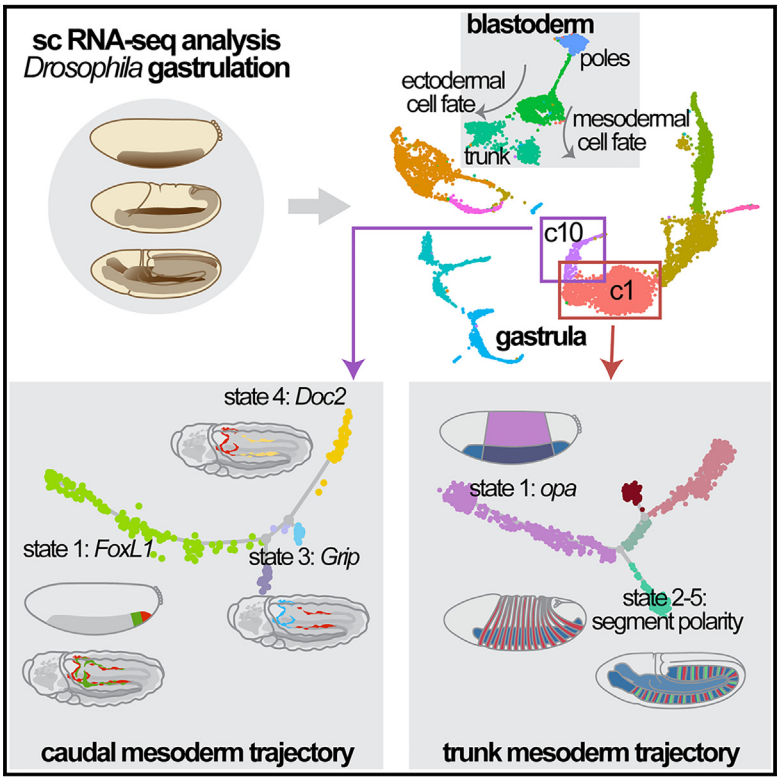
Resource

Cell Reports

# Single-cell transcriptomics illuminates regulatory steps driving anterior-posterior patterning of *Drosophila* embryonic mesoderm

## Graphical abstract




## Highlights

- Single-cell transcriptomic analysis of *Drosophila* gastrulation using live cells
- Cells spanning 90 min of development were profiled to illuminate temporal transitions
- Spatial information captured includes differences in trunk vs. poles and germ layers
- Mesoderm patterning along the anterior-posterior axis and progenitors uncovered

## Authors

Jingjing Sun, Chen Zhang, Fan Gao, Angelike Stathopoulos

## Correspondence

angelike@caltech.edu

## In brief

Sun et al. used single-cell RNA-seq to investigate how overall gene expression progresses to accommodate cellular changes and coordinate the morphogenetic movements associated with *Drosophila* gastrulation. Identifying regulatory steps involved in patterning the mesoderm, this study provides insights into the mechanisms regulating EMT and collective cell migration during morphogenesis.

Sun et al., 2023, Cell Reports *42*, 113289
October 31, 2023 
https://doi.org/10.1016/j.celrep.2023.113289

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

# Single-cell transcriptomics illuminates regulatory steps driving anterior-posterior patterning of *Drosophila* embryonic mesoderm

Jingjing Sun,[1] Chen Zhang,[1] Fan Gao,[2,3] and Angelike Stathopoulos[1,4,*]
[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA
[2]Bioinformatics Resource Center, Beckman Institute, California Institute of Technology, Pasadena, CA 91125, USA
[3]Present address: Lyterian Therapeutics, South San Francisco, CA 94080, USA
[4]Lead contact
*Correspondence: angelike@caltech.edu
https://doi.org/10.1016/j.celrep.2023.113289

## SUMMARY

Single-cell technologies promise to uncover how transcriptional programs orchestrate complex processes during embryogenesis. Here, we apply a combination of single-cell technology and genetic analysis to investigate the dynamic transcriptional changes associated with *Drosophila* embryo morphogenesis at gastrulation. Our dataset encompassing the blastoderm-to-gastrula transition provides a comprehensive single-cell map of gene expression across cell lineages validated by genetic analysis. Subclustering and trajectory analyses revealed a surprising stepwise progression in patterning to transition zygotic gene expression and specify germ layers as well as uncovered an early role for ecdysone signaling in epithelial-to-mesenchymal transition in the mesoderm. We also show multipotent progenitors arise prior to gastrulation by analyzing the transcription trajectory of caudal mesoderm cells, including a derivative that ultimately incorporates into visceral muscles of the midgut and hindgut. This study provides a rich resource of gastrulation and elucidates spatially regulated temporal transitions of transcription states during the process.

## INTRODUCTION

Quantitatively capturing the transcription state of an embryonic cell is essential for uncovering its identity, while tracking the dynamic transcriptional changes within the lineage is imperative for understanding the differentiation trajectory as well as for deciphering the gene regulatory networks associated with cell-fate specification.[1] We here apply single-cell sequencing technology in conjunction with conventional genetic tools and methods to investigate the molecular mechanisms driving the *Drosophila* embryo developmental progression from blastula to gastrula. By focusing on the mesoderm, we uncovered the transcription states that account for progenitor cell specification and the regulation that underlies epithelial-to-mesenchymal transition (EMT)[2,3] during gastrulation.

Gastrulation constitutes a crucial stage during development when metazoan animals acquire cellular diversity and set up basic body plans.[4–6] Initialized with totipotency, embryonic cells are specialized, forming three germ layers, rearranged through complex morphogenetic movements, and eventually shaped into a body architecture with various rudimentary organs. Transcription changes very rapidly prior to gastrulation during early embryogenesis. The initial 13 rounds of nuclear division in early embryogenesis take place in a syncytial embryo without cytokinesis, and all gap phases (i.e., G1 and G2) are bypassed.[7] Maternal factors are degraded, and zygotic transcription takes over gradually through two waves of zygotic gene activation (ZGA).[8–10] The major wave of gene expression associated with ZGA coincides with the 14$^{th}$ nuclear cycle (nc14), during which time cell membranes grow between the blastoderm nuclei and cellularization takes place.[11–13] Spatially localized gene transcription in the ventral domain (e.g., the expression of *snail* [*sna*] and *twist* [*twi*]) associated with the presumptive mesoderm at mid-nc14 is the first sign of lineage diversification during embryogenesis. By the end of nc14, mesoderm cells are specified downstream of Toll signaling at the ventral blastoderm, and invagination of those cells leads to mesoderm formation. Developmental time is the major variable that distinguishes gene expression programs of cell populations prior to the emergence of germ layers. However, once cell movements begin during gastrulation, it is generally thought that gene expression programs do not exhibit widespread changes until the cells reach their destination and initiate differentiation programs. Whether this is indeed the case remains unknown, and the interplay of transcriptional programs guiding these complex cellular transitions is also poorly explored. We hypothesized that the single-cell approach might be able to capture the dynamic transcriptional changes in migrating cells as well as elucidate the gene regulatory landscape during this important transition phase.

Gastrulation starts at stage (st) 6, following the accomplishment of dorsal-ventral (DV) patterning at st 5, as *sna* and *twi* genes function together to control the multifaceted cellular

changes that initiate and drive presumptive mesoderm internalization, ultimately resulting in a tube-like structure forming at the ventral side of the gastrula at st 7.[14–16] Subsequently, at st 8, the tube disassembles as mesoderm cells in the trunk lose apicobasal polarity, detach from each other, enter into the first mitotic cycle after nc14, and through an EMT acquire a migratory ability supporting their movement in the dorsal direction.[17,18] Concomitantly, these cells move in the posterior direction together with the ectoderm during germband extension (GBE) to double the length of the embryonic trunk along the anterior-posterior (AP) axis.[5] The coupling between EMT and GBE, two pronounced morphogenetic cellular movements at gastrulation, signifies the cross-regulation between the DV- and AP-axis patterning programs, respectively. Single-cell technology provides us with a unique opportunity to investigate the genetic links connecting EMT with GBE[19] with high spatiotemporal resolution by tracking gene expression trends in trunk mesoderm cells.

Several developmental timer genes, or timing factors, have been shown to control embryonic AP patterning at gastrulation,[20,21] but their role in supporting patterning and specification of progenitor cells within the mesoderm has not been characterized. These include *odd paired* (*opa*), which is broadly expressed in the trunk region at nc14 and acts as a pioneer factor to modulate chromatin accessibility and thereby regulate the spatiotemporal dynamics of zygotic gene expression.[22,23] Among Opa targets are the pair-rule and segment polarity genes,[20,22,24] which function in a cascade of segmentation genes to cooperatively pattern the *Drosophila* embryo trunk along the AP axis. The posterior pole, or "tail," of the embryo is patterned after the trunk under the regulation of the terminal system by the gene *tailless* (*tll*). This process requires *opa* and at least two additional timing factors, *caudal* (*cad*) and *Dichaete* (*D*).[21,25] The caudal mesoderm contains multipotent progenitors for muscle cells and Malpighian tubules (MTs), the *Drosophila* kidney equivalent.[26–28] In contrast to mesoderm cells in the trunk (which adopt different cell fates in response to inductive signals originating in the ectoderm), caudal mesoderm cells are migratory precursor cells that remain undifferentiated until they are incorporated into organs. While much focus has been on AP patterning of the ectoderm, the mechanism for subdivisions within the mesoderm remains unclear, including how the posterior founder populations and segmentation gene expression are regulated. Gene expression information at a single-cell resolution also holds the potential to uncover the transcription signatures within a trajectory of cell states that lead to the lineage commitment of posteriorly located progenitor cells.

## RESULTS

### Single-cell transcriptomic profiling and lineage identification of *Drosophila* embryos during gastrulation

To investigate transcriptional changes during gastrulation, we collected embryos from 3 to 5 hours after egg laying, representing st 5–10, and processed them for single-cell RNA sequencing (scRNA-seq) analysis (Figure 1A; see STAR Methods). The uniform manifold approximation and projection (UMAP) dimension reduction with 20 principal components (PCs) resulted in 12 unique cell clusters, c1–c12 (numbered based on cell population size ranking of the cluster, Figures 1B and S1B), which can be identified by the associated marker genes (Figure 1C and Table S1). Cells sampled using this approach covered the expected developmental stages (blastoderm and gastrula) and cell lineages, including mesoderm, endoderm and ectoderm anlage, primordia, and a sampling of differentiating cell types. Gene Ontology (GO) enrichment analysis of marker genes associated with each cluster supports the view that cells found in clusters have specialized functions (Figure S1A).

Aiming to investigate the transcription programs specifically within the mesoderm, we used a lineage tracer consisting of *UAS-mCD8GFP*[29] reporter and *twi(2xPE)-GAL4*[30] driver to follow *GFP* reporter gene expression in the embryo starting at st 5 and compared it with the expression of gene *twi* in mesoderm lineages (Figure 1D). Levels of *twi* expression are highest in undifferentiated mesoderm c1 and c10 (Figure 1H). *mCD8-GFP* is also enriched in c1 and c10; however, it is more strongly expressed in c3 and c8, representing more differentiated trunk mesoderm and specialized head mesoderm cell lineages (Figure 1H). These results further confirm the identities of the mesoderm clusters (i.e., *twi>mCD8GFP* largely follows the transcriptional activity of endogenous gene *twi*) but with a delay imposed by indirect expression through use of GAL4/UAS (i.e., c3 is older than c1).

Our data provide a comprehensive single-cell map to track gene expression of different cell lineages during gastrulation. For example, the head mesoderm cluster (c8) marked by *glial cells missing* (*gcm*), a zinc-finger transcription factor required for plasmatocyte differentiation, is well separated from the rest of the mesoderm lineages (i.e., c1, 3, 4, 10, 12) (Figure 1E compared with 1D, 1H). This is consistent with its head-specific expression as shown by *in situ* hybridization from st5 to st8 (Figure 1F) as well as its expression pattern predicted by the *Drosophila* Virtual Expression eXplorer (DVEX), an online tool published with a previous single-cell transcriptomic study of embryos at st 6[31] (Figure S1C). Our analyses also demonstrate a high degree of cell-type coverage within the embryo (i.e., median ~15,000 unique molecular identifiers [UMIs] per cell for our wild-type #1 [WT#1] sample), and deeper sequencing resulted in a single-cell transcriptome dataset containing high-information content that will illuminate even low-abundance transcripts. For example, with fewer than 10,000 cells covering multiple developmental stages, we were able to detect rare cell types of mesoderm origin such as crystal cells, which constitute only ~5% of the population[32] in comparison with *gcm*-marked plasmatocytes, which represent 90%–95% of the *Drosophila* blood cells. Despite their low abundance, we can identify the crystal cell marker *Prophenoloxidase 1* (*PPO1*) within c8 of our dataset (Figures 1G and 1H).

The marker genes were identified using Seurat differential expression tests for this single-cell transcriptome dataset (Figure 1C; Table S1) and were used to assign an identity to each of the 12 cell clusters based on previous studies and gene annotations at the Berkeley *Drosophila* Genome Project (BDGP) database and FlyBase (see STAR Methods). This dataset (live WT#1) is the focus of our study. In repeat experiments (live WT#2 and fixed WT#3), also with integrated dataset, cell clusters with similar lineage identities (Figures S2D and S2F) were recovered (see STAR Methods). As *Drosophila* single cells were isolated from a staged

CellPress
OPEN ACCESS

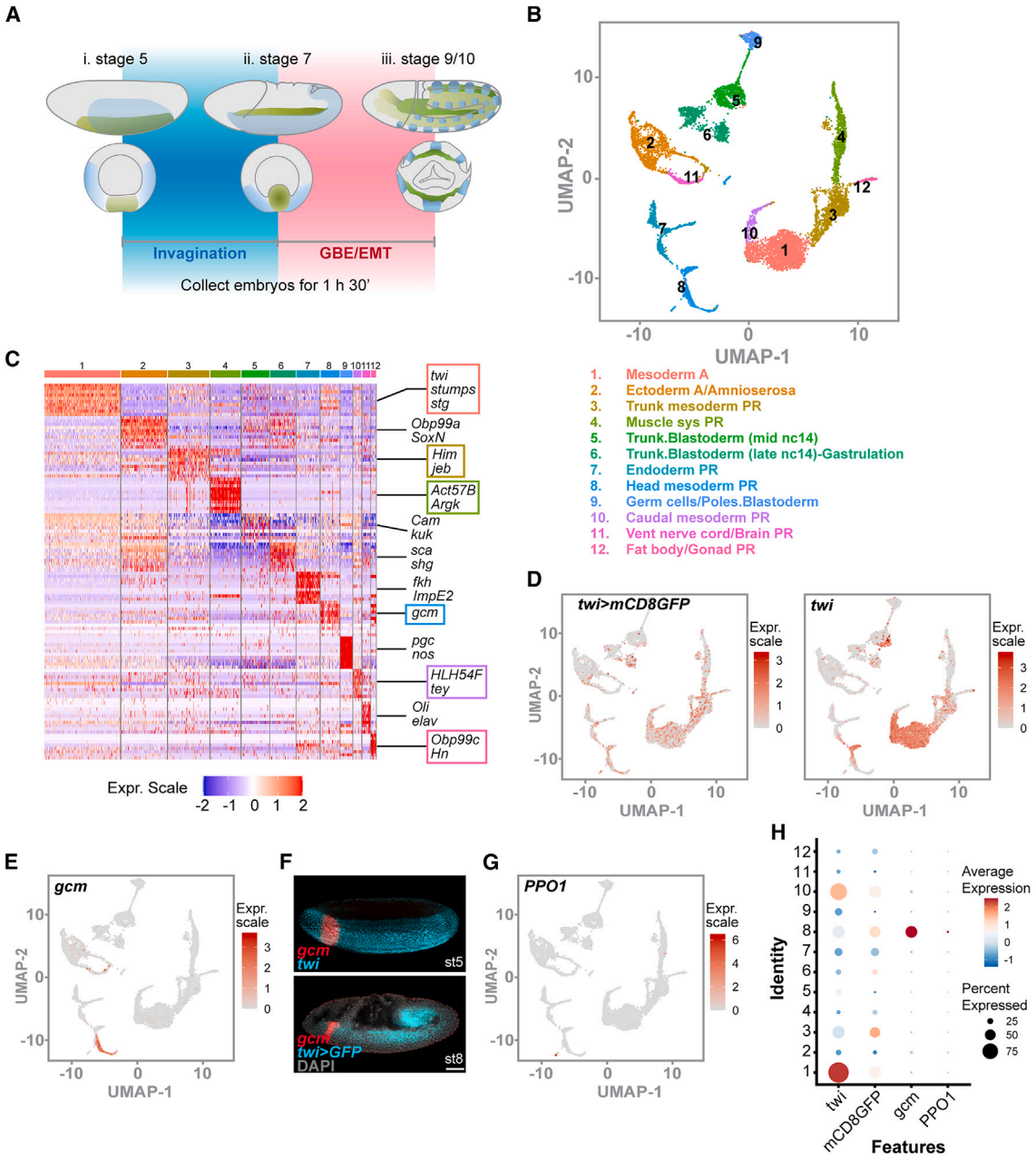



**Figure 1. scRNA-seq analysis of *Drosophila* embryos at gastrulation highlights distinct mesoderm cell lineages**

(A) Diagram illustrates timing of embryo collection and expected stages at which cells were isolated (see STAR Methods).

(B) The UMAP dimensionality reduction analysis revealed 12 cell clusters that were annotated based on their associated marker genes. Annotation follows Berkeley Drosophila Genome Project (BDGP) database guidelines: A, anlage (i.e., rudimentary part/organ); PR, primordium (i.e., organ in earliest recognizable state).

(C) Heatmap of the top 10 marker genes expressed in each cluster with examples labeled on the right. Mesoderm lineage markers are indicated within the boxes (see also Table S1).

(D) Expression of *GFP*, from *twi(2xPE)GAL4*-driven UAS-*mCD8GFP*, and *twi* genes. Red color intensity showing the expression levels or the number of reads at a global scale (normalized by the entire cell population).

(E–G) Detection of head mesoderm cell lineages. UMAP plot and *in situ* hybridization (red, F) of the head mesoderm marker *gcm*. *twi* (cyan, top) and *twi>GFP* (cyan, bottom) transcripts were labeled simultaneously with *gcm* at st 6 and 8, respectively, to indicate mesoderm lineage. DAPI (gray) was used for counter-staining. Scale bar, 50 μm. (G) Presence of *PPO1*-positive crystal cells in the head mesoderm cluster.

(H) Dotplot of *twi*, *mCD8GFP* driven by *twi(2xPE)GAL4*, *gcm*, and *PPO1* expression in all clusters. Expression levels and percentage of cells in the cluster (y axis) that express indicated genes (x axis) are shown. See also Figure S1 and Table S1.

embryo collection allowing approximately 90-min difference in age, we expect the clusters to segregate based on gene expression programs that differ with respect to both developmental timing as well as spatial registry within the embryo. We therefore examined marker genes using *in situ* hybridization to confirm the categorization (i.e., Figure 1B) and to provide additional insights into the spatiotemporal dynamics in transcription.

To further support data reproducibility, we integrated our dataset (WT#1) with a published transcriptome dataset from stage 12 embryos[33] (Figures S1D and S1E) and compared the mesoderm cell lineages contributed from these two sources. We found that the markers for more differentiated mesoderm cell clusters representing somatic, visceral muscle (s.c2, s.c10), fat body (s.c5), and hemocytes (s.c.12) are highly consistent, whereas the clusters associated with younger mesoderm cells, as expected, predominantly stem from our data focused on gastrulation (s.c4 and s.c17) (see Figure S1F).

## Establishing domains and germ layers in the blastoderm embryos

Marker genes were used to map cell clusters to germ layers and domains. To start, we focused on the cells relating to the blastoderm embryo, which encompasses nc14 within st 5. Nc14 is the longest nc, ~60 min in length at 22°C.[9] During early- to mid-nc14, the embryos exhibit an increase in zygotic transcription while maternal factors continue being degraded, and cellularization is completed by the end of nc14. Previous studies have examined gene expression dynamics at nc14 and characterized expression changes for select marker genes with high temporal resolution.[9,34] For example, by mid-nc14, maternally deposited transcripts of gene *deadhead* (*dhd*) are depleted, *heartless* (*htl*) and *grainyhead* (*grh*) are switched on, and *sna* and *twi* levels peak, while expression of *Cadherin-N* (*CadN*), activated at early nc14, continues to increase until gastrulation (Figures 2C and 2D).[9,35,36] Transcripts of early zygotic gene *nullo* and *bottleneck* (*bnk*) both play an essential part in cellularization and are completely eliminated from the embryo by late nc14.[37,38] Interestingly, *bnk* is only retained in the poles at mid-nc14 (Figure 2C).[38] We used this prior knowledge of dynamic gene expression trends to relate clusters c5, c6, and c9 to cells in the blastoderm embryo (Figure 2A). c9 comprises cells that express *dhd*, *bnk*, and *nullo*, likely derived from embryos at mid-nc14 (youngest possible stage based on our collection timing); c6 clearly contains older cells—likely of late nc14 to stage 6—due to higher levels of *CadN* and lack of *dhd*, *bnk*, and *nullo* (compare c9 vs. c6; Figures 2B and S2A). c5 appears to be a transitional state at which all above-mentioned genes are detectable but at low levels (Figure 2B). Therefore, based on their expression profiles, c9 and c5 cells might be at a similar stage and both derived from mid-nc14 embryos, whereas c6 cells are "older" and c5 → c6 represents a temporal progression in transcriptional states from cellularization to the beginning of gastrulation.

In addition, several markers for c6 are spatially localized to regions that give rise to either ectoderm or mesoderm, suggesting that the difference in expression between c9/c5 vs. c6 relates to specification of the three germ layers. *Short gastrulation* (*sog*), a marker for the ectoderm, is expressed in lateral regions (Figures 2G and 2I); cells expressing *sog* in the UMAP plot are represented on the left side of c6 and are most abundant in c2 that constitutes ectoderm/amnioserosa (Figure 2J *sog*). In contrast, *htl*, like *twi*, is expressed in ventral presumptive mesoderm cells (Figure 2I); *htl*-expressing cells are represented on the right side of c6 and most abundantly present in mesodermal lineages including c1, c3, c8, and c10 (Figure 2J *htl*; see also Figure 1D *twi*). Lastly, *DNaseII* is an example of a terminally expressed gene (i.e., anterior and posterior poles; Figure 2I); cells expressing *DNaseII* are present in the center of c6 and predominantly in c7, enriched for genes associated with endoderm development (Figure 2J *DNaseII*). The distinct localization of ectodermal and mesodermal markers in c6 (e.g., Figure 2J *sog*/ectoderm to the left vs. *htl*/mesoderm to the right of UMAP plots) suggests that germ layers arise at c6 during late nc14. This is also supported by trajectory analysis of the transcriptome of cells comprising c5, c6, and c9, which demonstrates that three transcriptional states emerge: state 1 is associated with maternal and pole cell markers (*dhd*, *pgc*), state 2 with ectodermal markers (*Lac*, *grh*), and state 3 with mesodermal ones (*llp4*, *twi*) (Figures 2E, 2F, and S2B; Table S3.1).

To further validate our conclusions regarding germ-layer specification timing, we subjected the c5 and c6 clusters to subclustering in an unbiased manner and identified five and three subclusters, respectively (Figures S2K and S2L; see STAR Methods). The three subclusters associated with c6 each relate to one of the three germ layers: ectoderm (c6 subcluster1: c6.1, marked by *Lac*, *grh*), mesoderm (c6.2, marked by *CadN*, *htl*), or endoderm (c6.3, marked by *DNaseII*, *Pdp1*) (Figures 2K and 2L; see also Figure S2M and Table S3.4). This result confirms germ layers are established in late nc14. In contrast, only one of the five subclusters associated with c5, c5.5, is associated with germ-layer-specific markers corresponding to the presumptive mesoderm (Figures S2J–S2L; Table S3.3), supporting the idea that c5 represents mostly uncommitted younger cells compared with c6, likely from mid-nc14 embryos.

As c9 is associated with marker genes that relate to pole cells (germline) as well as terminal patterning (Figures 2B and S2A; Table S1), we used subclustering to provide additional insights into its identity. This subclustering produced three classes (Figure S2H; Table S3.2). c9.1 expresses high levels of *dhd*, maternal transcripts known to be mostly degraded by mid-nc14 but retained in pole cells[39,40]; therefore, these are likely pole cells (Figures S2G and S2I). c9.2 is marked by expression of *tailless* (*tll*), *ribbon* (*rib*), and *zerknüllt* (*zen*), all known to be expressed in the anterior and posterior poles at mid-nc14[41] (Figures S2G and S2H). c9.3 appears to be anterior biased as it is enriched for anterior ectoderm marker *CG42342* as well as *Psc* and *CG45782*, both genes known to be expressed in the future head structure at the embryonic anterior (Figures S2G and S2I).[42–45] The fact that genes expressed at the poles of the embryo (both anterior and posterior; Table S1) are enriched in c9 cells while c5 is marked by *homothorax* (*hth*), only expressed in the trunk blastoderm, with little to no caudal genes expression (e.g., *tll* and *cad*) (Figures S2G–S2L; Table S1), suggests that the distinction between c5 and c9 likely reflects the spatial differences of cells in the embryo (trunk vs. poles) translated into differences in their transcriptional states.

CellPress
OPEN ACCESS

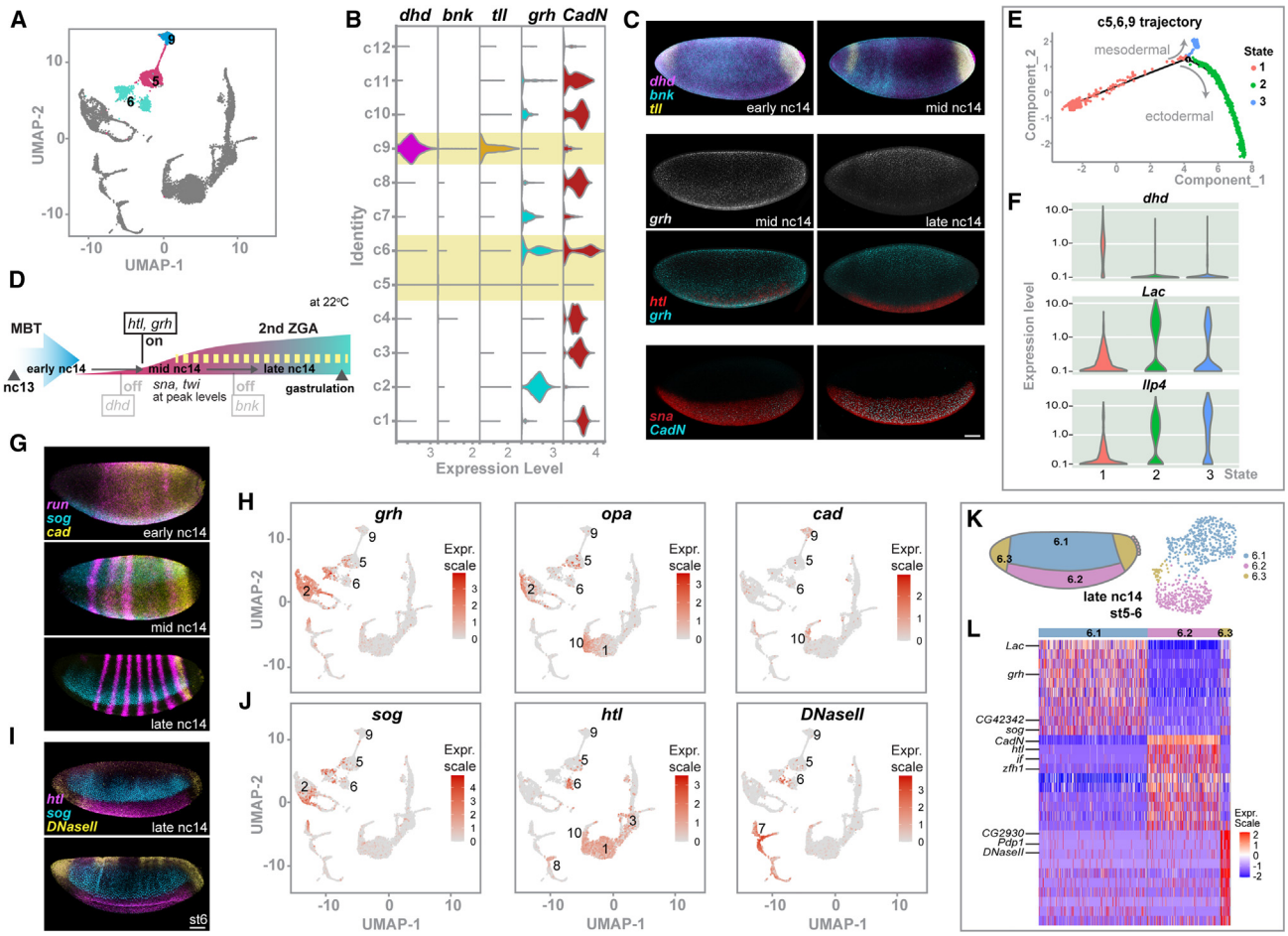



**Figure 2. Profiling transcription in the blastoderm at single-cell resolution reveals distinctive transcriptional programs among ectoderm, endoderm, and mesoderm lineages at nc14**

(A–D) Blastoderm embryos contribute to three clusters in the UMAP plot: c9, c5, c6 (A). Violin plots (B) and *in situ* hybridization (C) showing representative gene expression in the blastoderm embryos. Schematic representation of transcription changes during nc14 at 22°C; yellow dashed line indicates the expected age of embryos in the collection that by 195 min should be at or older than mid-nc14 (D).

(E and F) Trajectory analysis of cells specifically from c9, c5, and c6 generates 1–3 cell states (E) that are enriched for *dhd*, *Lac*, and *Ilp4*, respectively (F).

(G—J) *In situ* hybridization showing representative marker gene expression in different AP domains or germ layers of the blastoderm embryo during nc14: trunk (*run*) vs. posterior poles (*cad*) (G); ectoderm (*sog*) vs. mesoderm (*htl*) vs. endoderm (*DNaseII*) (I). UMAPs of selected marker genes highlighting timing factors (H) and germ layer-specific genes (J). Red color intensity indicates the relative expression levels at a global scale.

(K and L) UMAP plot (K) and heatmap (L) of c6 cells that are subdivided into three subclusters (c6.1, c6.2, and c6.3). Diagram illustrates a cellularized blastoderm embryo with color-coded germ layers that are labeled with their associated subcluster. Scale bars, 50 μm. See also Figure S2 and Tables S3–S5.

From our subclustering analysis, we identified additional steps in global gene expression related to activities of timing factors that contribute to embryonic patterning in the blastoderm. These factors, including *opa*, *cad*, and *grh*, function cooperatively to control gene expression at specific developmental time points.[22,23,25] It is thought that the initial specification of anterior-trunk-posterior programs along the AP axis relates to the localized expression of *opa* (trunk) and *cad* (posterior)[21,23,25](Figures 2G and 2H). *grh*,[46] expressed in both the trunk and poles, is associated with cells in the ectoderm and endoderm anlagen but excluded from mesoderm cells, suggesting a role for it in patterning along the DV axis (Figures 2C and 2H).[47,48] This single-cell approach provides a platform for systematic identification of genes co-expressed with such timing factors that also contribute to supporting a particular event in cellular differentiation and morphogenesis. The association of timing factors *grh*, *opa*, and *cad* with c5 and/or c9 (mid-nc14) while germ-layer emergence is associated with c6 (late nc14) supports the view that these clusters relate to transcriptional changes driving germ layer formation through the integration of the two patterning systems (AP and DV) (c6; Figures 2K, 2L, and S2M).

## Transcription factors responsible for ectoderm patterning are expressed in the mesoderm at stage 7 and play a role to repress endoderm fate in the trunk region

As our goal was to follow the gene expression program of the mesoderm germ layer during gastrulation, we next examined

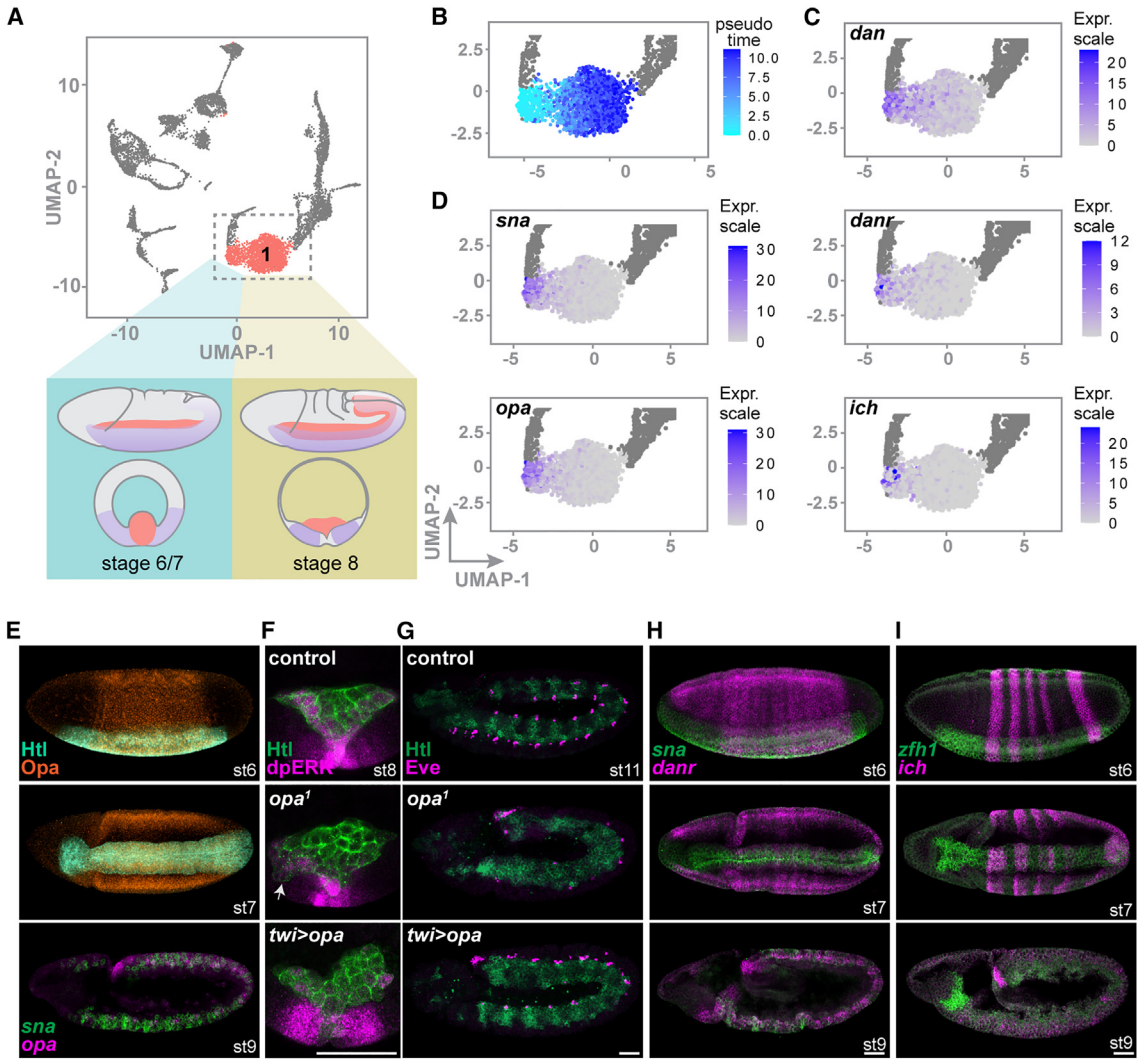



**Figure 3. Transcription factors responsible for ectoderm patterning are expressed in the mesoderm as EMT is initiated and play a role to repress endoderm fate in the trunk region**

(A) UMAP highlighting trunk mesoderm cells of c1 during gastrulation. Diagram depicting the morphological changes in both whole-mount and cross-sectioned views.

(B) UMAP of c1 in pseudotime with darker shade of blue indicating later developmental time points.

(C and D) Cropped UMAP plots of featured gene expression within c1; cells in view but associated with clusters other than c1 shown in dark gray. Purple intensity indicates the relative expression levels.

(E–G) Opa's role in the mesoderm. Expression of Opa protein (orange; E) and *opa* transcripts (purple, E) from st 6–9. Mesoderm is labeled by an anti-Htl antibody (green; E–G) or riboprobes to *sna* (green; E, bottom). Diphosphorylated ERK (dpERK) antibody staining (purple) indicates active receptor tyrosine kinase (RTK) signaling (purple, F). Loss of *opa* or overexpression of *opa* in the mesoderm results in asymmetry in mesoderm EMT (transverse plane, F; arrow indicates ectopic protrusions associated with $opa^1$ mutant) as well as a loss of pericardial cells (anti-Eve, purple) at st 11 (G).

(H and I) Expression of *danr* and *ich* (purple) in the mesoderm (labeled with riboprobes to *sna* and *Zn finger homeodomian 1* (*zfh1*) in green) detected using riboprobes. Stages are labeled at the bottom right. Scale bars, 50 μm. See also Figure S3.

the most prominent *twi*-positive cluster, c1 (Figure 3A). This cluster contains trunk mesoderm cells that relate to st 6–8, when mesoderm cells either are invaginating to form a tube (st 6–7) or going through an EMT (st 8). The presence of *sna* and *opa* transcripts serves as a reference for staging cells in c1. The expression of both genes is initiated prior to gastrulation and continues until st 7. By st 8, however, their transcription levels have decreased significantly in the mesoderm and proteins are no longer detectable by antibody staining (e.g., Figures 3D, 3E, and S3A).[49] Trajectory analysis (see STAR Methods) confirmed that cells that express *sna* and *opa* are positioned on one end (left side) of the pseudotime trajectory (Figure 3D compared with 3B; see also Figure 4A) and, in contrast, factors known to be expressed later are positioned on the other end (right side). St 8 (~50% of embryos in the collection) cells are expected to be the major component of c1, which is enriched for the GO

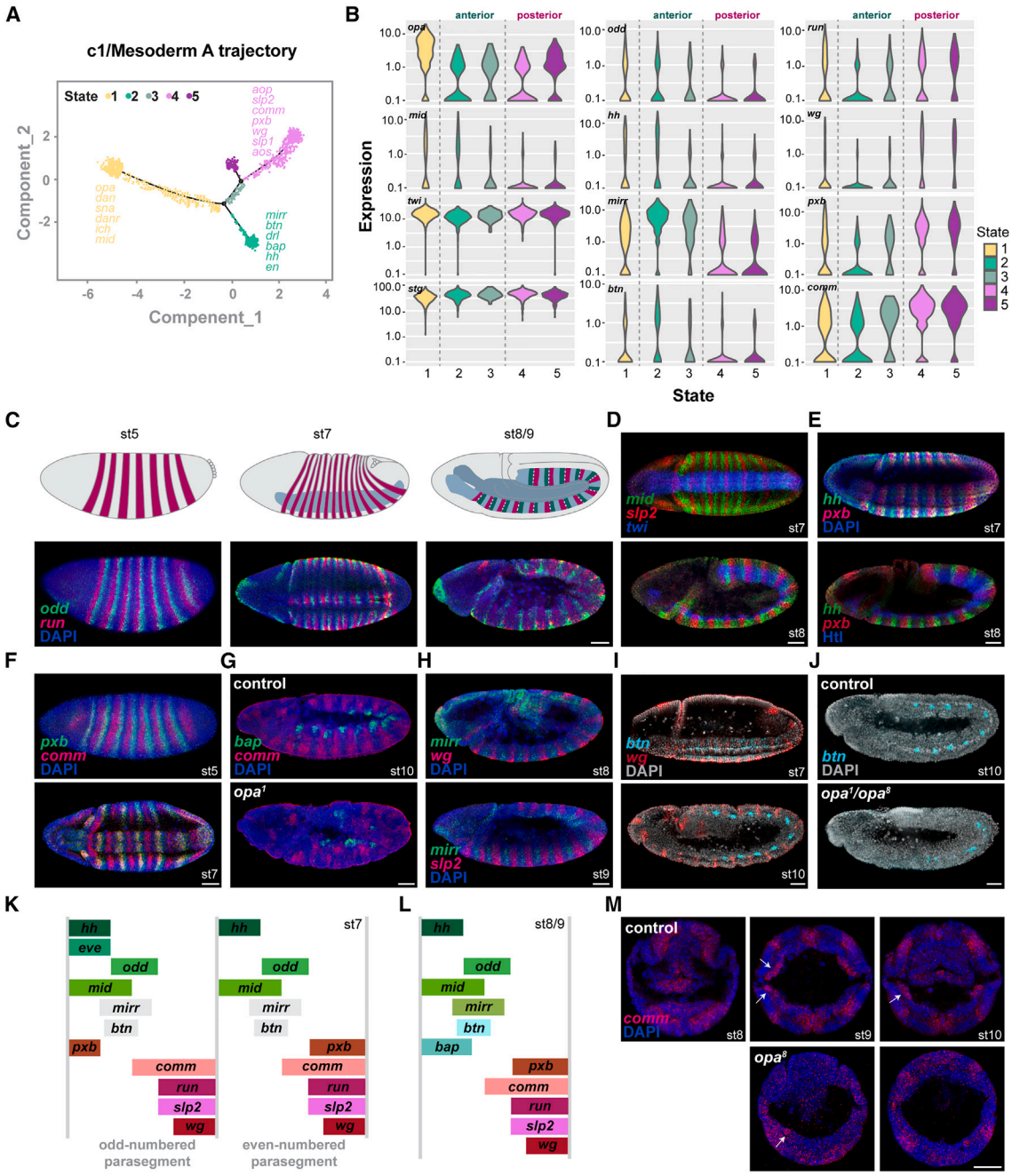



**Figure 4. Subclustering and trajectory analysis of cluster 1 mesoderm cells suggests that the transcription program establishing segment polarity follows EMT**

(A and B) Monocle trajectory analysis reveals five distinct cellular states (c1.1–c1.5) within c1 cells with marker genes shown (A). Violin plots of selected marker gene expression in cells at different states (B).

(C) Mesoderm-centered view of segmentation showing *odd/run* pair-rule gene expression during the process. Mesoderm is drawn in dark blue, gene expression in a segmented pattern is labeled in fuchsia from st5 to st7, while, at st8/9, the anterior parasegment expression is labeled in green (e.g., *hh*) and the posterior is labeled in fuchsia (e.g., *wg*). White dashed lines indicate the posterior/anterior boundary of the parasegments at st8/9.

(D–J) Selected marker genes of cellular state 2/3 (anterior parasegment) vs. state 4/5 (posterior parasegment) confirmed by *in situ* hybridization. Genes marking the anterior parasegment are in green while genes expressed in the posterior parasegment are in red. In $opa^{1}$ mutants, *comm* and *bap*, normally expressed in complementary domains, are detected in irregular patterns (G), while *btn* expression is greatly reduced (J).

(K and L) Diagram depicting the segmented expression of selected marker genes from cellular state 2–5 (including *mid*) at st7 (L) and st8/9 (M).

(M) Cross sections of st8-10 embryos (transverse plane) showing localized expression of *comm* in the dorsal migrating mesoderm cells (arrows) in control vs. $opa^{8}$ mutants. Probe sets for *in situ* hybridization and stage of the embryos are indicated. Scale bars, 50 μm. See also Figure S4 and Table S6.

terms relating to cell cycle (Figures 1C and S1A). Therefore, c1, containing cells from st 6/7 to st 8, represents mesoderm cells undergoing EMT and the expression of *opa*, only at the earlier stage, points to a regulatory role for it at gastrulation.

Such a role for *opa* in the trunk mesoderm (e.g., st 7–10) had not been demonstrated previously. It has been shown that *opa* is required in the mesoderm at a later stage during the specification of visceral mesoderm (st 11 and beyond).[50] We found that, in $opa^{1}$ mutants, EMT is aberrant (Figures 3F and S3C). Mesoderm patterning is also affected, as almost all pericardial cell precursors marked by Even-skipped (Eve) fail to be specified (Figures 3G and S3D). *opa* mutants do also exhibit ectodermal patterning defects[20,24] that could indirectly affect the mesoderm (e.g., misplaced Pyramus [Pyr] and reduced number of *slp1* stripes; Figure S3B). Disrupting Opa function by RNAi using the *twi(2xPE)GAL4* does not effectively knock down its expression or affect mesoderm patterning as only minor defects in EMT are observed (Figure S3E). Therefore, we cannot unequivocally conclude an autonomous requirement for Opa in supporting mesodermal patterning. However, ectopic expression of Opa in the mesoderm interferes with these same processes (Figure 3G), suggesting that precise control of Opa activity within the mesoderm proper is important to support normal mesoderm EMT and cell differentiation.

Within *opa* and *sna* marked cell state 1 (of the five states identified within c1 by trajectory analysis; see Figure 4A), we were surprised to find additional transcription factors with unknown functions in mesoderm development. These include *distal antenna-related* (*danr*) and *ichor* (*ich*), detected in the trunk but excluded from the posterior and largely also from the anterior domains (Figures 3C, 3H, and 3I). As state 1 represents the younger st 6–7 mesoderm cells (discussed above; compare Figures 3C–3B), these factors may relate to a trunk-specific gene expression program that acts at gastrulation. Furthermore, they exhibit similar expression trends to *opa* when visualized by UMAP (Figure S3F; compare with Figure 2H, *opa*). As such, these genes may be co-regulated and involved in common functions and developmental processes with *opa*,[51,52] acting to support gene expression in the embryonic trunk. The coordinated expression of this cohort of trunk-enriched transcription factors immediately prior to EMT is suggestive of a regulatory role for these genes in mesoderm cell organization and behavior at EMT (see section "discussion").

## Mesoderm cell transcription programs establish segment polarity following EMT

The segmented body plan in insects is built upon the modular organization of similar units, called parasegments, that are arranged serially along the AP axis.[53] In *Drosophila*, segmentation has been well studied in the ectoderm. Briefly, following the initial formation of the AP axis that takes place before cellularization, three classes of segmentation genes are activated sequentially in the blastoderm. First, expression of the gap genes in broad bands leads to the activation of pair-rule genes in a series of seven transverse stripes along the AP axis (Figure 4C). The combination of both stimulating and repressive inputs from the pair-rule gene products then establishes the expression patterns of the segment polarity genes.[54] These genes, such as *engrailed* (*en*) and *wingless* (*wg*), play a conserved role in defining the boundary between each parasegment, which falls at the anterior edge of the *en* domain or the posterior edge of the *wg* domain[25,55] (Figure 4C).

As discussed above, ordering c1 cells along a developmental trajectory identifies five distinct cell states, with the *opa*-marked cell state 1 (c1.1) representing cells at the earlier stage on one end and cell states 2–5 (st 8) with two additional branches on the other (Figure 4A). Consistent with *opa*'s role in regulating the segmentation gene network, markers of cell states 2–5 arise later in development within the trunk region and include various pair-rule and segment polarity genes (Figures 4A, 4B, S4A, and S4B; Table S6). Violin plots for selected markers indicate that cell states 2 and 3 (c1.2 and c1.3), associated with high expression of *en* and *hedgehog* (*hh*), comprise cells located at anterior parasegments. In comparison, cell states 4 and 5 (c1.4 and c1.5), associated with high expression of *wg* and pair-rule genes *sloppy paired 1/2* (*slp1/2*), comprise cells located at posterior parasegments (Figure 4B *hh*, *wg*; Figure S4B: *en*, *slp2*).

Several lesser-studied segmentally expressed genes were also uncovered. For instance, T-box transcription factor *midline* (*mid*) is localized anterior to *slp2* in 14 transverse stripes at st 7 (Figure 4D) and reportedly plays a role in maintaining segment polarity in the ectoderm by repressing *wg*.[56] *mid* is a marker of c1.1, whereas *slp2* is a marker for c1.2 and c1.3 (Figures 4B, S4A, and S4B). At later stages, *mid* has been found to contribute to somatic muscle morphogenesis and heart development,[57,58] consistent with a potential function in mesoderm patterning. In addition, we found that *pxb*, which encodes a transmembrane protein and was identified as an attenuator of Hh signaling in imaginal discs,[59] is expressed in a highly dynamic pattern within the mesoderm from st 7 to st 8. At st 7, transcripts of *pxb* overlap with *hh* in the anterior odd-numbered parasegments but are excluded from the anterior region of even-numbered parasegments (Figures 4E and 4K). During st 8, the seven *pxb* stripes double to form 14 stripes and shift to the posterior region of the parasegments (Figures 4E and 4L). The doubling of *pxb* stripes in the mesoderm follows the ectoderm in time and is in register with the ectodermal stripes spatially. We also detected the *mirror* (*mirr*) gene, initially discovered as a regulator that defines the DV boundary in the eye,[60] in 14 stripes positioned anterior to *wg* and *slp2* from st 8 (Figures 4H and S4C). *mirr* encodes one of the three Iroquois homeobox transcription factors that contribute to the differentiation of cardiac progenitors (Figure S4D).[61] Finally, we also observe *commissureless* (*comm*), well documented for its role in axon guidance in the nervous system[62,63] and recently shown to regulate ectodermal patterning during germband elongation (GBE) through enriching Myosin II at the compartmental boundaries,[64] already transitioned to single-segment periodicity in expression (i.e., 14 stripes) by st 7 and localizing to the posterior parasegments (Figures 4F and 4K). Intriguingly, *comm* expression is restricted to the dorsal mesoderm in WT embryos (arrows, Figure 4M), in line with its ectodermal expression domain, which also present in stripes from st 9 (Figure 4G). It is surprising that the most prominent transcription program associated with mesoderm cells at st 8, when EMT ensues, contains expression of segment polarity genes with demonstrated roles in the ectoderm.

Embryos with mutations in *mirr* or *comm* exhibit abnormalities in pericardial cell distribution, a phenotype typically associated with mesoderm patterning defects (Figure S4D). In addition, *opa* mutants exhibit disorganized and reduced expression of these genes; for example, *comm*, in the mesoderm and ectoderm (Figure 4G)[65] and *bap* as well as *btn* in the mesoderm (Figures 4G and 4J). It is possible that the observed mesoderm patterning defects in *opa* mutant embryos (Figures 3F, 3G, and S3B–S3E)[50,66] relate to misregulation of segment polarity genes. Therefore, these data point to a regulatory program supported by Opa's action in the trunk that actively establishes segmented gene expression patterns in mesoderm cells following EMT, possibly to facilitate proper cell migration.

We found that two mesoderm-specific segment polarity genes, *bagpipe* (*bap*) and *buttonless* (*btn*), are expressed exclusively in the mesoderm in segmental patterns. *bap*, known to support a visceral muscle fate later during differentiation,[65,67] is restricted to the dorsal anterior parasegments (Figures 4G, S4B, and S4C) prior to the completion of mesoderm migration.[66,67] Transcripts of *btn* are first detectable in the invaginated tube at st 7 (Figures 4I and S4E) with some low-level expression in the procephalic and ventral ectoderm. A segmented pattern for *btn*, arranged periodically at each parasegment boundary anterior to *wg* stripes, does not become apparent until st 9 (Figures 4I and S4C). *btn* has been shown to be required for the differentiation of dorsal median (DM) cells, which are 20 cells located along the midline that contribute to axon guidance of transverse and median nerves.[68] These results suggest that the later segmentally expressed genes likely function to promote differentiation of specific mesoderm lineages.

## Differential gene expression analysis reveals ecdysone signaling acts in the mesoderm as EMT ensues

Trunk mesoderm cells of c1 are at a critical transitional stage: these cells are already specified but have not yet received inductive signals for differentiation. Division and migration are the most prominent cellular changes associated with these cells. In contrast, c3 likely consists of cells of st 9 mesoderm as it is marked by genes implicated in muscle differentiation (e.g., *Him*, *meso18E*, *jeb*) and Notch signaling (e.g., *E(spl) complex*)[69,70] (Figures 5A and S5H–S5J; Table S1).

We found that genes relating to ecdysone signaling are enriched in cells comprising c1 (i.e., st 6–8) but downregulated in c3 (Figure S5C; Table S1). c1 markers include Halloween genes *phantom* (*phm*, *Cyp306a1*),[71,72] *dissembled* (*dib*, *Cyp302a1)*,[73] *shroud* (*sro*),[74] and *Niemann-Pick type C-2a* (*Npc2a)*, a homolog of human NPC2 linked to Niemann-Pick C type 2 neurodegenerative disease[75] (Figures 5B–5D and S5A–S5C), all of which encode enzymes in the biosynthesis pathway of ecdysone. The expression levels of these genes correlate with titers of ecdysone, which in its active form 20-hydroxyecdysone (20E) affects transcription by binding to nuclear receptor EcR.[76,77] Ecdysteroidogenic transcription factors are known targets of ecdysone signaling and often act in a positive feedback loop to ramp up the production of ecdysone through upregulation of *phm* and *dib*.[78] Indeed, we found that expression of a dominant negative EcR receptor construct (*EcR.B2.DN*) blocks *phm* transcription by st 8 (Figure S5D).

Therefore, a role for ecdysone signaling in supporting gastrulation during early embryogenesis was investigated. Previous studies have shown that it supports border cell migration and follicle cell differentiation during oogenesis as well as dorsal closure, head involution, and tracheal and midgut morphogenesis during mid- to late-embryonic development.[79–84] Although a *lacZ* reporter fused to the hormone binding domain (E region) of EcR was shown to respond to exogenous 20E and drives expression in the visceral mesoderm at the end of gastrulation (i.e., stage 10),[79,85,86] a functional role for ecdysone at this stage has not been determined. The requirement of EcR for oogenesis prohibits use of a null allele to analyze its zygotic function, as inactivation of EcR in the ovary results in a multitude of defects in oogenesis, including some embryos produced with aberrant egg chamber polarity.[82] However, expression of the dominant negative allele (*EcR.B2.DN*) using different genetic cross strategies can work to downregulate signaling in a temporal manner.[87] Through this approach, we found that maternal loading of *EcR.B2.DN* (using the *MTD-GAL4* to support expression in the ovary of first, F1, generation) disrupts AP patterning in the second (F2) generation (Figures 5E and S5E–S5G), in line with EcR's appreciated role during oogenesis.[81,82] We then observed that early zygotic expression of *EcR.B2.DN* in F1 embryos results in abnormal mesoderm EMT at st 8 (Figure 5F) without affecting AP patterning (Figure 5E). These results suggest ecdysone signaling is important for embryonic development, specifically, to support cell movement and division at EMT during gastrulation (see section "discussion"). We further hypothesize that dysregulated ecdysone signaling might partly account for the EMT and patterning defects in *opa* mutants (Figures 3F and S3C). This is supported by the finding that *phm* expression is lost in trunk mesoderm in $opa^{8}$ mutants (Figure 5G), which would be expected to result in reduced ecdysone signaling. Together, our data suggest that *EcR* acts downstream of Opa to, in part, support the transcriptional changes necessary for EMT of trunk mesoderm cells.

## Lineage commitment dynamics for cells arising from the posterior trunk mesoderm

We next turned from the trunk mesoderm to investigate the caudal mesoderm, a less-well-characterized cell population. In the blastoderm embryo, only ~40 cells located posterior to the trunk mesoderm comprise the caudal mesoderm, which encodes a basic helix-loop-helix (bHLH) transcription factor, *HLH54F*.[26] *HLH54F* expression is associated with cluster 10 in the UMAP (c10; Figures 6A and S6A, *HLH54F*; Table S1) and by *in situ* hybridization in st 5/6 embryos, and it is expressed in a region posterior to the *htl* expression domain in the trunk mesoderm (Figures 6D and S6B, *HLH54F*). Consistent with their generally non-overlapping patterns in the embryo, *htl*- and *HLH54F*-expressing cells predominantly occupy distinct spaces in the UMAP of c10 (Figures 6C and 6D). This result also suggests that c10 contains at least two progenitor populations: one enriched for *HLH54F* not expressing *htl*, while the other, inversely, is enriched for *htl* not expressing *HLH54F*. To provide insight into the identities of cells in this cluster and gene expression changes associated with the caudal mesoderm, we conducted a Monocle trajectory analysis of c10 that uncovered five cellular states (Figures 6B and S6C; Table S7).

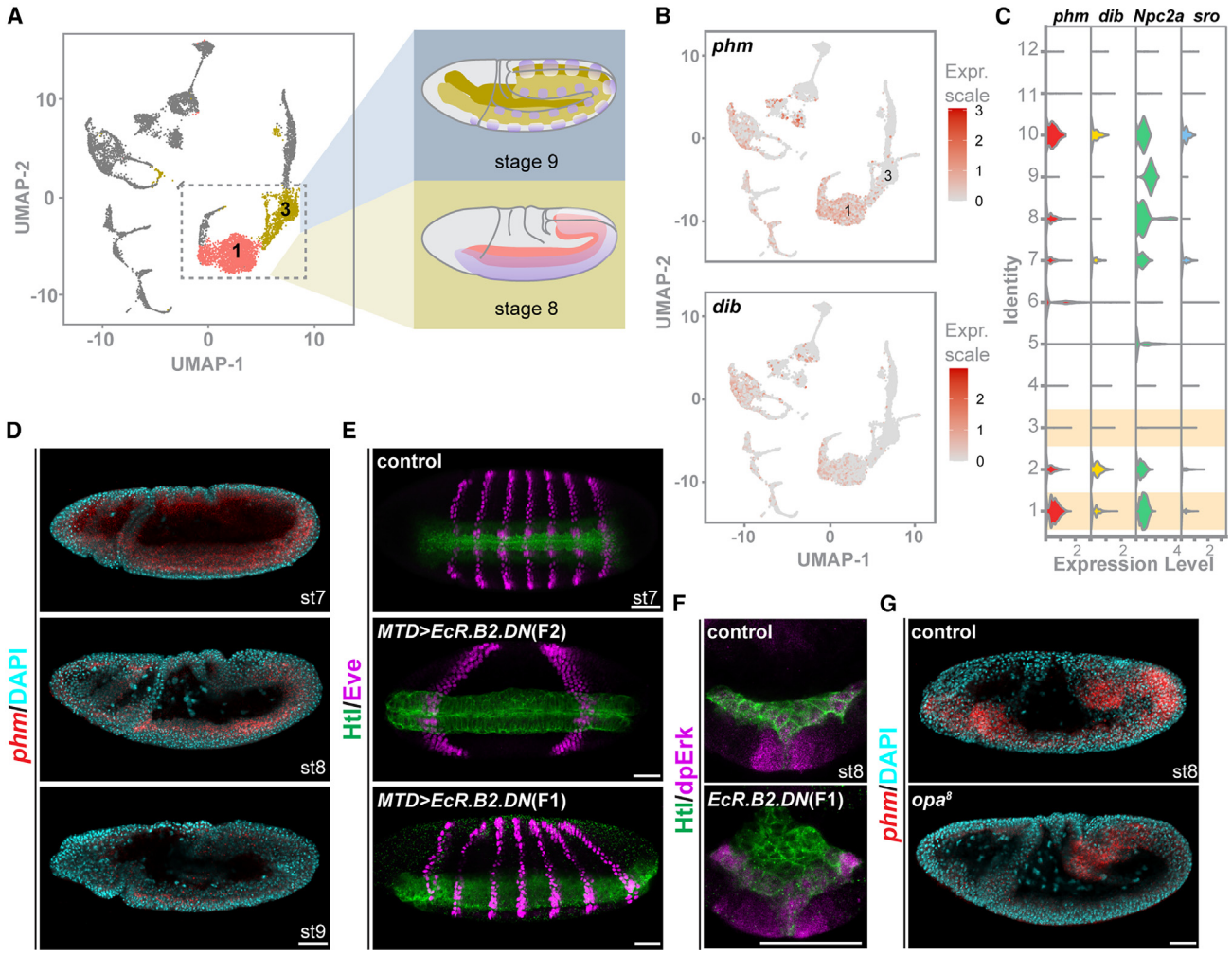



**Figure 5. Genes involved in ecdysteroid hormone biosynthesis are upregulated in the mesoderm during EMT**

(A) Cells from embryos of st6–8 and st9 comprise c1 and c3, respectively. Therefore, comparison of gene expression trends between these two clusters reveals transcription changes during and after mesoderm epithelial-to-mesenchymal transition (EMT).

(B and C) Expression of selected genes involved in ecdysteroid biosynthesis pathways as shown in UMAP (B) or violin (C) plots. Red intensity indicates the relative expression levels among all clusters. Mesoderm cell clusters are highlighted.

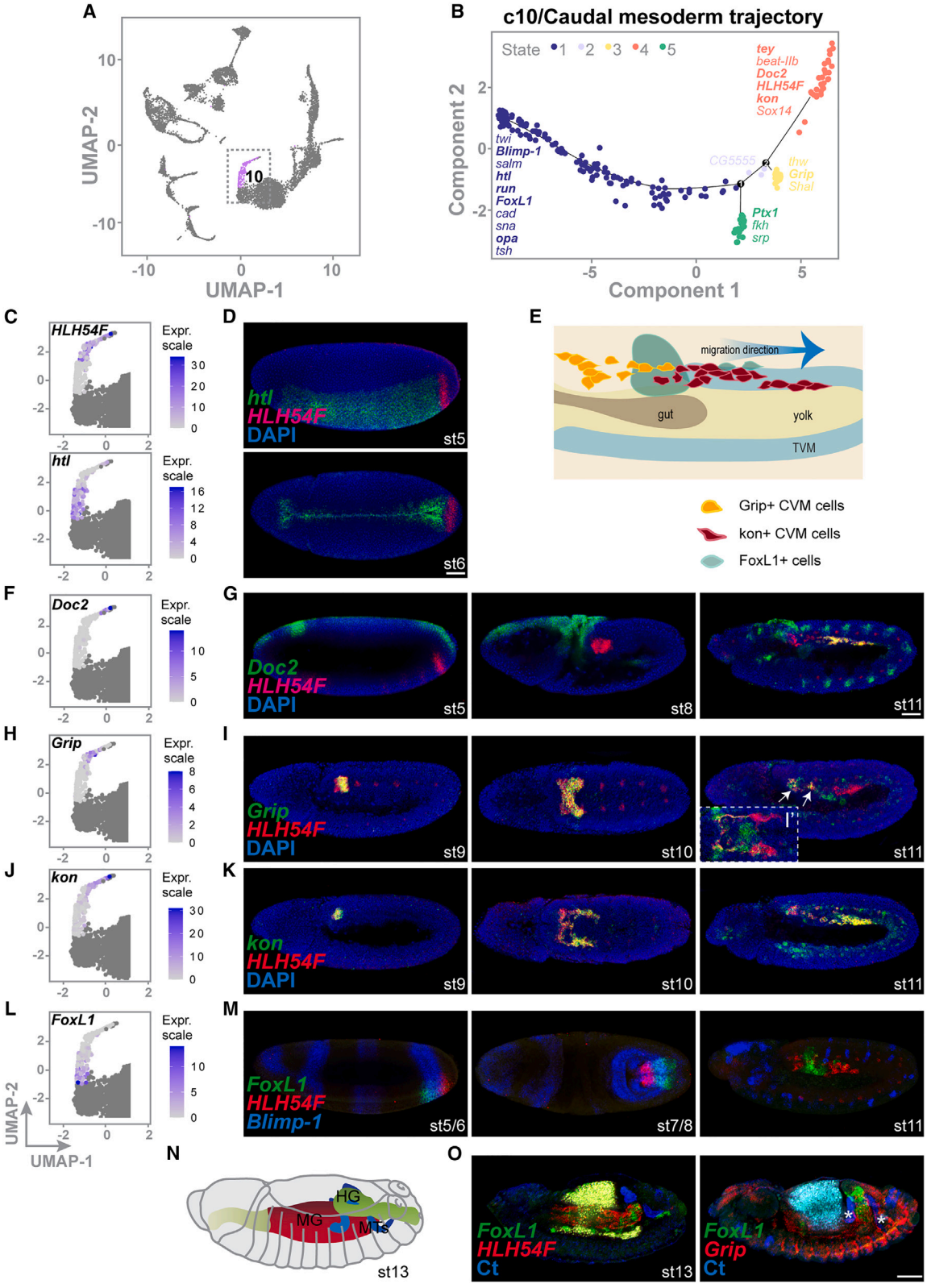
(D) *In situ* hybridization with *phm* riboprobe confirms its expression in the mesoderm is no longer detected post EMT (st9).

(E and F) Disrupting ecdysone signaling results in patterning defects in the second offspring generation F2 (E) or a mesoderm-spreading phenotype in the first offspring generation F1 (F).

(G) *phm* is downregulated within the mesoderm in *opa*$^{8}$ mutants, suggesting a role for *opa* in regulating ecdysteroid signaling in this tissue. Scale bar, 50 μm. See also Figure S5.

Among these five cellular states associated with c10, two cell states (c10.4 and c10.5) could be assigned to known cell types based on marker gene expression. c10.4 likely represents caudal visceral mesoderm (CVM) cells, a migratory population that expresses *HLH54F*. Cells in this subcluster also express other well-characterized CVM markers including *kon-tiki* (*kon*), *Dorsocross2* (*Doc2*), and *teyrha-meyrha* (*tey*) (Figures 6B, 6F, 6G, 6J, 6K, and S6F, *tey*). The presumptive CVM cells are specified by the end of nc14 (st 5) as a progenitor population.[26] Following invagination, these cells become internalized while also rapidly moving posteriorly during GBE. By st 9, they are situated in a pocket between the ectoderm and hindgut/posterior midgut, where they begin migrating out on top of the trunk visceral muscle (TVM) cells to eventually give rise to the longitudinal muscles of the midgut.[26,88,89] Intriguingly, *Doc2* expression is located at the very tip of the UMAP for c10 (Figure 6F). By *in situ* hybridization, *Doc2* is not expressed within the CVM until st 10, as demonstrated by colocalization with *HLH54F* (Figure 6G), suggesting that cells located at the tip of the c10 UMAP are older. On the other hand, cells of c10.5 are enriched with transcription factors expressed in the posterior endoderm and hindgut, e.g., *Ptx1* (Figures S6C and S6D, *Ptx1*), *fkh*, *Kr*, and *srp*, suggesting that c10.5 represents a cell type born at the mesoderm/endoderm border. MTs arise from exactly

CellPress
OPEN ACCESS




*(legend on next page)*

this location in ring-shaped ectodermal primordia shared with the hindgut.[28,90] MT cells can be recognized from st 10 by the expression of Cut (Ct), a homeobox-containing transcription factor functioning to control the evagination of MT primordia at the hindgut-midgut junction.[91,92] Several markers for c10.5 are implicated in the development of MTs,[93–95] making it likely that c10.5 contains precursor cells for MTs.

Cells in states 2 and 3 (c10.2 and c10.3; Figure 6B; Table S7) are most closely related to CVM cells according to the trajectory analysis. However, none of the top markers for c10.2 and c10.3 have previously characterized functions in the caudal mesoderm. Because c10.2 only contains one gene, we focused our efforts on c10.3 by further investigating the expression of its associated markers. For instance, *Glutamate receptor-interacting protein* (*Grip*) (Figures 6H, S6A, S6C, and S6D, *Grip*), encoding a scaffold protein known to regulate muscle attachment,[96,97] was found initially co-expressed with *HLH54F* at st 9 (Figure 6I). However, by st 11, the expression patterns of *Grip* and *HLH54F* have diverged: the majority of the *HLH54F*-positive cells at the front of the migratory cohort have turned off *Grip* and instead initiate expression of *Doc2* (Figures 6I and 6I′, and compare with 6G). In contrast to *Grip*, c10.4 marker *kon* also co-localizes with *HLH54F* at st 8–9 (Figures 6J, 6K, and S6D) but transitions to an opposite pattern to *Grip* at st 11: stronger in the front but diminishing at the back. *Doc2*-, *kon*-, and *Grip*-positive cells become migratory, with *Doc2*- and *kon*-coexpressing cells located at an anterior position and *Grip*-positive cells lagging behind (Figure 6J). Therefore, c10.3 likely represents a younger CVM precursor population (st 8–9) relative to c10.4, which constitutes the older CVM cell population from st 10. Upon the commitment of the *HLH54F*-expressing cells toward a muscle cell fate at st 11, the continuing expression of *Grip* defines cells that lag behind, marking a new migratory cell type that is distinct from the *Doc2*- and *kon*-positive CVM cells (Figures 6E, 6G, 6I, and 6K) (see section "discussion"). *Grip*-positive cells ultimately are incorporated into both the longitudinal muscle of midgut and the hindgut (Figures 6N and 6O).

Last, cell state 1 (c10.1) is associated with genes expressed as early as st 5, containing cells marked by *sna*, *opa*, *Blimp-1*, *salm*, as well as *htl* likely representing progenitor cell types of st 6/7 (Figures 6B and S6A–S6D). Among markers for c10.1 is a forkhead box (Fox) transcription factor, *Forkhead box L1* (*FoxL1*) (Figure 6L). Fox proteins are involved in fate determination of different mesoderm cell populations during organogenesis.[98–100] *FoxL1* has been shown to contribute to muscle development during late embryogenesis.[101] We found that *FoxL1* transcripts can be first detected in embryos by *in situ* hybridization at st 5, in a caudal region anterior to the *HLH54F*-positive CVM precursors (left panel of Figure 6M). *FoxL1* expression remains distinct from *HLH54F* throughout gastrulation (Figures S6B and S6E), while it partially overlaps with *Blimp-1* in the caudal mesoderm until st 8 (middle panels of Figures 6M and S6D). Following gastrulation, *FoxL1* is expressed in a population that co-migrates with CVM precursors at st 11 (right panels of Figure 6M). These cells ultimately locate to the anterior and posterior midgut as well as hindgut in a circular pattern around the basal surface of the endoderm, in contrast to the midgut longitudinal muscle fate of CVM cells (Figures 6N and 6O). Therefore, our analysis demonstrates that these caudal mesoderm cells in c10.1, marked by *FoxL1*, *opa*, and *Blimp-1*, are a muscle progenitor cell type that is different from the progenitors that give rise to CVM cells. These *FoxL1*-expressing progenitor cells arise early at the cellular blastoderm stage, stay localized to a region in proximity to CVM precursors, then migrate along with them (assuming a position at the back of the migrating collective) until becoming incorporated into the developing gut.

With the trajectory analysis identifying cellular states associated with distinct transcription signatures (Figure 6B) and our subsequent investigation of marker genes by *in situ* hybridization (Figures 6D, 6G, 6I, 6K, 6M, and 6O), we show that single-cell profiling can help tease apart the temporal and spatial differences in gene expression within closely related cell populations, providing insights into cell fate decisions toward lineage differentiation.

## Transcription programs specify, regionalize, then segment the mesoderm cells along the AP axis

This single-cell study highlighted transcription programs that control morphogenesis from blastula to gastrula stages. Our results favor the view that the blastoderm (nc14) embryo is first carved into three domains: trunk vs. anterior and posterior poles (Figure 7A). Additionally, centered on the patterning of mesoderm, it provides insights into both temporal and spatial gene regulation that contributes to the specification (Figure 7C) and subdivision of the mesoderm cell population (diagram,

**Figure 6. The single-cell transcriptomic analysis provides both temporal and spatial information on the lineage commitment of the posterior trunk mesoderm**

(A and B) Trajectory analysis reveals five cellular states along with marker genes (B) that represent both temporal and spatial changes in transcription in c10 (boxed area, A). States 1–5 refer to c10.1–c10.5.

(C, D, F, and G) Cropped UMAP of *HLH54F* (C), *htl* (C), and *Doc2* (F) are shown with expression scales. *In situ* hybridization using hybridization chain reaction (HCR) probes shows the *HLH54F*-marked CVM primordium (red) is specified at the blastoderm stage, distinct from the trunk mesoderm that expresses *htl* (green) (D), and migrating CVM cells express *Doc2* later in development (st10/11) (G).

(E and H–K) Diagram illustrating the position of *HLH54F*-expressing CVM cells located at the front (red) or back (yellow) of the migrating collective at st10–11, and depicting the position of *FoxL1*-labeled muscle precursors (green) that co-migrate with CVM (E). Expression of state 3 marker *Grip* and state 4 marker *kon* shown by UMAP (H and J) and *in situ* hybridization (I, I′, and K) colocalizing with *HLH54F*. Grip+ cells occupy the back of CVM migrating cohort (arrows, I; magnified view of another embryo showing dorsal view in inset, I′).

(L and M) State 1 cells are marked by *FoxL1* and *Blimp-1*, which label a distinct region of the c10 UMAP plot compared to *HLH54F* (compare L with C). Expression patterns of transcription repressors *Blimp-1* (blue) and *FoxL1* (green) partially overlap, but both are distinct from *HLH54F* (red) at st5–8 (M).

(N and O) Fate of the caudal mesoderm cells. Diagram illustrates that caudal mesoderm cells contribute to midgut (MG), hindgut (HG), and malpighian tubules (MTs) at st 13 (N). *FoxL1* (green), *HLH54F*, and *Grip* (red) expression detected by *in situ* hybridization with anti-Cut antibody labeling MTs (blue, asterisks in O). Genes and stages assayed are indicated. DAPI (blue, D, F, I, and K) for counterstaining. Scale bars, 50 μm. See also Figure S6 and Table S7.

CellPress
OPEN ACCESS

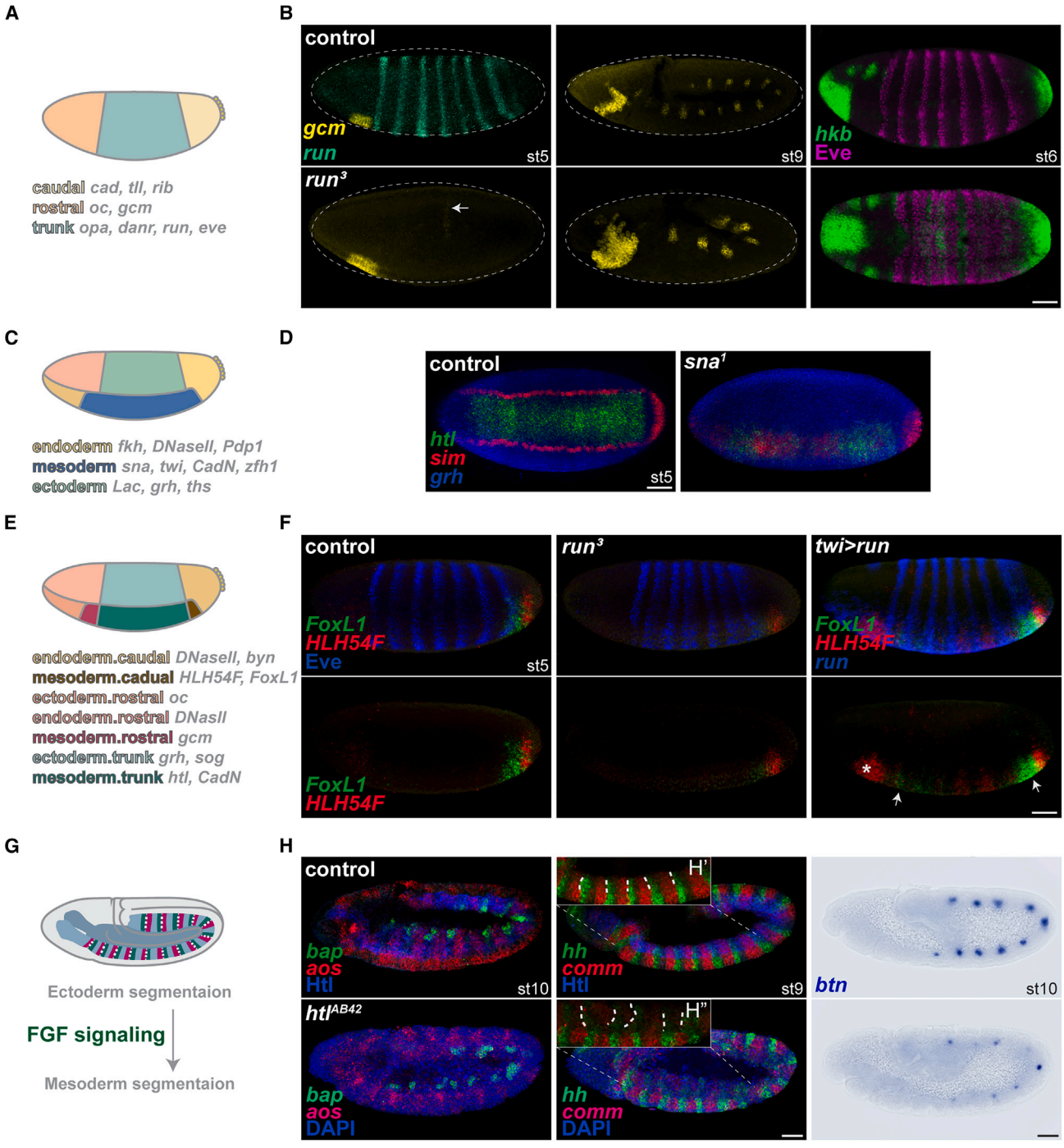



**Figure 7. The temporally ordered transcription programs establish the spatial domains and tissue identities during early embryonic development**

(A and B) The early blastoderm embryo is regionalized by transcription factors that specify the head, trunk, and tail, colored in beige, green, and yellow in the embryo diagram (A). *gcm*, *hkb*, and Eve are affected in *run*$^{3}$ mutants (B).

(C and D) Three germ layers are specified during nc14 as mesodermal (blue) and endodermal (both anterior and posterior, yellow) transcription programs are clearly distinct from the ectoderm tissue (orange and green) (C). Defects in *htl*, *grh*, and *sim* expression are observed in *sna*$^{1}$ mutants (D).

(E and F) The embryo is further regionalized by the end of nc14 as the mesoderm is divided into head/rostral (fuchsia), trunk (teal), and posterior/caudal (brown) mesoderm domains (E). Loss- and gain-of-function perturbations of *run* have opposite effects on the posterior mesoderm gene expression (F).

(G and H) Mesoderm segmentation follows the ectoderm, which ensures that mesoderm cells migrate and differentiate normally (G). Htl antibody staining and DAPI counterstaining (both blue) were used to visualize mesoderm cells. The white dashed lines in magnified views indicate mesoderm cells located at para-segment boundaries; compare H′ and H′′. Genes detected and stages are labeled. Scale bars, 50 μm. See also Figure S7.

Figure 7E), which takes place prior to gastrulation. This process is followed by further specialization and lineage commitment of cells located in the posterior trunk region and segmentation of the trunk mesoderm with continuing expression of pair-rule and segment polarity genes from nc14 onwards (Figure 7G). To provide additional genetic evidence for this framework, we assayed mutants of patterning genes to examine their role in supporting such a stepwise process.

To test how disrupting the trunk gene expression program affects transcription in the poles and later on the posterior trunk cell fate, we utilized a loss-of-function allele of *runt* (*run*), a pair-rule transcription factor known to act early at the blastula stage to regulate gene expression.[102] *run* is initially expressed broadly in the trunk at early nc14 then quickly becomes restricted to five to seven transverse stripes from mid- to late-nc14 (Figure 2G).[102] When *run* expression is broad, it is thought to inhibit anterior cell fates.[102,103] In $run^3$ mutants, we observed an expansion of *gcm* expression in the head mesoderm at nc14/st 5, becoming more apparent at st 9 (Figure 7B). The stripes associated with pair-rule gene *eve* appear compressed into a smaller area within the trunk in *run* mutants compared with WT embryos (Figure 7B, right). Furthermore, *hkb*, normally restricted to the poles prior to stage 8, is ectopically expressed in the trunk at st 6 in *run* mutants (Figure 7B, right). *opa*, on the other hand, appears to be dispensable for normal gene expression in the trunk (Figures S7B and S7C). While *run* is acting, loss of *opa* function does not lead to an obvious change in the *gcm* domain (Figure S7A).[20] However, *opa* is required later at gastrulation. For instance, not only does Opa promote gene expression within the trunk,[20,22,24] but it also represses terminal gene expression from the trunk (Figure S7D). These results suggest that *run*, functioning in a similar way but earlier than *opa*, is a critical component of the trunk gene regulatory network that restricts the expression of terminal genes to the anterior and posterior poles.

We used $sna^1$ mutants to test how specification of one germ layer influences the other germ layers. *sna* is required for the specification of mesoderm as well as mesectoderm in two rows of cells abutting it. Loss of *sna* results in the loss of mesoderm cell fate as *htl* expression is not maintained (Figure 7D). In addition, the gene *single-minded* (*sim*) is expressed only at low levels within patches in ventral regions of the embryo, indicating a failure to define mesectoderm. Last, *grh* expression (normally restricted to the ectoderm) expands into the ventral domain overlapping with *htl* in the $sna^1$ mutant (Figure 7D). These observations indicate that disrupting the transcription program active in the mesoderm alters gene expression in other tissue types and that the emergence of mesoderm cells at mid-nc14, driven by *sna* and *twi*, is a crucial programmatic step toward gastrulation.

*sna* activity is followed by division of the mesoderm into head, trunk, and posterior domains. To assay how posterior mesoderm cells respond to dysregulation of the trunk program, we again turned to *run*, which is a marker of the c10.1 cell state that likely represents progenitor cells that give rise to visceral muscles (Figures S7E and S7G). *run* pair-rule expression pattern at st 5/6 is dynamic (Figure S7F). At st 5, the *run* seventh stripe at the posterior overlaps with both *FoxL1* and *HLH54F* domains. However, just a bit later at st 6 when the *run* pattern encompasses 14 stripes, one stripe arises at the posterior that is in line with the *HLH54F* positive domain (Figure S7F). This suggests that *run* may act to support both these distinct cell populations. In $run^3$ mutants, *FoxL1* levels are greatly reduced (Figure 7F), whereas overexpressing *run* in the mesoderm using a *twi-GAL4* driver causes a concomitant increase in levels of *FoxL1* and *HLH54F* in the trunk as well as the anterior (arrows and asterisk, Figure 7F). *opa* is also enriched in c10.1 cells that express *FoxL1* (Figures 6B and S6C), but neither *FoxL1* or *HLH54F* levels are affected by loss of *opa*, consistent with a recent study in which no posterior patterning defects were detected by computer simulation.[21] Together, these results suggest that mesoderm cell fates are tightly controlled within the trunk and that the size of a lineage precursor population is determined by cooperative regulations involving multiple transcription factors (see section "discussion").

Our dataset also points to a segment polarity program in the mesoderm potentially contributing to their proper migration as a collective (Figures 4K–4M). The fibroblast growth factor (FGF) signaling pathway is essential to mesoderm development by coordinating cell division, cell-shape change, and movement during gastrulation as well as by promoting cardiac and somatic muscle differentiation later during organogenesis.[49,104] Using a loss-of-function $htl^{AB42}$ allele,[105] we tested whether FGF signaling is required for the maintenance of segment polarity within the mesoderm by examining the localized expression of markers identified in the c1 trajectory analysis (Figures 4A and 7H). Indeed, *bap* is a known FGF target and, as expected, expressed in irregular patches. However, *aos*, *hh*, and *comm* also all appear to be expressed at lower levels in $htl^{AB42}$ mutants (Figure 7H). Intriguingly, *hh* and *comm* stripes in the mesoderm are uneven in the mutant within the mesoderm (Figure 7H; compare Figure 7H″ with 7H′), suggesting that mesoderm cells fail to move in a coordinate manner with the ectoderm during GBE. Signaling pathways including FGF serve as bridges and messengers from the ectoderm to the mesoderm to ensure that the two germ layers are patterned concomitantly.[66,106] We hypothesize that the continual expression of segment polarity genes in the mesoderm is crucial for supporting proper collective behavior during mesoderm cell migration.

## DISCUSSION

By investigating gene expression trends using conventional genetic tools assisted by single-cell transcriptomic information, this study contributes to our understanding of how gene expression programs are orchestrated in a stepwise manner to lay down the basic body plan and its modular units to coordinate complex morphogenetic processes.

One of the advantages of our approach is that it illuminates several sequentially acting transcription programs that are associated with gene expression changes during early *Drosophila* embryogenesis, encompassing the transition from blastula to gastrula stages. In contrast to our initial assumption that we would see cells differentiated by early patterning programs relating to germ layer formation, we were surprised to find that AP position, specifically the trunk versus terminal programs, was more influential when grouping blastoderm cells according

to their transcriptional states. This result is consistent whether or not cells are fixed (i.e., f.c6 separates from f.c3; Figure S2D). Furthermore, our data that cells in the posterior pole, which express *cad* and *tll*, are preferentially associated with a "younger" cell state in the blastoderm (e.g., Figures 2B, 2E, 2F, S2A, S2C, and S2E) may relate to delay of maternal transcript degradation from the poles and/or to increased complexity of zygotic transcriptome associated with cells in the trunk versus termini. While it is also possible that products of germ plasm mRNAs enriched at the posterior pole act to repress transcription and cause a delay in maternal-to-zygotic-transition (MZT),[107,108] further investigation is required to demonstrate their role in regulating gene expression outside germ cells and examine how they affect AP patterning in the blastoderm embryo. Recent studies have suggested that *Drosophila* embryos, despite being long germband embryos, have retained properties of an ancient short-germband program in which posterior cells act with multipotent potential to generate segments in a delayed manner relative to segment specification in the trunk.[21,25] Our data support this view that patterning of termini is delayed and suggest that single-cell sequencing experiments can provide deep insights into how transcription timing is coordinated within different domains of the blastoderm, or, as manifested by transcription progression, how developmental speeds are differentially regulated in cells at different locations.

While this dataset serves as a resource for understanding the stepwise progression of all three germ layers, we specifically focused on the patterning and specialization of mesoderm lineages in our in-depth analyses, finding that allocation/divergence of the terminal lineages (rostral and caudal mesoderm) from early trunk mesoderm is accomplished prior to the morphogenetic movements associated with gastrulation, namely invagination, EMT, and GBE.

In the trunk, Opa continues (from the blastula stage) to act as a timing factor at the initial phase of EMT during gastrulation. Previously shown to drive segmental patterning in the ectoderm, we find that Opa is broadly expressed in the trunk mesoderm at EMT, preceding and likely required for the transcription of a number of segmentally expressed genes enriched within the same cluster. We show that several of these segmentation genes are implicated in patterning and ultimately also required for differentiation of mesoderm cell lineages through analysis of mutant phenotypes (e.g., *comm* and *mirr*; see Figures 4N and S4D). Other markers that were not investigated include *Toll* (*Tl*) and *derailed* (*drl*) in cell states 2 (c1 trajectory analysis) and *anterior open* (*aop*) and *argos* (*aos*) in state 4 (Figure 4A; Table S5). It is unlikely that Opa acts alone in regulating transcription timing, and potential roles for additional transcription factors were also highlighted by our results (e.g., Figure 3). Therefore, future experiments would focus on interrogating the mesoderm-specific requirement for *opa* during gastrulation and characterizing the contributions of these other factors to mesoderm patterning.

The unexpected segmentation gene expression we observed downstream of Opa is intriguing, as it signifies that part of the Opa-dependent transcription program also relates to promoting planar polarization of the migrating mesoderm cells. With the integrated dataset, an *opa*, *sna*-enriched cell cluster (i.c11) and two separate clusters corresponding to the anterior and posterior parasegment identities (i.c4 and i.c5, respectively) emerge from those cells undergoing EMT (Figures S4F and S4G). We suggest that the Toll genes, as well as other segment polarity genes, play a role similar to their function in patterning the ectoderm during GBE,[109] regulating mesoderm morphogenesis relating either to establishment of cell polarity (planar, apicobasal, or front-back) and/or in the regulation of spatially distinct actomyosin activities to keep the relative position of mesoderm cells in check during migration. EMT, the process through which mesoderm cells acquire their migratory ability, is tightly coupled with GBE[110] and these two processes may be regulated by the same segmentation gene network controlling AP patterning. Notably, mutations in segmentation genes affecting ectodermal planar cell polarity also lead to mesoderm spreading defects.[65,66] Furthermore, mesoderm cells likely lose polarity information during division as EMT occurs (Figure 4M) and have to regain planar polarity (i.e., segmental polarity) to support coordinate posterior movement following GBE of the ectoderm. Such a process allows mesoderm cells to position themselves in association with ectodermal lineages in order to form the proper ratio of different muscle types.

Surprisingly, a gene expression program associated with ecdysone biosynthesis also plays a role as trunk mesoderm cells prepare for/undergo EMT (Figures 5D and 5E). This finding is interesting in that it suggests that a hormonal/metabolic cue through ecdysone influences EMT at gastrulation in *Drosophila* embryos. However, it remains unclear what tissues provide the source for embryonic ecdysone (E) and when ecdysone signaling is first required during embryogenesis. An even earlier role supported by maternal products of the pathway could indirectly affect the movements of mesoderm cells. Future experiments are needed to address those questions. For neural crest cell delamination, recent studies have identified that metabolic triggers are important for regulating gene expression.[111] It is possible that the *Drosophila* embryo similarly does not initiate the GBE/EMT program unless the proper metabolic environment is present. Since cell movement requires a significant expenditure of energy, this link between metabolism/hormones and cell movements is understandable.

This study also provides insight into how multipotent mesodermal progenitor cells are carved out early in embryogenesis through the combinatorial input of transcription factors, including timing factors. Our developmental trajectory analysis identified transcription factors, expressed in spatially distinct domains, that allocate the posterior ventral regions of the blastoderm to ultimately give rise to neighboring migrating cell types. We show that *FoxL1*, along with several other transcription regulators (e.g., *Blimp-1*, *salm*; see c10.1; Figure 6B), marks a progenitor cell type located anterior to CVM in the caudal mesoderm. Mesoderm cells in the caudal region are also alternatively named for the cell types they ultimately contribute to. For example, CVM corresponds to longitudinal visceral muscle (LVM), while the muscle cells in the hindgut are referred to as hindgut visceral muscles (HVMs).[112,113] The FoxL1+ cells therefore represent a progenitor population established at stage 5 that likely contributes to the HVM as well as TVM. The transcription program responsible for the specification and later migration of these caudal mesoderm cell types are distinct from the trunk and relate to the terminal program (i.e., *torso*, *tll*, as well as timing factor *cad*). However, genes functioning in the trunk can still have an impact. For instance,

*run*, previously shown to repress *ocelliless* in the anterior,[102,103] also functions in the posterior blastoderm as part of the transcription regulatory network to activate expression of *FoxL1* and possibly *HLH54F* in the progenitor populations.

While it is appreciated that CVM cells appear to also contribute to the MTs, the *Drosophila* kidney equivalent,[114] direct evidence arguing for a secondary developmental outcome for *HLH54F*-positive mesoderm cells had not been demonstrated. Analyzing gene expression at single-cell resolution enabled us to discover subtle differences in transcriptomic signature even for patterns composed of a small number of cells. Here, we provide molecular support for the finding that CVM indeed contributes to MTs in addition to midgut longitudinal muscles. A possible hindgut fate is linked to the expression of *Grip*. Specifically, we found that there is a deterministic transcription program involving maintenance of *Grip* expression in a subset of CVM cells that accounts for how a subset of cells located at the back of the CVM migrating collective become MTs vs. other tissues and those located at the front become longitudinal muscles. Without the insight of Monocle trajectory analysis to further subdivide cells within clusters, we would not be aware that the CVM migrating collective has a different gene expression program for cells located at the front vs. back of the migrating collective.

In summary, by highlighting regulatory steps in transcriptional progression at single-cell resolution, this study contributes to our understanding of how animals control and coordinate morphogenetic processes and paves the way for future mechanistic analysis of gene regulatory networks and signaling pathways that account for the specification of multipotent progenitor cell types as well as collective cell migration. The single-cell transcriptomic map generated in this study also serves as a resource for future research focusing on gastrulation and, in particular, mesoderm development.

### Limitation of the study

In this study, we chose to focus analysis on scRNA-seq data acquired from live cells isolated by embryo homogenization. As we wanted to decrease processing time and because the isolation procedure leads to a great deal of material loss, we could only infer the developmental stages of the embryos within the ~90-min interval based on the gene expression information (*in situ* hybridization data) and comparisons with temporal information provided by previous studies. In addition, the following assumptions were made for data analysis: first, that cells do not age, or rather transcription does not progress, once cells are placed on ice; and second, that the isolation procedure, including physical (i.e., Douncer) and chemical (i.e., Accumax) methods used, as well as incubation with the magnetic beads during dead cell removal, do not introduce transcriptional changes. Furthermore, for cell types such as the caudal mesoderm, cell numbers are low, and such analyses in the future could benefit from enriching rare cell populations.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Fly stocks and husbandry
- METHOD DETAILS
  - Embryos collection, cell isolation and library construction
  - Generation of additional datasets
  - Cluster annotation and validation
  - Fixation, *in situ* hybridization, and immunostaining
  - Hybridization chain reaction (HCR)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Read processing, filtering, doublets removal and visualization
  - Data integration and clustering
  - Gene Ontology enrichment analysis
  - Subclustering, differential gene expression, and trajectory analysis

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.113289.

## ACKNOWLEDGMENTS

We are grateful to Deborah Andrew, Lauren Anllo, Peter Gergen, and Michael O'Connor for sharing fly stocks and antibodies; Siyu Chen and Jeff Park of Caltech SPEC, Michael O'Connor, Frank Macabenta, and Vince Stepanik for helpful discussions and technical support; and Isaryhia Rodriguez and Life Science Editors for comments on the manuscript. This study was supported by funding from National Institutes of Health grants R35GM118146 and R01HD100189 to A.S.

## AUTHOR CONTRIBUTIONS

A.S. and J.S. conceived the project. J.S. planned the experimental approach. A.S. directed the project. J.S. performed all experiments and F.G. performed the computational work. J.S., C.Z. and F.G. analyzed the data with input from A.S. The manuscript was written by J.S. and A.S. with input from F.G. and C.Z.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: March 28, 2023
Revised: August 29, 2023
Accepted: September 29, 2023

## REFERENCES

1. Calderon, D., Blecher-Gonen, R., Huang, X., Secchia, S., Kentro, J., Daza, R.M., Martin, B., Dulja, A., Schaub, C., Trapnell, C., et al. (2022). The continuum of embryonic development at single-cell resolution. Science *377*, eabn5800.

2. Lim, J., and Thiery, J.P. (2012). Epithelial-mesenchymal transitions: insights from development. Development *139*, 3471–3486.

3. Nieto, M.A., Huang, R.Y.-J., Jackson, R.A., and Thiery, J.P. (2016). EMT: 2016. Cell *166*, 21–45.

4. Stathopoulos, A., and Levine, M. (2004). Whole-genome analysis of Drosophila gastrulation. Curr. Opin. Genet. Dev. *14*, 477–484.

5. Leptin, M. (1995). Drosophila gastrulation: from pattern formation to morphogenesis. Annu. Rev. Cell Dev. Biol. *11*, 189–212.

6. Wolpert, L. (1992). Gastrulation and the evolution of development. Dev. Suppl. *116*, 7–13.

7. Farrell, J.A., and O'Farrell, P.H. (2014). From egg to gastrula: how the cell cycle is remodeled during the Drosophila mid-blastula transition. Annu. Rev. Genet. *48*, 269–294.

8. Thomsen, S., Anders, S., Janga, S.C., Huber, W., and Alonso, C.R. (2010). Genome-wide analysis of mRNA decay patterns during early Drosophiladevelopment. Genome Biol. *11*, R93.

9. Sandler, J.E., and Stathopoulos, A. (2016). Quantitative Single-Embryo Profile of *Drosophila* Genome Activation and the Dorsal–Ventral Patterning Network. Genetics *202*, 1575–1584. https://doi.org/10.1534/genetics.116.186783.

10. De Renzis, S., Elemento, O., Tavazoie, S., and Wieschaus, E.F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. PLoS Biol. *5*, e117.

11. Schejter, E.D., Rose, L.S., Postner, M.A., and Wieschaus, E. (1992). Role of the zygotic genome in the restructuring of the actin cytoskeleton at the cycle-14 transition during Drosophila embryogenesis. Cold Spring Harbor Symp. Quant. Biol. *57*, 653–659.

12. Lecuit, T., and Wieschaus, E. (2000). Polarized insertion of new membrane from a cytoplasmic reservoir during cleavage of the Drosophila embryo. J. Cell Biol. *150*, 849–860.

13. Foe, V.E., and Alberts, B.M. (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. J. Cell Sci. *61*, 31–70.

14. Martin, A.C., Kaschube, M., and Wieschaus, E.F. (2009). Pulsed contractions of an actin-myosin network drive apical constriction. Nature *457*, 495–499.

15. Parks, S., and Wieschaus, E. (1991). The Drosophila gastrulation gene concertina encodes a G alpha-like protein. Cell *64*, 447–458.

16. Kölsch, V., Seher, T., Fernandez-Ballester, G.J., Serrano, L., and Leptin, M. (2007). Control of Drosophila Gastrulation by Apical Localization of Adherens Junctions and RhoGEF2. Science *315*, 384–386. https://doi.org/10.1126/science.1134833.

17. Edgar, B.A., Sprenger, F., Duronio, R.J., Leopold, P., and O'Farrell, P.H. (1994). Distinct molecular mechanism regulate cell cycle timing at successive stages of Drosophila embryogenesis. Genes Dev. *8*, 440–452.

18. Grosshans, J., and Wieschaus, E. (2000). A genetic link between morphogenesis and cell division during formation of the ventral furrow in Drosophila. Cell *101*, 523–531.

19. Irvine, K.D., and Wieschaus, E. (1994). Cell intercalation during Drosophila germband extension and its regulation by pair-rule segmentation genes. Development *120*, 827–841.

20. Clark, E., and Akam, M. (2016). Odd-paired controls frequency doubling in segmentation by altering the pair-rule gene regulatory network. Elife *5*, e18215. https://doi.org/10.7554/eLife.18215.

21. Clark, E., Battistara, M., and Benton, M.A. (2022). A timer gene network is spatially regulated by the terminal system in the embryo. Elife *11*, e78902. https://doi.org/10.7554/eLife.78902.

22. Soluri, I.V., Zumerling, L.M., Payan Parra, O.A., Clark, E.G., and Blythe, S.A. (2020). Zygotic pioneer factor activity of Odd-paired/Zic is necessary for late function of the segmentation network. Elife *9*, e53916. https://doi.org/10.7554/eLife.53916.

23. Koromila, T., Gao, F., Iwasaki, Y., He, P., Pachter, L., Gergen, J.P., and Stathopoulos, A. (2020). Odd-paired is a pioneer-like factor that coordinates with Zelda to control gene expression in embryos. Elife *9*, e59610. https://doi.org/10.7554/eLife.59610.

24. Baumgartner, S., and Noll, M. (1990). Network of interactions among pair-rule genes regulating paired expression during primordial segmentation of Drosophila. Mech. Dev. *33*, 1–18.

25. Clark, E., Peel, A.D., and Akam, M. (2019). Arthropod segmentation. Development *146*, dev170480. https://doi.org/10.1242/dev.170480.

26. Ismat, A., Schaub, C., Reim, I., Kirchner, K., Schultheis, D., and Frasch, M. (2010). HLH54F is required for the specification and migration of longitudinal gut muscle founders from the caudal mesoderm of Drosophila. Development *137*, 3107–3117.

27. Singer, J.B., Harbecke, R., Kusch, T., Reuter, R., and Lengyel, J.A. (1996). Drosophila brachyenteron regulates gene activity and morphogenesis in the gut. Development *122*, 3707–3718.

28. Lengyel, J.A., and Liu, X.J. (1998). Posterior gut development in Drosophila: a model system for identifying genes controlling epithelial morphogenesis. Cell Res. *8*, 273–284.

29. Lee, T., and Luo, L. (1999). Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. Neuron *22*, 451–461.

30. Marqués, G., Bao, H., Haerry, T.E., Shimell, M.J., Duchek, P., Zhang, B., and O'Connor, M.B. (2002). The Drosophila BMP type II receptor Wishful Thinking regulates neuromuscular synapse morphology and function. Neuron *33*, 529–543.

31. Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The embryo at single-cell transcriptome resolution. Science *358*, 194–199.

32. Tattikota, S.G., Cho, B., Liu, Y., Hu, Y., Barrera, V., Steinbaugh, M.J., Yoon, S.-H., Comjean, A., Li, F., Dervis, F., et al. (2020). A single-cell survey of Drosophila blood. Elife *9*, e54818. https://doi.org/10.7554/eLife.54818.

33. Seroka, A., Lai, S.-L., and Doe, C.Q. (2022). Transcriptional profiling from whole embryos to single neuroblast lineages in Drosophila. Dev. Biol. *489*, 21–33.

34. Pérez-Mojica, J.E., Enders, L., Walsh, J., Lau, K.H., and Lempradl, A. (2023). Continuous transcriptome analysis reveals novel patterns of early gene expression in Drosophila embryos. Cell Genom. *3*, 100265.

35. Oda, H., Tsukita, S., and Takeichi, M. (1998). Dynamic behavior of the cadherin-based cell-cell adhesion system during Drosophila gastrulation. Dev. Biol. *203*, 435–450.

36. Casal, J., and Leptin, M. (1996). Identification of novel genes in Drosophila reveals the complex regulation of early gene activity in the mesoderm. Proc. Natl. Acad. Sci. USA *93*, 10327–10332.

37. Lecuit, T., Samanta, R., and Wieschaus, E. (2002). slam encodes a developmental regulator of polarized membrane growth during cleavage of the Drosophila embryo. Dev. Cell *2*, 425–436.

38. Schejter, E.D., and Wieschaus, E. (1993). bottleneck acts as a regulator of the microfilament network governing cellularization of the Drosophila embryo. Cell *75*, 373–385.

39. Nakamura, A., Amikura, R., Mukai, M., Kobayashi, S., and Lasko, P.F. (1996). Requirement for a noncoding RNA in Drosophila polar granules for germ cell establishment. Science *274*, 2075–2079.

40. Wang, C., and Lehmann, R. (1991). Nanos is the localized posterior determinant in Drosophila. Cell *66*, 637–647.

41. Pignoni, F., Baldarelli, R.M., Steingrímsson, E., Diaz, R.J., Patapoutian, A., Merriam, J.R., and Lengyel, J.A. (1990). The Drosophila gene tailless is expressed at the embryonic termini and is a member of the steroid receptor superfamily. Cell *62*, 151–163.

42. Martin, E.C., and Adler, P.N. (1993). The Polycomb group gene Posterior Sex Combs encodes a chromosomal protein. Development *117*, 641–655.

43. Xu, Y., and Wang, T. (2019). LOVIT Is a Putative Vesicular Histamine Transporter Required in Drosophila for Vision. Cell Rep. *27*, 1327–1333.e3.

44. Dominguez, C., Zuñiga, A., Hanna, P., Hodar, C., Gonzalez, M., and Cambiazo, V. (2016). Target genes of Dpp/BMP signaling pathway revealed by transcriptome profiling in the early D.melanogaster embryo. Gene *591*, 191–200.

45. Albright, A.R., Stadler, M.R., and Eisen, M.B. (2022). Single-nucleus RNA-sequencing in pre-cellularization Drosophila melanogaster embryos. PLoS One *17*, e0270471.

46. Maurange, C., Cheng, L., and Gould, A.P. (2008). Temporal transcription factors and their targets schedule the end of neural proliferation in Drosophila. Cell *133*, 891–902.

47. Jacobs, J., Atkins, M., Davie, K., Imrichova, H., Romanelli, L., Christiaens, V., Hulselmans, G., Potier, D., Wouters, J., Taskiran, I.I., et al. (2018). The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. Nat. Genet. *50*, 1011–1020.

48. Garcia, M., Nahmad, M., Reeves, G.T., and Stathopoulos, A. (2013). Size-dependent regulation of dorsal-ventral patterning in the early Drosophila embryo. Dev. Biol. *381*, 286–299.

49. Sun, J., and Stathopoulos, A. (2018). FGF controls epithelial-mesenchymal transitions during gastrulation by regulating cell division and apicobasal polarity. Development *145*, dev161927. https://doi.org/10.1242/dev.161927.

50. Mendoza-García, P., Hugosson, F., Fallah, M., Higgins, M.L., Iwasaki, Y., Pfeifer, K., Wolfstetter, G., Varshney, G., Popichenko, D., Gergen, J.P., et al. (2017). The Zic family homologue Odd-paired regulates Alk expression in Drosophila. PLoS Genet. *13*, e1006617.

51. Yin, W., Mendoza, L., Monzon-Sandoval, J., Urrutia, A.O., and Gutierrez, H. (2021). Emergence of co-expression in gene regulatory networks. PLoS One *16*, e0247671.

52. Marco, A., Konikoff, C., Karr, T.L., and Kumar, S. (2009). Relationship between gene co-expression and sharing of transcription factor binding sites in Drosophila melanogaster. Bioinformatics *25*, 2473–2477.

53. Martinez-Arias, A., and Lawrence, P.A. (1985). Parasegments and compartments in the Drosophila embryo. Nature *313*, 639–642.

54. Kornberg, T.B., and Tabata, T. (1993). Segmentation of the Drosophila embryo. Curr. Opin. Genet. Dev. *3*, 585–594.

55. Patel, N.H. (1994). Developmental evolution: insights from studies of insect segmentation. Science *266*, 581–590.

56. Formaz-Preston, A., Ryu, J.-R., Svendsen, P.C., and Brook, W.J. (2012). The Tbx20 homolog Midline represses wingless in conjunction with Groucho during the maintenance of segment polarity. Dev. Biol. *369*, 319–329.

57. Kumar, R.P., Dobi, K.C., Baylies, M.K., and Abmayr, S.M. (2015). Muscle cell fate choice requires the T-box transcription factor midline in Drosophila. Genetics *199*, 777–791.

58. Miskolczi-McCallum, C.M., Scavetta, R.J., Svendsen, P.C., Soanes, K.H., and Brook, W.J. (2005). The Drosophila melanogaster T-box genes midline and H15 are conserved regulators of heart development. Dev. Biol. *278*, 459–472.

59. Inaki, M., Kojima, T., Ueda, R., and Saigo, K. (2002). Requirements of high levels of Hedgehog signaling activity for medial-region cell fate determination in Drosophila legs: identification of pxb, a putative Hedgehog signaling attenuator gene repressed along the anterior-posterior compartment boundary. Mech. Dev. *116*, 3–18.

60. McNeill, H., Yang, C.H., Brodsky, M., Ungos, J., and Simon, M.A. (1997). mirror encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the Drosophila eye. Genes Dev. *11*, 1073–1082.

61. Mirzoyan, Z., and Pandur, P. (2013). The Iroquois complex is required in the dorsal mesoderm to ensure normal heart development in Drosophila. PLoS One *8*, e76498.

62. Keleman, K., Rajagopalan, S., Cleppien, D., Teis, D., Paiha, K., Huber, L.A., Technau, G.M., and Dickson, B.J. (2002). Comm sorts robo to control axon guidance at the Drosophila midline. Cell *110*, 415–427.

63. Justice, E.D., Barnum, S.J., and Kidd, T. (2017). The WAGR syndrome gene PRRG4 is a functional homologue of the commissureless axon guidance gene. PLoS Genet. *13*, e1006865. https://doi.org/10.1371/journal.pgen.1006865.

64. Sharrock, T.E., Evans, J., Blanchard, G.B., and Sanson, B. (2022). Different temporal requirements for tartan and wingless in the formation of contractile interfaces at compartmental boundaries. Development *149*, dev200292. https://doi.org/10.1242/dev.200292.

65. Lee, H.H., and Frasch, M. (2000). Wingless effects mesoderm patterning and ectoderm segmentation events via induction of its downstream target sloppy paired. Development *127*, 5497–5508.

66. Azpiazu, N., Lawrence, P.A., Vincent, J.P., and Frasch, M. (1996). Segmentation and specification of the Drosophila mesoderm. Genes Dev. *10*, 3183–3194.

67. Azpiazu, N., and Frasch, M. (1993). tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila. Genes Dev. *7*, 1325–1340.

68. Chiang, C., Patel, N.H., Young, K.E., and Beachy, P.A. (1994). The novel homeodomain gene buttonless specifies differentiation and axonal guidance functions of Drosophila dorsal median cells. Development *120*, 3581–3593.

69. Lee, H.-H., Norris, A., Weiss, J.B., and Frasch, M. (2003). Jelly belly protein activates the receptor tyrosine kinase Alk to specify visceral muscle pioneers. Nature *425*, 507–512.

70. Tapanes-Castillo, A., and Baylies, M.K. (2004). Notch signaling patterns Drosophila mesodermal segments by regulating the bHLH transcription factor twist. Development *131*, 2359–2372.

71. Warren, J.T., Petryk, A., Marqués, G., Parvy, J.-P., Shinoda, T., Itoyama, K., Kobayashi, J., Jarcho, M., Li, Y., O'Connor, M.B., et al. (2004). Phantom encodes the 25-hydroxylase of Drosophila melanogaster and Bombyx mori: a P450 enzyme critical in ecdysone biosynthesis. Insect Biochem. Mol. Biol. *34*, 991–1010.

72. Niwa, R., Matsuda, T., Yoshiyama, T., Namiki, T., Mita, K., Fujimoto, Y., and Kataoka, H. (2004). CYP306A1, a cytochrome P450 enzyme, is essential for ecdysteroid biosynthesis in the prothoracic glands of Bombyx and Drosophila. J. Biol. Chem. *279*, 35942–35949.

73. Chávez, V.M., Marqués, G., Delbecque, J.P., Kobayashi, K., Hollingsworth, M., Burr, J., Natzle, J.E., and O'Connor, M.B. (2000). The Drosophila disembodied gene controls late embryonic morphogenesis and codes for a cytochrome P450 enzyme that regulates embryonic ecdysone levels. Development *127*, 4115–4126.

74. Niwa, R., Namiki, T., Ito, K., Shimada-Niwa, Y., Kiuchi, M., Kawaoka, S., Kayukawa, T., Banno, Y., Fujimoto, Y., Shigenobu, S., et al. (2010). Non-molting glossy/shroud encodes a short-chain dehydrogenase/reductase that functions in the "Black Box" of the ecdysteroid biosynthesis pathway. Development *137*, 1991–1999.

75. Huang, X., Warren, J.T., Buchanan, J., Gilbert, L.I., and Scott, M.P. (2007). Drosophila Niemann-Pick type C-2 genes control sterol homeostasis and steroid biosynthesis: a model of human neurodegenerative disease. Development *134*, 3733–3742.

76. Niwa, Y.S., and Niwa, R. (2016). Transcriptional regulation of insect steroid hormone biosynthesis and its role in controlling timing of molting and metamorphosis. Dev. Growth Differ. *58*, 94–105.

77. Uyehara, C.M., and McKay, D.J. (2019). Direct and widespread role for the nuclear receptor EcR in mediating the response to ecdysone in Drosophila. Proc. Natl. Acad. Sci. USA *116*, 9893–9902.

78. Moeller, M.E., Danielsen, E.T., Herder, R., O'Connor, M.B., and Rewitz, K.F. (2013). Dynamic feedback circuits function as a switch for shaping a maturation-inducing steroid pulse in Drosophila. Development *140*, 4730–4739.

79. Chavoshi, T.M., Moussian, B., and Uv, A. (2010). Tissue-autonomous EcR functions are required for concurrent organ morphogenesis in the Drosophila embryo. Mech. Dev. *127*, 308–319.

80. Yoo, B., Kim, H.-Y., Chen, X., Shen, W., Jang, J.S., Stein, S.N., Cormier, O., Pereira, L., Shih, C.R.Y., Krieger, C., et al. (2021). 20-hydroxyecdysone (20E) signaling regulates amnioserosa morphogenesis during *Drosophila* dorsal closure: EcR modulates gene expression in a complex with the AP-1 subunit, Jun. Biol. Open *10*, bio058605. https://doi.org/10.1242/bio.058605.

81. Hackney, J.F., Pucci, C., Naes, E., and Dobens, L. (2007). Ras signaling modulates activity of the ecdysone receptor EcR during cell migration in the Drosophila ovary. Dev. Dynam. *236*, 1213–1226.

82. Carney, G.E., and Bender, M. (2000). The Drosophila ecdysone receptor (EcR) gene is required maternally for normal oogenesis. Genetics *154*, 1203–1211.

83. Jang, A.C.-C., Chang, Y.-C., Bai, J., and Montell, D. (2009). Border-cell migration requires integration of spatial and temporal signals by the BTB protein Abrupt. Nat. Cell Biol. *11*, 569–579.

84. Kozlova, T., and Thummel, C.S. (2003). Essential roles for ecdysone signaling during Drosophila mid-embryonic development. Science *301*, 1911–1914.

85. Koelle, M.R., Talbot, W.S., Segraves, W.A., Bender, M.T., Cherbas, P., and Hogness, D.S. (1991). The Drosophila EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. Cell *67*, 59–77.

86. White, K.P., Rifkin, S.A., Hurban, P., and Hogness, D.S. (1999). Microarray analysis of Drosophila development during metamorphosis. Science *286*, 2179–2184.

87. Staller, M.V., Yan, D., Randklev, S., Bragdon, M.D., Wunderlich, Z.B., Tao, R., Perkins, L.A., DePace, A.H., and Perrimon, N. (2013). Depleting Gene Activities in Early Drosophila Embryos with the "Maternal-Gal4–shRNA" System. Genetics *193*, 51–61.

88. Kusch, T., and Reuter, R. (1999). Functions for Drosophila brachyenteron and forkhead in mesoderm specification and cell signalling. Development *126*, 3991–4003.

89. Sun, J., Macabenta, F., Akos, Z., and Stathopoulos, A. (2020). Collective Migrations of Drosophila Embryonic Trunk and Caudal Mesoderm-Derived Muscle Precursor Cells. Genetics *215*, 297–322.

90. Beyenbach, K.W., Skaer, H., and Dow, J.A.T. (2010). The developmental, molecular, and transport biology of Malpighian tubules. Annu. Rev. Entomol. *55*, 351–374.

91. Liu, S., and Jack, J. (1992). Regulatory interactions and role in cell type specification of the Malpighian tubules by the cut, Krüppel, and caudal genes of Drosophila. Dev. Biol. *150*, 133–143.

92. Hatton-Ellis, E., Ainsworth, C., Sushama, Y., Wan, S., VijayRaghavan, K., and Skaer, H. (2007). Genetic regulation of patterned tubular branching in Drosophila. Proc. Natl. Acad. Sci. USA *104*, 169–174.

93. Vorbrüggen, G., Constien, R., Zilian, O., Wimmer, E.A., Dowe, G., Taubert, H., Noll, M., and Jäckle, H. (1997). Embryonic expression and characterization of a Ptx1 homolog in Drosophila. Mech. Dev. *68*, 139–147.

94. Harbecke, R., and Janning, W. (1989). The segmentation gene Krüppel of Drosophila melanogaster has homeotic properties. Genes Dev. *3*, 114–122.

95. Reuter, R. (1994). The gene serpent has homeotic properties and specifies endoderm versus ectoderm within the Drosophila gut. Development *120*, 1123–1135.

96. Swan, L.E., Wichmann, C., Prange, U., Schmid, A., Schmidt, M., Schwarz, T., Ponimaskin, E., Madeo, F., Vorbrüggen, G., and Sigrist, S.J. (2004). A glutamate receptor-interacting protein homolog organizes muscle guidance in Drosophila. Genes Dev. *18*, 223–237.

97. Swan, L.E., Schmidt, M., Schwarz, T., Ponimaskin, E., Prange, U., Boeckers, T., Thomas, U., and Sigrist, S.J. (2006). Complex interaction of Drosophila GRIP PDZ domains and Echinoid during muscle morphogenesis. EMBO J. *25*, 3640–3651.

98. Zhu, X., Ahmad, S.M., Aboukhalil, A., Busser, B.W., Kim, Y., Tansey, T.R., Haimovich, A., Jeffries, N., Bulyk, M.L., and Michelson, A.M. (2012). Differential regulation of mesodermal gene expression by Drosophila cell type-specific Forkhead transcription factors. Development *139*, 1457–1466.

99. Golson, M.L., and Kaestner, K.H. (2016). Fox transcription factors: from development to disease. Development *143*, 4558–4570.

100. Herman, L., Todeschini, A.-L., and Veitia, R.A. (2021). Forkhead Transcription Factors in Health and Disease. Trends Genet. *37*, 460–475.

101. Hanlon, C.D., and Andrew, D.J. (2016). Drosophila FoxL1 non-autonomously coordinates organ placement during embryonic development. Dev. Biol. *419*, 273–284. https://doi.org/10.1016/j.ydbio.2016.09.007.

102. Chen, H., Xu, Z., Mei, C., Yu, D., and Small, S. (2012). A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes. Cell *149*, 618–629.

103. Tsai, C.C., Kramer, S.G., and Gergen, J.P. (1998). Pair-rule gene runt restricts orthodenticle expression to the presumptive head of the Drosophila embryo. Dev. Genet. *23*, 35–44.

104. Stathopoulos, A., Tam, B., Ronshaugen, M., Frasch, M., and Levine, M. (2004). pyramus and thisbe: FGF genes that pattern the mesoderm of Drosophila embryos. Genes Dev. *18*, 687–699.

105. Gisselbrecht, S., Skeath, J.B., Doe, C.Q., and Michelson, A.M. (1996). heartless encodes a fibroblast growth factor receptor (DFR1/DFGF-R2) involved in the directional migration of early mesodermal cells in the Drosophila embryo. Genes Dev. *10*, 3003–3017.

106. Mbodj, A., Gustafson, E.H., Ciglar, L., Junion, G., Gonzalez, A., Girardot, C., Perrin, L., Furlong, E.E.M., and Thieffry, D. (2016). Qualitative Dynamical Modelling Can Formally Explain Mesoderm Specification and Predict Novel Developmental Phenotypes. PLoS Comput. Biol. *12*, e1005073.

107. Trcek, T., and Lehmann, R. (2019). Germ granules in Drosophila. Traffic *20*, 650–660.

108. Vastenhouw, N.L., Cao, W.X., and Lipshitz, H.D. (2019). The maternal-to-zygotic transition revisited. Development *146*, dev161471. https://doi.org/10.1242/dev.161471.

109. Paré, A.C., Vichas, A., Fincher, C.T., Mirman, Z., Farrell, D.L., Mainieri, A., and Zallen, J.A. (2014). A positional Toll receptor code directs convergent extension in Drosophila. Nature *515*, 523–527.

110. McMahon, A., Supatto, W., Fraser, S.E., and Stathopoulos, A. (2008). Dynamic analyses of Drosophila gastrulation provide insights into collective cell migration. Science *322*, 1546–1550.

111. Bhattacharya, D., Azambuja, A.P., and Simoes-Costa, M. (2020). Metabolic Reprogramming Promotes Neural Crest Migration via Yap/Tead Signaling. Dev. Cell *53*, 199–211.e6.

112. Secchia, S., Forneris, M., Heinen, T., Stegle, O., and Furlong, E.E.M. (2022). Simultaneous cellular and molecular phenotyping of embryonic mutants using single-cell regulatory trajectories. Dev. Cell *57*, 496–511.e8. https://doi.org/10.1016/j.devcel.2022.01.016.

113. San Martin, B., Bate, M., and Bate, M. (2001). Hindgut visceral mesoderm requires an ectodermal template for normal development in Drosophila. Development *128*, 233–242. https://doi.org/10.1242/dev.128.2.233.

114. Denholm, B., Sudarsan, V., Pasalodos-Sanchez, S., Artero, R., Lawrence, P., Maddrell, S., Baylies, M., and Skaer, H. (2003). Dual origin of the renal tubules in Drosophila: mesodermal cells integrate and polarize to establish secretory function. Curr. Biol. *13*, 1052–1057.

115. Stepanik, V., Sun, J., and Stathopoulos, A. (2020). FGF Pyramus Has a Transmembrane Domain and Cell-Autonomous Function in Polarity. Curr. Biol. *30*, 3141–3153.e5.

116. Prazak, L., Iwasaki, Y., Kim, A.-R., Kozlov, K., King, K., and Gergen, J.P. (2021). A dual role for DNA binding by Runt in activation and repression of sloppy paired transcription. Mol. Biol. Cell *32*, ar26.

117. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. *47*, W191–W198.

118. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

119. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29.

120. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. *8*, 329–337.e4.

121. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982. https://doi.org/10.1038/nmeth.4402.

122. Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A., and Trapnell, C. (2017). Single-cell mRNA quantification and differential analysis with Census. Nat. Methods *14*, 309–315.

123. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

124. Kosman, D., Mizutani, C.M., Lemons, D., Cox, W.G., McGinnis, W., and Bier, E. (2004). Multiplex detection of RNA expression in Drosophila embryos. Science *305*, 846.

125. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*, dev165753. https://doi.org/10.1242/dev.165753.

126. Slaidina, M., Gupta, S., Banisch, T.U., and Lehmann, R. (2021). A single-cell atlas reveals unanticipated cell type complexity in ovaries. Genome Res. *31*, 1938–1951.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Rabbit anti-Odd paired | This study | N/A |
| Guinea pig anti-Heartless | Stepanik et al.[115] | N/A |
| Rabbit anti-Phantom (Phm) | Dr. Michael O'Connor (Minneapolis, MN, USA) | N/A |
| Mouse anti-Even-skipped | Developmental Studies Hybridoma Bank (DSHB) | 2B8; RRID: AB_528230 |
| Goat anti-GFP | Rockland Immunochemicals | Cat# 600-101-215; RRID: AB_218182 |
| Rabbit anti-RFP | MBL International | Cat# PM005; RRID: AB_591279 |
| Rabbit anti-Beta-Galactosidase | MP Biomedicals | Cat# 559761; RRID: AB_2687418 |
| Monoclonal Anti-MAP Kinase, Activated | Sigma-Aldrich | Cat# M9692; RRID: AB_260729 |
| Mouse anti-Biotin (Z021) | Thermo Fisher | Cat# 03–3700; RRID: AB_2532265 |
| Rabbit anti-FITC polyclonal antibody | Thermo Fisher | Cat# A-889; RRID: AB_221561 |
| Sheep anti-Digoxigenin polyclonal antibody | Thermo Fisher | Cat# PA1-85378; RRID: AB_930545 |
| Anti-Digoxigenin-AP, Fab Fragments Antibody | Sigma-Aldrich | Cat# 11093274910; RRID: AB_2734716 |
| Alexa Fluor 488 goat anti-mouse | Molecular Probes | Cat# A-21202; RRID: AB_141607 |
| Alexa Fluor 488 goat anti-guinea pig | Molecular Probes | Cat# A-11073; RRID: AB_2534117 |
| Alexa Fluor 488 donkey anti-guinea pig | Jackson Immuno Research Labs | Cat# 706-545-148; RRID: AB_2340472 |
| Alexa Fluor 555 donkey anti-rabbit | Molecular Probes | Cat# A-31572; RRID: AB_162543 |
| Alexa Fluor 555 donkey anti-sheep | Molecular Probes | Cat# A-21436; RRID:AB_2535857 |
| Alexa Fluor 647 donkey anti-rabbit | Molecular Probes | Cat# A-31573; RRID: AB_2536183 |
| Alexa Fluor 647 donkey anti-mouse | Molecular Probes | Cat# A-31571; RRID: AB_162542 |
| Chemicals, peptides, and recombinant proteins | | |
| Boitin RNA labeling mix | Roche | 55612420 |
| DIG RNA labeling mix | Roche | 57127420 |
| Fluorescein RNA labeling mix | Roche | 59973820 |
| T7 RNA polymerase | Roche | 13644022 |
| Critical commercial assays | | |
| Chromium Next GEM Single Cell 3′ v3.1 (Dual index) | 10X Genomics | 11684795910 |
| Dead Cell Removal Kit | Miltenyi Biotec | 130-090-101 |
| LIVE/DEAD Cell Imaging kit (488/570) | Thermo Fisher | R37601 |
| Accumax® | Innovative Cell Technologies | #AM-105 |
| Hybridization Chain Reaction | Molecular Instruments | N/A |
| Deposited data | | |
| Single-cell RNA-seq data | This study | GEO: GSE222660 |
| Visualization of UMAP graph | This study | Mendeley data: https://doi.org/10.17632/s25d8y9pvz.1 |
| Experimental models: Organisms/strains | | |
| *D. melanogaster*: $htl^{AB42}$/*TM3,ftz-lacZ* | Bloomington Drosophila Stock Center (BDSC) | #5370 |
| *D. melanogaster*: *yw;*$opa^{1}$*, slp1-DESE-lacZ,/TM3,twi-GFP* | BDSC | #3312 |
| *D. melanogaster*: $opa^{8}$/*TM3,twi-GFP* | BDSC | #5335 |
| *D. melanogaster*: $run^{3}$/*FM7c,ftz-lacZ* | BDSC | #56499 |
| *D. melanogaster*: *yw;UAS-run[[232]/CyO* | Prazak et al.[116] | N/A |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| *D. melanogaster*: *UAS-opa[D10]* | Prazak et al.[116] | N/A |
| *D. melanogaster: Adh$^{n7}$sna$^{1}$cn$^{1}$vg$^{1}$/CyO* | BDSC | #25127 |
| *D. melanogaster*: *y[1] w[*]; P{w[+mC] =GAL4-twi.2xPE}3* | BDSC | #58804 |
| *D. melanogaster*: *y[1] w[*]; P{w[+mC] =GAL4-twi.2xPE}2* | BDSC | #2517 |
| *D. melanogaster: P{ry[+t7.2]=hsFLP}1, y[1] w[1118]; P{w[+mW.hs]=FRT(w[hs])}G13 P {w[+mC]=UAS-mCD8::GFP.L}LL5* | BDSC | #5131 |
| *D. melanogaster*: *y[1] w[*]; P{w[+mC]=UAS-mCD8::GFP.L}LL5, P{UAS-mCD8::GFP.L}2* | BDSC | #5137 |
| *D. melanogaster: P{w[+mC]=otu-GAL4::VP16.R}1, w[*]; P{w[+mC]=GAL4-nanos.NGT}40; P{w[+mC]=GAL4::VP16-nanos.UTR}CG6325[MVD1]* | BDSC | #31777 |
| *D. melanogaster: y w; P(mat-tub-Gal4) mat67; P(mat-tub-Gal4)mat15* | Staller et al.[87] | line 2318 |
| *D. melanogaster*: *w[*]; P{w[+mC]=UAS-EcR.A.W650A}TP5* | BDSC | #9451 |
| *D. melanogaster*: *w[*]; P{w[+mC]=UAS-EcR.B2.W650A}TP5* | BDSC | #9449 |
| *D. melanogaster*: *y[1] sc[*] v[1] sev[21]; P{y [+t7.7] v[+t1.8]=TRiP.HMS01185}attP2/ TM3, Sb[1]* | BDSC | #34706 |
| *D. melanogaster*: *y[1] w[*]; P{y[+t7.7] w [+mC]=E(spl)m5-HLH-GFP.FPTB}attP40* | BDSC | #66445 |
| *D. melanogaster*: *y[1] w[*]; P{y[+t7.7] w [+mC]=E(spl)mdelta-HLH-GFP.FPTB} attP40* | BDSC | #68191 |
| *D. melanogaster*: *y[1] w[*]; P{y[+t7.7] w [+mC]=mirr-GFP.FPTB}attP40* | BDSC | #68183 |
| *D. melanogaster*: *y[1] w[*]; PBac{y[+mDint2] w[+mC]=btn-GFP.FPTB}VK00037* | BDSC | #56154 |
| **Oligonucleotides** | | |
| HCR probes[117] (Table S8) | Molecular Instruments | Table S8 |
| PCR primers used to generate riboprobes (Table S8) | This paper | Table S8 |
| **Recombinant DNA** | | |
| *danr* cDNA clone RE72284 | Drosophila Genomics Resource Center (DGRC) | RRID: DGRC_10042 |
| *ich* cDNA clone RE65372 | Drosophila Genomics Resource Center (DGRC) | RRID: DGRC_1135960 |
| *mid* cDNA clone LP04777 | Drosophila Genomics Resource Center (DGRC) | RRID: DGRC_1337803 |
| **Software and algorithms** | | |
| Zeiss Zen 3.0 (Blue edition) | Zeiss | RRID: SCR_013672 |
| Adobe Photoshop v24.1 | www.adobe.com | RRID: SCR_014199 |
| Adobe Illustrator v27.1.1 | www.adobe.com | RRID: SCR_010279 |
| Cell Ranger v7.0 | 10x Genomics | RRID: SCR_017344 |
| Seurat[118] v3.2.0, v4.1.0[119] | https://satijalab.org/seurat/ | RRID: SCR_016341 |
| DoubletFinder[120] | https://github.com/chris-mcginnis-ucsf/DoubletFinder | RRID: SCR_018771 |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Monocle2[121–123] | http://cole-trapnell-lab.github.io/monocle-release/docs/ | RRID: SCR_016339 |
| g:Profiler[117] | https://biit.cs.ut.ee/gprofiler/gost | RRID: SCR_006809 |
| Other | | |
| Qubit™ dsDNA HS Assay | Thermo Fisher | Q32854 |
| Bioanalyzer High Sensitivity DNA Analysis | Agilent | 5067–4626 |
| HiSeq X | Illumina | N/A |
| NovaSeq 6000 | Illumina | N/A |
| R scripts and markdown files | This study | https://github.com/StathopoulosLab/Single_Cell_RNAseq_Analysis_supplementary; https://zenodo.org/badge/latestdoi/668423939 |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Angelike Stathopoulos (angelike@caltech.edu).

### Materials availability

*Drosophila* strains and other reagents generated in this study will be available upon request from the lead contact, or the commercial sources listed in the key resources table.

### Data and code availability

- The raw sc-RNA sequencing data generated in this study has been submitted to NCBI database (https://www.ncbi.nlm.nih.gov/geo/) and is available under accession number GEO: GSE222660
- A github repository was generated and publicly available with the analysis codes for all datasets (https://github.com/StathopoulosLab/Single_Cell_RNAseq_Analysis_supplementary or https://zenodo.org/badge/latestdoi/668423939)
- Additional Rds objects, R script and installation instructions for visualizing RNA sequencing UMAP graphs are deposited into Mendeley Data Respository (Mendeley data: https://doi.org/10.17632/s25d8y9pvz.1)
- Any additional information required to reanalyze the data shown in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Fly stocks and husbandry

All stocks were kept at 22°C in the standard medium. Experimental crosses were kept in cages with apple juice agar plates supplemented with yeast paste. yw were used to show the wild-type expression pattern, and heterozygous embryos from the mutant collections were used as the controls. Embryos overexpressing run were generated by crossing twi(2PE)-GAL4 females (BDSC#58804, #2517) with UAS-run (gift from Peter Gergen) males. EcR dominant negative zygotic mutants (F1) were generated by crossing MTD-GAL4 (BDSC#31777) virgin females with UAS-EcR.B2.W650A (BDSC#9449) males. Two step crosses were used to obtain the maternal mutants (F2). First, MTD-GAL4 males with UAS-EcR.B2.W650A virgin females, virgins collected from the progeny then crossed with UAS-EcR.B2.W650A males.

## METHOD DETAILS

### Embryos collection, cell isolation and library construction

*Drosophila* embryos produced by crossing *twi(2PE)-GAL4* females (BDSC#58804, #2517) and *UAS-mCD8GFP* (BDSC#5131, #5137) males were collected on apple agar plates at 22°C. Two sessions of 45 min pre-lay collections were performed in order to clear the older embryos held by the females. After collecting for 1 h and 30 min, embryos were aged for 3 h and 15 min. A similar timed embryo collection followed by immunostaining with a GFP antibody (1:2000, Rockland) was done in advance to estimate the percentage of embryos at a specific developmental stage obtained, indicating that our collection resulted in an embryonic sample in which the majority (~50%) were at st8, ~30% at st5-7, ~20% at st9-10 and very few older than st10 (~1%). This window allowed us to study both

the patterning of embryos that contributes to germ layer designation (st6-7), germband elongation (GBE) and mesoderm epithelial-to-mesenchymal transition (EMT) (stage 8), as well as the initial differentiation of tissues (stages 9–10).

Prior to cell isolation, Accumax was thawed, and the Douncer was cleaned. Schneider's medium, PBS, Douncer and low-binding tubes were kept on ice. Embryos were dechorionated with 50% bleach/water, rinsed well, then immediately transferred to the Douncer-containing cold media. Gross dissociation was done with the loose pestle by gently moving it up and down ~15 times. After brief centrifugation at 300 RCF to pellet the cells at 4°C, 5min room temperature Accumax treatment was performed to further dissociate the cells. Following washing in cold media twice, cells were filtered through a 30μm MACS cell strainer (Miltenyi Biotec). Lastly, the dead cell removal kit (Miltenyi Biotec) was used to clean up the dying cells and cell debris. The final cell stock was resuspended in cold PBS with 1% BSA, then promptly counted in order to be diluted to an ideal concentration (~900 cells/μl) for the GEM generation following the recommended protocol from 10x Genomics (CG000315 RevC). The LIVE/DEAD cell imaging kit (Invitrogen) was used in test runs to evaluate the viability of the cells (typically >95% after dead cell removal) in the final single-cell suspension. The Chromium Next GEM Single Cell 3′ *v*3.1 kit (10x Genomics) with dual indexes was used for library construction and sequenced on an Illumina HiSeq X platform (Fulgent, Temple City, CA).

Care was taken to expedite processing such that homogenization to running the Chromium Controller occurred in about one and a half hours. Alternatively, protocols that homogenize and then fix cells can have lengthy processing steps. It is possible that particular gene expression bias could be associated with different protocols. This is also why we chose to validate our clusters using *in situ* hybridization or antibody staining of fluorescent protein-tagged genes; in every case (for over 40 genes examined), we detected expression in cells of the tissue type we had expected based on the clustering data. We did find that the mesoderm cell population was overrepresented (4865 out of the original 9845 cells for live wild-type#1, Figure S1B), which may relate to the fact that mesoderm cells are mesenchymal, are detached from each other during gastrulation, and therefore may be more likely to survive the cell isolation process. This was fortuitous, as our goal was to study patterning and signaling dynamics associated with the mesoderm at gastrulation.

### Generation of additional datasets

Live WT#1 is the focus of this study, unless otherwise noted. Two additional wild-type datasets (live WT#2 and fixed WT#3) were generated as biological replicates. Embryos collection, cell isolation and dead cell removal were performed as described above. Methanol fixation and rehydration was done largely following the suggested protocol from 10x Genomics (CG000136 RevB). Briely, 400μL of chilled methanol was added to 100μL of cell suspension drop by drop while gently mixing to prevent clumping. Cells were fixed on ice for 15 min before being transferred to −20°C for short term storage. For rehydration, cells were first placed on ice for 15 min to equilibrated to 4°C, centrifuged at 1000 RCF for 5 min at 4°C and then resuspended in rehydration buffer (1% BSA, 0.5u/μl RNase inhibitor in PBS). One repeat sample of live cell suspension (WT#2) and one sample from dehydrated methanol-fixed cells (WT#3) were processed together using the Chromium Next GEM Single Cell 3′ *v*3.1 kit (10x Genomics). The constructed libraries were sequenced on an Illumina NovaSeq 6000 platform.

### Cluster annotation and validation

Berkeley *Drosophila* Genome Project (BDGP) database was used for cluster annotation. With the expression patterns for ~8600 genes in staged embryos detected by *in situ* hybridization, it also provides standard vocabularies to annotate gene expression patterns during embryonic development. Following the list of markers (Tables S1, S4, and S5), we checked the top 20 genes enriched in each cluster, especially those with more restricted expression domains and time windows. The identity of the cluster was also verified with *in situ* hybridization or antibody staining.

### Fixation, *in situ* hybridization, and immunostaining

Embryos were fixed following the standard protocols. Riboprobes for *wg*, *hh* and *run* were made from cDNA cloned into the pBluescript vector and reverse transcribed with T7 RNA polymerase (Roche). For, *danr* (RE72284), *ich* (RE65372) and *mid* (LP04777), full-length cDNA clones from DGRC were used as templates. The rest of RNA probes used in this study were made by amplifying a 400–1100 bp region from a genomic DNA sample extracted from wild-type flies, with the reverse primers containing a T7 promoter sequence to facilitate reverse transcription. The antisense RNA probes labeled with Digoxiygenin, Fluorescein and Biotin (Roche) were used in combination with primary antibodies (1:400) of different origin (Thermo Fisher) to detect the *in vivo* expression of target genes.[124] Alexa Fluor (488, 555, 647) secondary antibodies (1:500; Molecular Probes) were used for fluorescent signal amplification and detection. Anti-DIG-alkaline phosphatase (1:400; Sigma) with its NBT/BCIP substrates were used to detect *btn* expression in $htl^{AB42}$ mutant background. Standard protocol for antibody staining was used. Dilutions for the primary antibodies were as such: guinea pig anti-Htl (1:800; in house), rabbit anti-Opa (1:200; in house, see below), goat anti-GFP (1:1000; Rockland immunochemicals), mouse anti-Eve (1:80; DSHB), rabbit anti-β-Gal (1:800; MP Biomedicals), rabbit anti-Phtm and guinea pig anti-Sro (1:500, both kind gifts from Dr. Michael O'Connor, Minneapolis, USA). Same secondary antibodies and dilution were used. Purified MBP-Opa amino acids Q61-T410, 6xHis protein was used to immunize a rabbit (Pocono Rabbit Farms and Lab, Canadensis, PA). Subsequently anti-Opa IgG was affinity purified from serum and concentrated by ammonium sulfate precipitation.

### Hybridization chain reaction (HCR)

Hybridization chain reaction (HCR)[125] was performed as described previously[126] with modifications to initial steps to account for embryonic as opposed to ovarian tissue. These probes were used in the study: *eGFP*, *lacZ*, *grh*, *sna*, *htl*, *CadN*, *HLH54F*, *FoxL1*, *Blimp-1*, *Doc2*, *kon*, *Grip*, *run*, *hkb*, *fkh*, *cad*, *sog*, *DNaseII*, *dhd*, *bnk* and *sim*.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Read processing, filtering, doublets removal and visualization

Illumina sequencing reads (294 million raw read pairs) were mapped to customized *Drosophila* dm6-transgene reference and quantified using Cell Ranger *v*7.0 (10x Genomics). A total of 9,845 single cells were recovered from the Cell Ranger pipeline with a mean value of 29,901 reads per cell and a median value of 2,910 genes and 15,251 unique molecular identifiers (UMIs) per cell. The scRNA-seq output matrix was imported to Seurat *v*3.2.0[118] for analysis. Cells with more than 250 but less than 100,000 unique molecular identifiers (UMIs), less than 7000 unique genes, and less than 10% ribosomal reads were kept for downstream analysis (9834 cells). Gene expression counts were log normalized, and log transformed.

Standard data processing was performed as recommended including NormalizeData, FindVariableFeatures (with method set to 'vst'), ScaleData, RunPCA, RunUMAP (with dims set to 1:30), FindNeighbors (with reduction set to PCA and dims set to 1:30). DoubletFinder[120] was used for doublet detection and removed 707 cells.

After the above filtering step (718 cells were removed), the matrix of a total of 9,127 cleaned single cells was subject to normalization, variable gene identification (method as 'vst', nfeatures as 2000), scaling, and PCA analysis. As PCA elbow plot showed the plateau after the first 20 PCA components, dims were set to 1:20 for FindNeighors and RunMAP functions. With resolution set as 0.2 for FindClusters function, a total of 12 unique clusters were identified. Cell cluster marker genes were identified using FindAllMarkers function with the default settings. Top10 markers for each cluster were collected for generating expression heatmap (Table S1).

### Data integration and clustering

The standard Seurat integration pipeline[118] using Log Normalization method was used to integrate the presented dataset with two other wild-type datasets (repeat live WT, WT#2, and methanol-fixed WT, WT#3, shown in Figure S2, Tables S4, and S5). Before integration, all three WT dataset were filtered using Seurat package *v*4.1.0[119] based on UMIs, unique gene features, and ribosomal reads. Cells with more than 250 but less than 100,000 unique molecular identifiers (UMIs), less than 7000 unique genes, and less than 10% ribosomal reads were kept for downstream analysis. DoubletFinder was used for doublet detection and removed 10% of the cells from the WT repeat and fixed WT datasets. The top 2000 variable gene features (FindVariableFeatures, method = 'vst') in each dataset were used for the SelectIntegrationFeatures function and FindIntegrationAnchors function. The three datasets are then integrated with the IntegrateData function using the LogNormalization method and a default PCA dimension of 1:30. A standard workflow of ScaleData, RunPCA, RunUMAP (with dims set to 1:30), FindNeighbors (with reduction set to PCA and dims set to 1:30) were run on the combined dataset. Finally, the UMAP was visualized with 30 PCs.

The same integration pipeline as described above but with Seurat *v*3.2.0 was used to integrate the presented dataset with a previously published dataset (GSE202987: GSM6145582)[33] obtained from stage 12 embryos shown in Figure S1 and Table S2. The stage 12 dataset was previously obtained by Seroka et al. through the 10x Chromium platform, sequences aligned through 10x CellRanger *v*3.1, and analyzed using Seurat package *v*3.1. Before integration, the loaded stage 12 dataset was filtered using Seurat package *v*3.2.0 based on UMIs, unique gene features, and ribosomal reads. Cells with more than 250 but less than 100,000 unique molecular identifiers (UMIs), less than 7000 unique genes, and less than 25% ribosomal reads were kept for downstream analysis (19465 cells). The filtered presented dataset (filter parameters described above) with 9,127 were used to combine with the filtered stage 12 dataset with 19465 cells. The standard integration workflow as described above was performed with a PCA dimension of 1:20. The UMAP was visualized with 20 PCs.

### Gene Ontology enrichment analysis

GO enrichment analysis was performed using the cell signature genes predicted by Seurat. The gene list was analyzed using R version of g:Profiler (https://biit.cs.ut.ee/gprofiler)[117] for drosophila GO: Biological Pathways and FDR threshold set at 0.05.

### Subclustering, differential gene expression, and trajectory analysis

For cell subclustering analysis, the raw gene account matrix of a specific cluster identified in the global analysis was subject to Seurat clustering analysis using the same protocols (20 PCA, 0.5 resolution setting) that have been applied to global analysis. DEG/marker gene analysis (Table S3) was performed using FindAllMarkers function.

Pseudotime trajectory analysis for individual cell clusters was performed with Monocle2[121–123] following the online tutorial. Top 200 (for analysis with cell clusters 5,6, and 9) and top 500 variable genes (for cluster 1 and 10) within the specific cluster detected by Seurat VariableFeatures function were used as ordering_gene for Monocle2 setOrderingFilter function (Tables S3, S6, and S7). Dimension reduction was performed using the DDRTree method. Pseudotime values of individual cells were projected onto a UMAP plot for visualization.