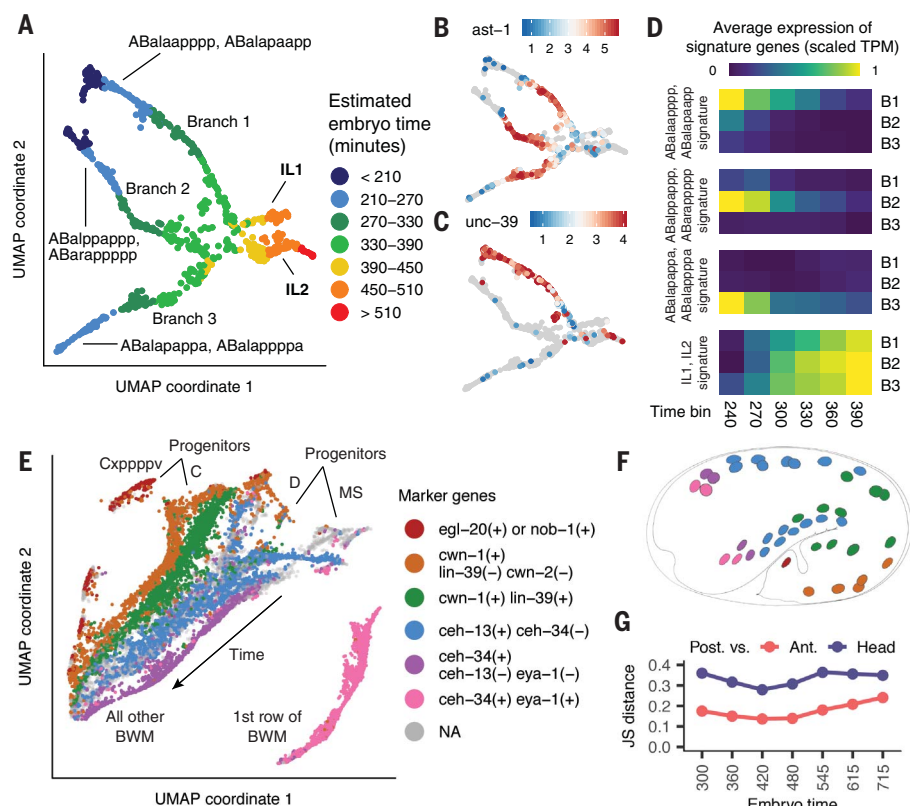


Fig. 4. Full versus incomplete convergence of lineages producing common cell types. (A) UMAP of 854 IL1/IL2 neurons and progenitors colored by estimated embryo time (cells selected on the basis of annotations in Fig. 3A and fig. S15). (B) IL1/IL2 UMAP colored by *ast-1* expression level (log2 size-factor-normalized UMI counts). (C) IL1/IL2 UMAP colored by expression of *unc-39*, a gene specific to branch 1. (D) Heatmap showing the average expression level of lineage-specific and terminal cell type-specific genes over time for each of the three branches. (E) Figure S5A shows a UMAP of BWM and mesoderm cells. This panel is a zoomed-in view of that UMAP, including only 17,520 BWM cells, which are grouped into “bands” based on marker gene expression patterns (here, a cell is considered to express a gene if it or at least two of five of its nearest neighbors have nonzero reads for the gene). (F) Physical positions of cells in each BWM band [colors matched to (E)] in the embryo at 430 min. [Adapted from figure 8B of (16)] (G) Transcriptome Jensen-Shannon (JS) distance for posterior [orange and green bands in (E)] BWM versus row 2 (blue band) or row 1 (pink band) head BWM over time. Heterogeneity between BWM subsets persists throughout development and may reflect functional differences.

difference was not large (Fig. 5A). In AB6 (mid-gastrulation; 100-cell stage) and AB7 (late gastrulation; 200-cell stage), the transcriptomes of sister cells become more similar than in AB5, whereas those of distant relatives become more divergent, resulting in a strong correlation between transcriptome distance and lineage distance. In AB8 (350-cell stage), most epidermal cells exit the cell cycle and begin terminal differentiation, whereas neuron and glia progenitors continue for one or two additional cell divisions. AB8 thus features a bimodal distribution of transcriptome JS distances: Terminal epidermal cells become highly distinct from neuron and glia progenitors, but cells within each group are more similar (fig. S28). Finally, most neuron and

glia progenitors in AB8 produce two terminal daughters in AB9 that have distinct cell fates and a much weaker lineage-transcriptome correlation than in earlier generations.

Together, these statistics suggest that progenitor cells develop strong expression signatures of their lineage identity and that these signatures are rapidly lost or overshadowed by new expression at the time of the terminal division. An analysis of cells from the mesoderm (MS lineage) replicated the trends observed in the ectoderm (fig. S29A).

To summarize the strength of the lineage-transcriptome correlation in a cell generation as a single number, we developed a statistic analogous to the concept of pseudo- R^2 in generalized

linear regression models (45). Consistent with the above analysis, we find that the extent to which lineage predicts the transcriptome increases throughout gastrulation, peaks at 55% in AB7, and then falls to 18% after terminal differentiation in AB9 (Fig. 5B). Next, we asked how much of the total pseudo- R^2 for one cell generation was attributable to gene expression signatures associated with each preceding cell generation. For cells in AB5 to AB8, the largest contributor to pseudo- R^2 was the identity of their ancestor in the AB3 generation (fig. S30). This is notable because many of the clades formed at AB3 share a broadly defined tissue fate. For example, the clade founded by the cell ABala produces only neurons and glia, whereas the clade founded by the cell ABarp produces mostly (but not exclusively) epidermal cells. The second-largest lineage signal was from the identity of a cell's parent in the preceding generation (i.e., the tendency of sister cells to be more similar than cousins). Thus, both broad and fine-grained structure in the lineage contribute toward shaping the transcriptome.

To investigate the potential regulatory mechanisms that differentiate sister cells, we identified TFs that distinguish each cell in AB5 to AB9 from its sister. The median number of these "lineage signature TFs" per cell increased over time, ranging from 1.5 in AB5 to 14 in AB9 (Fig. 5C). A substantial number of lineage signature TFs (~40 to 50%) had expression selectively maintained in only one of a cell's two daughters (Fig. 5D). In other words, TFs that distinguish a cell from its sister in one generation are frequently reused to distinguish that cell's daughters from each other. Sister cells are also differentiated by the expression of new TFs not present in their parents. The proportion of lineage signature TFs that are newly expressed ranged from 33 to 61% and increased over time in AB6 to AB9 (Fig. 5E). Temporal dynamics of lineage signature TFs was similar in the mesoderm (fig. S29).

Taken together, these results highlight the incremental nature of cell fate decisions: Every terminal cell is the result of a series of lineage bifurcations, each of which, on average, involves multiple differentially expressed TFs.

Global patterns of gene expression and transcriptome specialization

Hierarchical clustering of expression levels in all annotated lineages and cell types (tables S7 and S8) provides a global view of expression dynamics for all genes in our dataset. A heatmap of preterminal lineage expression profiles (fig. S31) does not reveal large clusters of genes specific to particular lineages, other than one cluster of genes specific to the early C and D lineages. Similarly, most marker genes used for lineage annotation are not part of large clusters of co-expressed genes. The clusters that do form are composed of early tissue-specific genes. The lack of cluster structure in the heatmap suggests that differential fates for tissue sublineages are specified by relatively small sets of genes. By contrast, a heatmap of terminal cell type expression

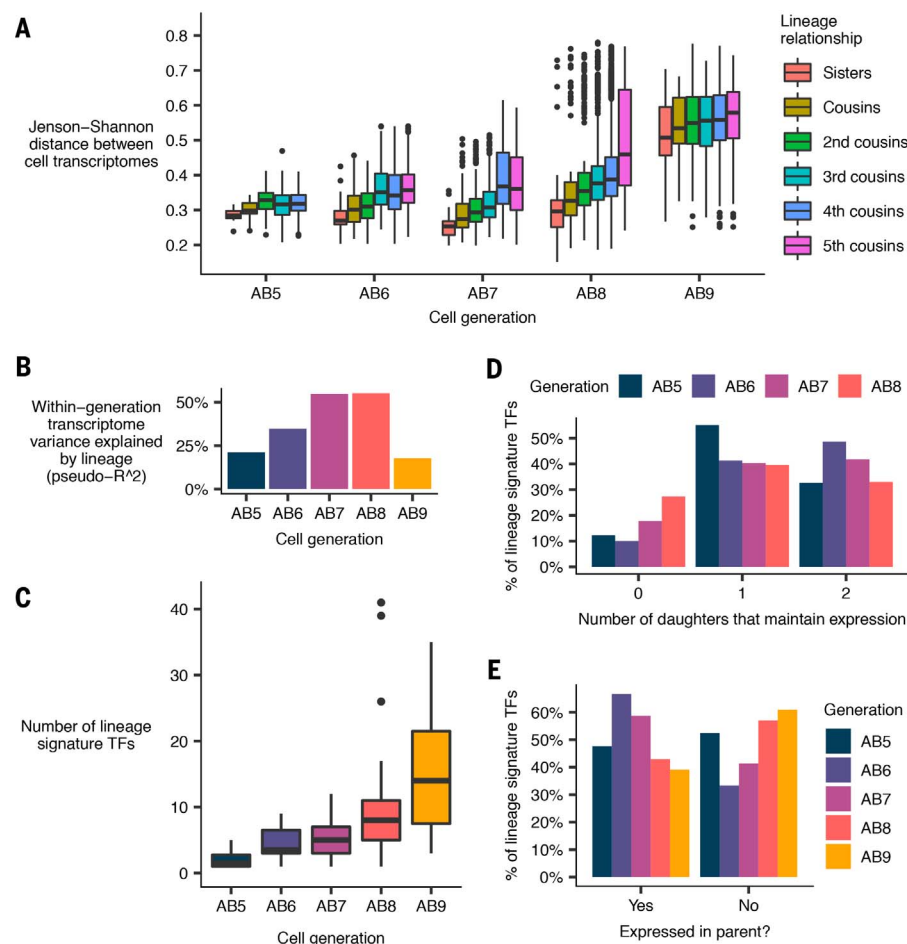


Fig. 5. Correlation between cell lineage and transcriptome in the ectoderm. (A) JS distance between the transcriptomes of pairs of ectodermal cells (AB lineage), faceted by cell generation and lineage distance. AB5 refers to the cell generation produced by five divisions of the AB founder cell, and likewise for generations AB6 to AB9. The transcriptome of a given anatomical cell is defined as the average gene expression profile of all sc-RNA-seq cells annotated as that anatomical cell. Pairs of bilaterally symmetric cells are excluded from the statistics. (B) Estimates of the extent to which lineage predicts the transcriptome in AB5 to AB9. (C) Distribution of the number of lineage signature TFs (i.e., TFs that distinguish a cell from its sister) for all cells in AB5 to AB9. The outlier points in AB8 are instances where a terminal epidermal cell is a sister of a neuroblast. (D) Proportion of lineage signature TFs for a cell in a given generation that have expression maintained in zero to two of the cell's daughters in the subsequent generation. (E) Proportion of lineage signature TFs for which expression in a given cell was maintained from the cell's parent versus newly activated after the parent's division.

profiles (fig. S32) has more obvious structure. Cells in each major tissue express ~500 to 1500 tissue-enriched genes. There is little reuse of tissue-enriched genes between tissues other than the hypodermis, which shares many genes with the glia and intestine. Neuron subtypes and other specialized cells (such as the hmc or M cell) are typically distinguished from other cells within their tissue by expression of <20 to 300 genes. Finally, there are substantial temporal changes in expression, especially in muscle and hypodermis.

We observed substantial variation between cells with regard to the Gini coefficient, which measures how unequally different genes are expressed in a given cell type (fig. S33A). Hypodermis, seam cells, and the pharyngeal gland express small sets of cell type-specific genes at very high levels (high Gini coefficient), whereas the intestine and germ line feature diverse gene expression patterns (low Gini coefficient). In several cell types, such as those of the pharyngeal gland, increases in Gini coefficient over time coincide with decreases in the number of TFs expressed per cell (fig. S33B). Families of TFs also exhibit differential expression patterns over time and across lineages. Nuclear hormone receptors (NHRs) are, on average, activated later in development than other TF families, such as forkhead and homeodomain TFs (fig. S33C). The hypodermis and intestine express many distinct NHRs, whereas expression of Sox family TFs is largely restricted to neurons, glia, and pharynx (fig. S33D).

An RShiny app to explore and extend on our analysis

We developed VisCello to distribute single-cell analyses and provide interactive visualizations (fig. S34). It is available as a web application (<https://cello.shinyapps.io/celegans/>) and can also be installed as an R package (<https://github.com/qinzhuzhu/VisCello.celegans>). VisCello hosts dimensionality reductions (e.g., UMAPs), cell annotations, and marker gene tables for the different subsets of the data described in this manuscript. Users can visualize gene expression on UMAP or principal components analysis (PCA) plots, on a lineage tree diagram, or as box or violin plots grouped by cell type or lineage. The plots are interactive, allowing users to zoom in on subsets of cells, define new cell annotation groups, and run differential expression analysis and GO/KEGG (Gene Ontology/Kyoto Encyclopedia of Genes and Genomes) enrichment with these newly defined groups. The program state can be downloaded and shared, facilitating collaboration. VisCello can also be used to host and disseminate other single-cell datasets, including data from the *C. elegans* 1- to 16-cell stage (8) and L2 stage (2) (<https://github.com/qinzhuzhu/VisCello>).

Discussion

The cells of *C. elegans* are limited in number and invariant in lineage and cell fate, making it feasible to conduct comprehensive, whole-organism investigations. Yet within this limited reper-

toire of cells exists a wide diversity of cell types that work together to produce complex anatomical structures and behaviors. This study and our previous work (2, 17) have shed light on the molecular basis for the specification of these cell types but are only the first step toward a comprehensive understanding of the molecular basis of development. We hope that this resource will help guide future projects in the *C. elegans* community.

In contrast to developmental sc-RNA-seq datasets from other species, this dataset links gene expression trajectories to the exact cell lineages to which they correspond, allowing cell differentiation steps to be associated with specific cell division events. Thus, our data provide a quantitative portrait of Waddington's landscape (46) for a whole organism. However, the abruptness of many cell fate decisions in *C. elegans*, with many distinct terminal cell types becoming distinguished only in the final embryonic cell division, contrasts with the smooth landscape in Waddington's illustrations and warrants further investigation.

We observe convergence of gene expression patterns in many instances where distinct cell lineages produce identical or related cell types. Data from a recent atlas of mouse organogenesis (13) suggests that this phenomenon is also prevalent in vertebrates. For example, myocytes in the mouse atlas are produced by two convergent trajectories, and excitatory neurons are produced by several trajectories.

Our analysis highlights two important challenges for efforts to reconstruct the cell lineages of other organisms using sc-RNA-seq. First is the difficulty of accurately connecting developmental trajectories that start after the convergence of lineages with similar cell fates to trajectories that span earlier stages of development. A naive interpretation of the UMAP projection of the full dataset (Fig. 1A) could lead to inferred trajectories that are inconsistent with the correct lineage (for example, incorrectly concluding that hypodermis and seam cells are produced from a common ancestor that previously diverged from the progenitors of neurons). Second is the difficulty of constructing continuous trajectories for lineages that undergo abrupt changes in gene expression. In our data, progenitor cells that give rise to glia, excretory cells, and nonciliated neurons were typically disconnected from their terminal daughters in UMAP space (figs. S25 and S26), reflecting the fact that many of these lineages commit to a terminal fate only after their final cell division.

Because of these challenges, we anticipate that constructing end-to-end trajectories of vertebrate organogenesis will require the integration of sc-RNA-seq with experimental lineage-tracing methods (47). It will also require improved computational methods that can model heterogeneity among poorly differentiated progenitor cells and highly differentiated cell types in an integrated manner.

Between this study, our previous study of the L2 stage (2), and earlier studies of the 1- to 16-cell stage embryos (8, 19), a large portion of the early

C. elegans life cycle has now been profiled by single-cell transcriptomics. However, more datasets will be needed to complete missing stages, including other larval stages and the adult soma and germ line. In the future, single-cell profiling of different strains or species will be a useful approach to examine the evolution of cell types and their expression programs. All of these datasets will ideally be integrated into a single visualization platform, such as VisCello, to allow full tracking of cell trajectories from fertilization through the end of life. A greater challenge will be to discover the precise mechanisms that produce transcriptomic outputs. Single-cell transcriptome analysis of mutants will likely need to be integrated with new single-cell multi-omic technologies (48) to bring mechanistic studies to a whole-organism scale.

Materials and methods summary

Sample preparation

C. elegans embryos were prepared by standard hypochlorite treatment methods from populations of synchronized early adult worms grown at 20°C. Embryos were dissociated immediately or aged in egg buffer before dissociation. Eggshells were removed by chitinase digestion, embryos were dissociated by manual shearing, and single cells were isolated by filtration or centrifugation. sc-RNA-seq libraries were generated with 10X Genomics v2 chemistry and standard protocols and were sequenced on Illumina NextSeq instruments.

Computational analysis

The sc-RNA-seq data were processed with the 10X Genomics Cell Ranger pipeline, aligning reads to a modified version of the WormBase (26) WS260 reference transcriptome that had transcript 3' untranslated regions extended by 0 to 500 base pairs. The data were then visualized using dimensionality reduction methods. Single-cell transcriptomes were first projected into 50 to 100 dimensions (depending on the analysis) using PCA and then projected into two or three dimensions using the UMAP algorithm (22). Cells in the UMAP space were clustered using the Louvain algorithm (25). For each cell, the age of the embryo from which it came (i.e., the embryo time) was estimated by correlating its transcriptome with a bulk RNA-seq time series (27). Cells were then manually annotated with their corresponding cell type and lineage, as described in the main text. The annotation process was guided by the UMAP projections, Louvain clusters, and embryo time estimates but also relied heavily on fluorescent reporter imaging data from the EPIC (17) and WormBase (26) databases.

A full description of the materials and methods used in this work is provided in the supplementary materials.

REFERENCES AND NOTES

1. X. Qiu et al., Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* 14, 309–315 (2017). doi: 10.1038/nmeth.4150; pmid: 28114287

2. J. Cao *et al.*, Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017). doi: [10.1126/science.aam8940](https://doi.org/10.1126/science.aam8940); pmid: 28818938
3. C. T. Fincher, O. Wurtzel, T. de Hoog, K. M. Kravarik, P. W. Reddien, Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* **360**, eaag1736 (2018). doi: [10.1126/science.aag1736](https://doi.org/10.1126/science.aag1736); pmid: 29674431
4. M. Plass *et al.*, Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* **360**, eaag1723 (2018). doi: [10.1126/science.aag1723](https://doi.org/10.1126/science.aag1723); pmid: 29674432
5. Tabula Muris Consortium, Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018). doi: [10.1038/s41586-018-0590-4](https://doi.org/10.1038/s41586-018-0590-4); pmid: 30283141
6. A. Zeisel *et al.*, Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018). doi: [10.1016/j.cell.2018.06.021](https://doi.org/10.1016/j.cell.2018.06.021); pmid: 30096314
7. A. Seb  Pedr  s *et al.*, Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell* **173**, 1520–1534.e20 (2018). doi: [10.1016/j.cell.2018.05.019](https://doi.org/10.1016/j.cell.2018.05.019); pmid: 29856957
8. S. C. Tintori, E. Osborne Nishimura, P. Golden, J. D. Lieb, B. Goldstein, A Transcriptional Lineage of the Early C. elegans Embryo. *Dev. Cell* **38**, 430–444 (2016). doi: [10.1016/j.devcel.2016.07.025](https://doi.org/10.1016/j.devcel.2016.07.025); pmid: 27554860
9. N. Karaikos *et al.*, The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **358**, 194–199 (2017). doi: [10.1126/science.aan3235](https://doi.org/10.1126/science.aan3235); pmid: 28860209
10. J. A. Briggs *et al.*, The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780 (2018). doi: [10.1126/science.aar5780](https://doi.org/10.1126/science.aar5780); pmid: 29700227
11. D. E. Wagner *et al.*, Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018). doi: [10.1126/science.aar4362](https://doi.org/10.1126/science.aar4362); pmid: 29700229
12. J. A. Farrell *et al.*, Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018). doi: [10.1126/science.aar3131](https://doi.org/10.1126/science.aar3131); pmid: 29700225
13. J. Cao *et al.*, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019). doi: [10.1038/s41586-019-0969-x](https://doi.org/10.1038/s41586-019-0969-x); pmid: 30787437
14. B. Pijuan-Sala *et al.*, A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019). doi: [10.1038/s41586-019-0933-9](https://doi.org/10.1038/s41586-019-0933-9); pmid: 30787436
15. J. E. Sulston, H. R. Horvitz, Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977). doi: [10.1016/0012-1606\(77\)90158-0](https://doi.org/10.1016/0012-1606(77)90158-0); pmid: 838129
16. J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983). doi: [10.1016/0012-1606\(83\)90201-4](https://doi.org/10.1016/0012-1606(83)90201-4); pmid: 6684600
17. J. I. Murray *et al.*, Multidimensional regulation of gene expression in the C. elegans embryo. *Genome Res.* **22**, 1282–1294 (2012). doi: [10.1101/gr131920.111](https://doi.org/10.1101/gr131920.111); pmid: 22508763
18. M. Sarov *et al.*, A genome-scale resource for in vivo tag-based protein function exploration in C. elegans. *Cell* **150**, 855–866 (2012). doi: [10.1016/j.cell.2012.08.001](https://doi.org/10.1016/j.cell.2012.08.001); pmid: 22901814
19. T. Hashimshony, F. Wagner, N. Sher, I. Yanai, CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012). doi: [10.1016/j.celrep.2012.08.003](https://doi.org/10.1016/j.celrep.2012.08.003); pmid: 22939981
20. X. Qiu *et al.*, Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017). doi: [10.1038/nmeth.4402](https://doi.org/10.1038/nmeth.4402); pmid: 28825705
21. T. Hashimshony, M. Feder, M. Levin, B. K. Hall, I. Yanai, Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature* **519**, 219–222 (2015). doi: [10.1038/nature13996](https://doi.org/10.1038/nature13996); pmid: 25487147
22. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426* [stat.ML] (9 February 2018).
23. E. Becht *et al.*, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018). doi: [10.1038/nbt.4314](https://doi.org/10.1038/nbt.4314); pmid: 30531897
24. See supplemental note 1 in the supplementary materials for a discussion of the term “trajectory”.
25. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *arXiv:0803.0476* [physics.soc-ph] (4 March 2008).
26. R. Y. N. Lee *et al.*, WormBase 2017: Molting into a new stage. *Nucleic Acids Res.* **46**, D869–D874 (2018). doi: [10.1093/nar/gkx998](https://doi.org/10.1093/nar/gkx998); pmid: 29069413
27. C. L. Araya *et al.*, Regulatory analysis of the C. elegans genome with spatiotemporal resolution. *Nature* **512**, 400–405 (2014). doi: [10.1038/nature13497](https://doi.org/10.1038/nature13497); pmid: 25164749
28. G. Broitman-Maduro *et al.*, The NK-2 class homeodomain factor CEH-51 and the T-box factor TBX-35 have overlapping function in C. elegans mesoderm development. *Development* **136**, 2735–2746 (2009). doi: [10.1242/dev.038307](https://doi.org/10.1242/dev.038307); pmid: 19605496
29. W. C. Spencer *et al.*, A spatial and temporal map of C. elegans gene expression. *Genome Res.* **21**, 325–341 (2011). doi: [10.1101/gr.114595.110](https://doi.org/10.1101/gr.114595.110); pmid: 21177967
30. J. L. Richards, A. L. Zacharias, T. Walton, J. T. Burdick, J. I. Murray, A quantitative model of normal *Caenorhabditis elegans* embryogenesis and its disruption after stress. *Dev. Biol.* **374**, 12–23 (2013). doi: [10.1016/j.ydbio.2012.11.034](https://doi.org/10.1016/j.ydbio.2012.11.034); pmid: 23220655
31. M. Hu *et al.*, Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11**, 774–785 (1997). doi: [10.1101/gad.11.6.774](https://doi.org/10.1101/gad.11.6.774); pmid: 9087431
32. P. Laslo *et al.*, Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell* **126**, 755–766 (2006). doi: [10.1016/j.cell.2006.06.052](https://doi.org/10.1016/j.cell.2006.06.052); pmid: 16923394
33. M. Thomson *et al.*, Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**, 875–889 (2011). doi: [10.1016/j.cell.2011.05.017](https://doi.org/10.1016/j.cell.2011.05.017); pmid: 21663792
34. E. W. Brunskill *et al.*, Single cell dissection of early kidney development: Multilineage priming. *Development* **141**, 3093–3101 (2014). doi: [10.1242/dev.110601](https://doi.org/10.1242/dev.110601); pmid: 25053437
35. W. Wang *et al.*, A single-cell transcriptional roadmap for cardiopharyngeal fate diversification. *Nat. Cell Biol.* **21**, 674–686 (2019). doi: [10.1038/s41556-019-0336-z](https://doi.org/10.1038/s41556-019-0336-z); pmid: 31160712
36. O. Hobert, A map of terminal regulators of neuronal identity in *Caenorhabditis elegans*. *WIREs Dev. Biol.* **5**, 474–498 (2016). doi: [10.1002/wdev.233](https://doi.org/10.1002/wdev.233); pmid: 27136279
37. S. Yu, L. Avery, E. Baude, D. L. Garbers, Guanylyl cyclase expression in specific sensory neurons: A new family of chemosensory receptors. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3384–3387 (1997). doi: [10.1073/pnas.94.7.3384](https://doi.org/10.1073/pnas.94.7.3384); pmid: 9096403
38. E. R. Troemel, A. Sagasti, C. I. Bargmann, Lateral signaling mediated by axon contact and calcium entry regulates asymmetric odorant receptor expression in C. elegans. *Cell* **99**, 387–398 (1999). doi: [10.1016/S0092-8674\(00\)81525-1](https://doi.org/10.1016/S0092-8674(00)81525-1); pmid: 10571181
39. O. Hobert, K. Tessmar, G. Ruvkun, The *Caenorhabditis elegans* lim-6 LIM homeobox gene regulates neurite outgrowth and function of particular GABAergic neurons. *Development* **126**, 1547–1562 (1999). pmid: 10068647
40. J. T. Pierce-Shimomura, S. Faumont, M. R. Gaston, B. J. Pearson, S. R. Lockery, The homeobox gene lim-6 is required for distinct chemosensory representations in C. elegans. *Nature* **410**, 694–698 (2001). doi: [10.1038/35070575](https://doi.org/10.1038/35070575); pmid: 11287956
41. B. J. Lesch, C. I. Bargmann, The homeodomain protein hmbx-1 maintains asymmetric gene expression in adult C. elegans olfactory neurons. *Genes Dev.* **24**, 1802–1815 (2010). doi: [10.1101/gad.1932610](https://doi.org/10.1101/gad.1932610); pmid: 20713521
42. M. Harterink *et al.*, Neuroblast migration along the anteroposterior axis of C. elegans is controlled by opposing gradients of Wnts and a secreted Frizzled-related protein. *Development* **138**, 2915–2924 (2011). doi: [10.1242/dev.064733](https://doi.org/10.1242/dev.064733); pmid: 21653614
43. K. Brunschwig *et al.*, Anterior organization of the *Caenorhabditis elegans* embryo by the labial-like Hox gene *ceh-13*. *Development* **126**, 1537–1546 (1999). pmid: 10068646
44. T. Hirose, B. D. Galvin, H. R. Horvitz, Six and Eya promote apoptosis through direct transcriptional activation of the proapoptotic BH3-only gene *egl-1* in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15479–15484 (2010). doi: [10.1073/pnas.1010023107](https://doi.org/10.1073/pnas.1010023107); pmid: 20713707
45. See supplementary materials and methods (section titled “Pseudo-R² statistic used in Fig. 5B and Fig. S29B”).
46. C. H. Waddington, *The Strategy of the Genes* (Allen & Unwin, 1957).
47. L. Kester, A. van Oudenaarden, Single-Cell Transcriptomics Meets Lineage Tracing. *Cell Stem Cell* **23**, 166–179 (2018). doi: [10.1016/j.stem.2018.04.014](https://doi.org/10.1016/j.stem.2018.04.014); pmid: 29754780
48. J. Packer, C. Trapnell, Single-Cell Multi-omics: An Engine for New Quantitative Models of Gene Regulation. *Trends Genet.* **34**, 653–665 (2018). doi: [10.1016/j.tig.2018.06.001](https://doi.org/10.1016/j.tig.2018.06.001); pmid: 30007833
49. S. J. Husson *et al.*, Impaired processing of FLP and NLP peptides in carboxypeptidase E (EGL-21)-deficient *Caenorhabditis elegans* as analyzed by mass spectrometry. *J. Neurochem.* **102**, 246–260 (2007). doi: [10.1111/j.1471-4159.2007.04474.x](https://doi.org/10.1111/j.1471-4159.2007.04474.x); pmid: 17564681
50. T. R. Sarafi-Reinach, P. Sengupta, The forkhead domain gene *unc-130* generates chemosensory neuron diversity in C. elegans. *Genes Dev.* **14**, 2472–2485 (2000). doi: [10.1101/gad.832300](https://doi.org/10.1101/gad.832300); pmid: 11018015
51. Q. Zhu, J. I. Murray, K. Tan, J. Kim, qinZhu/VisCello.celegans: VisCello.celegans v1.1.0 release, Zenodo (2019); <https://doi.org/10.5281/zenodo.3262314>.
52. Q. Zhu, J. I. Murray, K. Tan, J. Kim, qinZhu/VisCello: VisCello v1.0.0, Zenodo (2019); <https://doi.org/10.5281/zenodo.3262313>.
53. P. Weisdepp *et al.*, Data from: A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution, Dryad (2019); <https://doi.org/10.5061/dryad.7tg31p7>.

ACKNOWLEDGMENTS

We thank members of the Murray, Waterston, and Kim laboratories as well as B. Lehner and M. Sundaram for providing critical comments on the manuscript. We also thank A. Zacharias, D. Vafeados, M. Corson, R. Terrell, L. Gevirtzman, and P. Weisdepp for their contributions to the EPIC database. **Funding:** This work was funded by NIH grants U41HG007355 and R01GM072675 (R.H.W.) and R35GM127093 and R21HD085201 (J.I.M.). This work was also funded in part by Commonwealth of Pennsylvania Health Research Formula Funds and RMIHG010023 (J.K.), U2C CA233285 (K.T.), the William H. Gates Chair of Biomedical Sciences (R.H.W.), and the Allen Discovery Center for Lineage Tracing (J.S.P. and C.T.). The opinions expressed in this article are the authors' own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the U.S. government.

Author contributions: J.S.P., C.H., J.K., R.H.W., and J.I.M. conceived and designed the study; C.H., P.S., E.P., H.D., and D.S. performed the experiments; J.P., Q.Z., R.H.W., and J.I.M. did the analyses; C.T., J.K., R.H.W., and J.I.M. supervised the analyses; J.K., J.I.M., and K.T. supervised the development of VisCello; and J.S.P., Q.Z., J.K., R.H.W., and J.I.M. wrote the paper. **Competing interests:** The authors have no competing interests. **Data and materials availability:** The raw data have been deposited with the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) under accession code GSE126954. Source code of VisCello (with C. elegans data) has been deposited at Github (<https://github.com/qinZhu/VisCello.celegans>) and Zenodo (51). Source code of VisCello for hosting other single-cell data has been deposited at Github (<https://github.com/qinZhu/VisCello>) and Zenodo (52). Gene expression movies used in the annotation but not previously published have been deposited at Dryad (53). This article was prepared while HD was employed by the University of Pennsylvania.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6459/eaax1971/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S35
Tables S1 to S16
References (54–68)
Data S1

1 March 2019; accepted 21 August 2019
Published online 5 September 2019
[10.1126/science.aax1971](https://doi.org/10.1126/science.aax1971)

A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution

Jonathan S. Packer, Qin Zhu, Chau Huynh, Priya Sivaramakrishnan, Elicia Preston, Hannah Dueck, Derek Stefanik, Kai Tan, Cole Trapnell, Junhyong Kim, Robert H. Waterston and John I. Murray

Science **365** (6459), eaax1971.

DOI: 10.1126/science.aax1971 originally published online September 5, 2019

Identifying terminal nematode cells

Single-cell RNA sequencing provides the power to identify the developmental trajectory of an organism. However, identifying the temporal lineage of cell development can be difficult without large-scale analyses. Packer *et al.* sequenced more than 80,000 cells from embryos of the roundworm *Caenorhabditis elegans* to determine the expression of genes directing the development of terminal cell types. Because all somatic cells in a *C. elegans* individual have been mapped, the authors are able to connect gene expression with cell lineages over time during development, noting stark transitions in some cases.

Science, this issue p. eaax1971

ARTICLE TOOLS

<http://science.sciencemag.org/content/365/6459/eaax1971>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2019/09/04/science.aax1971.DC1>

REFERENCES

This article cites 64 articles, 26 of which you can access for free
<http://science.sciencemag.org/content/365/6459/eaax1971#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works