



science.sciencemag.org/cgi/content/full/science.aax1971/DC1

Supplementary Materials for

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†

*These authors contributed equally to this work.

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

Published 5 September 2019 on *Science* First Release

DOI: 10.1126/science.aax1971

This PDF file includes:

- Materials and Methods
- Supplementary Text
- Figs. S1 to S35
- Captions for Tables S1 to S16
- Caption for Data S1
- References

Other Supplementary Material for this manuscript includes the following:

(available at science.sciencemag.org/cgi/content/full/science.aax1971/DC1)

- Tables S1 to S6, S9, S12, S13, S15, and S16 (.xlsx)
- Tables S7, S8, S10, S11, and S14 (.tsv.gz)
- Data S1 (.gtf.gz)

1 **Materials and Methods**

2

3 **Sample preparation**

4

5 To obtain a broad range of embryo ages, including early stages, roughly synchronized *C.*
6 *elegans* adults (N2 strain) were obtained by releasing embryos with standard hypochlorite
7 treatment and letting the L1 larvae hatch and undergo growth arrest on unseeded plates. Starved
8 L1s were transferred to NGM plates seeded with *E. coli* OP50 bacteria. Embryos were released
9 from these synchronized young adults using hypochlorite treatment followed by three washes
10 with L15-10 media. To generate cell suspensions, embryos were then treated with 0.5 mg/ml
11 chitinase at room temperature until the shells were dissolved (30-40 minutes at ~22 degrees C)
12 followed by dissociation of the cells using a 3 ml syringe fitted with a 21 gauge 1¼ inch needle
13 until >80% of embryos were disrupted. The cell suspension was then passed through a 10 µm
14 filter, washed in phosphate buffered saline (PBS) and finally resuspended in PBS. An estimated
15 14,000 cells were loaded immediately onto a 10X Chromium instrument. The trypan blue
16 negative viable cell count was estimated using a hemocytometer and was >84% for all samples.

17 To sample later stages more deeply, more tightly synchronized embryo populations (used
18 for the 300-minute, 400-minute, and 500-minute time series shown in **Fig. 1B**) were obtained
19 through two cycles of bleaching adult worms (strain VC2010, a strain derived from N2 that has
20 been completely sequenced). On the first round of synchronization, populations of mixed stage
21 embryos recovered by hypochlorite treatment of mixed populations were hatched overnight in
22 egg buffer (118 mM NaCl, 48 mM KCl, 3 mM CaCl₂, 3 mM MgCl₂, 5 mM HEPES pH 7.2)
23 with gentle shaking. The hatched L1s were plated onto 150 mm peptone rich NGM plates seeded

24 with *E. coli* NA22 at no more than 100,000 worms per plate. When worms reached the adult
25 stage, the number of embryos inside the adults was monitored until most had about 4 embryos on
26 each gonad arm. The adult worms were collected and treated with hypochlorite to release
27 embryos. The embryos were again allowed to hatch in the absence of food at 20 °C for 12 hours
28 yielding a more tightly synchronized population of L1 worms. Around 250,000 L1 larvae were
29 plated onto four 100 mm petri plates seeded with NA22 bacteria and allowed to develop at 20
30 °C. As the worms reached the young adult stage, the population was closely monitored. When
31 about 20-30% of the adults had a single embryo in either arm of the gonad, worms were
32 subjected to hypochlorite treatment. The time hypochlorite was added to the worms was
33 considered t = 0 (see Warner *et al.* (54) for typical age distributions). The capture time was taken
34 as when the cells were loaded onto the 10x Chromium instrument. The embryos were allowed to
35 develop in egg buffer until one hour prior to capture time. The embryos were collected by
36 centrifugation, resuspended in 0.5 ml egg buffer and 1 ml chitinase (1 U/ml), and transferred to
37 30 mm petri dishes. The degradation of eggshell was monitored; after ~20 min (when about half
38 the eggs had lost the shell), the suspension was transferred to a 15 ml falcon tube and centrifuged
39 at 200 g for 5 min. The chitinase solution was aspirated; a solution of 200 ul pronase (15 mg/ml)
40 together with 0.5 ml egg buffer was added to the embryo pellet. The vitelline membrane was
41 disrupted and the cells released by repeated passage through 21 guage 1¼ inch needle attached to
42 a 1 ml syringe. When sufficient single cells were observed, the reaction was stopped by adding 1
43 ml of egg buffer containing 1% BSA. Cells were separated from intact embryos by centrifuging
44 the pronase treated embryos at 150 g for 5 min at 4 °C. The supernatant was transferred to a 1.5
45 ml microcentrifuge tube and centrifuged at 500 g for 5 min at 4 °C. The cell pellet was washed
46 twice with egg-buffer containing 1% BSA.

47 Single cell capture and library preparation followed 10X Genomics published
48 protocols. For each channel, 14,000 *C. elegans* cells were mixed with reverse transcriptase
49 reaction solution and loaded immediately onto the capture chip to minimize the time that the
50 cells spent in the reverse transcription cocktail. The exception was the first 500 minute sample,
51 when three channels were loaded with 14,000, 4,666, and 1,555 cells respectively.

52

53 **Read mapping and gene expression quantification**

54

55 The single cell RNA-seq data was processed using the 10X Genomics CellRanger
56 pipeline. Reads were mapped to the *C. elegans* reference transcriptome from WormBase (26),
57 version WS260. We noticed that many 3' UTR annotations in the reference transcriptome were
58 too short, causing genic reads to be called as intergenic, affecting gene expression quantification.
59 To address this, we also mapped reads to modified versions of the WS260 transcriptome in
60 which all 3' UTRs were extended by either 100, 200, 300, 400, or 500 bp (these 3' UTR
61 extensions were cut short if the extended UTR would overlap with a downstream gene).

62 We then defined a set of criteria that specified for each gene whether it was beneficial to
63 extend the 3' UTR for that gene, and if so, by how much. For each gene, we counted the number
64 of reads across the entire dataset mapped to that gene for each version of the reference. We
65 computed the ratio of the read counts from the 500 bp 3' UTR extended reference to the baseline
66 reference. If this ratio was < 1.2, or if the total read count for the gene in the 500 bp 3' UTR
67 extended reference was < 20, we used the baseline 3' UTR annotation for that gene. Otherwise,
68 we used the shortest 3' UTR extension (100, 200, 300, 400, or 500 bp) that gave at least 90% of
69 the read count gain that was given by the 500 bp 3' UTR extension.

70 We repeated this process with reads from our previous study on L2 worms (2). If a gene
71 met our criteria for extending the 3' UTR based on embryo reads, we used the extension length
72 determined by the embryo reads. If a gene did not meet our criteria for extending the 3' UTR
73 based on embryo reads but did meet the criteria based on L2 stage reads, we used the extension
74 length determined by the L2 stage reads. After deciding on how much to extend each gene's 3'
75 UTR, we made a final reference transcriptome incorporating all of the per-gene 3' UTR
76 extension lengths. We then used this final reference transcriptome as input to the CellRanger
77 pipeline to generate gene-by-cell UMI count matrices.

78 Our final reference transcriptome is available as **Additional Data File 1**. It should be
79 suitable for future studies of *C. elegans* embryos or early larva. However, neither the embryonic
80 or L2 datasets contain reads from developed germline cells (sperm and oocytes) or developed
81 somatic gonad cells (e.g. spermatheca), so our reference will not properly extend 3' UTRs for
82 genes specific to these cell types.

83

84 **Criteria for distinguishing cells from empty droplets**

85

86 The default barcode filtering algorithm in the 10X CellRanger pipeline can fail for
87 experiments where the cells profiled are highly variable in size, resulting in a non-normal
88 distribution of UMIs per cell. This is the case for our data. The total volume of the *C. elegans*
89 embryo remains constant as cells divide within it, making cells of later generations smaller than
90 those from earlier generations. Additionally, some cell types are more prone to damage and
91 mRNA leakage than others. Neurons in particular usually have lower UMI counts than other cell
92 types. To account for these factors, we manually set UMI count thresholds to distinguish cell

93 barcodes from empty droplet barcodes on a sample-by-sample basis, based on the knee plots
94 reported by CellRanger. The UMI count thresholds ranged for 700-1100.

95 While performing downstream analyses, we noticed that several neuronal, glial, rectal,
96 and excretory cell types were missing from our data. We discovered that this was due to cells
97 with extra low UMI counts (< 700 UMIs) being excluded by our UMI count thresholds.
98 Lowering the UMI count threshold for all cells, however, would include low-quality, potentially
99 damaged cells for other cell types where the average UMIs/cell is higher. To integrate the low-
100 UMI count cells, we:

- 101 1. made a set of all cells with UMI count ≥ 500 (vs. the previous threshold of 700)
 - 102 2. ran UMAP dimensionality reduction (described below) on this set of cells
 - 103 3. identified clusters of cells corresponding to neurons (using the pan-neuronal marker
104 genes *sbt-1* and *egl-21*) or glia, rectal, and excretory cells (using a variety of markers; see
105 **Table S1**)
 - 106 4. made new UMAPs from just neurons, just glia and excretory cells, or just rectal cells
 - 107 5. filtered putative doublets (i.e. cells also expressing markers of non-neuronal cell types in
108 the neuron UMAP, or cells also expressing markers of non-glia/hypodermal cell types in
109 the glia UMAP)
 - 110 6. made whitelists of the remaining cells
- 111 These whitelisted low-UMI count cells were then included when generating the final tissue
112 UMAPs presented in this paper (**Figs. 3A, S9-11, S13**). They are not included in the original
113 global UMAP (**Fig. 1A**).

114

115

116 **Dimensionality reduction**

117

118 For each dimensionality reduction (both for the global analysis of all cells and the tissue
119 specific analyses), the first step was to perform PCA and adjust the PCA results to correct for
120 batch effects. We performed PCA on the size-factor corrected, log transformed expression
121 matrix, typically with 50-100 PCs depending on the dataset.

122 For batch effect correction, we noted that the predominant source of batch effects in our
123 data appeared to be background contamination where RNA from lysed or damaged cells enters
124 droplets in the 10X sc-RNA-seq apparatus that contain intact cells, causing each cell to receive
125 reads from exogenous RNA. For each experimental sample, we computed the gene expression
126 distribution of this background RNA by summing the read counts for cell barcodes that had < 50
127 UMIs, i.e. empty droplets. We transformed the background RNA count vector for each sample as
128 if it were the count vector for a cell, and projected this vector into the PCA space computed from
129 real cells. We then computed the dot product of each real cell PCA coordinate vector with each
130 sample's background vector, calling this the "background loading" of a given cell for a given
131 sample (each cell actually comes from exactly one sample, but computing each cell's loading for
132 each sample's background made the next step mathematically/computationally simpler). Next,
133 we fit a linear regression model, real cell PCA coordinate matrix ~ cell background loadings, and
134 called its residuals the "background corrected PCA matrix." This background correction method
135 is similar to, but developed independently of, a recently published method (55).

136 We found that the UMAP (22, 23) algorithm, which provides a way to project the data
137 into a low-dimensional space, better maintains the topology of the dataset compared to the
138 commonly used t-SNE algorithm. In our dataset, UMAP often creates long, continuous

139 trajectories, while t-SNE clusters distinct cell types but does not clearly show the relationships
140 between them. UMAP and t-SNE have been compared in the context of sc-RNA-seq by Becht *et*
141 *al.* (23), but this paper focuses on the empirical performance of the algorithms and does not
142 explain precisely how and why the mathematical differences between the algorithms underlie
143 their qualitatively different results. We chose UMAP over t-SNE based on our subjective
144 evaluation of how the two algorithms' results compared to our expectations given the known *C.*
145 *elegans* lineage.

146 We reduced the dimensionality of the background corrected PCA matrix to 2 or 3
147 dimensions using UMAP, using the wrapper function for this algorithm provided by the Monocle
148 software package, version 3 alpha (the reduceDimension function). The UMAP parameters were:
149 metric = "cosine", min_dist = 0.1, n_neighbors = 20.

150 Lastly, cells in the UMAP space were clustered using the Louvain algorithm (25). The
151 Louvain algorithm is one of several algorithms that group nodes in a weighted, undirected graph
152 into clusters in a way that seeks to maximize a statistic called "modularity." Modularity is
153 essentially the difference between the total edge weight between nodes assigned to the same
154 cluster and the expectation of the total within-cluster edge weight if all edges were randomized.
155 Exact optimization of modularity is computationally intractable for large graphs, so the Louvain
156 algorithm uses a heuristic. In the context of our study, the graph used for the Louvain algorithm
157 is a k -nearest neighbor graph ($k = 20$) constructed from cell coordinates in UMAP space.

158 For more details on the UMAP and Louvain algorithms, we refer the reader to these web
159 resources:

- 160 • A talk by Leland McInnes describing the intuition behind UMAP:
161 <https://www.youtube.com/watch?v=nq6iPZVUxZU>.

- 162 • The section in the UMAP paper on arxiv (22) titled “A Computational View of
163 UMAP”.
164 • An explanation of the Louvain algorithm on Quora: <https://www.quora.com/Is-there-a-simple-explanation-of-the-Louvain-Method-of-community-detection>.
165
166

167 **Doublet identification**

168

169 We used two complementary methods to identify doublets. The first method involved
170 identifying clusters of doublets in iterated UMAP projections of the data on the basis of co-
171 expression of high-confidence cell type specific marker genes, reported in WormBase (26), for
172 >1 cell type (e.g. a cluster expressing the muscle markers *myo-3* and *pat-10* along with the
173 neuron markers *egl-21* and *sbt-1* was considered a muscle-neuron doublet cluster). We applied
174 this simple approach to a global UMAP of all cells and iterated UMAPs of tissues / related
175 groups of cells from the global UMAP (e.g. muscle, intestine, ciliated neurons, etc.).

176 The second approach involved logistic regression models, one for each broadly-defined
177 terminal cell type (e.g. body wall muscle, intestine, ciliated neurons, non-ciliated neurons, etc.),
178 that predict whether a cell is part of that cell type or not. We fit one such model for each broadly-
179 defined cell type and used the models to score each cell for the probability of it being a member
180 of each broadly-defined cell type. Cells that had ≥ 2 cell types with a $\geq 20\%$ predicted
181 probability of the cell being a member of that cell type were considered doublets. Clusters in the
182 UMAP projections that were enriched for cells considered doublets by these regression models
183 were manually examined, and in some cases manually filtered.

184 Due to the abundance of cell type specific marker genes, we estimate that we were able to
185 filter out almost all terminal cell type doublets. Residual expression of genes from one cell type
186 in a cluster corresponding to another cell type appears to be driven by background RNA
187 contamination, not doublets. Our approach is less likely to catch doublets between progenitor
188 cells that do not yet express marker genes of differentiated terminal cell types. For earlier-stage
189 embryos however, the cell dissociation protocol works more reliably than for late stage embryos,
190 so we expect the doublet rate to be close to the reported rate for the 10X Genomics Chromium
191 platform, which is low (~4.5% given ~9k cells loaded per lane).

192 While performing downstream analyses, we noticed that a few cell types were missing
193 from our data, including rectal epithelial and gland cells, the excretory duct and pore, and the T
194 cell. These were erroneously excluded by our doublet filter due to co-expressing genes that were
195 enriched in two or more tissues (e.g. co-expressing hypodermis-enriched genes with pharynx-
196 enriched genes). We used marker genes to identify these cells in a non-doublet-filtered global
197 UMAP, whitelisted them, and included them in the appropriate tissue UMAPs (**Figs 3A, S9-11,**
198 **S13**). These cells are not included in the global UMAP (**Fig. 1A**).

199

200 **Embryo time estimation**

201

202 For each cell, we estimated the age of the embryo that the cell came from (“embryo
203 time”) based on Pearson correlation of its transcriptome with bulk RNA-seq time series data
204 from Hashimshony *et al.* (21). Their data show that the majority of genes that change expression
205 over time in any given lineage are not lineage specific. Thus, we first defined a list of genes with
206 time-dependent expression patterns, requiring an auto-correlation greater than 0.6 and standard

207 deviation greater than 1.5 across bulk RNA-seq time points (units = log TPM). Pearson
208 correlation was then computed between log-scaled single cell and bulk data using only the time-
209 dependent genes. We observed for non-multiplet cells, the Pearson correlation across time shows
210 a strong peak pattern (**Fig. S1A**). Thus, by fitting a loess regression curve and finding its
211 maximal point, we were able to assign each cell with its most correlated bulk time point.

212 Embryo times estimated based on data from Hashimshony *et al.* (21) approximately agree
213 with embryo collection times from our experimental design (**Fig. S1B**), and also have a strong
214 correlation with embryo times estimated based on data from Boeck *et al.* (56) (**Fig. S1C**). To
215 further validate our embryo time estimates, we computed for each anatomical cell in the *C.*
216 *elegans* embryonic lineage the 5th percentile of the embryo times for the set of sc-RNA-seq cells
217 that we annotated as corresponding to that anatomical cell. This effectively estimates the birth
218 time of the anatomical cell. These cell birth time estimates correlated well with cell birth time
219 estimates derived from live imaging (30) (**Fig. S1D**).

220 In the Waterston lab samples, embryos were incubated for a specific amount of time after
221 hypochlorite treatment. However, each sample has some outlier cells with abnormally low
222 embryo time estimates, i.e. lower than the incubation time. There are several biological and
223 technical factors that could produce these outlier cells. The developmental rate of *C. elegans*
224 embryos can vary by over 2-fold depending on temperature, and may also be influenced by
225 differences in crowding, hypoxia, or the effects of hypochlorite and chitinase treatment.
226 Consistent with this, embryo times estimated using data from Boeck *et al* (56), which was
227 collected using methods more similar to those used in this study, were systematically later than
228 embryo times estimated using data from Hashimshony *et al.* (21) (**Fig. S1C**). Alternatively, some
229 cells may have embryo time estimates that are lower than the true developmental age of the

230 embryo they came from. Sparsity in the single cell data contributes to noise in the estimates.
231 Finally, the most extreme outlier embryo time estimates in each sample are for germline cells.
232 The germline maintains expression of many genes that turn off during early embryogenesis in all
233 other cells. This causes embryo time estimates based on correlation to bulk RNA-seq to be
234 inaccurate for this cell type.

235

236 **Per-cell background correction and filtering**

237

238 Our method for correcting for background RNA contamination, described in the section
239 above titled “Dimensionality reduction”, works solely on the level of PCA coordinates and does
240 not change the underlying gene-by-cell expression matrices. We used a separate background
241 correction method to adjust these gene expression matrices on a per-cell basis for purposes of
242 making plots of gene expression.

243 Our per-cell background correction method relies on a panel of cell-type specific marker
244 genes that are assumed, based on the literature (and confirmed empirically in our data), to be
245 specific to either hypodermis (including seam and P cells) or body wall muscle (BWM). The
246 hypodermis-specific genes were: *sqt-3*, *dpy-17*, *dpy-14*, *dpy-10*, *dpy-7*, *dpy-2*, *dpy-3*, *bus-8*, *wrt-2*,
247 and *noah-1*. The BWM-specific genes were: *pat-10*, *mlc-3*, *cpn-3*, *clik-1*, *ost-1*, *mlc-1*, *mlc-2*,
248 *ttn-1*, *unc-15*, and *myo-3*.

249 The gene expression distribution for the background contamination of each biological
250 sample was estimated by aggregating the reads for cell barcodes that had < 50 UMIs, which were
251 assumed to correspond to empty droplets in the 10X sc-RNA-seq apparatus. The expression level
252 of each gene in the panel was computed for each sample’s background, measured in transcripts

253 per million (TPM). Similarly, the expression level of each gene in the panel was computed for
254 each cell, also measured in TPM. The background fraction of a cell was estimated as the sum of
255 the expression of panel genes in the cell divided by the sum of the expression of panel genes in
256 the background distribution for the sample that cell came from. For cells annotated as
257 hypodermis, glia, or potential progenitors of those cell types, hypodermis-specific genes from the
258 panel were excluded from the computation. Likewise, for cells annotated as body wall muscle,
259 intestinal/rectal muscle, or a non-pharyngeal mesoderm cell type, as well as progenitors of those
260 cell types, BWM-specific genes from the panel were excluded from the computation. For all
261 other cells, all genes from the panel were used.

262 The median estimated background fraction across all cells in the dataset was 17.7%.
263 Putatively damaged cells with an estimated background fraction $\geq 75\%$ (8.3% of all cells, see
264 **Fig. S35A**) were filtered entirely from all subsequent plots and analyses. For the remaining cells,
265 the cells' gene expression profiles were corrected to subtract the contribution from background.
266 A cell's raw gene expression vector (UMI counts) was converted to transcripts per million by
267 dividing each entry by the sum and multiplying by one million. The background-corrected TPM
268 value for each gene was computed according to the formula:

269
270 background-corrected TPM =
271
$$\max(\text{raw TPM} - \text{background fraction} * \text{background TPM}, 0)$$

272 where background TPM is the expression of the given gene in the background distribution for
273 the biological sample that the cell came from. The background-corrected corrected TPM values
274 were then rescaled to once again sum to 1,000,000 and then converted back into (pseudo-)counts

276 based on the total UMI count of the cell. Fractional count values were rounded probabilistically
277 (i.e. a value of 2.7 was rounded to 3.0 with a 70% chance and to 2.0 with a 30% chance).

278 After background correction, cells with low background fractions and cells with high
279 background fractions have near-identical average gene expression profiles (**Fig. S35B**). This
280 indicates that non-background gene expression observed in high background cells is not
281 systematically biased compared to low background cells.

282

283 Computing aggregate gene expression profiles for cell types and lineages

284

285 To compute the aggregate gene expression profile for a cell type (**Table S7**) or lineage
286 (**Table S8**), we (1) subsetted the whole-dataset gene-by-cell gene expression matrix to include
287 just the cells annotated as the given cell type or lineage; (2) divided each column by the
288 corresponding cell’s size factor (a statistic computed by the Monocle software package equal to
289 the cell’s total UMI count divided by the geometric mean of all cells’ total UMI counts); (3) took
290 the mean of each row (gene); and (4) rescaled the resulting vector to sum to 1,000,000. This
291 results in a gene expression vector measured in transcripts per million (TPM).

292 We performed these computations using the original gene expression matrix, not
293 corrected for background RNA contamination. After computing the aggregate gene expression
294 vector for a cell type or lineage, we then corrected it for background contamination using the
295 same method described in the previous section (“Per-cell background correction and filtering”),
296 treating the aggregate vector as if it were the expression vector for a single cell. Compared to the
297 alternative option of correcting each cell for background first and then computing the aggregate

298 profile, aggregating first then correcting makes the estimate of the background fraction more
299 robust due to an increased sample size (number of reads).

300 Even after background correction, we noticed some residual, aberrant expression of genes
301 that should not be expressed in a given cell type. In several cases, this aberrant expression was
302 due to just one or two outlier cells within a given UMAP cluster. We suspected these outlier cells
303 included both doublets missed by our filtering procedure and “pseudo-doublets”, consisting of a
304 real cell plus debris from another cell. In order to reduce the impact of these outliers, in **Tables**
305 **S7-8** we report a “robust” estimate of the mean expression for a gene in a given cell type or
306 lineage (in addition to the “raw” estimate). This robust estimate excludes the highest-expressing
307 cell and the lowest-expressing cell for a given gene in a given cell type / lineage before
308 computing the mean expression.

309 The impact of outliers is greater for cell types represented by only a small number of cells
310 in our dataset, and for cell types that have a low average number of UMIs per cell. Estimates of
311 mean gene expression values are therefore less precise for these cell types. To estimate the
312 variance of our mean gene expression statistics, we used bootstrap resampling: for a cell type
313 with N cells in our dataset, we randomly sampled, with replacement, N cells from that set and
314 computed mean expression statistics from that sample. We repeated this process for 1,000
315 iterations for each cell type and computed bootstrap confidence intervals from the resulting
316 distribution of mean estimates. If one must make a statement that a gene is “expressed” in a
317 given cell type or lineage, we recommend using the criterion that the lower bound of the 95%
318 bootstrap confidence interval is >0 TPM.

319
320

321 **Differential expression analysis for Fig. 3D and Fig. S22**

322

323 We included four classes of transcription factors (TFs) in the heatmaps of **Fig. 3D** and
324 **Fig. S22**. Both figures consider differential expression of TFs between different ciliated neuron
325 lineages. For the division of a parent neuroblast into two daughter cells, the four TF classes of
326 interest were:

- 327 1. TFs enriched in one daughter vs. the parent and vs. the other daughter
328 2. TFs depleted in one daughter vs. the parent and vs. the other daughter
329 3. TFs enriched in the parent vs. both daughters and vs. other neuroblasts of the same cell
330 generation
331 4. TFs enriched in parent vs. other neuroblasts of the same cell generation; and in both
332 daughters vs. other terminal cells

333 We considered a TF “enriched” in cell set A vs. cell set B if the expression in A was at least 3-
334 fold higher than in B; and if the difference in expression was statistically significant with q-value
335 < 0.01 . We considered a TF “depleted” in cell set A vs. cell set B if it was “enriched” in B vs. A.
336 q-values were computed using the Monocle (version 3 alpha) function “differentialGeneTest”.
337 Differential expression tests were performed for all genes, not just TFs—the non-TF results were
338 discarded, but this was done to produce more conservative q-values compared to considering
339 only TF DE tests. Cells with embryo time >650 minutes were excluded from all comparisons.
340 Due to limited figure space, some TFs that matched the criteria of the four TF classes but had
341 low absolute expression levels were excluded from the figure heatmaps.

342

343

344 **Derivation of lineage specific and terminal cell type specific genes for Fig. 4D**

345

346 Lineage specific genes were derived by one vs. rest differential expression analysis on the
347 three input branches based on Louvain clustering results and annotations from **Fig. S15**, using
348 “sSeq” (57), as implemented in the cellrangerRkit package. Genes associated with IL1/IL2
349 terminal cell types were derived by comparing IL1/IL2 cells to all other ciliated neurons in **Fig.**
350 **3A**. For each of the gene sets, the average TPM across all genes in the set was computed for cells
351 from each of the three input branches, binned in 30 minute intervals up to 390 minutes, where
352 the branches can no longer be distinguished from each other in the UMAP. Values in each
353 heatmap were linearly rescaled to be within the range of 0 to 1.

354

355 **Pseudo- R^2 statistic used in Fig. 5B and Fig. S29B**

356

357 For each anatomical cell annotated in our dataset, we compute an aggregate gene
358 expression profile from all of the sc-RNA-seq cells that we annotated as corresponding to that
359 anatomical cell. This procedure is described in above section titled, “Computing aggregate gene
360 expression profiles for cell types and lineages.” The result is that each anatomical cell is
361 associated with a vector of relative gene expression values. We refer to this vector as the
362 anatomical cell’s “transcriptome.”

363 In **Fig. 5B** and **Fig. S29B**, we seek to estimate the extent to which the transcriptomes of
364 cells in a given generation of the AB or MS lineages are predicted by the lineage. To do this, we
365 have defined a statistic that measures how much more similar, on average, are the transcriptomes
366 of sister cells compared to random pairs of cells. Specifically, we compute:

367
368 average Jensen-Shannon divergence between the transcriptomes
369 of pairs of sister cells in the cell generation

370 1 - _____
371 average Jensen-Shannon divergence between the transcriptomes
372 of random pairs of cells in the cell generation

373

374 In the main text and figures, we refer to our statistic as a pseudo- R^2 statistic. The so-
375 called pseudo- R^2 statistics are a family of statistics that have been proposed in the context of
376 generalized linear regression models (58) and aim to have similar properties to the coefficient of
377 determination, R^2 , that is commonly used in the analysis of ordinary linear regression models.
378 Similarly, the statistic we have defined aims to have similar properties to R^2 , despite not being
379 mathematically comparable to it in a rigorous sense. Below, we discuss the similarities between
380 our pseudo- R^2 statistic and R^2 .

381

382 One of several equivalent definitions of R^2 for an ordinary linear regression model is:

383
384 mean squared error of the regression model's predictions
385 1 - _____
386 overall variance of the response variable

387

388 This formula for R^2 and our formula for pseudo- R^2 are both expressed in terms of a
389 fraction subtracted from one. The numerator in our formula for pseudo- R^2 , which we defined in
390 terms of the Jensen-Shannon divergence, can be re-expressed as the average prediction error of a
391 certain regression model, analogous to the numerator of regular R^2 .

392 Specifically, the numerator in our pseudo- R^2 is equivalent to the average prediction error of a
393 model that:

- 395 1. seeks to predict a cell's transcriptome based on the identity of its parent.
396 2. measures the deviation between its predicted transcriptome and the observed
397 transcriptome for a cell using Kullback-Leibler (KL) divergence.

398

399 This equivalence is a consequence of the following:

- 400 1. When tasked to predict the transcriptomes of two sister cells, a model that predicts a
401 cell's transcriptome based on the identity of its parent effectively guesses the midpoint of
402 the two sister cells' transcriptomes.
- 403
- 404 2. Therefore, if one measures the deviation between the model's predictions and the
405 observed transcriptomes using KL divergence, then the mean prediction error of the
406 model, when applied to pairs of sister cells, is simply the average KL divergence between
407 each cell's transcriptome and the midpoint of it and its sister's transcriptomes.
- 408
- 409 3. By the definition of Jensen-Shannon (JS) divergence, this is the same as the average JS
410 divergence between each pair of sister cells' transcriptomes, which is the numerator used
411 in our pseudo- R^2 .

412

413 The denominator of our formula for pseudo- R^2 , the average JS divergence between the
414 transcriptomes of random pairs of cells, is a measure of the overall variability in the
415 transcriptomic data. This is analogous to the denominator of regular R^2 , which is also a measure
416 of the overall variability (i.e. the variance) of the response variable in an ordinary linear
417 regression model.

418 Thus, both the numerator and the denominator in our formula for pseudo- R^2 are
419 qualitatively similar measurements to the numerator and denominator of regular R^2 .

420

421 **Methods used in Fig. S30**

422

423 In **Fig. S30**, we estimate the extent to which the ability of lineage to predict the
424 transcriptome in a given cell generation, “generation N”, is a consequence of gene expression
425 signatures associated with each of the preceding cell generations 1 to N-1. We compute the
426 overall ability of the lineage to predict the transcriptome in generation N using the pseudo- R^2
427 statistic described in the previous section. To compute the contribution of the parent generation
428 N-1 to the total pseudo- R^2 for generation N, we use the formula:

429

$$\frac{(\text{average JS divergence between cells that share a grandparent} - \text{average JS divergence between sisters})}{\text{average JS divergence between random pairs of cells}}$$

434

435 This formula evaluates how much more similar are cells that share a parent (i.e. sisters) than cells
436 that share a grandparent (i.e. cousins or sisters), and scales this relative to the average
437 dissimilarity of random pairs of cells in the same generation.

438

439 Generalizing this formula, we estimate the contribution of the generation $N - M$ as:
 440

$$\frac{(\text{average JS divergence between cells with lineage distance } \leq M+1 - \text{average JS divergence between cells with lineage distance } \leq M)}{\text{average JS divergence between random pairs of cells}}$$
 444

445 where the lineage distance between two cells is the number of cell divisions since their most
 446 recent common ancestor (1 for sisters, 2 for cousins, etc.).
 447

448 Using this formula, the sum of the contributions of each ancestor generation 1 to $N-1$
 449 simplifies to:
 450

451

$$\frac{(\text{average JS divergence between cells with lineage distance } \leq N-1 - \text{average JS divergence between cells with lineage distance } \leq 1)}{\text{average JS divergence between random pairs of cells}}$$
 454

455 All cells in generation N have lineage distance $\leq N-1$, so the first term in the numerator is equal
 456 to the average JS divergence between random pairs of cells (same as the denominator).
 457

458 Furthermore, the only cells with lineage distance ≤ 1 are sisters. Making these substitutions, we
 459 get:
 460

461

$$\frac{(\text{average JS divergence between random pairs of cells} - \text{average JS divergence between sisters})}{\text{average JS divergence between random pairs of cells}}$$
 464

465 Which simplifies to our original statistic for total pseudo- R^2 :
 466

467

$$1 - \frac{\text{average JS divergence between sister cells}}{\text{average JS divergence between random pairs of cells}}$$
 470

471 **Definition of lineage signature transcription factors for Fig. 5C-E and Fig. S29C-E**

472

473 For the analyses presented in **Fig. 5C-E** and **Fig. S29C-E**, we introduced a concept of
474 “lineage signature transcription factors.” For each anatomical cell that we annotated in our data,
475 we defined the set of lineage signature transcription factors associated with that cell to be set of
476 TFs that satisfy both of the following criteria:

- 477 1. The lower bound of the 95% bootstrap confidence interval for the average expression
478 level of the TF in the cell, as reported in **Table S8**, is >0 . In other words, the TF must be
479 robustly expressed in the cell.
- 480 2. The TF must be expressed at least 5-fold higher in the cell compared to its sister in the
481 lineage, with a differential expression q-value < 0.01 .

482

483 **Computing the adjusted Gini coefficient for Fig. S33A**

484

485 The Gini coefficient is biased by sample size (59). Therefore, to adjust for total UMI
486 count differences between cells, we first downsampled the data from each cell to a total of 500
487 UMIs (the minimum UMI count across all cells) using a multinomial distribution, with
488 probability equal to each gene’s UMI count divided by the total UMI count of the cell. We then
489 computed Gini coefficients for each cell using the downsampled data, and used the z-score of the
490 adjusted Gini coefficients to compare transcriptome inequality across cells.

491

492

493 **Comparison of data from this study to single cell RNA-seq data from Tintori *et al.*, 2016 (8)**

494

495 Due to technical limitations, we have data from relatively few cells prior to the 28-cell
496 stage. Therefore, we compared single cell RNA-seq profiles of cells from the 16-cell stage
497 collected by Tintori *et al.* (8) to their corresponding lineages or immediate descendants in our
498 dataset (**Fig. S19**). We downloaded normalized expression data (measured in reads per kilobase
499 of transcript per million mapped reads, RPKM) from Tintori *et al.* (8) and computed average
500 log₂ normalized expression levels for each of their annotated lineages. We then applied the same
501 log₂ transformation on our normalized gene expression data, and measured pairwise similarity
502 between the expression vectors for each lineage using Pearson correlation. To enrich for lineage-
503 specific signals, we computed correlation using gene sets that had been selected by Tintori *et al.*
504 (8) using an iterative PCA approach. Gene sets 7, 9, and 10 in supplemental document S1 of
505 Tintori *et al.* (8) were used to discriminate 16-cell stage lineages. Set 8 was excluded because
506 most germline (P4) specific genes are also differentially expressed over time throughout the
507 whole embryo and thus confound time with lineage. Intersecting genes from sets 7, 9, and 10
508 with genes detected in our data, we obtained a list of 593 genes that we then used to generate the
509 correlation matrix shown in **Fig. S19A**. Hierarchical clustering was performed on the correlation
510 matrix using the pheatmap package with default parameters (60).

511 To demonstrate that our data are consistent with Tintori *et al.* (8) at the level of single
512 cells, we repeated their PCA analysis and projected 16- and 28-cell stage cells from our dataset
513 onto the PCA space derived from their dataset (**Fig. S19B-E**). The distribution and orientation of
514 lineages in the PCA space was similar for our and their data. For example, the PCA in the top
515 sub-panel of **Fig. S19B** was computed using all 16-stage cells from Tintori *et al.* (8) and 421

516 genes (intersection of Set 7 and expressed genes in our data). In the bottom sub-panel of **Fig.**
517 **S19B**, we project 292 cells from the 16- and 28-cell stages from our dataset using the loading
518 matrix derived from the PCA of the Tintori *et al.* data (8). Germline (P4, Z2/Z3) and endoderm
519 lineage (Ex, Exx) cells from our data are located at the left and right-hand sections of the PCA
520 projection respectively, consistent with the pattern observed with cells from Tintori *et al* (8).

521

522 **Comparison of data from this study to data from Spencer *et al.*, 2011 (29)**

523

524 Spencer *et al.* (29) used microarrays to profile the transcriptomes of *C. elegans* cell types
525 obtained by fluorescent activated cell sorting. For each cell type they profiled, they derived a set
526 of genes that are enriched in that cell type compared to all other cells. We used these “signature”
527 gene sets to validate our cell type annotation and the robustness of our data. First, we
528 downloaded signature gene sets from cell types profiled at the embryonic stage from
529 https://www.vanderbilt.edu/wormdoc/wormmap/Enriched_genes.html. We then used the AUCell
530 package (61) to check for enrichment of Spencer *et al.* (29) signature genes in single cells from
531 our dataset. For each cell, AUCell ranks genes by expression level and computes a recovery
532 curve for each gene set. It then uses “Area Under the Curve” (AUC) as a measure of enrichment
533 of the gene set.

534 We found most Spencer *et al.* (29) signature genes have strong enrichment in the
535 corresponding cell types in our data (**Fig. S20**). Due to the method by which the Spencer *et al.*
536 (29) signature genes were derived—comparing one cell type to all other cells—most of the genes
537 are tissue-specific, not cell-type specific, so enrichment was in some cases also observed in a set
538 of several related cell types in our data.

539 Spencer *et al.* (29) signature genes for pharyngeal muscle were unusual in that they were
540 enriched in intestine cells from our dataset. Examining the pharyngeal muscle gene set, we
541 noticed it contains *elt-2* and *elt-7*, which are known to be endoderm specific (62). Checking this
542 gene set against expression patterns from Warner *et al.*, 2019 (54), we found that 18 out of the
543 top 20 genes are intestine specific/enriched. Therefore, we concluded the pharyngeal muscle
544 signature list is problematic and dropped the comparison from **Fig. S20**.

545

546 **Supplementary Text**

547

548 **Supplemental Note 1**

549

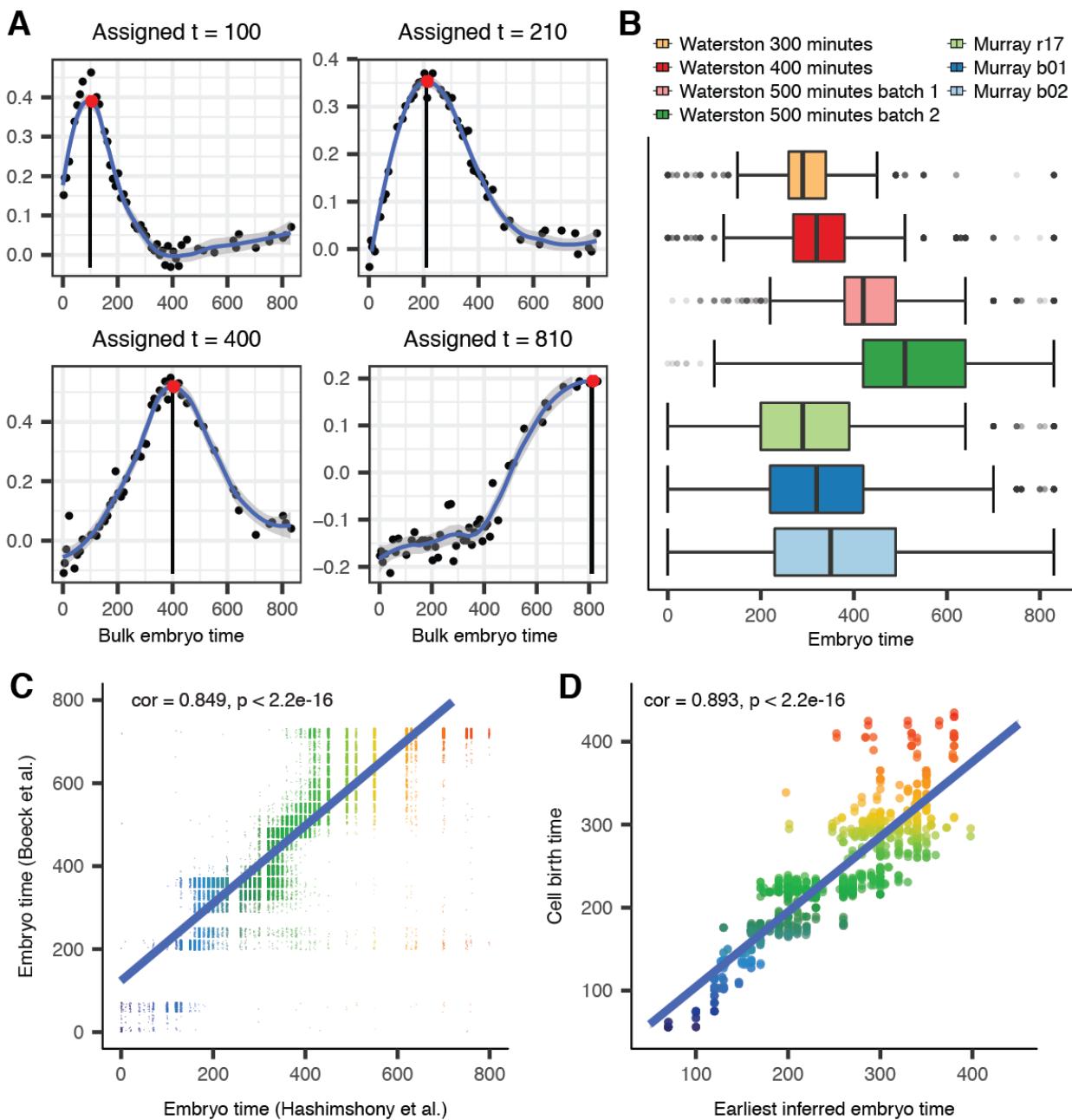
550 In the analysis of single cell RNA-seq data, the term “trajectory” is often used to refer to
551 a group of cells that 1) represent a specific cell lineage or cell type, and 2) can be ordered in a
552 way that reflects the cells’ progression over time from one transcriptomic state to another.

553 Algorithms for trajectory inference can construct such orderings by mapping high dimensional
554 gene expression data to a low dimension and fitting a graph structure to the data points in the low
555 dimension. The distance between a user-defined root vertex on the graph to the location of a cell
556 on the graph is called the cell’s “pseudotime,” and for a particular path along the graph, the
557 ordering of cells on that path by pseudotime defines a “trajectory” (the graph as a whole can be
558 considered a “branched trajectory”).

559 In this manuscript, we have not used any trajectory inference algorithm. Instead, we use
560 our embryo time estimates for each cell, which are computed based on correlation of the cell’s
561 transcriptome with a high-resolution bulk RNA-seq time series (see **Methods** and **Fig. S1**), as a
562 universal ordering for all cells in the dataset. We annotate cell types and lineages based on
563 marker genes from the literature (**Table S1, S4**) and clustering in UMAPs of our data. Ordering
564 cells with a common cell type or lineage annotation by embryo time defines a “trajectory”. In
565 most cases, trajectories defined in this way also form contiguous shapes in UMAPs, to which one
566 could fit a graph structure to if desired. Our universal ordering of cells by embryo time is a more
567 robust approach, however, since 1) it allows annotations from multiple UMAPs to be integrated
568 to define a single “trajectory” per cell type/lineage, and 2) the approach still works even when

569 cells of a common type/lineage are split into disparate, non-contiguous groups in a UMAP
570 (which may occur due to abrupt changes in gene expression or due to technical factors, such as
571 non-uniform sampling with respect to time).
572

573 **Supplementary Figures**



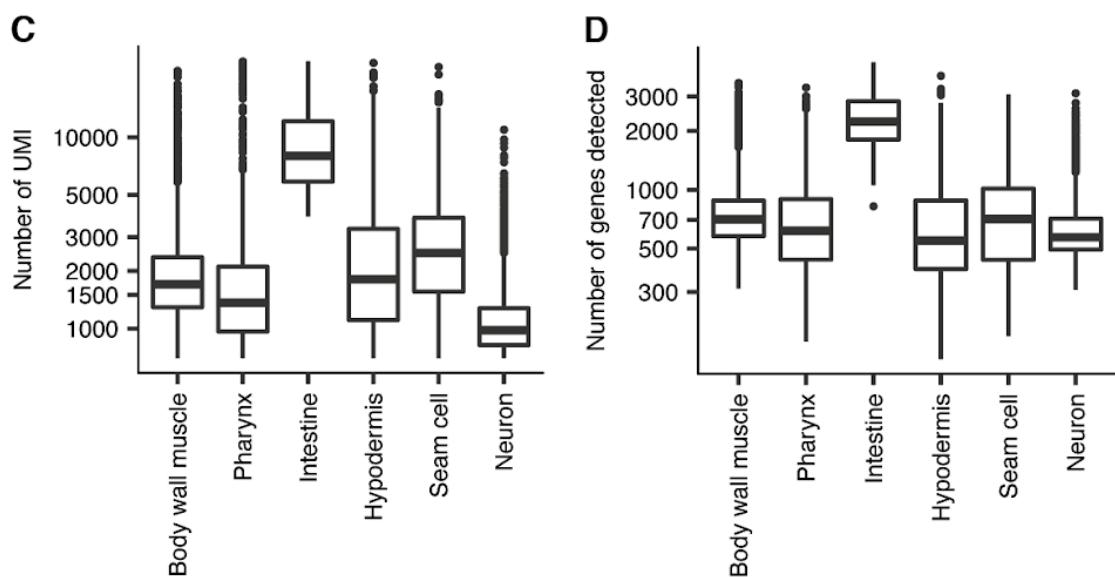
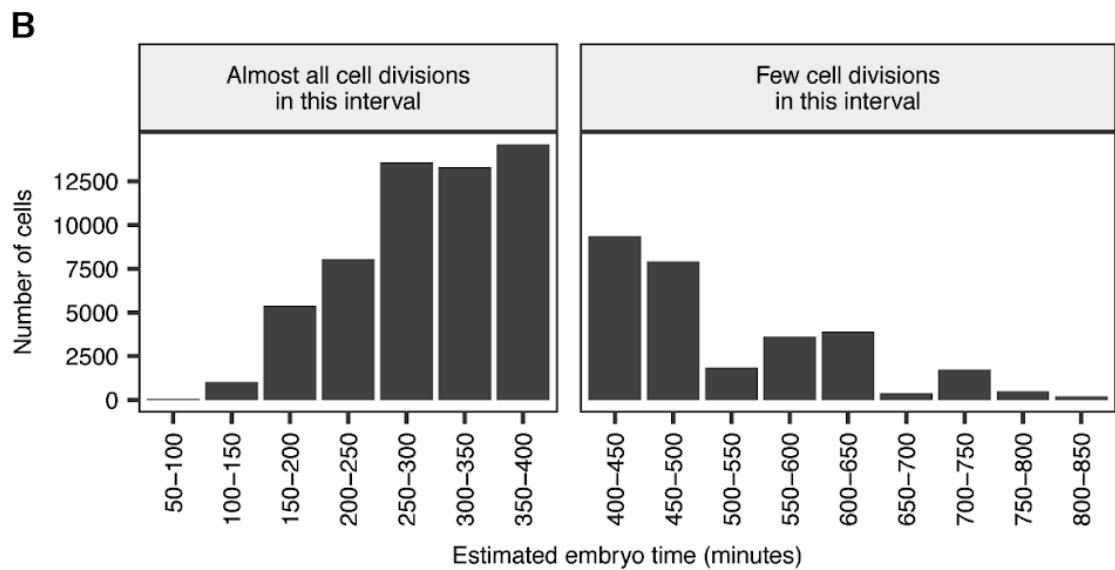
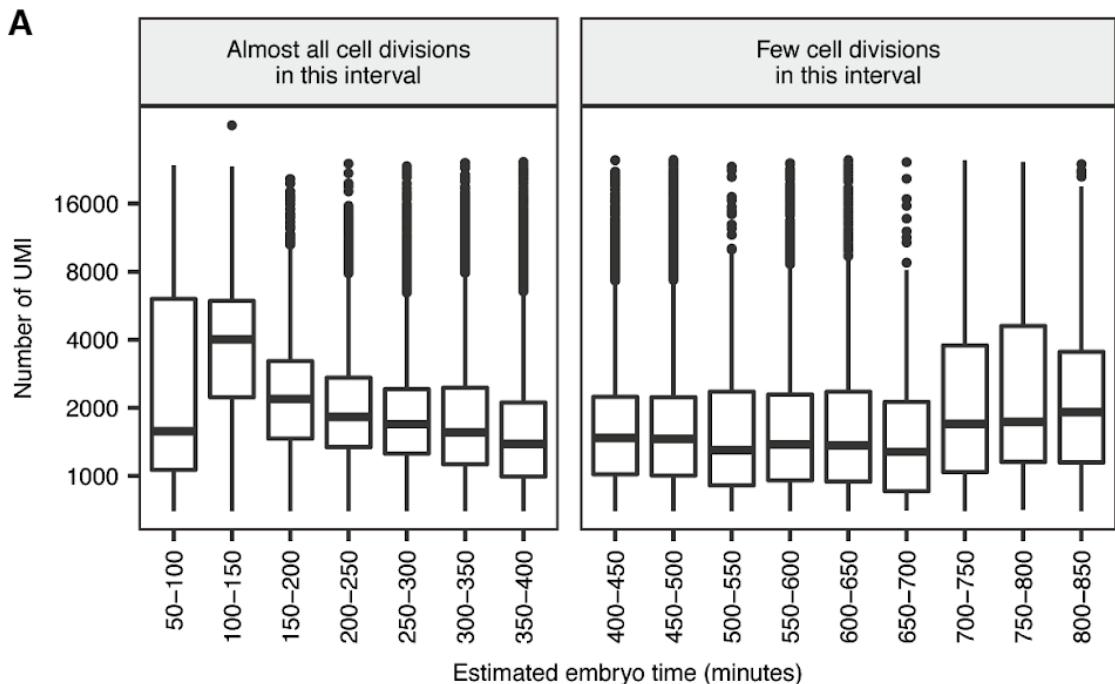
574

575 **Fig. S1. Method for estimating the age of the embryo that a sc-RNA-seq cell came from.**

576 Embryo times are measured in minutes post first cleavage. (A) Embryo times are estimated
577 based on Pearson correlation of a single cell's transcriptome to a bulk RNA-seq time series (see
578 **Methods**). Pointwise estimates of the correlation to each time point are smoothed using a loess

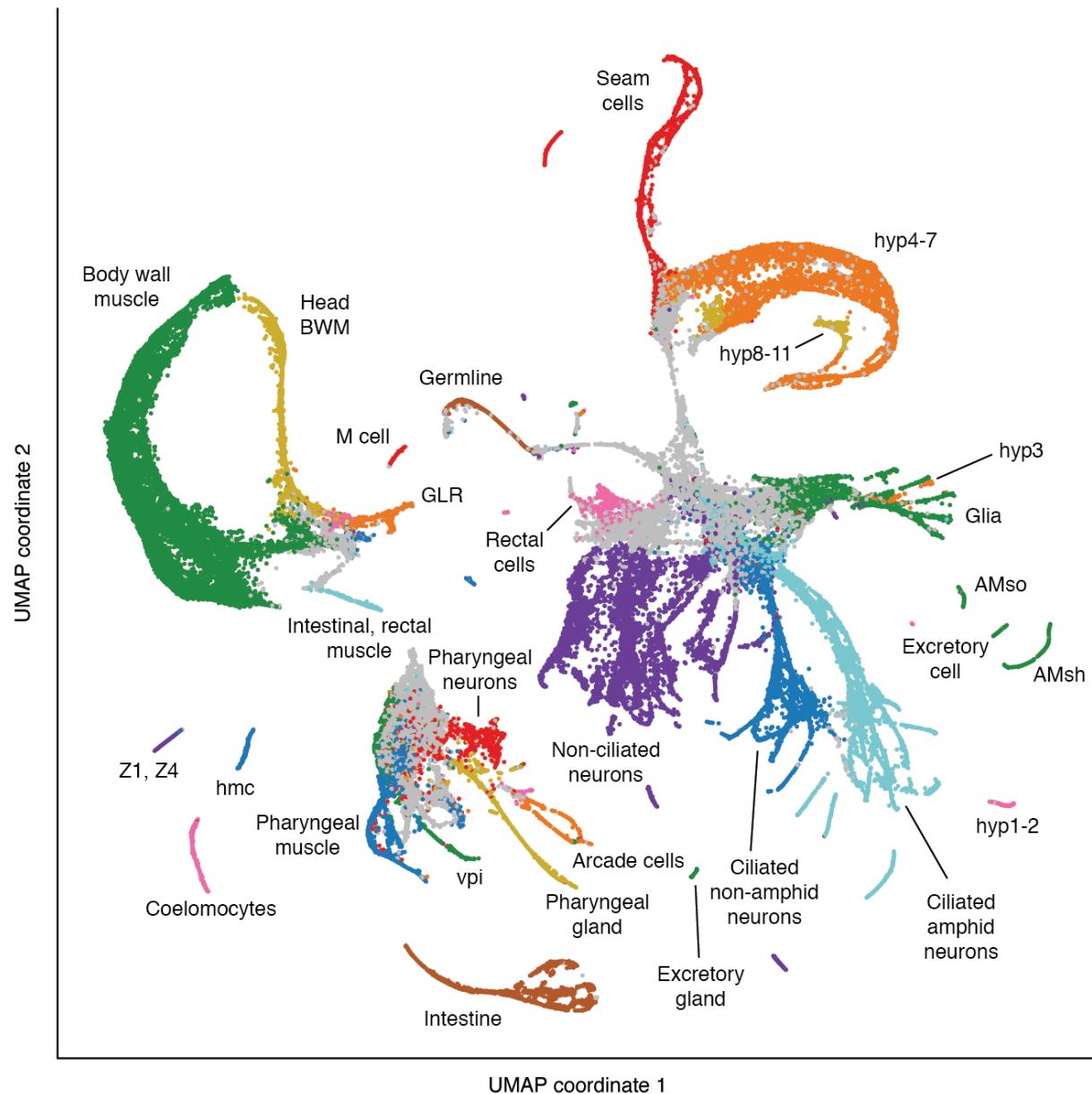
579 regression. **(B)** Distribution of estimated embryo times for each biological sample. The average
580 embryo time estimate in the Waterston lab sample correlates with the real time duration that the
581 embryos were incubated. Each sample contains some outlier cells with abnormally low embryo
582 times. Potential biological and technical causes for the presence of these outlier cells are
583 discussed in the Methods. **(C)** Correlation of embryo time estimates based on Hashimshony *et al.*
584 (21) to an alternate set of embryo time estimates based on Boeck *et al.* (56). Estimates based on
585 Hashimshony *et al.* (21) were used for all downstream analyses. **(D)** Correlation between cell
586 birth times estimated based on our lineage annotations (x-axis) with cell birth times computed
587 based on automated analysis of imaging data (y-axis) (30).

588



590 **Fig. S2. UMIs recovered per cell decreases with embryo age.** All Y-axes are log scaled. **(A)**
591 Distributions of number of UMIs recovered per cell, binned by estimated embryo age. Median
592 UMIs per cell decreases until ~400 minutes, after which almost all cell division has stopped.
593 Comparing each embryo time bin on the X-axis to the subsequent bin, e.g. comparing 100-150
594 minutes to 150-200 minutes, the decrease in median UMIs per cell is statistically significant for
595 each step from 100-400 minutes (Wilcoxon rank sum tests, all p-values < 2.2e-16). Note that our
596 quality control procedures exclude cells with < 700 UMIs (or < 500 UMIs for neurons), causing
597 the decrease in UMIs/cell to be understated, as the proportion of cells falling below the cutoff is
598 greater for later stage embryos. **(B)** Number of cells included in each time bin from panel A. **(C)**
599 and **(D)** Number of UMIs and genes detected for cells with embryo time in the range of 390-650
600 minutes, by tissue.

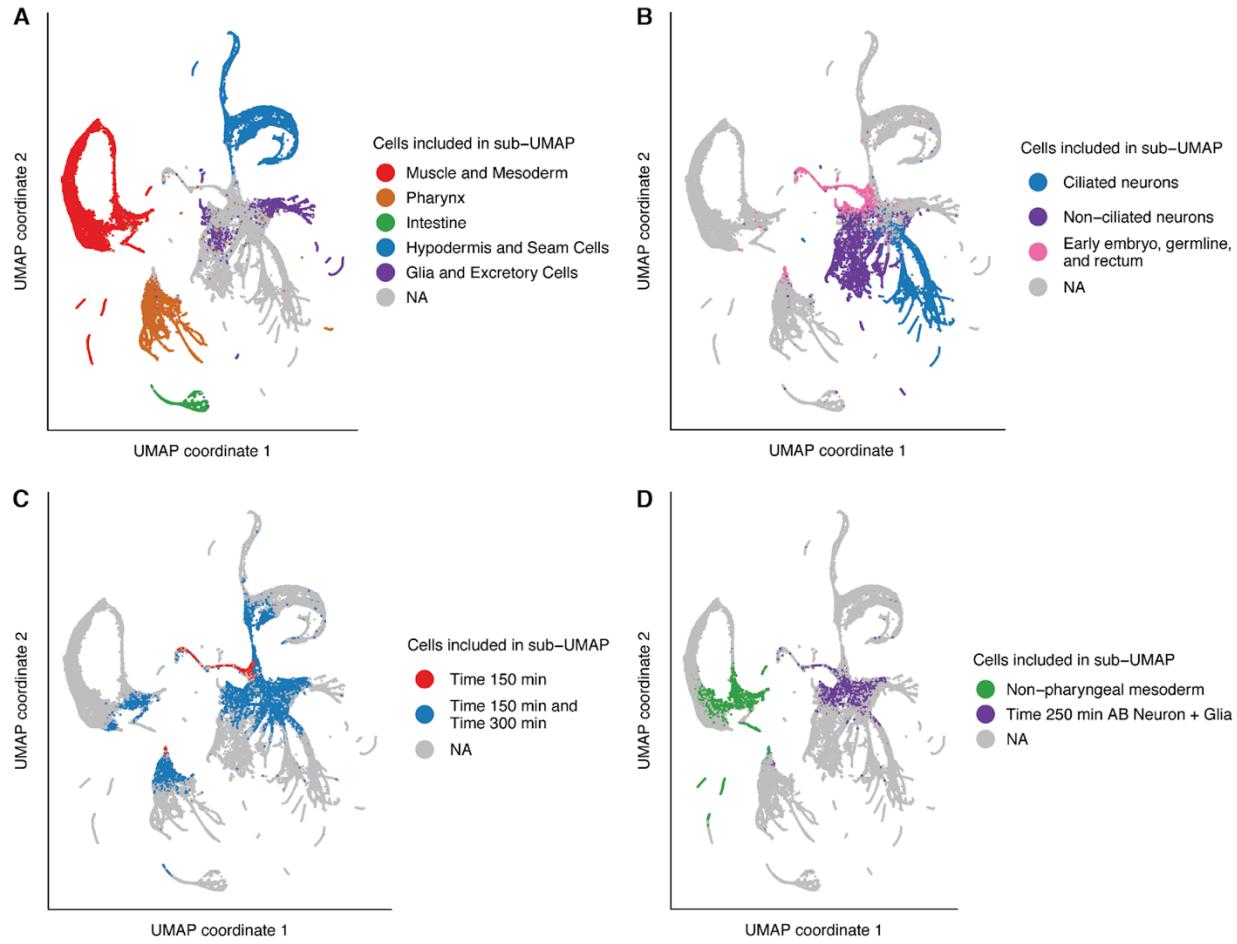
601



602

603 **Fig. S3. Cell type annotations for the global UMAP of 81,286 cells.** This plot shows more cell
 604 type annotations for the global UMAP from **Fig. 1A**. This UMAP does not include 4,738
 605 additional cells that were initially filtered, but were later whitelisted and included in downstream
 606 analyses (see **Methods**). For fine-grained annotations of cell types in each major tissue, see **Figs.**
 607 **3A and S5-13**. For fine-grained annotations of progenitor cell lineages, see **Fig. S14-17**.

608



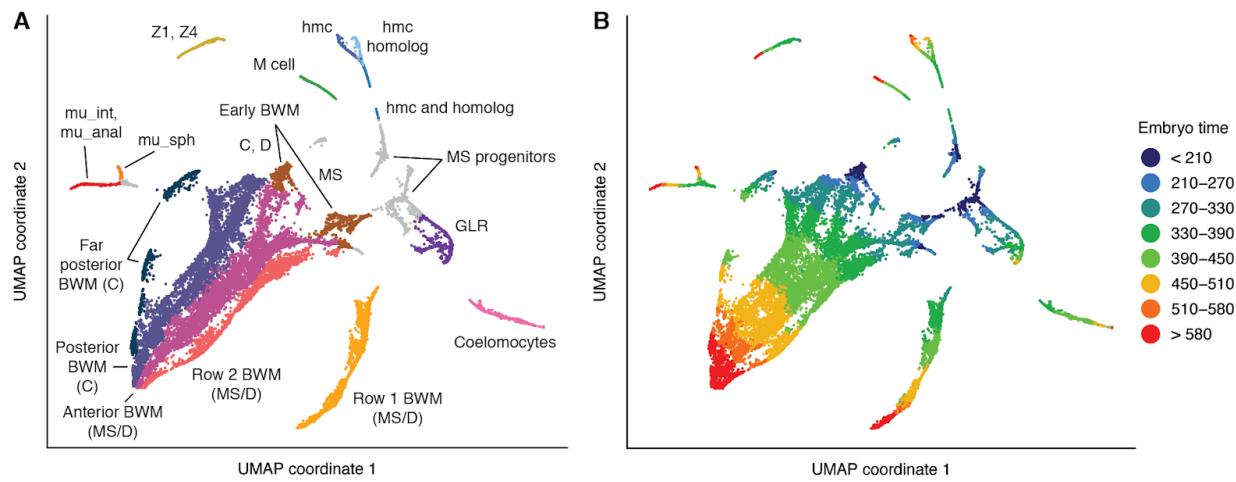
609

610 **Fig. S4. Cells included in each sub-UMAP.** Plots show which cells from the global UMAP
 611 (**Fig. S3**) are included in each sub-UMAP (**Figs. S5-17**), including UMAPs aimed at visualizing
 612 terminal cell types (**A, B**) and UMAPs focused aimed at visualizing progenitor lineages (**C, D**).
 613 Note that the actual assignment of cells to sub-UMAPs was performed based on a 3D version of
 614 the global UMAP (not shown). In (**C**), all cells included in the Time 150 min. sub-UMAP are
 615 also included in the Time 300 min. sub-UMAP.

616

617 **Note:** The figures below show UMAPs of muscle and the non-pharyngeal mesoderm (**Fig. S5**),
 618 pharynx (**Fig. S6**), intestine (**Fig. S7**), hypodermis and seam cells (**Fig. S8**), glia and excretory
 619 cells (**Fig. S9**), non-ciliated neurons (**Fig. S10**), touch receptor neurons (**Fig. S11**), germline
 620 (**Fig. S12**), and rectum (**Fig. S13**). A UMAP of ciliated neurons is shown in the main text (**Fig.**
 621 **3A**). UMAPs focused on annotating progenitor lineages are shown in **Figs. S14-17**.

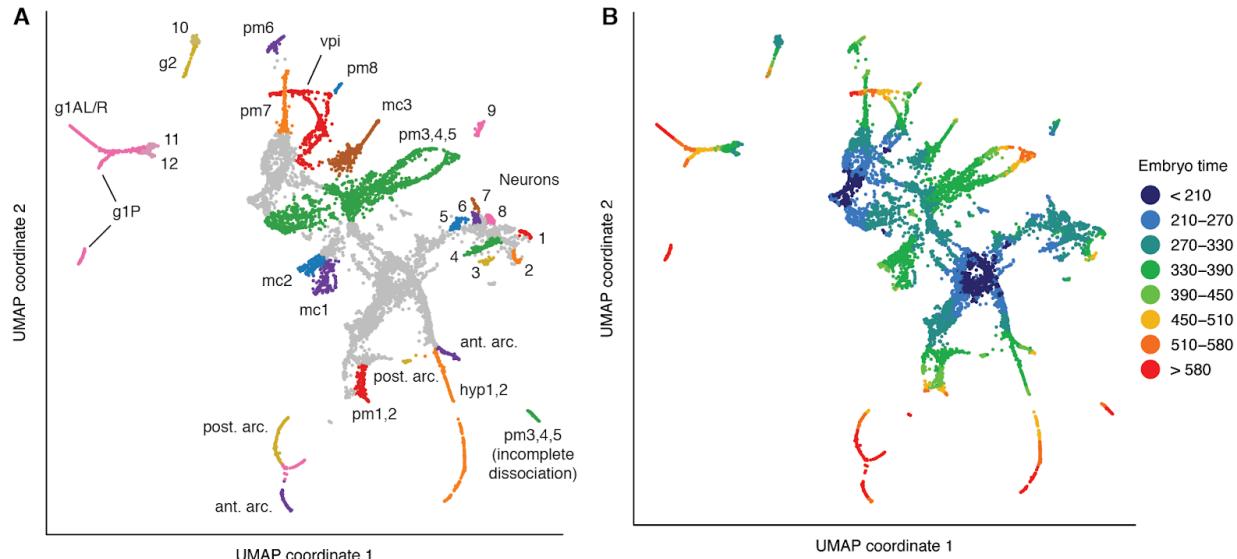
622



623

624 **Fig. S5. UMAP of 22,371 body wall muscle and non-pharyngeal mesoderm cells. (A)** Labels
 625 indicate cell types. See **Table S1** for marker genes used to annotate cell types. MS, C, and D
 626 indicate cell lineages. Abbreviations: BWM = body wall muscle, mu_int = intestinal muscle,
 627 mu_anal = anal depressor muscle, mu_sph = anal sphincter muscle, hmc = head mesodermal cell.
 628 **(B)** Colors show estimated embryo times (minutes post first cleavage) for each cell.

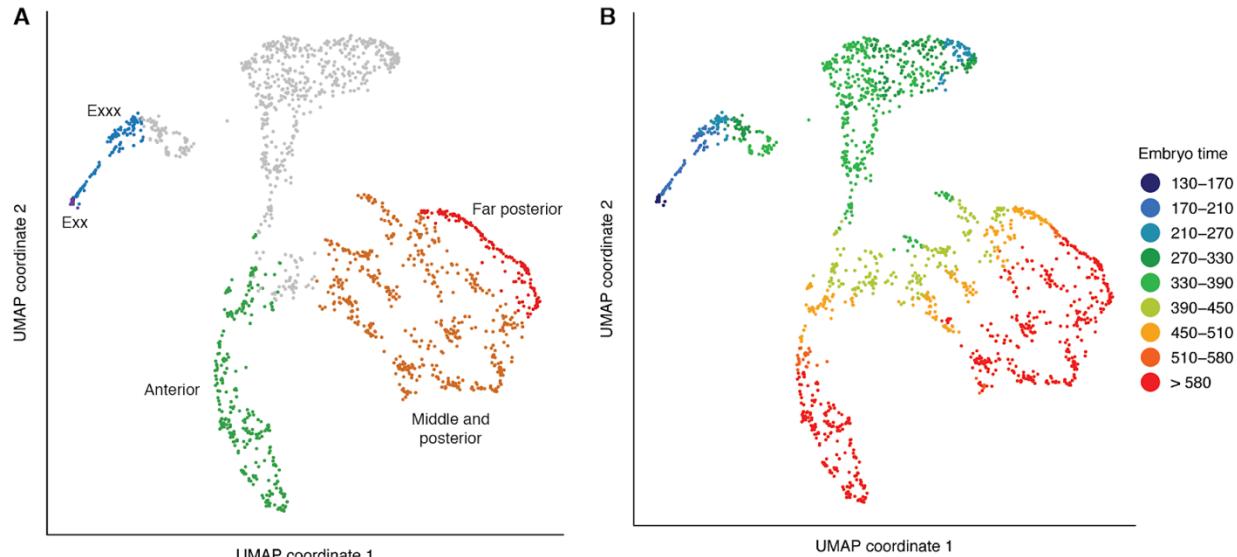
629



630

631 **Fig. S6. UMAP of 10,784 pharyngeal cells. (A)** Labels indicate cell types. See **Tables S1 and**
 632 **S4** for marker genes used to annotate cell types. Abbreviations: pm = pharyngeal muscle, mc =
 633 pharyngeal marginal cell, g1A/g1P/g2 = pharyngeal gland, vpi = pharyngeal-intestinal valve, hyp
 634 = hypodermis, ant. arc. = anterior arcade cells, post. arc. = posterior arcade cells. Anterior and
 635 posterior arcades from late embryos converge in the UMAP to a common transcriptomic profile
 636 (pink cells at the bottom of the plot). Numeric labels indicate: **1** parent of NSM **2** MC **3** parent of
 637 MI and pm1DR **4** grandparent of I2 **5** parent of M1 **6** parent of M2 and M3 **7** parent of M5 and
 638 I6 **8** parent of I1 **9** parent of M4 **10** parent of g2 **11** parent of g1P and I3 **12** parent of g1A. **(B)**
 639 Colors show estimated embryo times (minutes post first cleavage) for each cell.

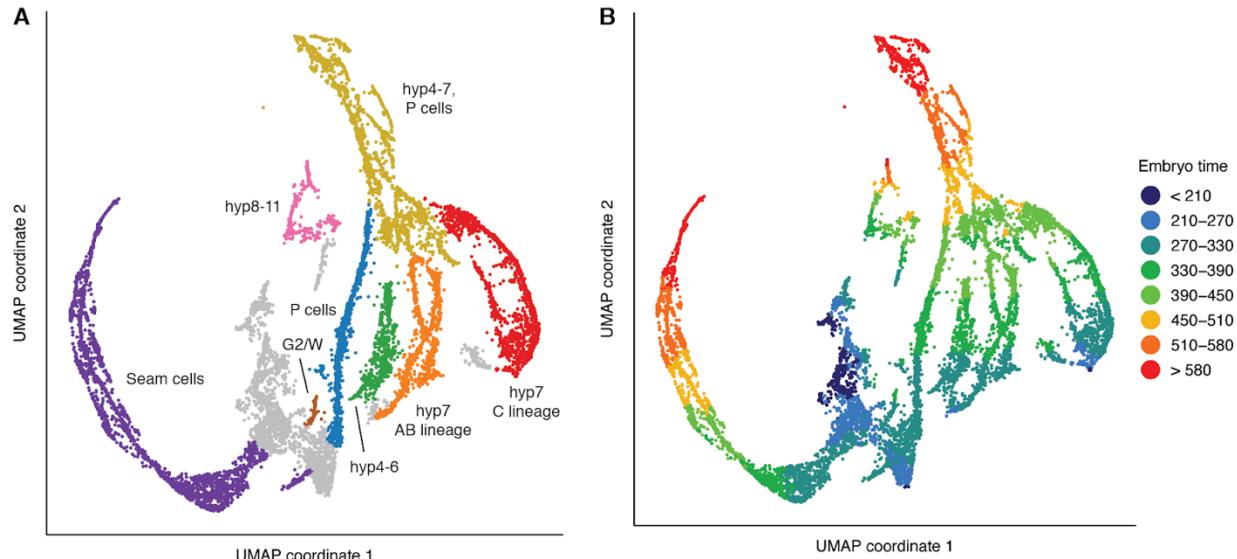
640



641

642 **Fig. S7. UMAP of 1,734 intestine cells.** **(A)** Labels indicate subsets of intestine cells and their
 643 relative position on the anterior-posterior axis. See **Table S1** for marker genes used to annotate
 644 cell types. **(B)** Colors show estimated embryo times (minutes post first cleavage) for each cell.

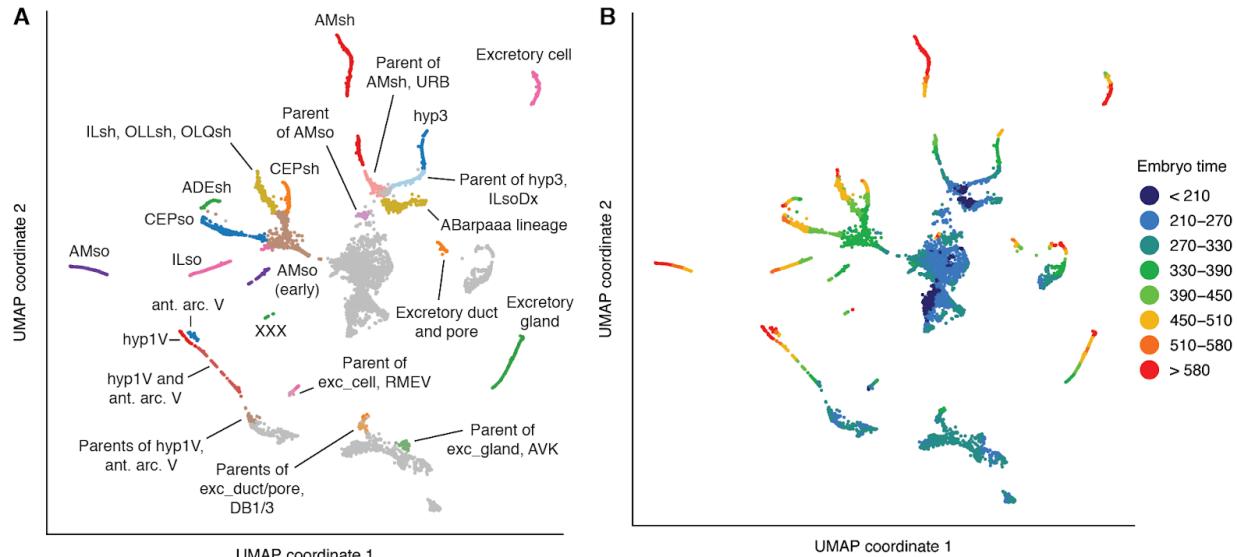
645



646

647 **Fig. S8. UMAP of 12,254 hypodermis and seam cells.** (A) Labels indicate cell types. See
 648 **Table S1** for marker genes used to annotate cell types. hyp1-3 are not included here. hyp1-2
 649 appear in the pharynx UMAP (**Fig. S6**), and hyp3 appears in the glia UMAP (**Fig. S9**), consistent
 650 with their cell lineage (hyp1-2 are sisters/cousins of arcade cells, and hyp3 are sisters of
 651 ILsoDx). (B) Colors show estimated embryo times (minutes post first cleavage) for each cell.

652

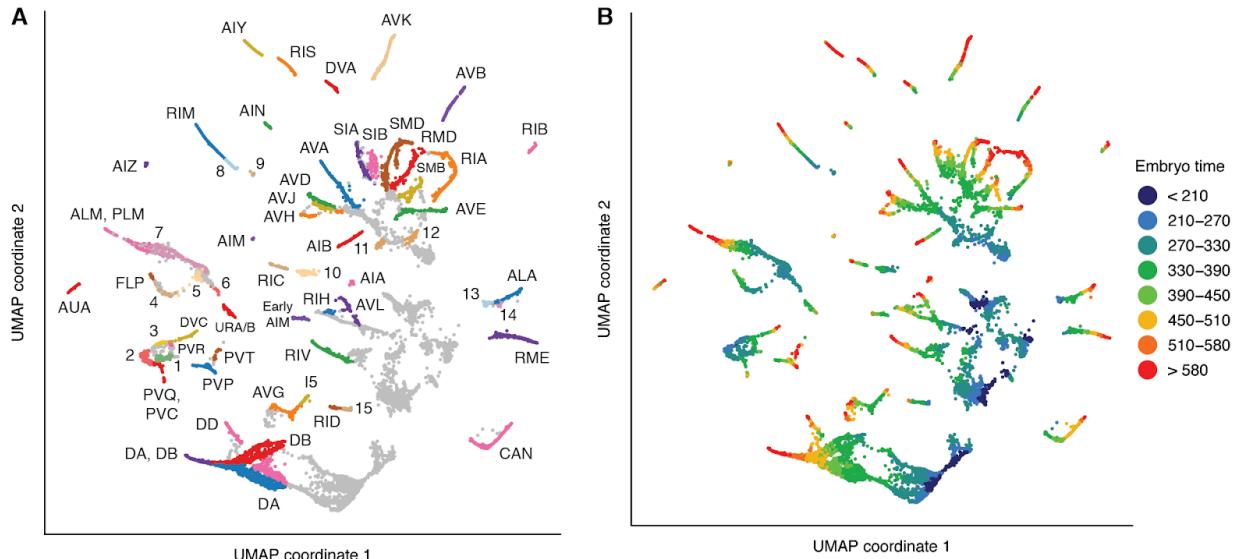


653

654 **Fig. S9. UMAP of 7,512 glia, excretory cells, and progenitors. (A)** Labels indicate cell types.

655 See **Tables S1 and S4** for marker genes used to annotate cell types. Some non-glia/excretory
 656 cells are also included in the UMAP, such as neuron/glia/rectal progenitors. The annotations of
 657 hyp1V and anterior arcade V are very tentative—evidence is described in a note in **Table S1. (B)**
 658 Colors show estimated embryo times (minutes post first cleavage) for each cell.

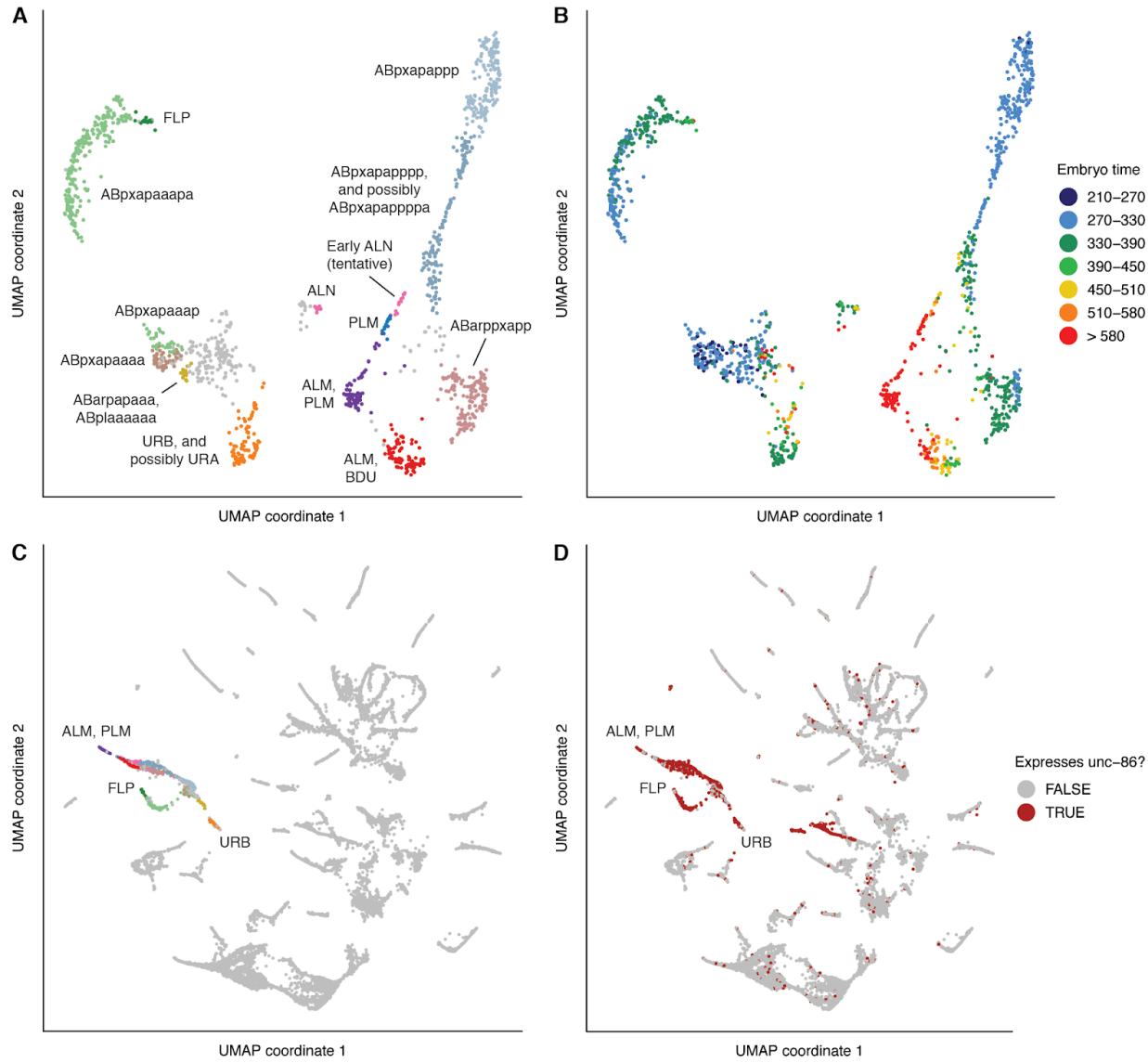
659



660

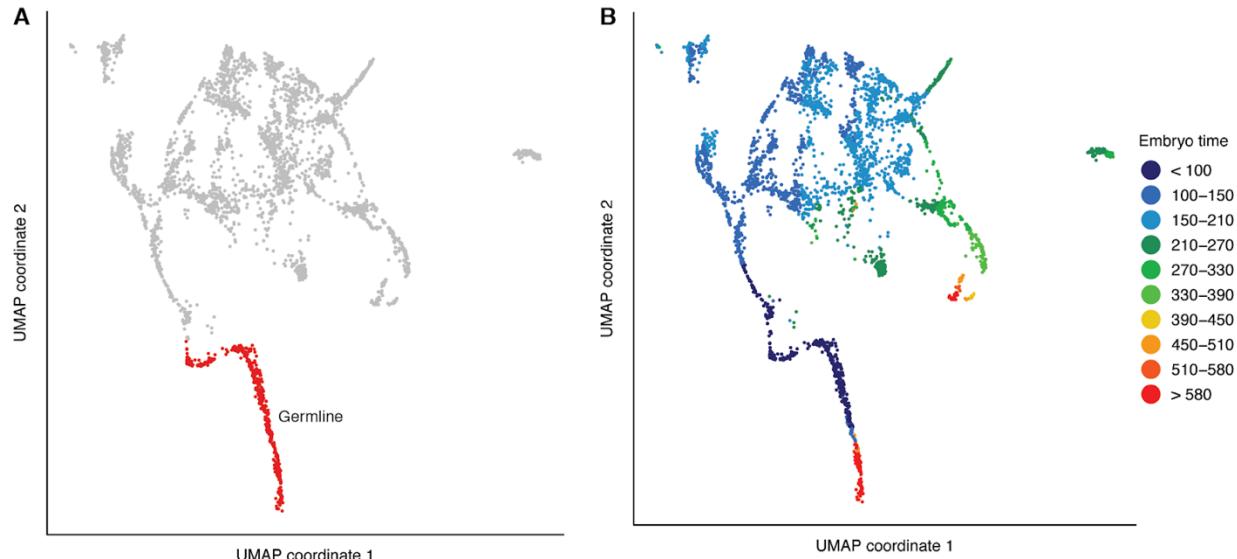
661 **Fig. S10. UMAP of 14,728 non-ciliated neurons and progenitors.** For a UMAP of ciliated
 662 neurons, see **Fig. 3A**. **(A)** Text labels indicate terminal cell types. Numeric labels indicate: **1**
 663 PVC-LUA neuroblast **2** parent of PVQ **3** parent of DVC **4** FLP-AIZ neuroblast **5** FLP-AIZ-RMG
 664 neuroblast **6** parent of URADx **7** progenitors of ALM, BDU, PLM, and ALN (see **Fig. S11** for a
 665 UMAP of the touch receptor lineages) **8** parent of RIM **9** AVG-RIR neuroblast **10** parent of RIC
 666 **11** parent of AVH **12** parent of RIA **13** ALA-RMED neuroblast **14** RMED, early after parent's
 667 division **15** parent of RID. See **Tables S1 and S4** for marker genes used to annotate cell types.
 668 **(B)** Colors show estimated embryo times (minutes post first cleavage) for each cell.

669



670

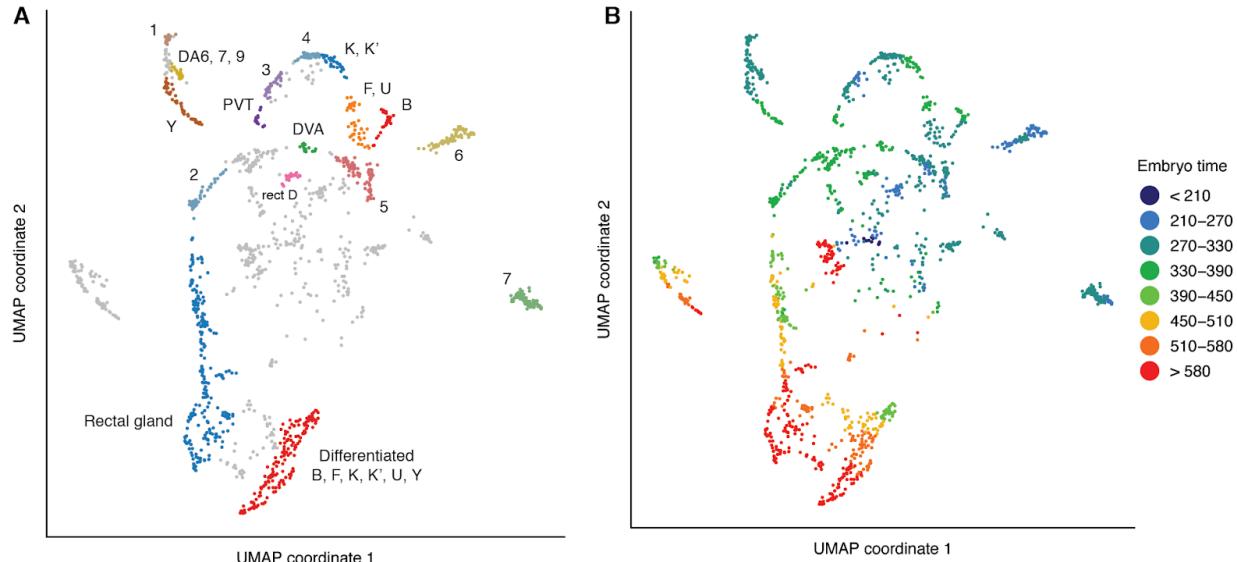
671 **Fig. S11. UMAP of 1,300 touch receptor neurons, URB neurons, and progenitors.** URB
 672 neurons are included because they cluster near the touch receptors in the UMAP of all non-
 673 ciliated neurons (**Fig. S10**). This is in part due to high *unc-86* expression. **(A)** Labels indicate cell
 674 type (for terminal cells) or lineage (for progenitors). **(B)** Colors show estimated embryo times
 675 (minutes post first cleavage) for each cell. **(C)** Location of cells shown in panel A on the UMAP
 676 of all non-ciliated neurons from **Fig. S10**. **(D)** Expression pattern of *unc-86* on the UMAP of all
 677 non-ciliated neurons. Both touch receptor lineages and URB express high levels of *unc-86*.



678

679 **Fig. S12. UMAP of 3,476 early embryo, germline, and rectal cells.** This UMAP was used only
 680 for its trajectory of germline development (500 cells). Other lineages that are included in this
 681 UMAP were better resolved in other UMAPs, shown below. **(A)** Germline cells highlighted in
 682 red. **(B)** Colors show estimated embryo times (minutes post first cleavage) for each cell. These
 683 estimates, which are based on correlation to a whole-embryo bulk RNA-seq time series, are
 684 inaccurate for germline cells, as genes that follow the same temporal dynamics for all somatic
 685 cells often have different expression dynamics in the germline.

686



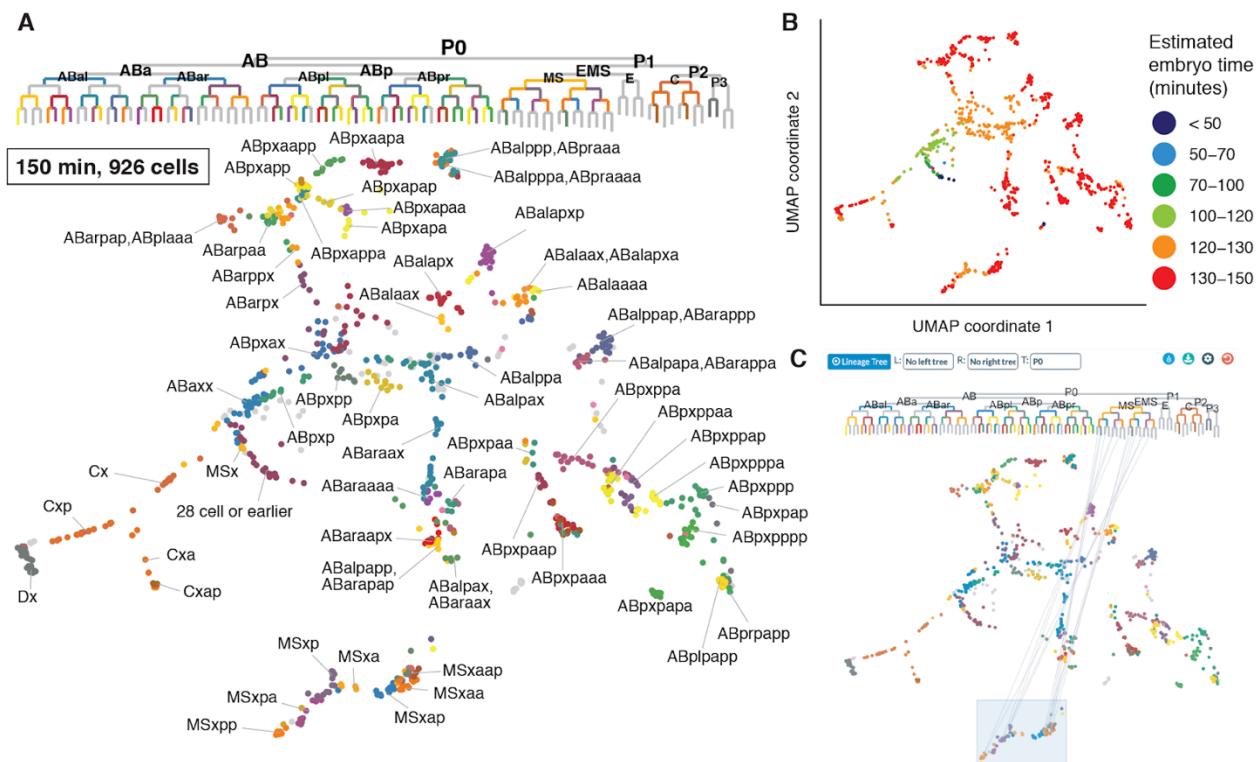
687

688 **Fig. S13. UMAP of 1,598 rectal cells and progenitors.** (A) Text labels indicate terminal cell
 689 types. Numeric labels indicate: **1** parents of (Y and DA7) and (DA6 and DA9). **2** parent of PVP
 690 and rect_V **3** parent of PVT and rect_D **4** parent of K and K' **5** parents of (B and DVA) and (F
 691 and U) **6** Parent of the tail spike cells and hyp10 **7** Parent of PHsh and hyp8/9. (B) colors show
 692 estimated embryo times (minutes post first cleavage) for each cell. The cluster of cells from late
 693 embryos (>580 minutes) in the center of the UMAP are AMsh (glia, not rectal cells) that were
 694 included in this UMAP by mistake.

695

696 **Note:** Figs S14-17 show a representative subset of the UMAPs that were used to annotate
 697 progenitor lineages. Several additional UMAPs can be visualized in VisCello. **Table S4** includes
 698 a list of which UMAP was used to annotate each lineage in our annotation set.

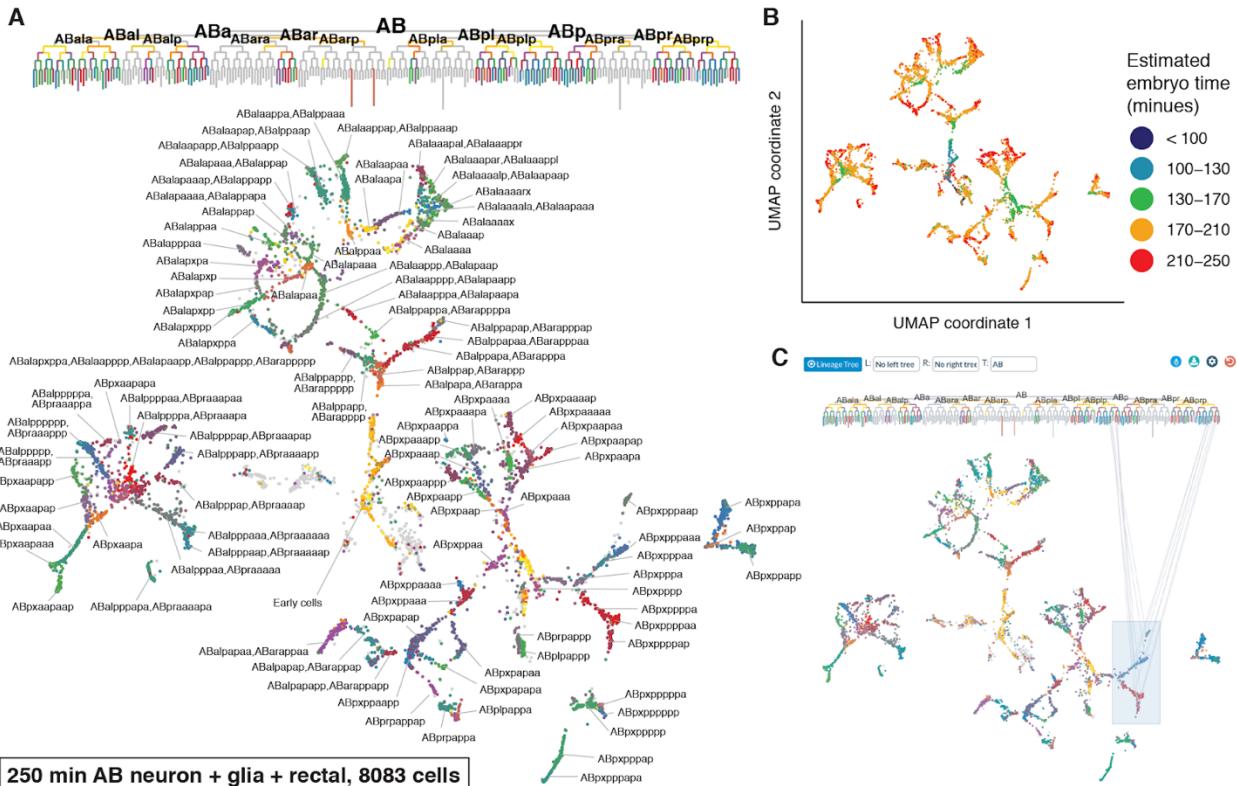
699



700

701 **Fig. S14. UMAP and detailed annotation of 926 cells from embryos < 150 minutes post first**
 702 **cleavage.** E lineage and germline cells are excluded from the UMAP and were analyzed
 703 separately (**Figs S7 and S12**). (A) Detailed labeling of lineages, co-visualized with the lineage
 704 tree. (B) Colors show estimated embryo times (minutes post first cleavage) for each cell. (C)
 705 Screenshot of an interactive co-visualization implemented in VisCello, highlighting the
 706 connection between MS lineage clusters and corresponding leaves in the lineage tree.

707



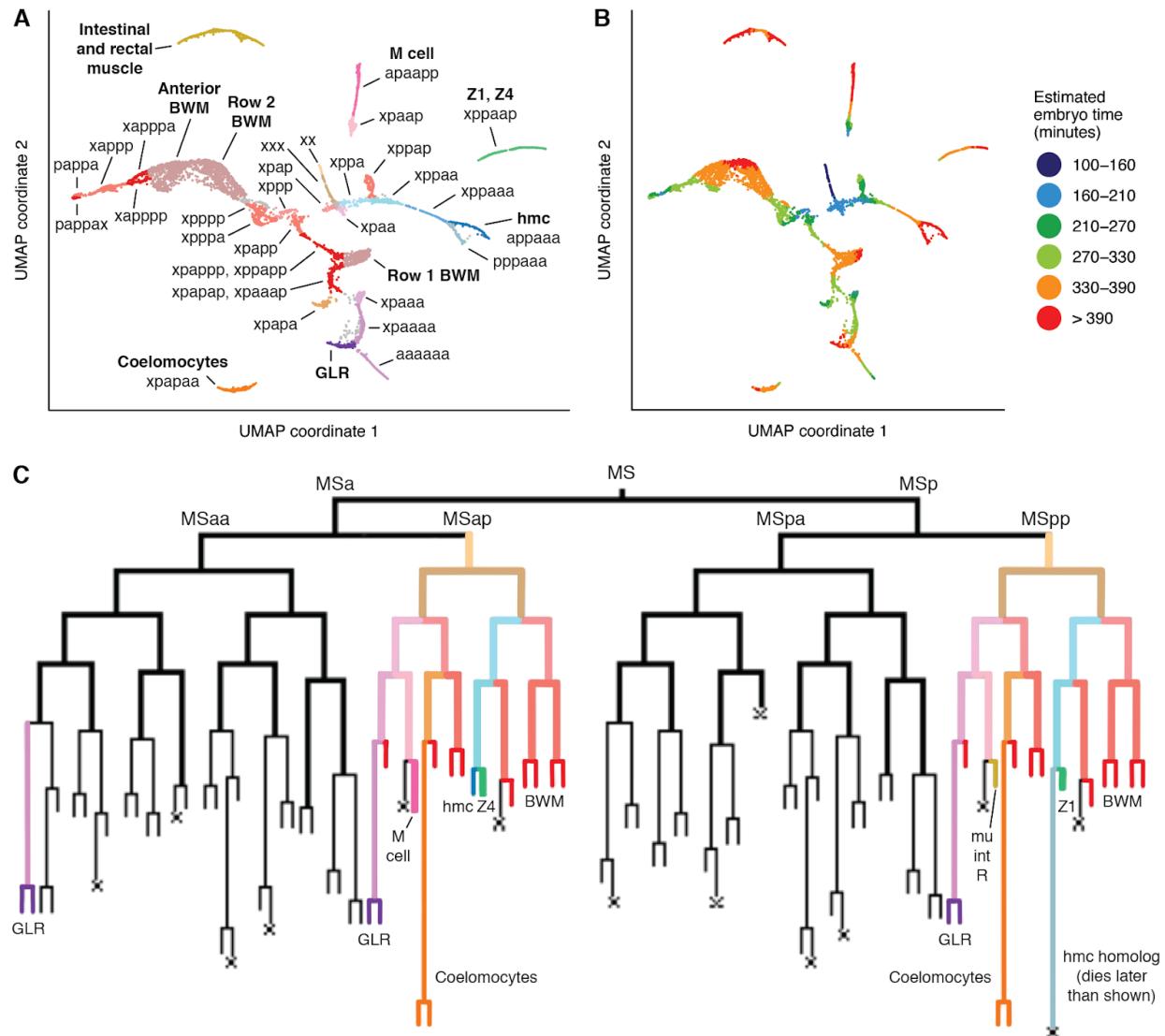
708

709 **Fig. S15. UMAP and detailed annotation of 8,083 AB lineage neuron/glia/rectal progenitor**
 710 **cells from embryos < 250 minutes post first cleavage.** This UMAP is derived from another
 711 UMAP that can be visualized in VisCello (“Time 250min lineage”). This UMAP includes only
 712 AB lineage cells that give rise to neurons, glia, and rectal cells. **(A)** Detailed labeling of lineages,
 713 co-visualized with the AB lineage tree. **(B)** Colors show estimated embryo times (minutes post
 714 first cleavage) for each cell. **(C)** Screenshot of an interactive co-visualization implemented in
 715 VisCello, highlighting the connection between ABpxppp lineage clusters and corresponding
 716 leaves in the lineage tree.

717

721 separately (**Figs S7 and S12**). **(A)** Detailed labeling of lineages, co-visualized with the lineage
722 tree. **(B)** Colors show estimated embryo times (minutes post first cleavage) for each cell. **(C)**
723 Screenshot of an interactive co-visualization implemented in VisCello, highlighting the
724 connection between the pharynx cluster in the UMAP and the corresponding leaves in the
725 lineage tree. All cells in the pharynx cluster are annotated as descendants of the ABalp, ABara
726 and MS lineages, consistent with previous observations that pharyngeal cells only arise from
727 these lineages.

728

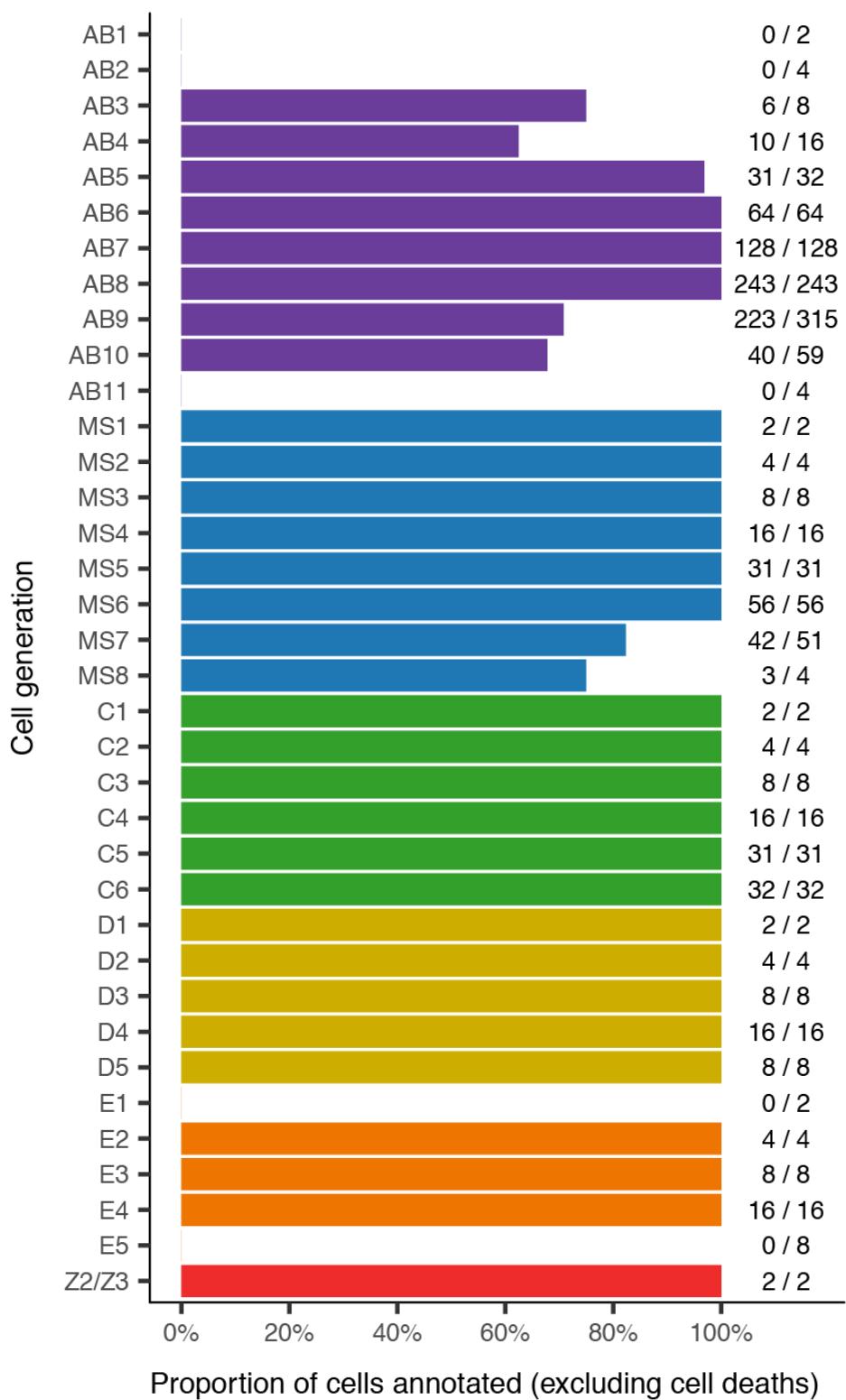


729

730 **Fig. S17. UMAP of 8,233 non-pharyngeal mesoderm cells, focused on the early lineage.** This
 731 UMAP includes the same cells as the muscle and mesoderm UMAP (Fig. S5), but excludes
 732 putative C and D lineage body wall muscle, MS lineage body wall muscle with estimated
 733 embryo time >400 minutes (post first cleavage), and coelomocytes with embryo time >400
 734 minutes. This UMAP serves as a representative example of a set of several UMAPs used to
 735 connect terminal cells to their immediate progenitors. Additional UMAPs can be viewed in
 736 VisCello. (A) Text labels indicate MS lineages (i.e. “xppa” = MSxppa). Bold text labels indicate
 737 cell types. MSxppapx was not conclusively identified, but is presumed to be included in the head

738 BWM cluster. **(B)** Estimated embryo time for each cell. **(C)** diagram of the MS lineage. Colored
739 sub-lineages match the colors of cell groups in panel (A).

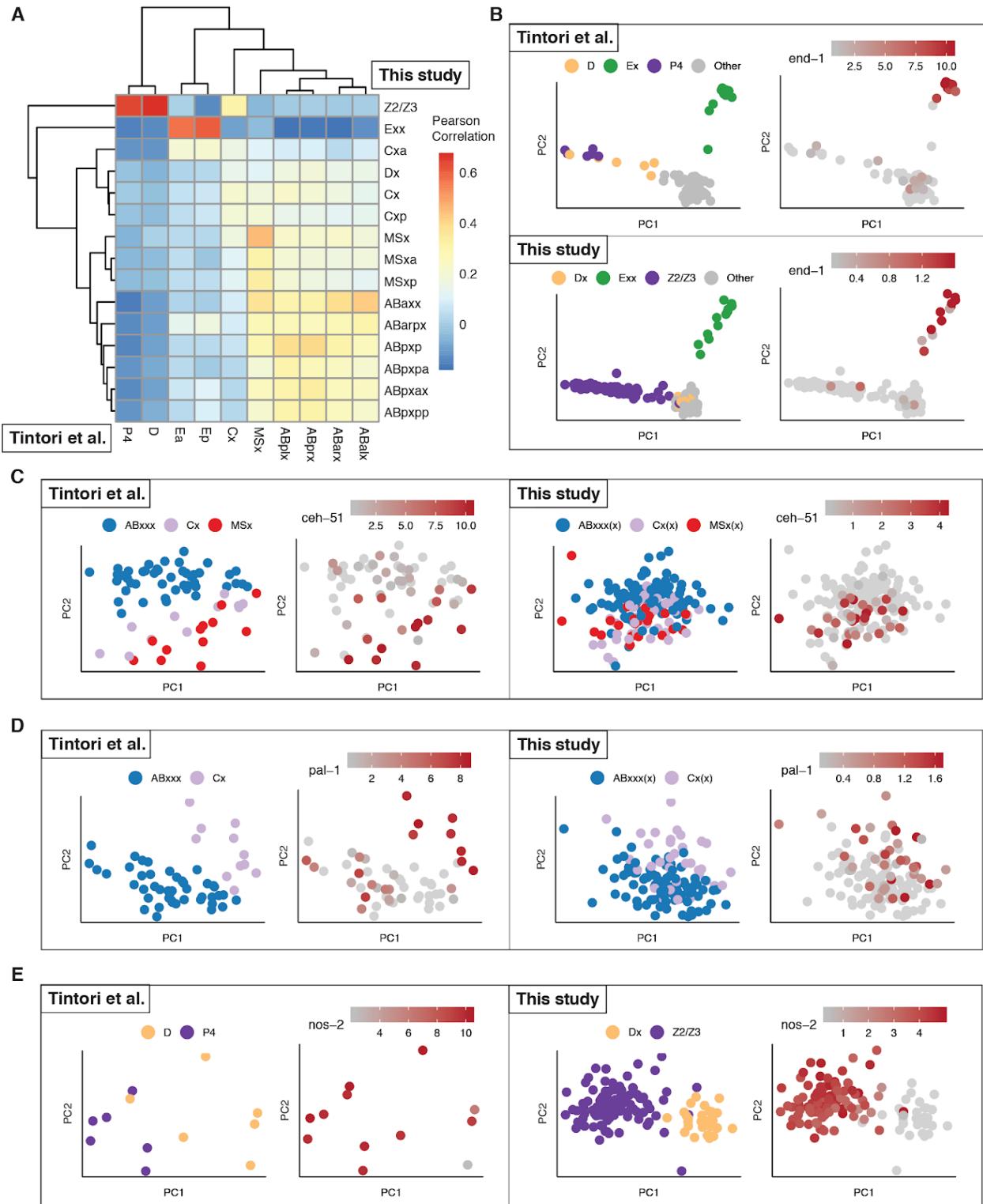
740



742 **Fig. S18. Summary of lineage annotations.** Each row corresponds to a subset of cells in the *C.*
743 *elegans* embryonic cell lineage. Row labels consist of one or two letters, which identify a broad
744 lineage (AB, MS, C, D, or E), and a number, which specifies the number of cell divisions since
745 the founding cell of the broad lineage. For example, “AB5” refers to the 32 cells produced by 5
746 divisions of the AB founder cell, and “C2” refers to the 4 cells produced by 2 divisions of the C
747 founder cell. The founder cells themselves are not included in the plot. The label “Z2/Z3” is an
748 exception to the nomenclature and refers to the two germline lineages, Z2 and Z3.

749 Bar lengths indicate the percent of cells within the specific lineage and cell generation
750 specified by the row label that are included in our annotations of our single cell RNA-seq
751 dataset. Lineages that undergo programmed cell death are excluded from the statistics. Numbers
752 to the right of the bars indicate the absolute number of lineages annotated and the total number of
753 lineages present within a particular cell generation.

754

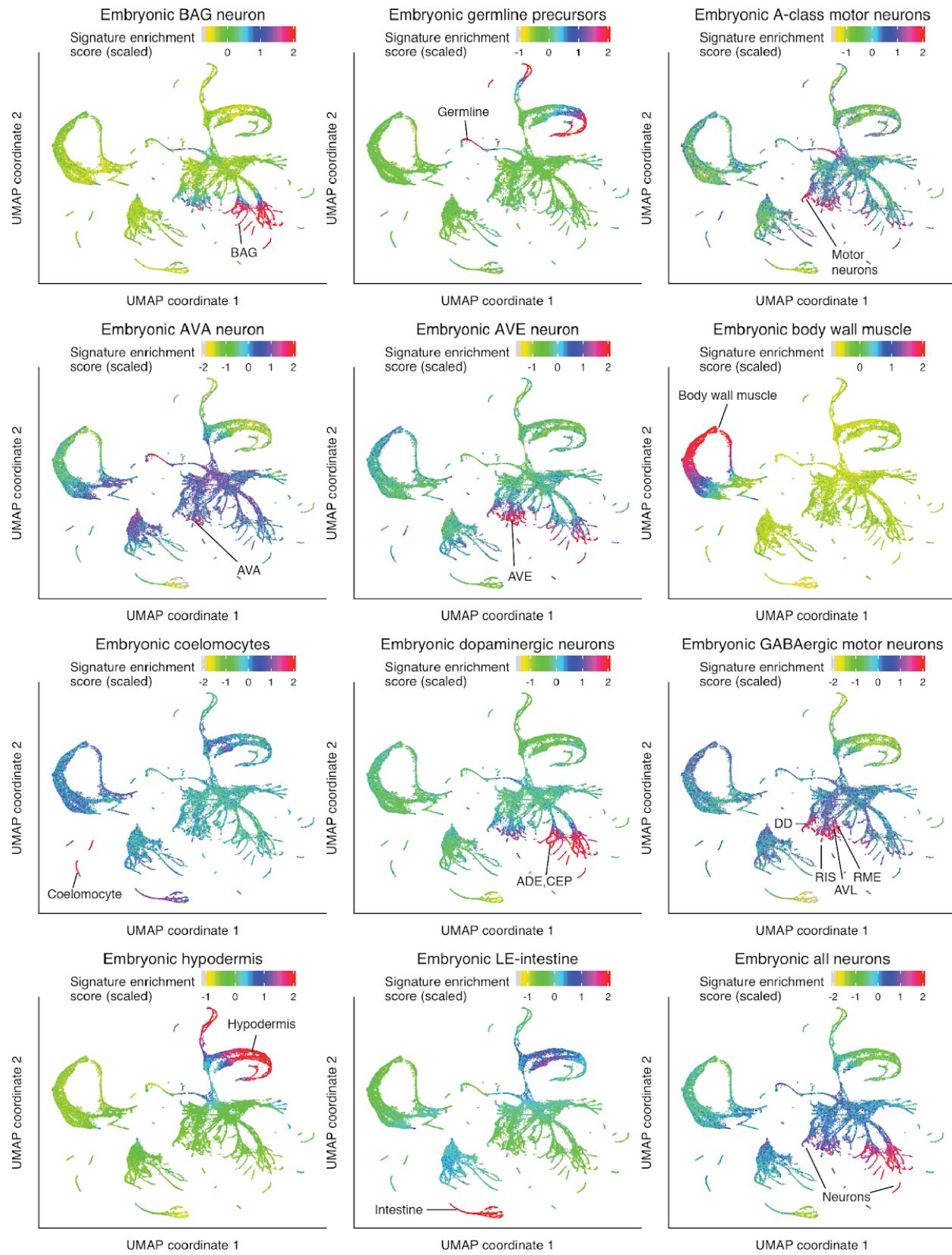


755

756 **Fig. S19. Comparison of data from this study to data from Tintori et al., 2016 (8).** Tintori et
757 al. (8) profiled the transcriptomes of single cells from the *C. elegans* 1- to 16-cell stages. (A)

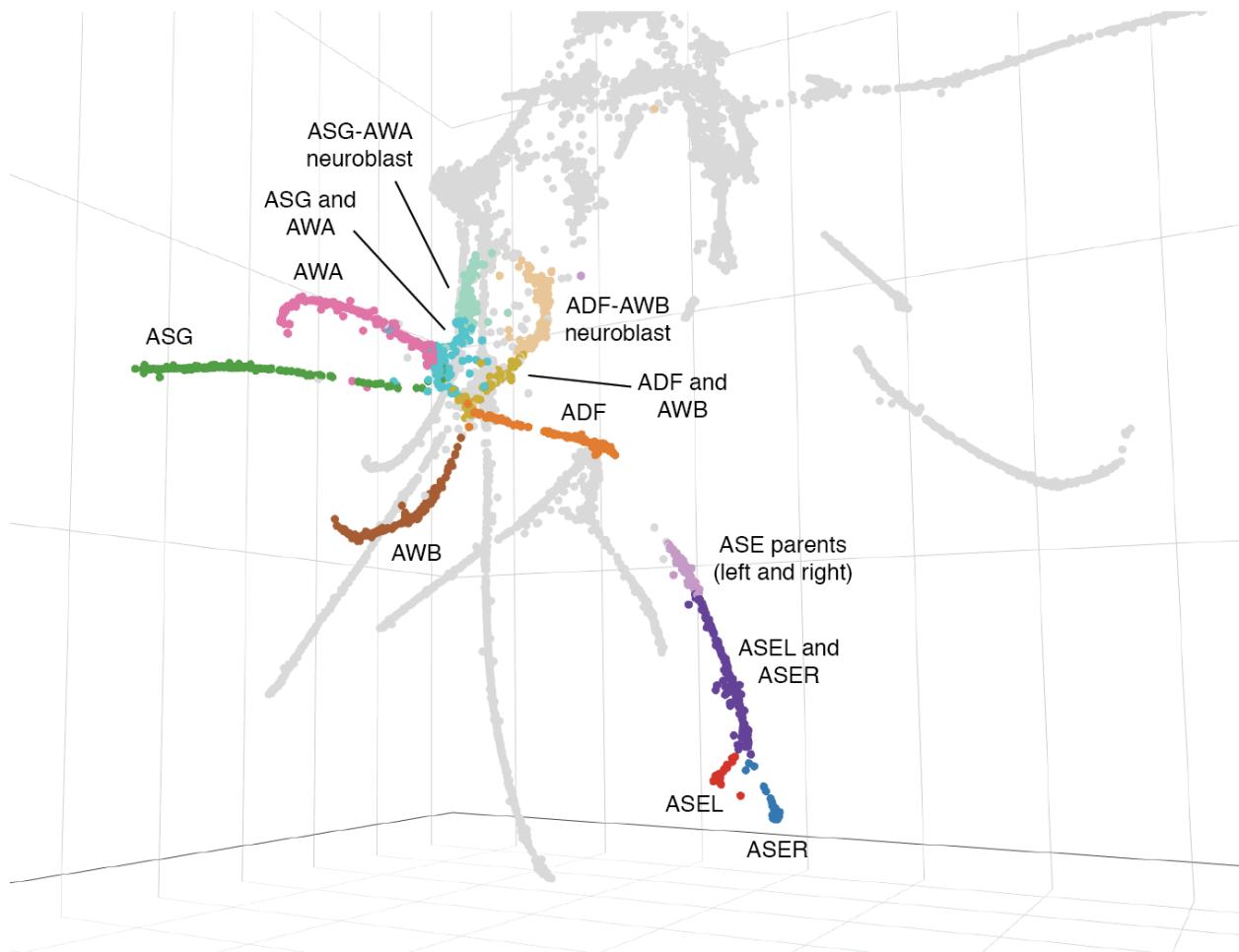
758 Heatmap showing Pearson correlations between the log2-scaled gene expression profiles of 16-
759 cell stage cells from Tintori *et al.* (8) vs. 16- and 28-cell stage cells from this study. Correlation
760 was computed using informative genes selected by an iterative PCA approach used by Tintori *et*
761 *al.* (8) (see **Methods**). **(B-E)** First sub-panel shows a PCA projection computed using 16-cell
762 stage cells from Tintori *et al.* (8), reproducing their original analysis. Second sub-panel shows a
763 projection of 16- and 28-cell stage cells from this study into the same PCA space. Each PCA
764 uses a different set of informative genes, as originally defined by Tintori *et al.* (8), to
765 discriminate particular lineages (see **Methods**). For each PCA, the gene expression level of a
766 selected lineage-specific marker gene was plotted. Gene expression is measured in log2 RPKM
767 for data from Tintori *et al.* (8), and log2 size-factor normalized UMI counts for data from this
768 study.

769



771 **Fig. S20. Comparison of data from this study to microarray data from Spencer *et al.*, 2011**
772 (29). Each panel shows a global UMAP of cells from this study, colored by a score that measures
773 the extent to which each single-cell transcriptome is enriched for genes from a particular gene set
774 reported by Spencer *et al.* (29). Signature gene sets from Spencer *et al.* (29) were downloaded
775 from https://www.vanderbilt.edu/wormdoc/wormmap/Enriched_genes.html. Each signature gene
776 set corresponds to genes that are enriched in a particular embryonic cell type compared to all
777 other cells in the Spencer *et al.* microarray data (29). Signature genes are therefore mostly tissue-
778 specific, rather than cell-type specific. Gene set enrichment scores were computed using the
779 AUCell package (61). Comparison with pharyngeal muscle was dropped because most of the
780 signature genes reported in Spencer *et al.* (29) for this cell type are intestine specific, as
781 confirmed by a third dataset (54). See **Methods** for more details.

782



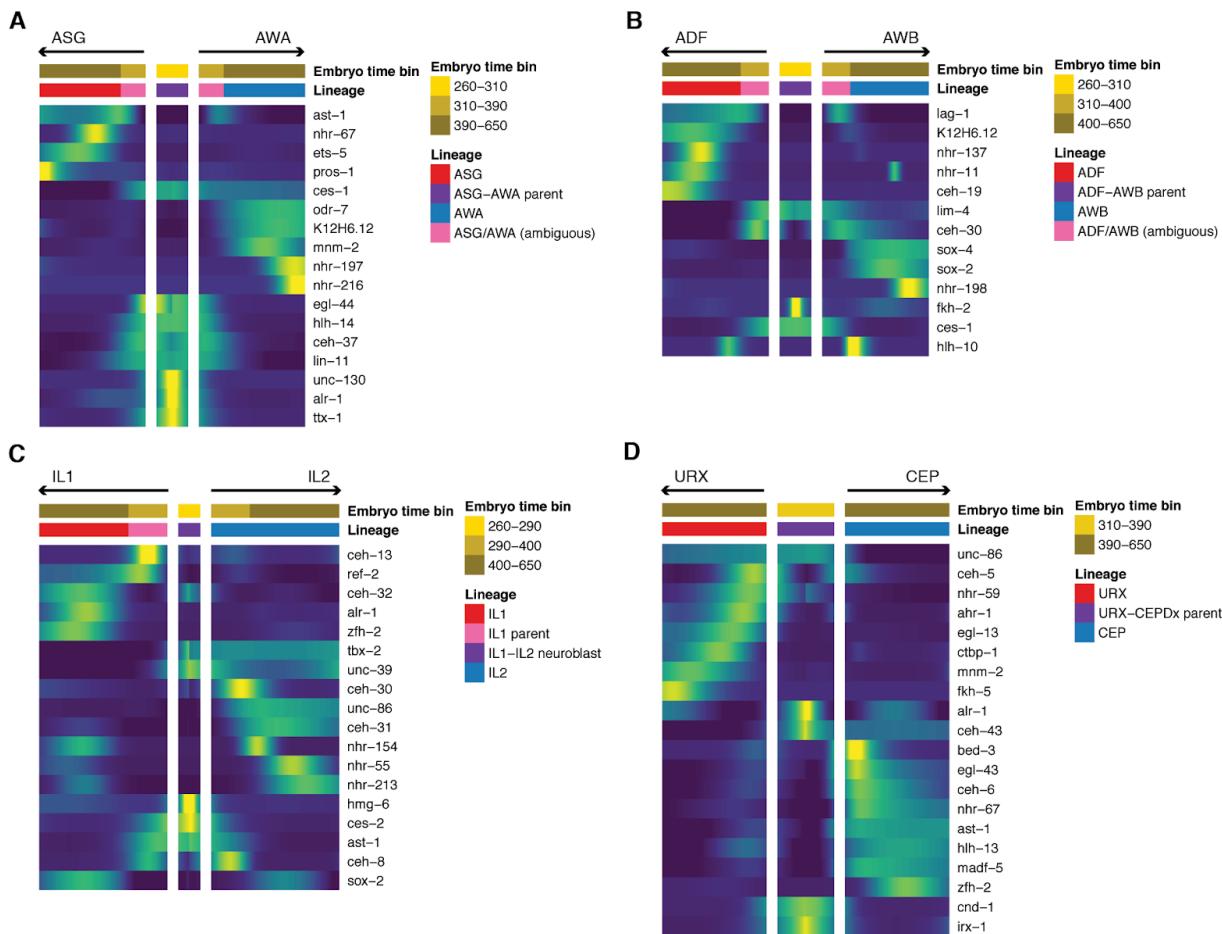
783

784 **Fig. S21. Ciliated neuron developmental trajectories are more continuous in a 3D UMAP.**

785 This plot is a 2D screenshot of part of a 3D UMAP of ciliated neuron cells, oriented to show
 786 specific lineage relationships. The cells are the same as in **Fig. 3A**; the only difference is
 787 projecting into 3D instead of 2D. Developmental trajectories connecting the ASG-AWA and
 788 ADF-AWB neuroblasts to their respective daughter cells are continuous in this UMAP space, as
 789 is the branching trajectory of the left and right ASE neurons (ASEL and ASER). In the ASG-
 790 AWA and ADF-AWB trajectories, there are sections that appear before the branch points in the
 791 UMAP, but based on our embryo time estimates are likely to be terminal cells and not the parent
 792 neuroblasts. These sections may contain both daughter cells of each trajectory after their birth

793 but before they differentiate. Cells in the “ADF and AWB” section co-express in the same cells
794 the marker genes *lag-1*, which persists only in ADF, and *lim-4*, which persists only in AWB;
795 however, their estimated embryo times span ~100 minutes after the parent cells’ division time.
796 Note that the grey, unannotated cells below the ADF trajectory are behind the ADF cells in 3D
797 space, as are the grey cells overlapping the AWB trajectory.

798

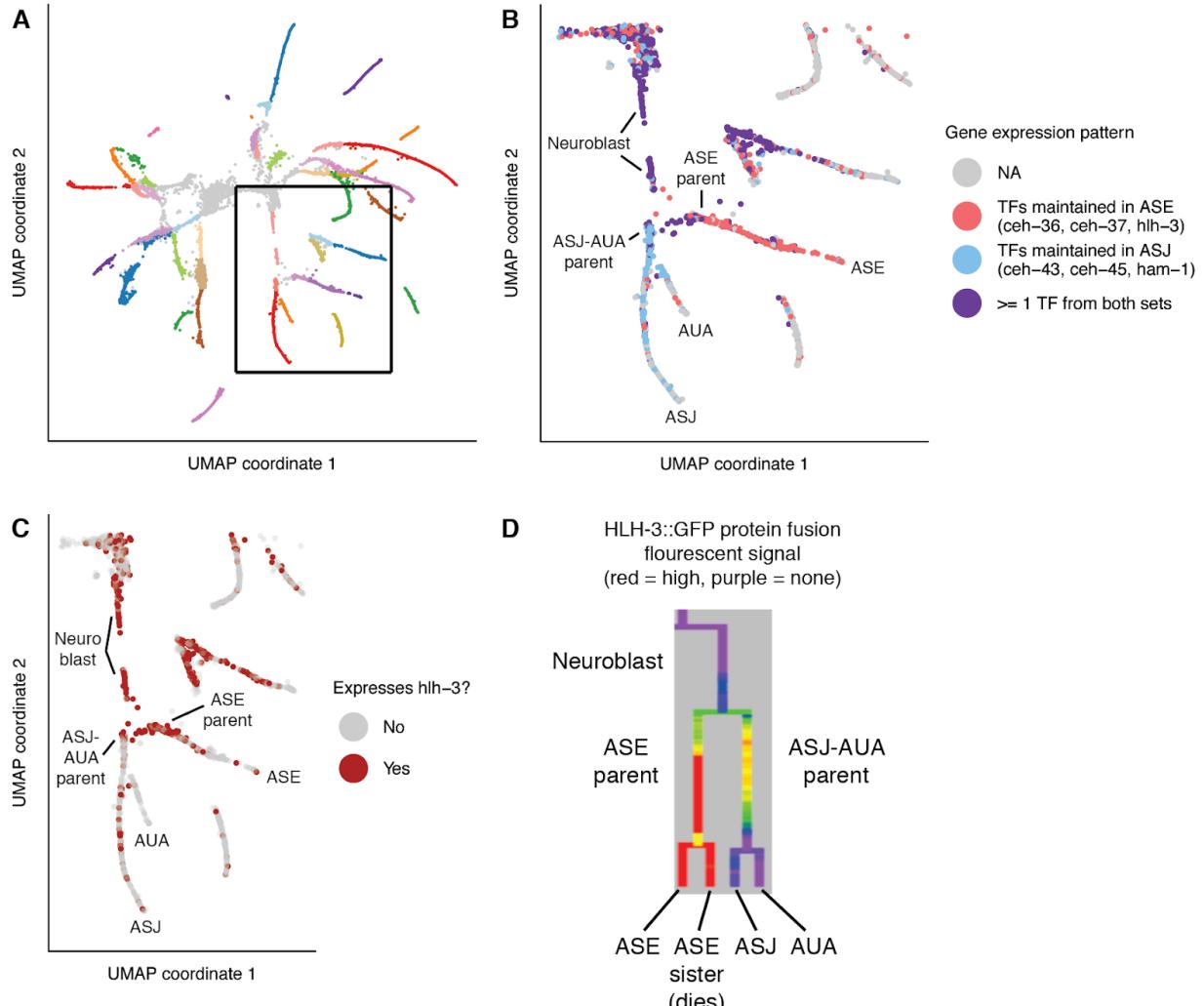


799

800 **Fig. S22. Differentially expressed transcription factors associated with ciliated neuron**
 801 **lineage branches.** Heatmaps showing patterns of differential transcription factor expression
 802 associated with branches in (A) the ASG-AWA lineage, (B) the ADF-AWB lineage, (C) the IL1-
 803 IL2 lineage, and (D) the URX-CEPDx lineage. A heatmap for the ASE-ASJ-AUA lineage is
 804 shown in Fig. 3D. Expression values are log-transformed, then centered and scaled by standard
 805 deviation for each row (gene). In each of the ASG-AWA and ADF-AWB lineages, there is a set
 806 of cells that are before the branch point of the trajectory in UMAP space (see Fig. S21), but
 807 based on embryo time estimates and marker gene expression patterns, are likely to be terminal
 808 cells. In the ADF-AWB lineage, these cells co-express *lag-1*, which is selectively retained in

809 ADF, and *lim-4*, which is selectively retained in AWB, suggesting that this cell set may include
810 undifferentiated, terminal ADF and AWB cells.

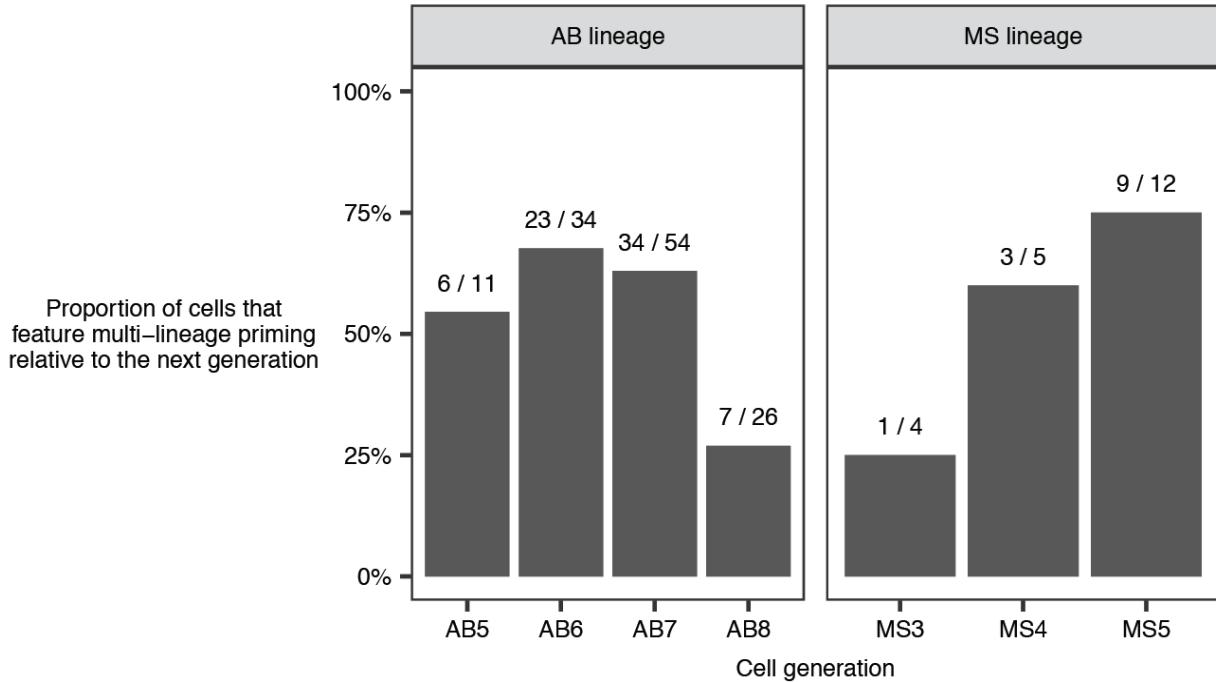
811



812

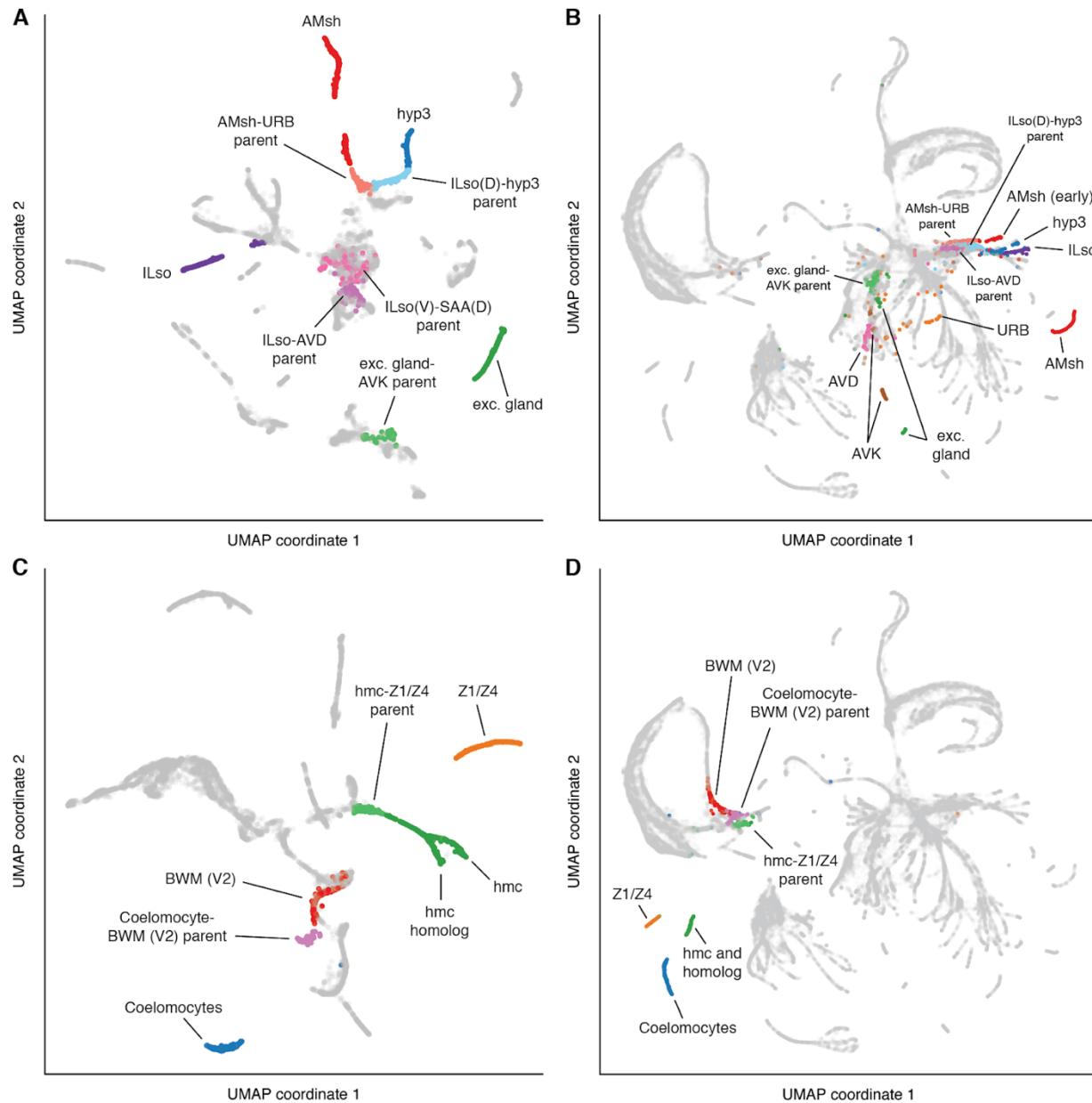
813 **Fig. S23. Multilineage priming in the ASE-ASJ-AUA lineage.** (A) Section of the ciliated
 814 neuron UMAP from **Fig. 3A** that is shown in panels B and C. This section includes the trajectory
 815 of the lineage that produces the ASE, ASJ, and AUA neurons (ABalpppppp/ABpraaappp). (B)
 816 Expression patterns for transcription factors that are expressed in the ASE-ASJ-AUA neuroblast
 817 and selectively maintained in only one of its daughters. Red and blue points indicate cells that
 818 express ≥ 1 TF for which expression is maintained only in the ASE lineage (red) or only the
 819 ASJ lineage (blue). Purple points indicate cells that express ≥ 1 TF from both sets. (C)
 820 Expression pattern of *hh-3*, which is expressed in the ASE-ASJ-AUA neuroblast and maintained
 821 in the ASE parent but not the ASJ-AUA parent. (D) Fluorescent signal from a HLH-3::GFP

822 protein fusion from EPiC (17) (series 20160301_hlh-3_OP650_L2). Red indicates high signal,
823 yellow/green indicate medium signal, blue indicates low signal, and purple indicates no signal.
824 Due to translation and the folding time of GFP, the fluorescent signal has a time lag compared to
825 the RNA expression in panel C. The presence of signal in the ASJ-AUA parent indicates that
826 HLH-3 protein does not undergo asymmetric localization during cell division; instead, it is
827 simply maintained in the ASE lineage and allowed to degrade in the ASJ-AUA lineage.
828



829

830 **Fig. S24. Prevalence of multilineage priming in *C. elegans*.** X-axis shows different cell
 831 generations of the ectoderm (AB lineage) and mesoderm (MS lineage). “AB5” refers to the
 832 generation produced by 5 divisions of the AB founder cell, and likewise for AB6-8 and MS3-5.
 833 Y-axis shows the proportion of lineages in a given generation that co-express at least one
 834 transcription factor (TF) that has expression selectively maintained in one daughter, and at least
 835 one TF that has expression selectively maintained in the other daughter (e.g. TF A expressed in
 836 parent and daughter 1, TF B expressed in parent and daughter 2). Lineages that satisfy these
 837 criteria are considered to exhibit “multilineage priming.” Text labels above each bar indicate the
 838 absolute number of lineages in each generation that exhibit multilineage priming (numerator) and
 839 the total number of lineages included in the analysis (denominator). Gene expression levels are
 840 taken from **Table S8**. Lineages that do not have exactly two, transcriptomically distinct
 841 daughters annotated in our dataset are excluded from the statistics. Cell generations that are not
 842 shown in this plot were excluded due to having a sample size of ≤ 3 lineages.



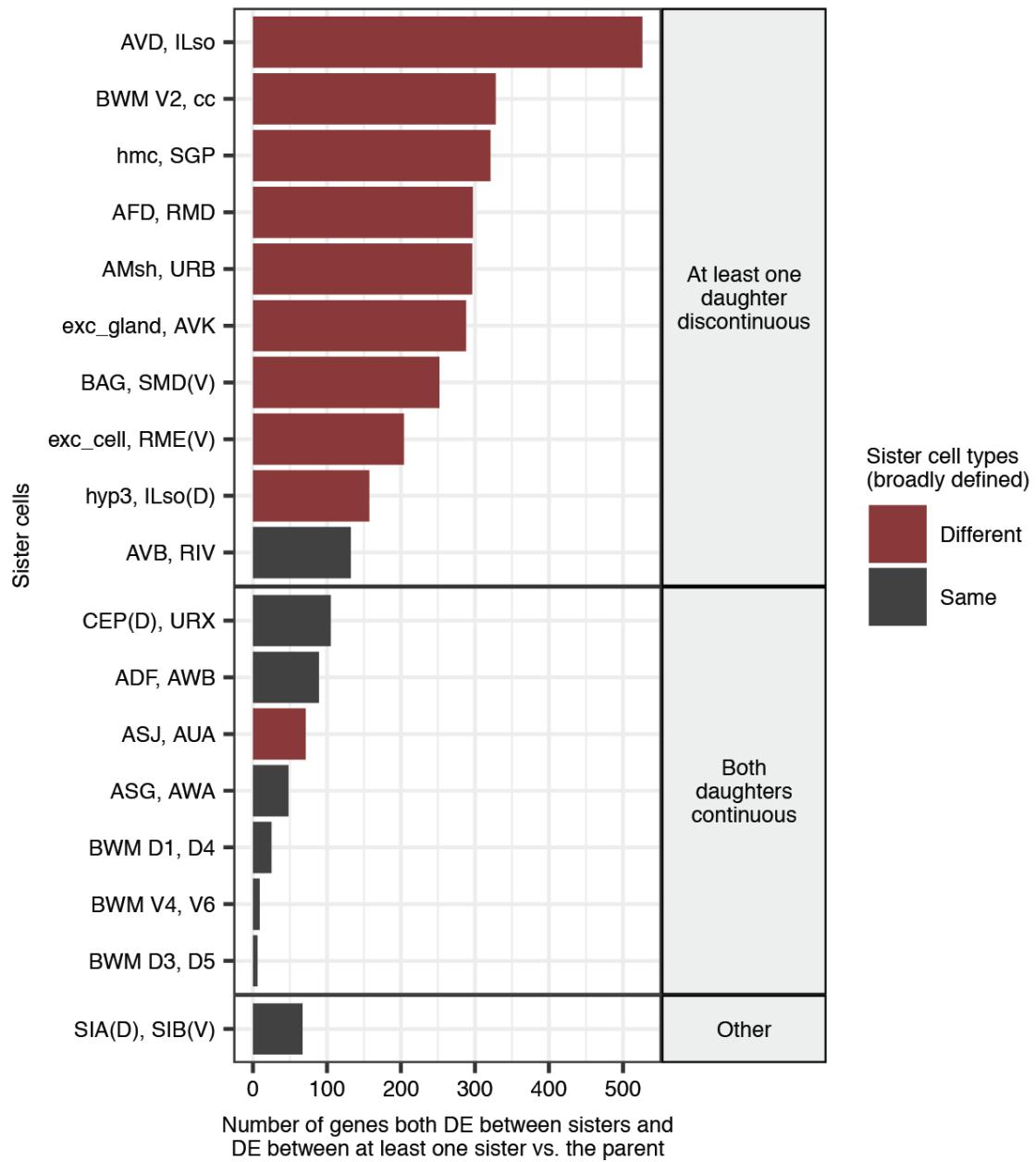
843

844 **Fig. S25. Examples of lineages that form discontinuous trajectories in UMAP space.**

845 (A) UMAP of 7,512 glia, excretory cells, and progenitors (same as Fig. S9). ILso glia are
 846 formed by three input lineages. Two input lineages, the ILso-AVD parent and the ILso(D)-hyp3
 847 parent, form discontinuous trajectories with terminal ILso. Some early terminal ILso cells are
 848 likely to be unannotated, so it is not clear if there is a continuous or discontinuous trajectory with
 849 the third input lineage, the ILso(V)-SAA(D) parent. (B) Global UMAP of 81,286 cells (same as

850 **Fig. 1A).** Annotated cell populations are the same as in panel A, plus additional neuron types.
851 The AVD, AVK, and URB neurons are sisters of glia/excretory cells, but form discontinuous
852 trajectories with their parents. **(C)** UMAP of 8,233 non-pharyngeal mesoderm cells (same as **Fig.**
853 **S17**). Coelomocytes and Z1/Z4 (the somatic gonad precursors) form discontinuous trajectories
854 with their parents. **(D)** Global UMAP, same as panel B. Annotated cell populations are the same
855 as in panel C.

856

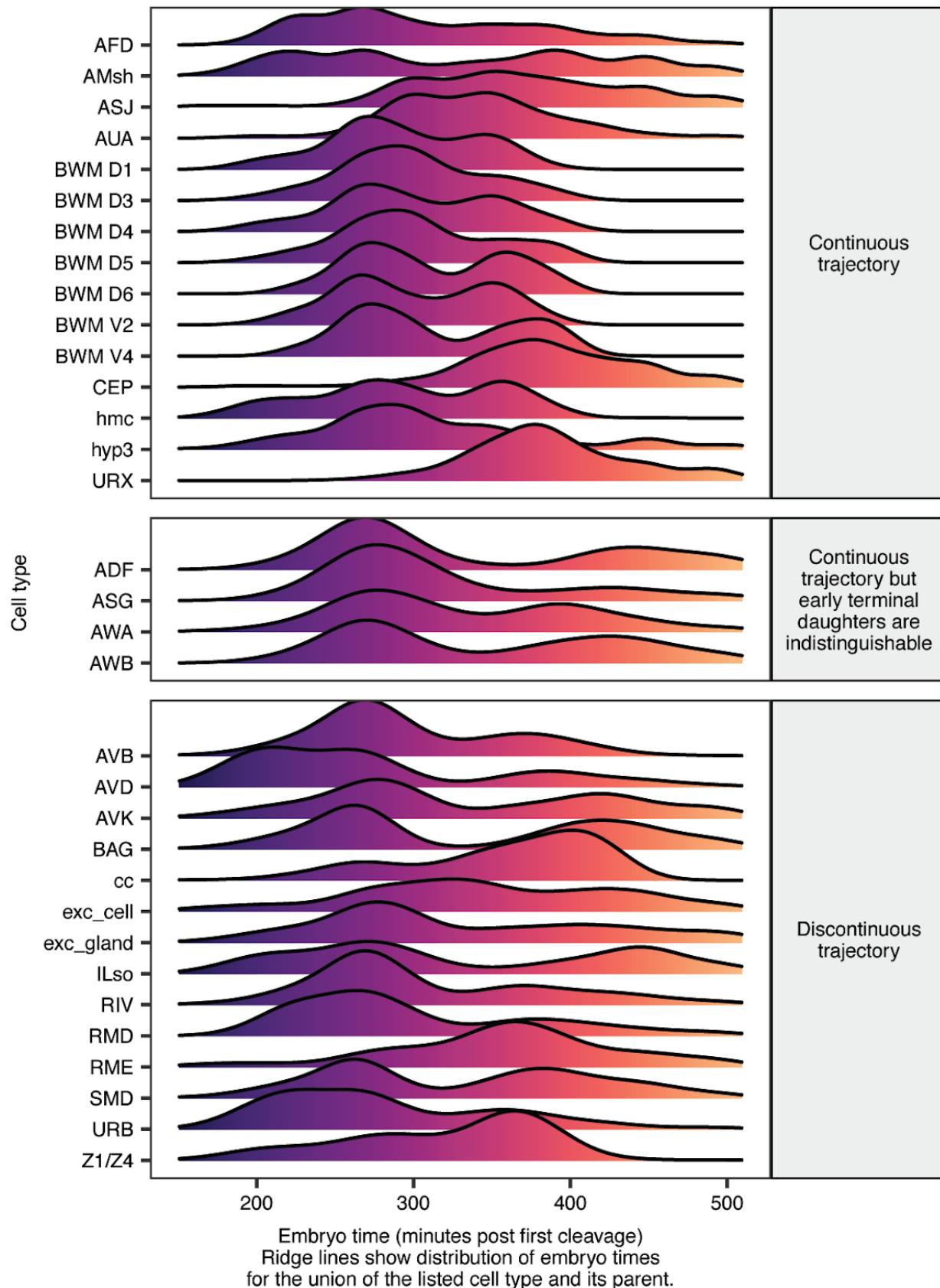


857

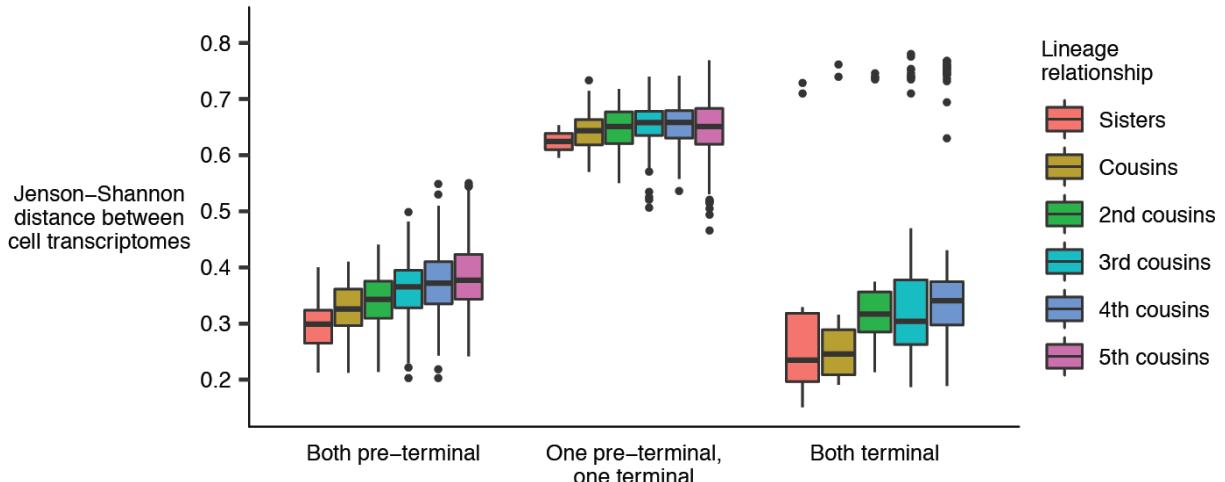
858 **Fig. S26. Counts of differentially expressed genes for lineages that form continuous vs.
859 discontinuous trajectories in UMAP space.** Each row (y-axis) corresponds to a pair of terminal
860 sister cells in the ectoderm (AB lineage, generations 9 and 10) or mesoderm (MS lineage,
861 generation 6). Bar length (x-axis) indicates the number of genes that are both differentially
862 expressed (fold difference > 3, q-value < 0.1) between the sister cells and also differentially
863 expressed (same thresholds) between at least one of the sisters and their parent. Genes that

864 satisfy these criteria are genes that are changing over time in a lineage-specific manner (and
865 therefore exclude broadly expressed genes). Before performing differential expression analysis,
866 the sc-RNA-seq cells that correspond to each of the listed anatomical cells and their parent were
867 downsampled to ensure that each comparison had approximately the same statistical power.
868 Rows are grouped based on whether or not the developmental trajectories formed by the sister
869 cells and their parent in UMAP space were discontinuous for at least one sister. Trajectories were
870 considered discontinuous only if the discontinuity was present in both the global UMAP (**Figs.**
871 **1A, S3**) and the relevant tissue UMAP (**Figs. 3A, S9-10, S17**). Rows are colored to indicate
872 whether or not the sister cells share the same broadly-defined cell type. For example, ASG and
873 AWA, two ciliated neurons, are considered to have the same broadly-defined cell type, while
874 AFD and RMD, a ciliated and non-ciliated neuron respectively, are considered to have different
875 broadly-defined cell types.

876



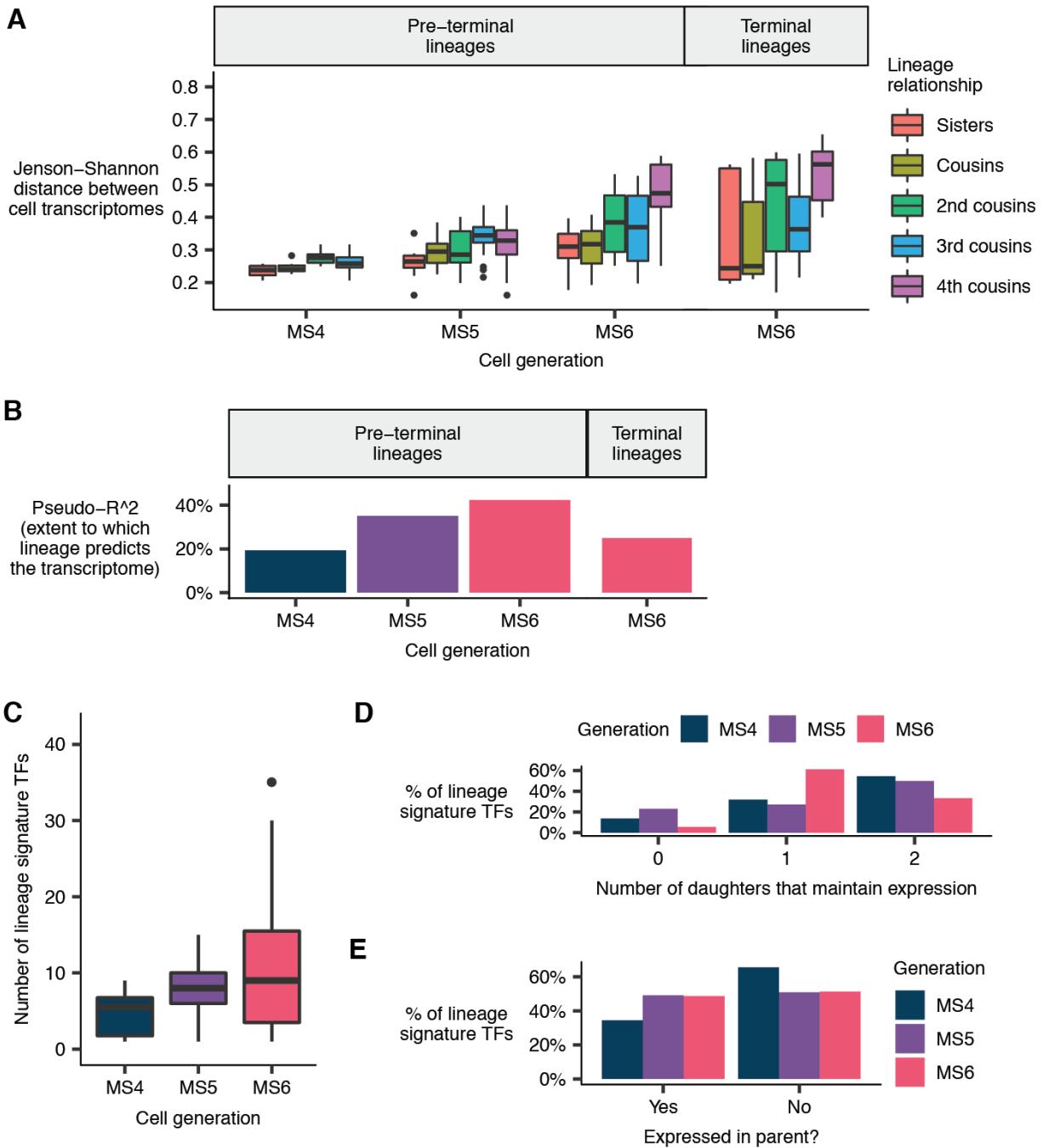
878 **Fig. S27. Embryo time distributions for trajectories included in Fig. S26.** Ridge plot shows
879 the distribution of estimated embryo times (minutes post first cleavage) for all of the sc-RNA-seq
880 cells annotated as one of the terminal cells listed in **Fig. S26**, or its parent. For example, the ridge
881 line for the row labeled AFD has the distribution of embryo times for all sc-RNA-seq cells
882 annotated as either AFD (lineage = ABalpppapav/ABpraaaapav) or the AFD-RMD parent
883 (lineage = ABalpppapa/ABpraaaapa). Rows are grouped based on whether or not the listed
884 terminal cell forms a discontinuous trajectory with its parent in UMAP space. Trajectories were
885 considered discontinuous only if the discontinuity was present in both the global UMAP (**Figs.**
886 **1A, S3**) and the relevant tissue UMAP (**Figs. 3A, S9-10, S17**).
887



888

889 **Fig. S28. Lineage distance vs. transcriptome distance in AB generation 8.** Jensen-Shannon
 890 (JS) distance between the transcriptomes of pairs of cells in AB8, the generation produced by 8
 891 cell divisions since the AB founder cell. Data is faceted by lineage distance and by whether the
 892 pair consists of two pre-terminal cells, one pre-terminal and one terminal cell, or two terminal
 893 cells. Most terminal epidermal cells in the AB lineage are produced in AB8, while most terminal
 894 neurons, glia, and pharyngeal cells are produced in the subsequent generation, AB9. The
 895 terminal epidermal cells in AB8 exit the cell cycle and begin to differentiate, resulting in a large
 896 transcriptome distance between them and neuron/glia/pharynx progenitor cells that remain in the
 897 cell cycle.

898



899

900 **Fig. S29. Correlation between cell lineage and the transcriptome in the mesoderm. (A)**

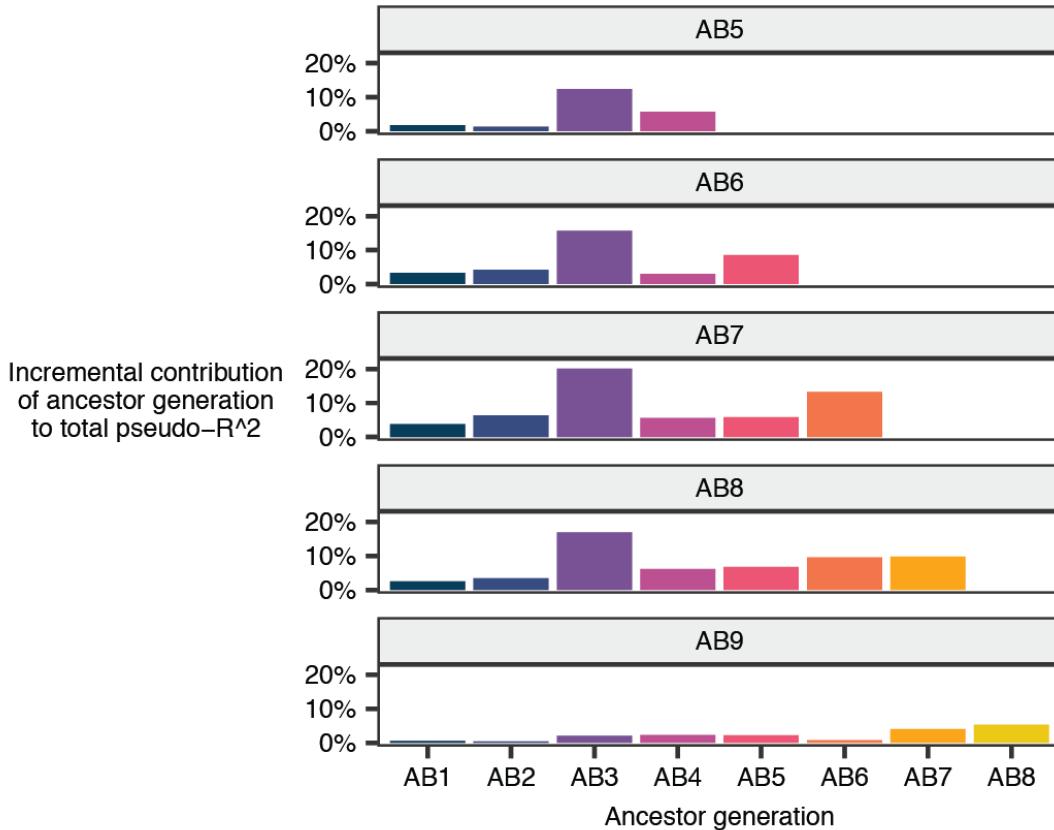
901 Jensen-Shannon (JS) distance between the transcriptomes of pairs of mesoderm cells (MS

902 lineage), faceted by cell generation and lineage distance. MS4 refers to the cell generation

903 produced by 4 divisions of the mesoderm founder cell (MS), and likewise for generations MS5-

904 6. The “transcriptome” of a given anatomical cell is defined as the average gene expression
905 profile of all sc-RNA-seq cells annotated as that anatomical cell. Pairs of bilaterally symmetric
906 cells are excluded from the statistics. The MS6 generation contains both terminal cells and pre-
907 terminal cells that are still dividing. The data for MS6 in the plot is faceted to separate these,
908 comparing only pairs of pre-terminal cells (left panel) or only pairs of terminal cells (right
909 panel). **(B)** Estimates of the extent to which lineage explains the transcriptome in MS4-6, using a
910 pseudo- R^2 statistic (see **Methods**). **(C)** Distribution of the number of “lineage signature
911 transcription factors”—TFs that distinguish a cell from its sister—for a cells in MS4-6. **(D)**
912 Proportion of lineage signature transcription factors for a cell in a given generation that have
913 expression maintained in 0, 1, or 2 of the cell’s daughters in the subsequent generation. **(E)**
914 Proportion of lineage signature TFs for which expression in a given cell was maintained from the
915 cell’s parent vs. newly activated after the parent’s division.

916



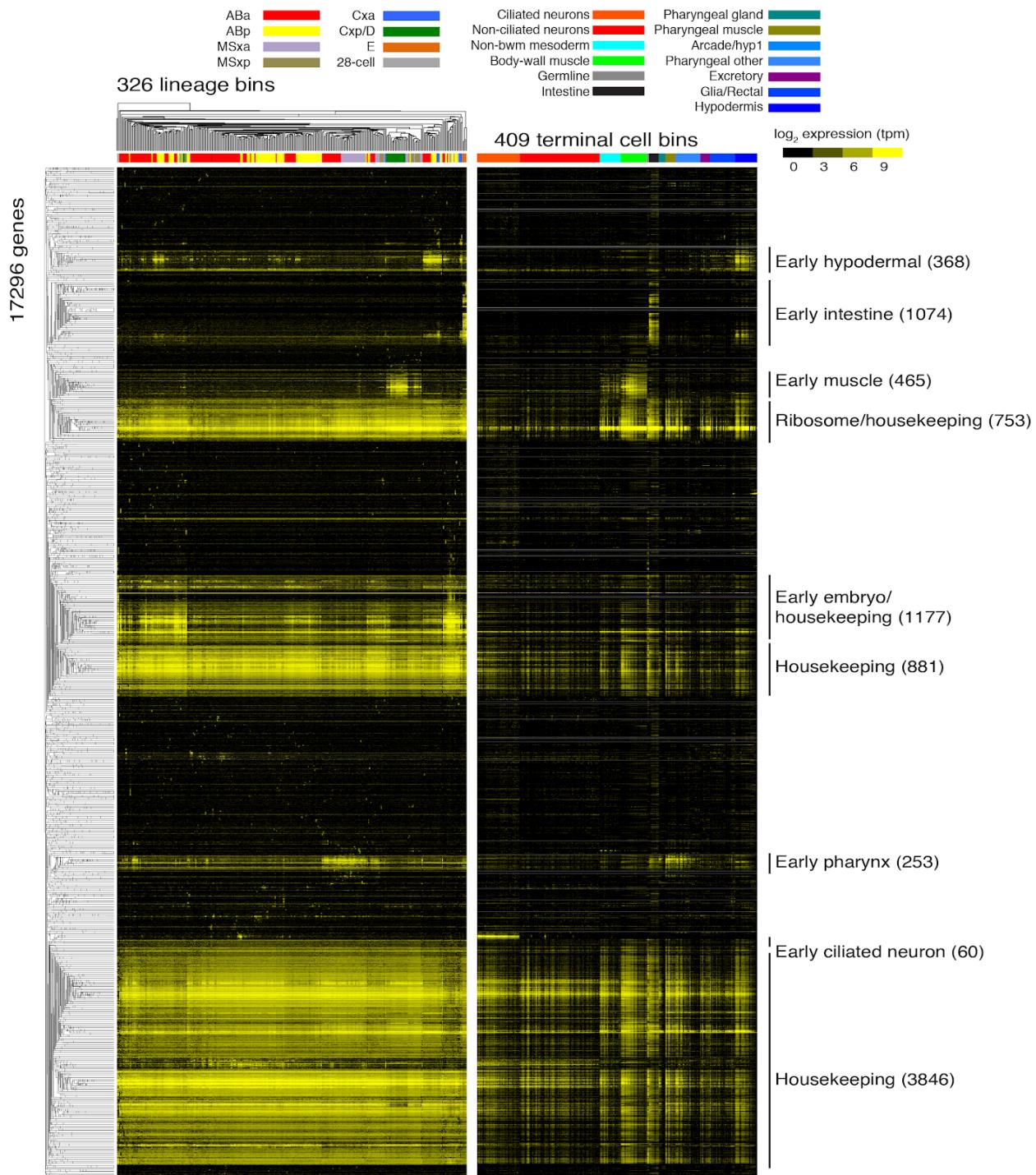
917

918 **Fig. S30. Both recent and distant ancestry contribute to the ability of the lineage to predict
919 a cell's transcriptome.**

920 In **Fig. 5B**, we used a pseudo- R^2 statistic to estimate the extent to which lineage predicts
921 the transcriptomes of cells within a given generation. Specifically, our pseudo- R^2 statistic
922 computes how much more similar are the transcriptomes of sister cells than those of random
923 pairs of cells (see methods section titled “Pseudo- R^2 statistic used in Fig. 5B and Fig. S29B”).

924 Here, we estimate how much of the similarity of sisters is specifically due to gene
925 expression signatures associated with their parent, and how much is due to gene expression
926 signatures associated with more distant ancestors. We describe how these estimates are
927 computed in the methods section titled “Methods used in Fig. S30”.

928 Each panel in the figure corresponds to a generation of the AB lineage. Each bar on the x-
929 axis corresponds to one of the generations that precede it. For example, AB5 is preceded by the
930 generations AB4, AB3, AB2, and AB1. The height of each bar represents the contribution of
931 gene expression signatures associated with that specific ancestor generation to the ability of the
932 lineage to predict the transcriptome in the descendant generation. The sum of the heights of all of
933 the bars in a panel is equal to the total pseudo- R^2 for the descendant generation (**Fig. 5B**).
934

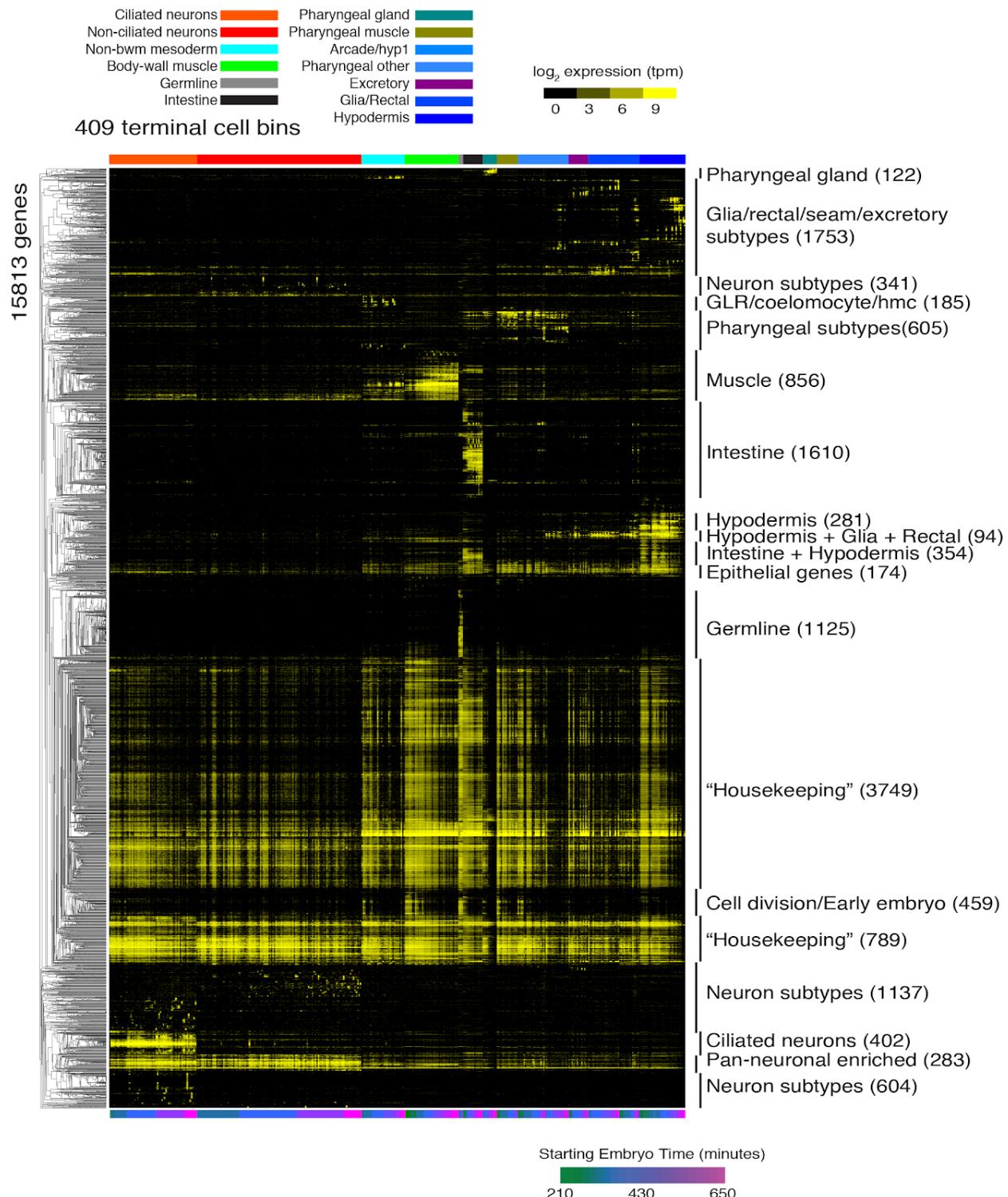


935

936 **Fig. S31. Hierarchical clustering of progenitor lineage transcriptomes.** This heatmap shows
937 the \log_2 expression (\log_2 transcripts per million) of all genes (rows) that are expressed in at least
938 one pre-terminal lineage (columns). Gene expression values are taken from **Table S8**. Genes and
939 lineages are ordered by hierarchical clustering. The right panel shows the expression values in

940 terminal cell bins (**Table S7**), with genes (rows) ordered by the clustering as generated from the
941 pre-terminal lineages and terminal cell bins (columns) ordered as in **Fig. S32**.

942



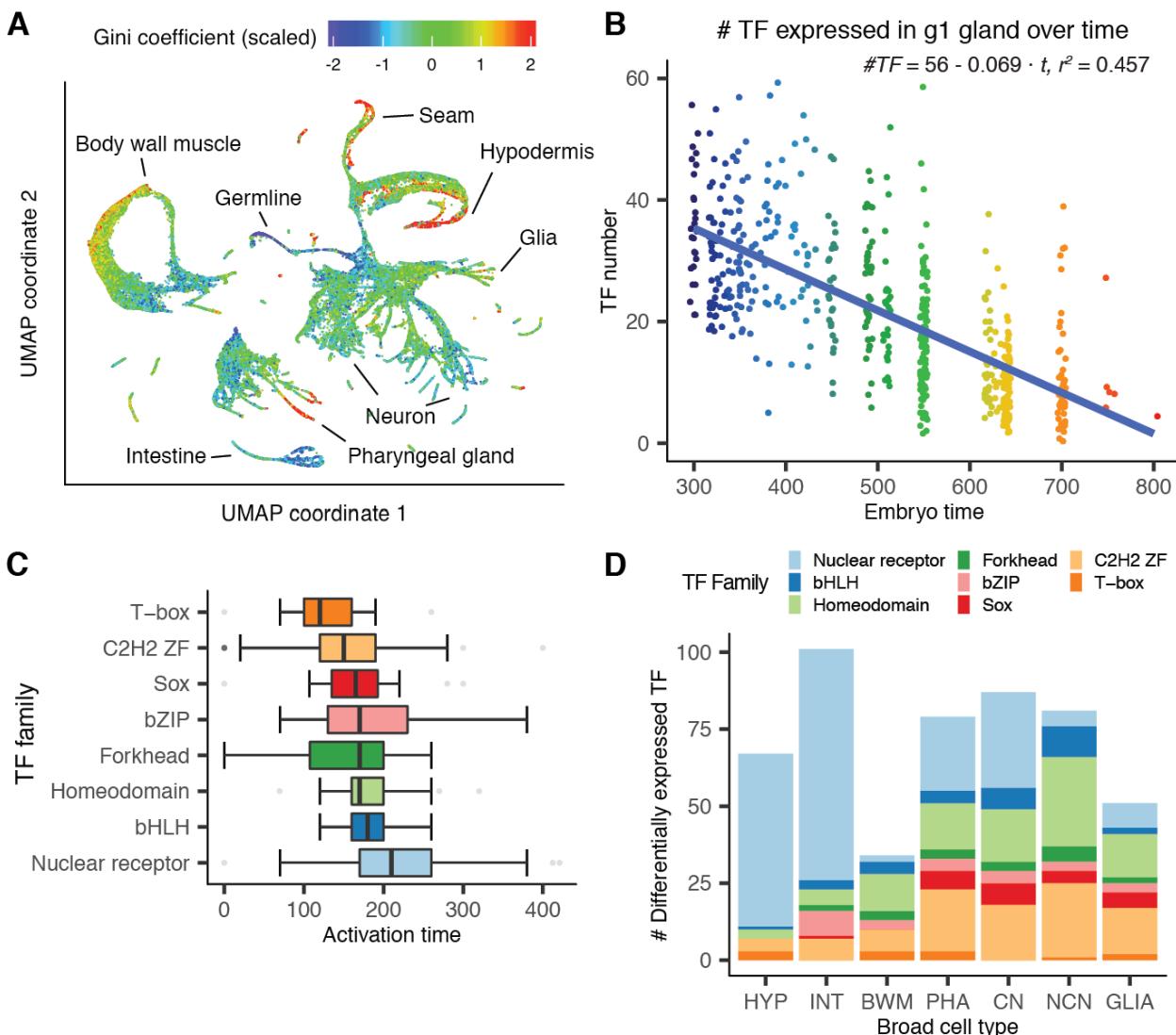
943

944 **Fig. S32. Hierarchical clustering identifies signatures of tissue and cell type differentiation.**

945 This heatmap shows the \log_2 expression (\log_2 transcripts per million) of all genes (rows) that are
 946 expressed in at least one terminal cell bin (columns). Gene expression values are taken from

947 **Table S7.** Genes are ordered by hierarchical clustering, and cell bins are ordered by tissues
948 (colored as in the legend), and within tissues by the beginning of the time bin in minutes (early to
949 late). Gene clusters are labeled by sites of predominant expression. Numbers in parentheses are
950 the number of genes in that cluster.

951



952

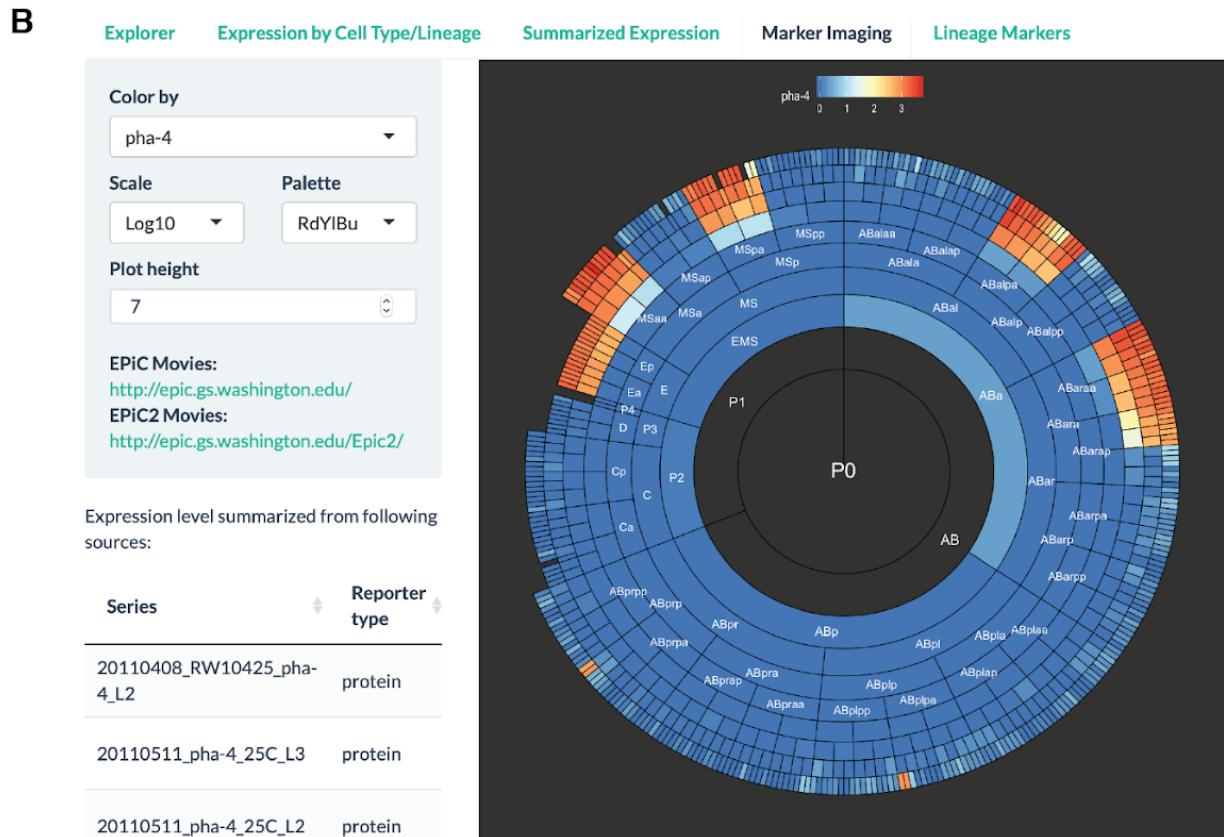
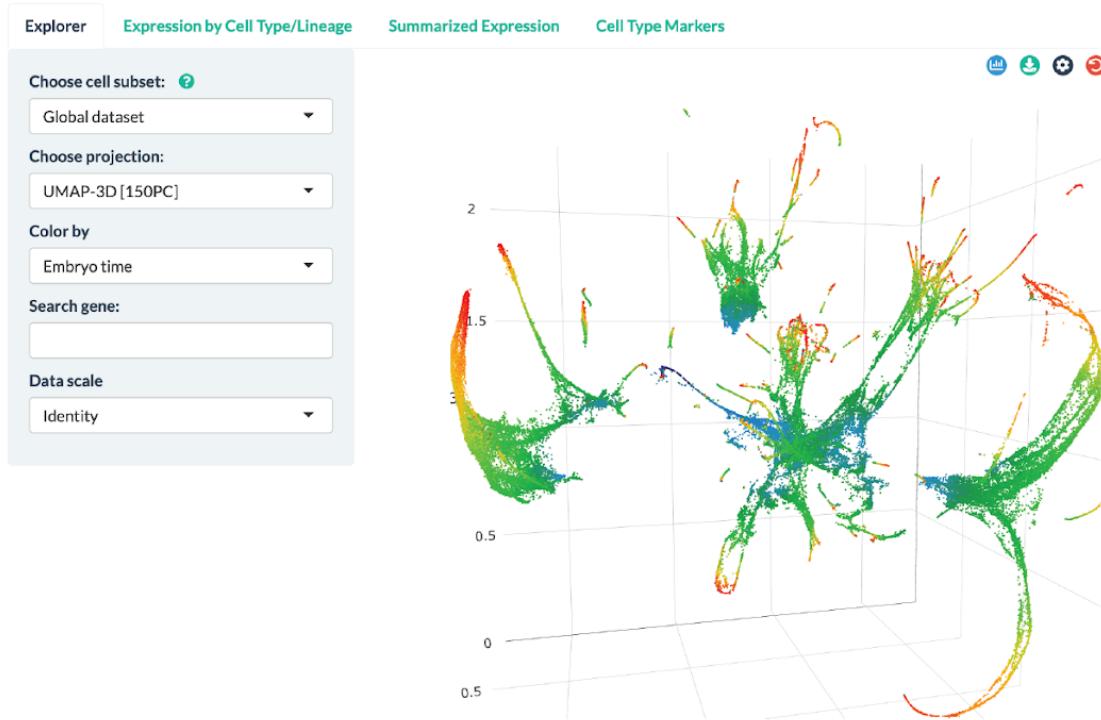
953 **Fig S33. Transcriptome specialization and transcription factor usage across cell types and**
 954 **time.** (A) A global UMAP with 81,286 cells colored by the Gini coefficient of their gene
 955 expression vector, adjusted to correct for sample size bias and scaled by converting to z-scores.
 956 High Gini coefficients indicate that a small set of genes produces a large fraction of cell mRNA
 957 content. (B) Number of TF expressed in g1 gland over time. Equation shows linear regression
 958 result. Points are colored by estimated embryo time. (C) Box plot showing TF activation times—
 959 the embryo time when a TF first becomes expressed—grouped by TF family. For each TF, its
 960 activation time is defined as the 5th percentile of the estimated embryo time values for cells that

961 express that TF. TF family annotations are taken from the CIS-BP database (63). Families that

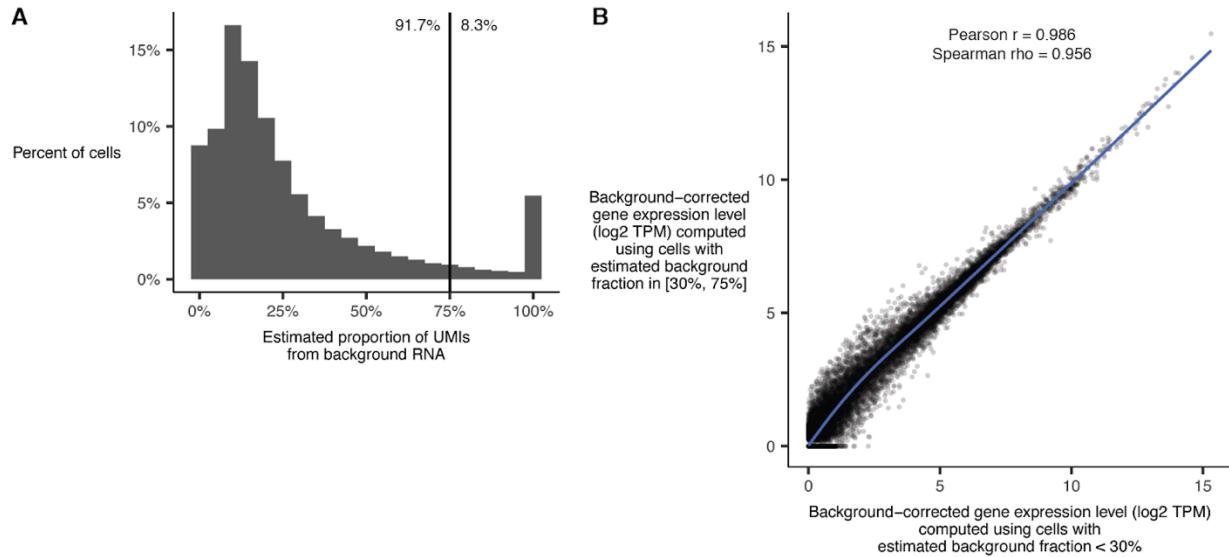
962 have fewer than 10 members detected in the current dataset were excluded from this plot. (**D**)

963 Number of differentially expressed TFs and TF family composition across broad cell types.

964



966 **Fig. S34. Screenshots of VisCello.** (A) Screenshot of the cell type explorer, which enables
967 interactive visualization of 2D and 3D UMAPs and PCA plots for different subsets of the data.
968 The view shown in the panel is a 3D UMAP for all cells colored by estimated embryo time.
969 Users can overlay gene expression, cell type, number of expressed genes and other statistics on
970 this plot. The cell type explorer also features box/violin plots for gene expression across cell
971 types, lineages or time, summarized gene expression tables, and marker gene tables. (B)
972 Screenshot of the early cell lineage explorer, which enables interactive visualization and
973 comparison of the sc-RNA-seq data and summarized live imaging data. Panel shows a
974 radiograph of average fluorescent intensity (log10 scaled) of *pha-4*, measured by live imaging.
975



976

977 **Fig. S35. Distribution of estimates for the proportion of UMIs in a cell that come from**
 978 **background RNA.** (A) The process for making the estimates is described in the methods section
 979 “Per-cell background correction and filtering”. Due to the sparsity of the single cell data, the
 980 estimates are noisy. Numbers to the left and right of the vertical line indicate the proportion of
 981 cells with estimated background fraction < or \geq 75%. Cells with background fraction \geq 75%
 982 are filtered from all downstream analyses. (B) After per-cell background correction, cells with
 983 low and high background fractions have near-identical average gene expression profiles. Plot
 984 shows average gene expression profiles (measured in transcripts per million) computed from
 985 non-head body wall muscle cells divided into two groups: cells with estimated background
 986 fraction < 30% (x axis) and cells with background fraction in the range [30%, 75%].

987

988 **Supplementary Tables**

989

990 Supplementary tables can be found in the online supplementary materials that accompany this
991 manuscript. Tables S1-6, S9, S12-13, and S15-16 are provided in an Excel file,
992 Supplementary_Tables.xlsx. Tables S7-S8, S10-11, and S14 are provided as separate gzip-
993 compressed tab separated value (TSV) files.

994

995 **Table S1. Marker genes for terminal cell type annotations.**

996 This table lists the marker genes that were used to annotate sc-RNA-seq cells with their
997 corresponding cell types. Expression patterns for marker genes were retrieved from Wormbase
998 (26) and EPiC (17).

999

1000 **Table S2. Terminal cell type annotation statistics.**

1001 This table lists statistics related to the cell type annotations, e.g. the number of cells annotated for
1002 each cell type, and median number of UMIs recovered per cell for each cell type.

1003

1004 **Table S3. List of EPiC movies used for lineage annotations.**

1005 This table lists unique identifiers and other information for each of the EPiC (17) movies used in
1006 the process of annotating sc-RNA-seq cells with their corresponding lineage identity. Movies are
1007 available on the EPiC website (<http://epic.gs.washington.edu/Epic2/>). Movies that have not
1008 previously been published are also available in Dryad (53).

1009

1010 **Table S4. Marker genes for lineage annotations.**

1011 This table lists the marker genes that were used to annotate sc-RNA-seq cells with their
1012 corresponding lineage identity. Expression patterns for marker genes were retrieved from
1013 Wormbase (26) and EPiC (17).

1014

1015 **Table S5. Pre-terminal lineage annotation statistics.**

1016 This table lists statistics related to the lineage annotations, e.g. the number of cells annotated for
1017 each lineage, and the median number of UMIs recovered per cell for each lineage.

1018

1019 **Table S6. Map of *C. elegans* anatomical cells to annotations defined in this study.**

1020 This table provides a list of each of the cells in the *C. elegans* embryonic lineage. For the subset
1021 of these cells that were located in our dataset, the table also lists a unique identifier for the
1022 lineage annotation in which the anatomical cell is included. For example, the cells MSaaaa and
1023 MSpaaa is included in a lineage annotation that we named “MSxaaa”.

1024

1025 **Table S7. Gene expression profiles for terminal cell types.**

1026 This table lists the average expression level of each gene in each annotated terminal cell type, for
1027 several different windows of embryo time.

1028

1029 **Table S8. Gene expression profiles for annotated cell lineages.**

1030 This table lists the average expression level of each gene in each lineage annotation.

1031

1032 **Table S9. Differentially expressed genes between all pairs of sister lineages.**

1033 This table lists genes that are differentially expressed between each pair of sister lineages
1034 included in our annotations.

1035

1036 **Table S10. Results from differential gene expression tests between all pairs of sister**
1037 **lineages.**

1038 This table lists the results of differential gene expression tests that compare the expression of
1039 genes between every pair of sister lineages annotated in our dataset. The data in this table is a
1040 superset of the data in Table S9. Table S9 only includes results that are statistically significant,
1041 while this table includes all results.

1042

1043 **Table S11. Results from differential gene expression tests between all pairs of parent vs.**
1044 **daughter lineages.**

1045 This table lists the results of differential gene expression tests that compare the expression of
1046 genes for every pair of parent vs. daughter lineages annotated in our dataset.

1047

1048 **Table S12. Marker genes for cell types at the L2 stage.**

1049 We re-annotated sc-RNA-seq data that we previously generated from L2 stage worms (2),
1050 identifying many more cell types than in the original analysis. This table lists the marker genes
1051 that were used in the re-annotation process. Expression patterns for marker genes were retrieved
1052 from Wormbase (26) and EPiC (17).

1053

1054 **Table S13. Annotation statistics for cell types at the L2 stage.**

1055 We re-annotated sc-RNA-seq data that we previously generated from L2 stage worms (2),
1056 identifying many more cell types than in the original analysis. This table lists statistics related to
1057 the new cell type annotations, e.g. the number of cells annotated for each cell type, and median
1058 number of UMIs recovered per cell for each cell type.

1059

1060 **Table S14. Gene expression profiles for cell types at the L2 stage.**

1061 We re-annotated sc-RNA-seq data that we previously generated from L2 stage worms (2),
1062 identifying many more cell types than in the original analysis. This table lists the average
1063 expression level of each gene in each cell type. The data for some cell types is additionally
1064 faceted by estimated progression through the molting cycle, e.g. whether a cell comes from an
1065 animal at the mid L2 stage vs. the late L2 stage.

1066

1067 **Table S15. Candidate terminal selectors of neuron types identified by differential gene
1068 expression analysis.**

1069 This table lists associations between neuron types and genes that satisfy all of the following
1070 criteria:

- 1071 1. The gene is a transcription factor and is robustly expressed in the neuron type (lower
1072 bound of 95% bootstrap confidence interval > 0).
- 1073 2. The gene is at least 5-fold enriched in the neuron type compared to its sister in the cell
1074 lineage, and this difference is significant ($q\text{-value} < 0.05$).
- 1075 3. Expression of the gene is maintained in the neuron type at the L2 stage (lower bound of
1076 95% bootstrap confidence interval for expression > 0).

1077

1078 **Table S16. Body wall muscle cells associated with each position-related group in Fig. 4C**
1079 **and 4D.**

1080 This table summarizes the marker gene expression patterns that we used to associate anatomical
1081 BWM cells to the transcriptomically-defined subsets shown in Fig. 4C and 4D.

1082

1083 **Additional Data Files**

1084

1085 An additional data file, **Additional Data File 1**, can be found in the online supplementary
1086 materials that accompany this manuscript.

1087

1088 **Additional Data File 1**

1089 This file (`Additional_Data_File_1_WS260_with_3p_UTR_extensions.gtf.gz`) is provided as a
1090 gzip-compressed GTF (gene transfer format) file. The file contains a modified version of the
1091 WormBase (26) reference transcriptome for *C. elegans* (version WS260), in which we have
1092 extended the 3' UTRs of some of the transcript annotations. See the section of the Materials and
1093 Methods titled “Read mapping and gene expression quantification”.

1094

References and Notes

1. X. Qiu, A. Hill, J. Packer, D. Lin, Y.-A. Ma, C. Trapnell, Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
[doi:10.1038/nmeth.4150](https://doi.org/10.1038/nmeth.4150) [Medline](#)
2. J. Cao, J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, J. Shendure, 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) [Medline](#)
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**, eaaq1736 (2018). [doi:10.1126/science.aaq1736](https://doi.org/10.1126/science.aaq1736) [Medline](#)
4. M. Plass, J. Solana, F. A. Wolf, S. Ayoub, A. Misios, P. Glažar, B. Obermayer, F. J. Theis, C. Kocks, N. Rajewsky, Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* **360**, eaqaq1723 (2018). [doi:10.1126/science.aaq1723](https://doi.org/10.1126/science.aaq1723) [Medline](#)
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) [Medline](#)
6. A. Zeisel, H. Hochgerner, P. Lönnerberg, A. Johnsson, F. Memic, J. van der Zwan, M. Häring, E. Braun, L. E. Borm, G. La Manno, S. Codeluppi, A. Furlan, K. Lee, N. Skene, K. D. Harris, J. Hjerling-Leffler, E. Arenas, P. Ernfors, U. Marklund, S. Linnarsson, 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) [Medline](#)
7. A. Sebé-Pedrós, B. Saudemont, E. Chomsky, F. Plessier, M.-P. Mailhé, J. Renno, Y. Loe-Mie, A. Lifshitz, Z. Mukamel, S. Schmutz, S. Novault, P. R. H. Steinmetz, F. Spitz, A. Tanay, H. Marlow, 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) [Medline](#)
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) [Medline](#)
9. N. Karaiskos, P. Wahle, J. Alles, A. Boltengagen, S. Ayoub, C. Kipar, C. Kocks, N. Rajewsky, R. P. Zinzen, 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) [Medline](#)
10. J. A. Briggs, C. Weinreb, D. E. Wagner, S. Megason, L. Peshkin, M. W. Kirschner, A. M. Klein, 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) [Medline](#)
11. D. E. Wagner, C. Weinreb, Z. M. Collins, J. A. Briggs, S. G. Megason, A. M. Klein, Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018). [doi:10.1126/science.aar4362](https://doi.org/10.1126/science.aar4362) [Medline](#)

12. J. A. Farrell, Y. Wang, S. J. Riesenfeld, K. Shekhar, A. Regev, A. F. Schier, Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018). [doi:10.1126/science.aar3131](https://doi.org/10.1126/science.aar3131) [Medline](#)
13. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, 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) [Medline](#)
14. B. Pijuan-Sala, J. A. Griffiths, C. Guibentif, T. W. Hiscock, W. Jawaid, F. J. Calero-Nieto, C. Mulas, X. Ibarra-Soria, R. C. V. Tyser, D. L. L. Ho, W. Reik, S. Srinivas, B. D. Simons, J. Nichols, J. C. Marioni, B. Göttgens, 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) [Medline](#)
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) [Medline](#)
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) [Medline](#)
17. J. I. Murray, T. J. Boyle, E. Preston, D. Vafeados, B. Mericle, P. Weisdepp, Z. Zhao, Z. Bao, M. Boeck, R. H. Waterston, Multidimensional regulation of gene expression in the *C. elegans* embryo. *Genome Res.* **22**, 1282–1294 (2012). [doi:10.1101/gr.131920.111](https://doi.org/10.1101/gr.131920.111) [Medline](#)
18. M. Sarov, J. I. Murray, K. Schanze, A. Pozniakovski, W. Niu, K. Angermann, S. Hasse, M. Rupprecht, E. Vinis, M. Tinney, E. Preston, A. Zinke, S. Enst, T. Teichgraber, J. Janette, K. Reis, S. Janosch, S. Schloissnig, R. K. Ejsmont, C. Slightam, X. Xu, S. K. Kim, V. Reinke, A. F. Stewart, M. Snyder, R. H. Waterston, A. A. Hyman, 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) [Medline](#)
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) [Medline](#)
20. X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H. A. Pliner, C. Trapnell, 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) [Medline](#)
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) [Medline](#)
22. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. [arXiv:1802.03426](https://arxiv.org/abs/1802.03426) [stat.ML] (9 February 2018).
23. E. Becht, L. McInnes, J. Healy, C. A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginkhoux, E. W. Newell, 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) [Medline](#)

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](https://arxiv.org/abs/0803.0476) [physics.soc-ph] (4 March 2008).
26. R. Y. N. Lee, K. L. Howe, T. W. Harris, V. Arnaboldi, S. Cain, J. Chan, W. J. Chen, P. Davis, S. Gao, C. Grove, R. Kishore, H.-M. Muller, C. Nakamura, P. Nuin, M. Paulini, D. Raciti, F. Rodgers, M. Russell, G. Schindelman, M. A. Tuli, K. Van Auken, Q. Wang, G. Williams, A. Wright, K. Yook, M. Berriman, P. Kersey, T. Schedl, L. Stein, P. W. Sternberg, 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) [Medline](#)
27. C. L. Araya, T. Kawli, A. Kundaje, L. Jiang, B. Wu, D. Vafeados, R. Terrell, P. Weissdepp, L. Gevirtzman, D. Mace, W. Niu, A. P. Boyle, D. Xie, L. Ma, J. I. Murray, V. Reinke, R. H. Waterston, M. Snyder, 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) [Medline](#)
28. G. Broitman-Maduro, M. Owraghi, W. W. K. Hung, S. Kuntz, P. W. Sternberg, M. F. Maduro, 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) [Medline](#)
29. W. C. Spencer, G. Zeller, J. D. Watson, S. R. Henz, K. L. Watkins, R. D. McWhirter, S. Petersen, V. T. Sreedharan, C. Widmer, J. Jo, V. Reinke, L. Petrella, S. Strome, S. E. Von Stetina, M. Katz, S. Shaham, G. Rätsch, D. M. Miller 3rd, 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) [Medline](#)
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) [Medline](#)
31. M. Hu, D. Krause, M. Greaves, S. Sharkis, M. Dexter, C. Heyworth, T. Enver, 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) [Medline](#)
32. P. Laslo, C. J. Spooner, A. Warmflash, D. W. Lancki, H.-J. Lee, R. Sciammas, B. N. Gantner, A. R. Dinner, H. Singh, 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) [Medline](#)
33. M. Thomson, S. J. Liu, L.-N. Zou, Z. Smith, A. Meissner, S. Ramanathan, 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) [Medline](#)
34. E. W. Brunskill, J.-S. Park, E. Chung, F. Chen, B. Magella, S. S. Potter, 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) [Medline](#)
35. W. Wang, X. Niu, T. Stuart, E. Jullian, W. M. Mauck 3rd, R. G. Kelly, R. Satija, L. Christiaen, 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) [Medline](#)

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) [Medline](#)
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) [Medline](#)
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) [Medline](#)
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). [Medline](#)
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) [Medline](#)
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) [Medline](#)
42. M. Harterink, D. H. Kim, T. C. Middelkoop, T. D. Doan, A. van Oudenaarden, H. C. Korswagen, 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) [Medline](#)
43. K. Brunschwig, C. Wittmann, R. Schnabel, T. R. Bürglin, H. Tobler, F. Müller, Anterior organization of the *Caenorhabditis elegans* embryo by the labial-like Hox gene ceh-13. *Development* **126**, 1537–1546 (1999). [Medline](#)
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) [Medline](#)
45. See supplementary materials and methods (section titled “Pseudo- R^2 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) [Medline](#)
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) [Medline](#)
49. S. J. Husson, T. Janssen, G. Baggerman, B. Bogert, A. H. Kahn-Kirby, K. Ashrafi, L. Schoofs, 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) [Medline](#)
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) [Medline](#)
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, D. Vafeados, L. Gevirtzman, M. Corson, R. Terrell, E. Preston, A. Zacharias, J. I. Murray, R. H. Waterston, Data from: A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution, Dryad (2019); <https://doi.org/10.5061/dryad.7tg31p7>.
54. A. D. Warner, L. Gevirtzman, L. W. Hillier, B. Ewing, R. H. Waterston, The *C. elegans* embryonic transcriptome with tissue, time, and alternative splicing resolution. *Genome Res.* **29**, 1036–1045 (2019). [doi:10.1101/gr.243394.118](https://doi.org/10.1101/gr.243394.118) [Medline](#)
55. M. D. Young, S. Behjati, SoupX removes ambient RNA contamination from droplet based single cell RNA sequencing data. bioRxiv 303727 [Preprint]. 20 April 2018. <https://doi.org/10.1101/303727>.
56. M. E. Boeck, C. Huynh, L. Gevirtzman, O. A. Thompson, G. Wang, D. M. Kasper, V. Reinke, L. W. Hillier, R. H. Waterston, The time-resolved transcriptome of *C. elegans*. *Genome Res.* **26**, 1441–1450 (2016). [doi:10.1101/gr.202663.115](https://doi.org/10.1101/gr.202663.115) [Medline](#)
57. D. Yu, W. Huber, O. Vitek, Shrinkage estimation of dispersion in Negative Binomial models for RNA-seq experiments with small sample size. *Bioinformatics* **29**, 1275–1282 (2013). [doi:10.1093/bioinformatics/btt143](https://doi.org/10.1093/bioinformatics/btt143) [Medline](#)
58. UCLA Statistical Consulting Group, FAQ: What are pseudo-R-squareds? (2006); <https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/>.
59. G. Deltas, The Small-Sample Bias of the Gini Coefficient: Results and Implications for Empirical Research. *Rev. Econ. Stat.* **85**, 226–234 (2003). [doi:10.1162/rest.2003.85.1.226](https://doi.org/10.1162/rest.2003.85.1.226)
60. R. Kolde, pheatmap: Pretty Heatmaps. CRAN (2019); <https://CRAN.R-project.org/package=pheatmap>.
61. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017). [doi:10.1038/nmeth.4463](https://doi.org/10.1038/nmeth.4463) [Medline](#)
62. E. M. Sommermann, K. R. Strohmaier, M. F. Maduro, J. H. Rothman, Endoderm development in *Caenorhabditis elegans*: The synergistic action of ELT-2 and -7 mediates the specification→differentiation transition. *Dev. Biol.* **347**, 154–166 (2010). [doi:10.1016/j.ydbio.2010.08.020](https://doi.org/10.1016/j.ydbio.2010.08.020) [Medline](#)

63. M. T. Weirauch, A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.-C. Lozano, M. Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F.-Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker, T. R. Hughes, Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014). [doi:10.1016/j.cell.2014.08.009](https://doi.org/10.1016/j.cell.2014.08.009) [Medline](#)
64. I. Abdus-Saboor, C. E. Stone, J. I. Murray, M. V. Sundaram, The Nkx5/HMX homeodomain protein MLS-2 is required for proper tube cell shape in the *C. elegans* excretory system. *Dev. Biol.* **366**, 298–307 (2012). [doi:10.1016/j.ydbio.2012.03.015](https://doi.org/10.1016/j.ydbio.2012.03.015) [Medline](#)
65. A. Ebbing, Á. Vértesy, M. C. Betist, B. Spanjaard, J. P. Junker, E. Berezikov, A. van Oudenaarden, H. C. Korswagen, Spatial Transcriptomics of *C. elegans* Males and Hermaphrodites Identifies Sex-Specific Differences in Gene Expression Patterns. *Dev. Cell* **47**, 801–813.e6 (2018). [doi:10.1016/j.devcel.2018.10.016](https://doi.org/10.1016/j.devcel.2018.10.016) [Medline](#)
66. A. Streit, R. Kohler, T. Marty, M. Belfiore, K. Takacs-Vellai, M.-A. Vigano, R. Schnabel, M. Affolter, F. Müller, Conserved regulation of the *Caenorhabditis elegans* labial/Hox1 gene ceh-13. *Dev. Biol.* **242**, 96–108 (2002). [doi:10.1006/dbio.2001.0544](https://doi.org/10.1006/dbio.2001.0544) [Medline](#)
67. G.-J. Hendriks, D. Gaidatzis, F. Aeschimann, H. Großhans, Extensive oscillatory gene expression during *C. elegans* larval development. *Mol. Cell* **53**, 380–392 (2014). [doi:10.1016/j.molcel.2013.12.013](https://doi.org/10.1016/j.molcel.2013.12.013) [Medline](#)
68. E. M. Hedgecock, J. G. Culotti, D. H. Hall, B. D. Stern, Genetics of cell and axon migrations in *Caenorhabditis elegans*. *Development* **100**, 365–382 (1987). [Medline](#)