

<https://doi.org/10.1038/s42003-025-08335-x>

# Genome-wide quantitative dissection of an arthropod segmented body plan at single-cell resolution

Check for updates

Takanori Akaiwa<sup>1,2,5</sup>, Hiroki Oda<sup>1,2</sup> & Yasuko Akiyama-Oda<sup>1,3,4</sup>

Developmental processes underlying the characteristic segmented body plans in arthropods vary widely. While *Drosophila* is well-studied, few other arthropod species offer platforms for comparable genomics at single-cell resolution. Here, we present high-quality quantitative data from single-nucleus RNA sequencing of spider *Parasteatoda tepidariorum* embryos at late stage 5 and stage 7, a critical period of emergence of segmental units along the anterior–posterior (AP) axis. Clustering analysis of the stage-7 dataset reconstructs an axial alignment of ectoderm cells, reflecting the differing cell states along the segmenting AP axis. This enables us to obtain genome-wide quantitative gene expression profiles along the reconstructed axis, which were used for unbiased and thorough molecular investigation of pattern elements employing statistical methods. Comprehensive gene-to-gene correlation analyses suggest distinct gene-regulatory interactions in different regions along the reconstructed axis. This study lays the foundation for exploring the origins of developmental diversity in the arthropod body plan.

Multicellular animals within the same phyletic group share characteristic features; however, the developmental processes that generate these features may be remarkably diverse. A prime example of this is the segmented body plan of arthropods, which is a conserved trait achieved through diverse cellular and molecular processes across species<sup>1,2</sup>. Understanding how such shared traits originated and diversified during evolution, particularly from a genomic perspective, is a crucial objective in evolutionary and developmental biology.

Segmentation along the anterior–posterior (AP) axis is one of the most well-known biological processes that has provided the opportunity for exploring the mechanisms of patterning from experimental and theoretical perspectives<sup>3–14</sup>. Genetic studies on segmentation in the *Drosophila melanogaster* embryo through comprehensive identification of the genes involved and their interactions have provided a foundational understanding of the mechanisms underlying arthropod segmentation<sup>15–17</sup>. Candidate-gene approaches, focusing on orthologs of *Drosophila* segmentation genes, have expanded the knowledge on segmentation to a wider range of arthropod species<sup>18–25</sup>. However, these approaches have revealed substantial variations in mechanisms and processes, despite the conserved outcome of periodic gene expression, such as the expression of *engrailed* in one stripe per segment. Genome studies have also shown variations in gene content

even within the same phyla<sup>26–30</sup>. To understand how animals, as genetic and cellular systems, could have diversified the processes of segmentation, model platforms are required for collecting and integrating the genome-wide, unbiased, cell-level, and quantitative data suitable for comparative studies.

The spider *Parasteatoda tepidariorum* offers a valuable alternative model system to *Drosophila*. Unlike the syncytial environment of the *Drosophila* embryo, where the transcription factor Bicoid plays a crucial role in establishing the global AP polarity<sup>31,32</sup>, patterning in the *Parasteatoda* embryo predominantly occurs in a cellular environment<sup>33,34</sup>, with the cell-cell signaling protein Hedgehog playing a similar role<sup>35</sup>. In the downstream patterning processes, segmental stripes arise via temporally repeated expression of genes in *Parasteatoda*<sup>21,25,36–38</sup>, in contrast to spatial regulation by region-specifically expressed gap genes through stripe-specific enhancers in *Drosophila*<sup>39–41</sup>. The dynamics of the temporally changing expression in *Parasteatoda* vary across body regions: bi-splitting in the head and oscillation in the posterior region, together with a non-temporally repeated dynamics, that is, tri-splitting in the thorax<sup>21,25,38,42</sup>. Furthermore, genes such as *Pt-msx1*, which have no known role in *Drosophila* segmentation, contribute to these dynamics of *Parasteatoda* stripe formation<sup>38</sup>. These molecular differences within the conserved patterning outcomes underscore the developmental diversity underlying the shared segmented body plan. With a

<sup>1</sup>JT Biohistory Research Hall, Takatsuki, Osaka, Japan. <sup>2</sup>Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka, Osaka, Japan. <sup>3</sup>Department of Microbiology and Infection Control, Faculty of Medicine, Osaka Medical and Pharmaceutical University, Takatsuki, Osaka, Japan.

<sup>4</sup>PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan. <sup>5</sup>Present address: Exploratory Research Center on Life and Living Systems (ExCELLS), National Institutes of Natural Sciences, Okazaki, Aichi, Japan. e-mail: [yasuko@brh.co.jp](mailto:yasuko@brh.co.jp)

well-annotated, chromosome-level genome assembly, available experimental tools and techniques, bioinformatics resources, and a developing mathematical modeling platform, *Parasteatoda* provides an excellent platform for in-depth investigation of the formation of the stripe pattern<sup>43–47</sup>. Given the importance of temporal transcriptional dynamics within each cell in *Parasteatoda* segmentation and the distinct molecular mechanisms involved, a single-cell-level genome-wide analysis is crucial for understanding the patterning processes.

Single-cell and single-nucleus RNA-sequencing (scRNA-seq and snRNA-seq, respectively) are powerful techniques that enable genome-wide profiling of transcriptional states in individual cells, providing valuable insights into the trajectories of cell differentiation during animal development<sup>48–57</sup>. However, integrating the spatial information with data obtained from dissociated cells and nuclei remains a challenge, despite the development of several integration techniques<sup>55,58–61</sup>. Our group and others have applied scRNA-seq and snRNA-seq techniques to *Parasteatoda* embryos to unravel the cell clusters associated with body regions and specific differentiation trajectories<sup>62–64</sup>. Furthermore, the previous data derived from a small number of sibling embryos at late stage 5 have successfully reconstructed the alignment of cells along the radius of the germ disc (the future AP axis) in a Uniform Manifold Approximation and Projection (UMAP) plot, reflecting gradually different transcriptional states<sup>62</sup>. These findings highlight the unique opportunity presented by the *Parasteatoda* embryo for leveraging scRNA-seq and snRNA-seq to comprehensively investigate spatial gene expression patterns, particularly those reflecting the AP polarity, and gain insights into the molecular mechanisms underlying the body axis patterning.

In this study, we expanded single-cell analysis to later-stage spider embryos exhibiting more complicated stripe patterns. We demonstrate that data obtained from snRNA-seq contained sufficient information to reconstruct the cellular states along the AP axis, accurately reproducing the observed stripe pattern. This reconstructed pattern was then used to generate quantitative gene expression profiles, which were subsequently utilized to identify gene groups exhibiting similar stripe patterns. Furthermore, novel cell states within the mesoderm, endoderm, and extraembryonic tissues were identified. This study establishes a novel platform for conducting genome-wide, single-cell-level, and quantitative analyses of the patterning processes underlying the shared features of arthropods.

## Results

### Single-cell and single-nucleus RNA-seq of germ-band cells

We conducted scRNA-seq and snRNA-seq using *Parasteatoda* embryos at stage 7 [51 h after egg laying (AEL)] in essentially the same way as previously described for embryos at late stage 5 (39 h AEL)<sup>62</sup> and analyzed the late stage-5 and stage-7 datasets together. There is a key transition associated with axis formation and segmentation. The late stage-5 germ disc, which consists of approximately 2000 cells, represents the radial symmetry of gene expression with AP polarity in the peripheral to central order, whereas the stage-7 nascent germ band, which consists of approximately 3500 cells, represents repetitive and nonrepetitive stripes of gene expression along the AP axis (Fig. 1A)<sup>25,38</sup>.

We constructed stage-7 scRNA-seq and snRNA-seq libraries using 15 and 20 carefully staged embryos, respectively (Fig. 1B; Supplementary Fig. 1) and obtained reliable data from 865 cells and 6239 nuclei by filtering out data stemming from potential doublets or low-quality cells and nuclei (Table 1; Supplementary Fig. 2, Supplementary Data 1). The relatively low recovery rate for cells in the scRNA-seq experiment was due to technical difficulties associated with cell dissociation and sedimentation during library construction. To record the precise developmental stages of the embryos used for library construction in the context of stripe-pattern formation, sibling embryos were simultaneously fixed and stained for *Pt-hh* and *Pt-msx1* transcripts (Supplementary Fig. 1). All five embryos from each sample were at very similar stages, displaying bi-splitting stripes in the head, tri-splitting stripes in the thorax, and stripes formed via oscillatory gene expression in the posterior region. Based on the gene expression pattern, the

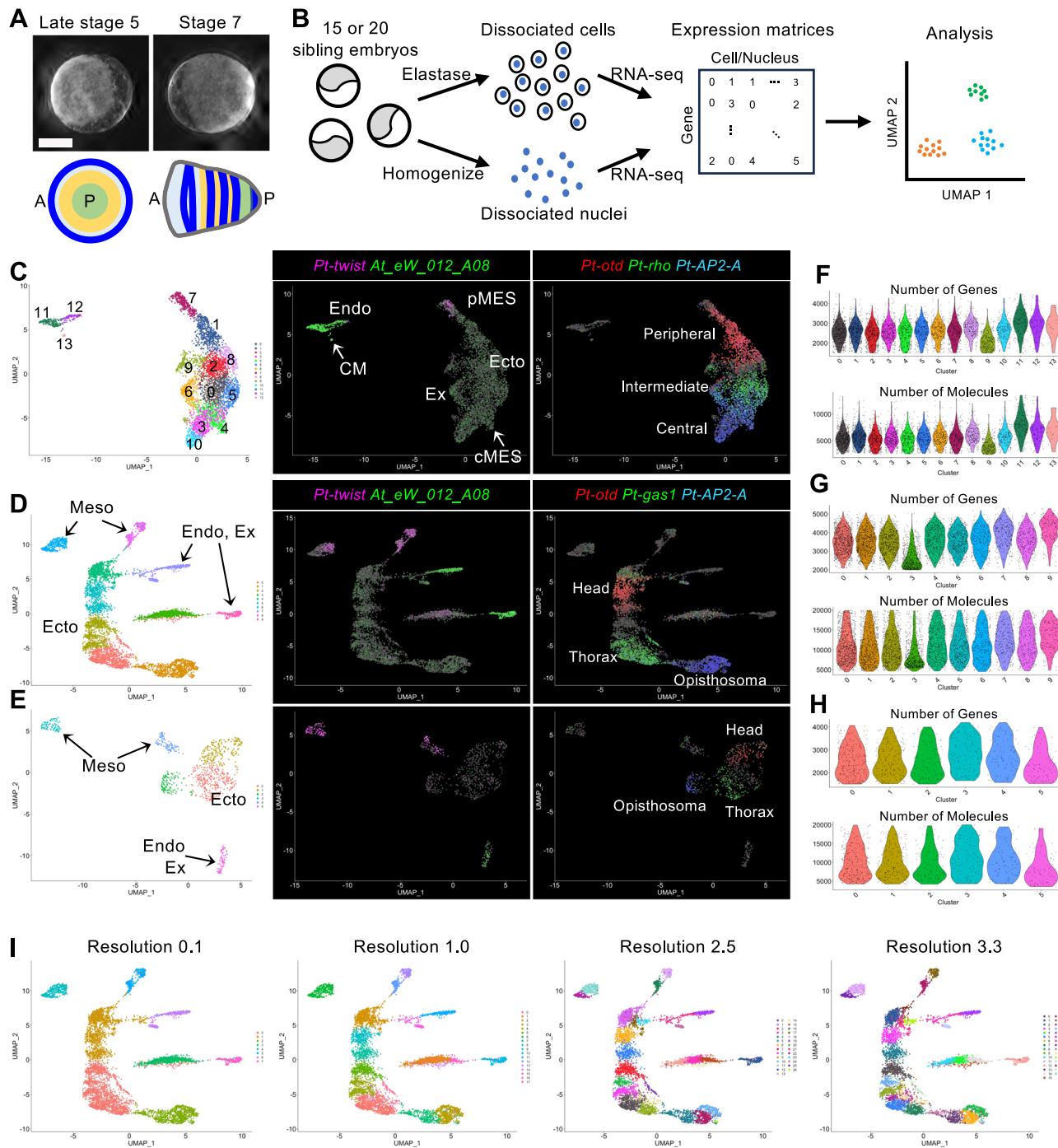
stages of embryos used for the single-cell preparation were slightly younger than those used for the single-nucleus preparation.

Clustering analysis was performed against the cells and nuclei from the three datasets (late stage-5\_nucleus, stage-7\_nucleus, and stage-7\_cell) using the Seurat package with various parameter settings. The representative results are displayed as colored dot plots in the 2-dimensional (2-D) space using UMAP, characterizing the cell states, and as violin plots characterizing the clustered cell populations (Figs. 1C–I, 2; Supplementary Figs. 3–5). These UMAP plots similarly exhibited the spatial separation of cells/nuclei into one large group and several smaller groups, irrespective of the dimension values used in the analyses (Supplementary Fig. 3). Here, the term “supercluster” of cells/nuclei is used to indicate a group of cells/nuclei plotted in a close area of the 2-D space of the UMAP throughout this study. Each supercluster consisted of multiple clusters and their numbers changed markedly depending on the resolution parameters (Fig. 1I). The mean number of genes detected per cell/nucleus in the clusters ranged between 2449 and 3010 in the late stage-5\_nucleus sample (except for cluster 9), between 3457 and 4125 in the stage-7\_nucleus sample (except for cluster 3), and between 2499 and 2827 in the stage-7\_cell sample (Fig. 1F–H). Only in the late stage-5\_nucleus cluster 9 and in the stage-7\_nucleus cluster 3 were the gene numbers lower than those in the other clusters, and the same trend was observed for the RNA counts. Furthermore, cluster markers for these and derived clusters were not specifically expressed, showing higher expression levels in other clusters or very faint expression in the respective clusters (Fig. 2, clusters 3, 13, and 15; Supplementary Fig. 4C). It is possible that these two clusters were pseudoclusters and were, therefore, excluded from further analyses.

The clusters of cells/nuclei in the UMAP plots were characterized by the expression of known marker genes (Table 2). The largest late stage-5\_nucleus supercluster consisted of the presumptive ectoderm, mesoderm (pMES and cMES cells), and extraembryonic cells, with a recapitulated AP order of gene expression domains observed in the germ disc (Fig. 1C) as previously reported<sup>62</sup>. The dorsal-inducing CM cells and endoderm cells formed a supercluster but were sharply separated from each other (Fig. 1C; Supplementary Fig. 4D). The largest stage-7\_nucleus supercluster consisted of only ectoderm cells, showing an elongated area of distribution with a recapitulated order of gene expression domains corresponding to the head-thorax-opisthosoma sequence (Fig. 1D). The higher the resolution, the more clusters emerged along the longitudinal axis of the supercluster (Fig. 1D, I). The largest stage-7\_cell supercluster consisted of only ectoderm cells with distinct clusters corresponding to the head, thorax, and opisthosoma but showed a less elongated distribution (Fig. 1E). No axial alignment of the ectoderm clusters was observed when the resolution was increased (Supplementary Fig. 5A). Two separate mesoderm superclusters were identified in the nucleus sample at stage 7 (Fig. 1D), whereas in the cell sample there were two separate mesoderm clusters (Fig. 1E), which were not subdivided into smaller clusters in clustering with higher resolution values (Supplementary Fig. 5A). Using an endoderm/extraembryonic cell marker (*At\_eW\_012\_A08*)<sup>36</sup>, two superclusters were identified in the single-nucleus sample (Fig. 1D) and one cluster in the single-cell sample (Fig. 1E). Overall, these results indicate that the cell states for the germ layers are well separated by stage 7 (51 h AEL), and that the axial alignment of the ectoderm clusters occurs in the 2-D space of the UMAP when using the nuclei dissociated from stage-7 embryos.

### Characterization of the axial alignment of ectoderm clusters in the single-nucleus UMAP plots

Next, we characterized the ectoderm cell array in the stage-7\_nucleus UMAP plots constructed at serial resolutions. Three, five, eight, eleven, and thirteen clusters were aligned along the axis at resolutions of 0.1, 0.5, 1.0, 2.5, and 3.3 plots, respectively (Fig. 1D, I). Of these resolutions, the resolution of 2.5 (Fig. 1I, the third plot and Fig. 2) was used for the comprehensive characterization of cell states, as the cluster number was comparable to the number of segmental units forming at stage 7. We identified marker genes for the ectoderm cell clusters at stage 7 (Supplementary Data 2) and stained



**Fig. 1 | Single-cell and single-nucleus analysis.** **A** Spider embryos at late stage 5 (39 h AEL) and stage 7 (51 h AEL). Microscopic images (top) and schematic illustrations showing characteristics of gene expression patterns in the germ disc and germ band at these stages. Bar = 200  $\mu$ m. **B** Illustration of scRNA-seq and snRNA-seq protocols. **C–E** UMAP plots of the late stage-5\_nucleus (C), stage-7\_nucleus (D), and stage-7\_cell (E) RNA-seq data. The clustering parameters of the dimension and the resolution are 1:45 and 1.4 (C), 1:50 and 0.5 (D), and 1:50 and 0.5 (E). Middle and right panels show the expression of mesodermal and endodermal

markers (middle) and body region markers (right). Gene names and colors are indicated. Endo endoderm, Ecto ectoderm, Meso mesoderm, pMES peripheral mesoderm, cMES central mesoderm, Ex extraembryonic tissue, CM cumulus mesenchymal cells. **F–H** Violin plots showing the number of genes (upper) and molecules (lower) detected per cell or nucleus of the clusters in the UMAP shown in (C–E). **I** Stage-7\_nucleus UMAP plots constructed with multiple resolution parameters, 0.1, 1.0, 2.5, and 3.3, at a dimension of 1:50.

the embryos for the transcripts of several selected markers (Fig. 3A). The staining revealed a perfect correlation between the AP positions of the gene expression domains in the germ band and the positions of the expressing cell population in the ectoderm cell array of the UMAP plot. This included the characteristic expression pattern of LOC107451631 (*Pt-osr-B*) in the two distinct domains. These observations indicate that the global polarity of the

embryonic AP axis in the ectoderm layer was reconstructed in the stage-7\_nucleus UMAP plot.

The aligned clusters of the stage-7\_nucleus and late stage-5\_nucleus datasets were examined and compared using the expression of the top-10 markers for each of the eleven stage-7 clusters and each of the seven late stage-5 clusters (Fig. 3B–G; Supplementary Data 2, 3). The late stage-5

**Table 1 | Summary of scRNA-seq and snRNA-seq**

Library	Stage-7 cell	Stage-7 nucleus	Late stage-5 nucleus
No. of embryos	15	20	20
No. of cells	865	6239	4820
Ave. genes/cell	2553	3456	2558
Ave. molecules/cell	9988	11,020	5561
Total genes	13,766	19,692	19,099

marker gene set revealed gradually differing cell states along the peripheral-central alignment of the late stage-5 germ disc (Fig. 3D)<sup>32</sup>. However, the same set of genes exhibited the position-dependent cell states less clearly in the stage-7 ectoderm, showing expression in multiple ectoderm clusters (Fig. 3E). Similarly, most of the stage-7 markers showed the highest expression levels in the respective clusters, but the levels were lower in the neighboring clusters as well as in more distant ones in the stage-7 data (Fig. 3G), which manifests gradually differing cell states along the axis. In the late stage-5 data, similar tendencies were recognizable in the expression of these markers but were much less prominent (Fig. 3F), with expression of some of them (e.g., markers of clusters 1, 6, 0, 7, and 10) showing very weak expression at late stage 5. These results indicate that the cell states at stage 7 are derived from those at late stage 5, and that newly expressed genes contribute to specifying the varied cell states during the development. In addition, the cell states at and near the stage-7 anterior and posterior termini (clusters 1/6 and clusters 4/7/10) were characterized by a relatively restricted expression of marker genes.

In contrast to the nucleus plots, the stage-7\_cell UMAP plot exhibited a less organized pattern of cell populations expressing the AP markers (Supplementary Fig. 5B, compared with Fig. 3A). The number of nuclei used for the stage-7\_nucleus plot ( $n = 6239$ ) was much larger than that used for the stage-7\_cell plot ( $n = 865$ ) (Table 1). To assess the effect of data volume on pattern reconstruction, we prepared 10 subsets of randomly selected 865 nuclei from the stage-7\_nucleus data and analyzed them for clustering. The resulting UMAP plots showed axially aligned ectoderm clusters in seven cases (trials 1–6, 8), but less prominent in the remaining cases (Supplementary Fig. 6). We could not conclude whether the inability to reconstruct the AP pattern using the stage-7\_cell dataset was due to insufficient data volume, the property of data derived from the cytoplasm, or both.

### Characterization of the mesoderm and endoderm clusters

We next characterized the cell states of the non-ectoderm clusters in detail using known and newly identified markers (Fig. 4; Supplementary Figs. 4, 7, 8). The late stage-5\_nucleus clusters 11, 12, and 13 formed the endoderm supercluster (Figs. 1C, 4A), and of these, cluster 13 was the CM cells (Supplementary Fig. 4D), as mentioned above. LOC107443634 (aug3\_g486)<sup>65</sup> marked only cluster 11, whereas LOC107436131 (the present study) marked both clusters 11 and 12 (Supplementary Fig. 4E). Multicolor fluorescence *in situ* hybridization (FISH) for these two genes revealed two cell states of endoderm cell populations beneath the germ disc (Supplementary Fig. 4F). The locations of the marked cells indicate that this difference in the cell state is not relevant to the classification of pEND and cEND cells, which were named after their birthplace (i.e., the peripheral and central sides of the germ disc<sup>36</sup>). The cells located on the surface and outside the germ disc area were also included in cluster 11 (Fig. 4A, B). The top-10 markers of cluster 12 were expressed not only in the cluster-12 cells but also in the cells of clusters 11 and 13, in contrast to the markers of clusters 11 and 13, which showed more specific expression (Supplementary Figs. 4C, 7C).

The stage-7\_nucleus clusters 12, 19, 21, and 24 were endoderm and extraembryonic cells (Figs. 1D, 2). Staining of embryos with specific markers revealed that the cells of cluster 12 were typical endoderm cells located between the surface layer and yolk mass, with a scattered distribution over the embryo (Fig. 4C, D; Supplementary Fig. 8A, B). This cluster was related to the late stage-5\_nucleus cluster 11, as was evident from the expression of

the top-10 markers (Supplementary Fig. 7D, E). The stage-7\_nucleus cluster 19 showed a slight similarity to the late stage-5\_nucleus endoderm clusters (clusters 11 and 12) in terms of marker expression (Supplementary Fig. 7D); however, staining of a specific marker revealed extraembryonic cells (Supplementary Fig. 8C). The stage-7\_nucleus clusters 21 and 24, forming a supercluster with the cluster 19, were determined to correspond to surface cells at and near the boundary between the embryonic and extraembryonic regions; cluster-24 cells were at the caudal terminal area and cluster-21 cells were at a more anterior area (Fig. 4D; Supplementary Fig. 8C, D). The cell states of these two clusters were similar to those of late stage-5\_nucleus cluster 6 (Supplementary Fig. 7D), which were emerging extraembryonic cells (Supplementary Fig. 4D).

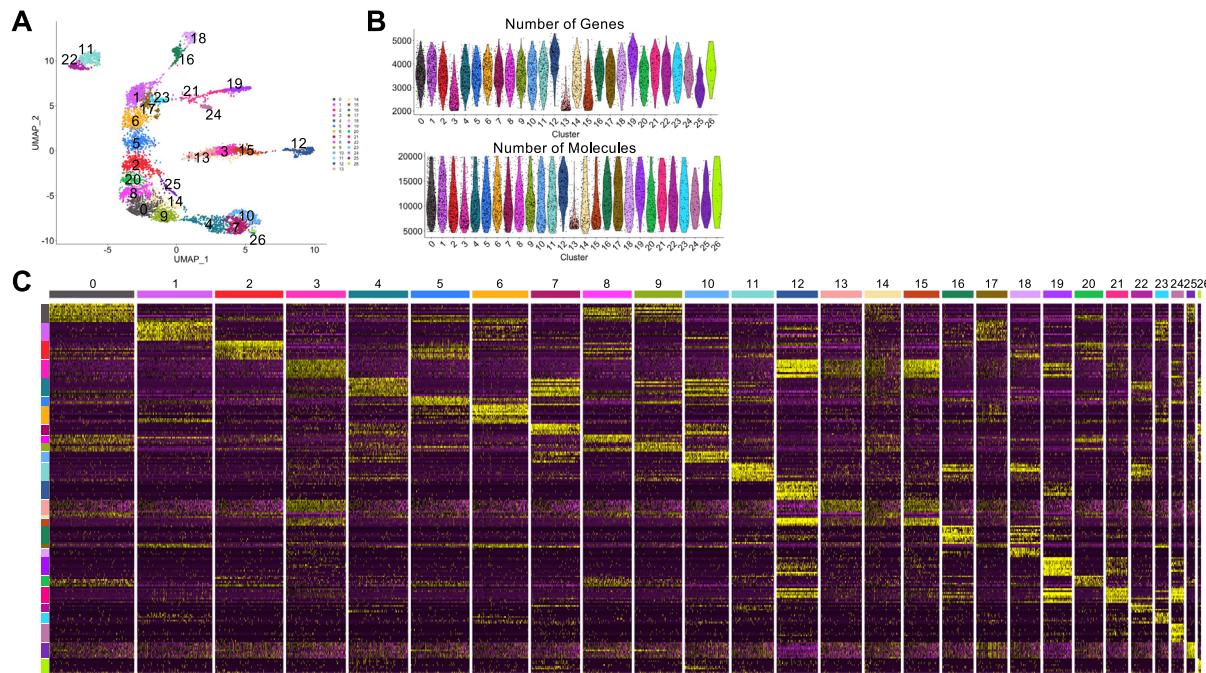
Two mesoderm superclusters in the stage-7\_nucleus UMAP consisted of two clusters each, namely clusters 16/18 and clusters 11/22 (Figs. 1D, 2). Staining of the marker gene expression revealed four subpopulations of the mesoderm, defined by the cell states of the clusters 16, 18, 11, and 22, in this order, from the anterior to the posterior (Fig. 4D; Supplementary Fig. 8E, F). The cell states of clusters 16 and 18 were relatively similar to those of late stage-5\_nucleus cluster 7, and the cell states of clusters 11 and 22 were relatively similar to those of late stage-5\_nucleus cluster 4 (Supplementary Fig. 7D), consistent with the previous descriptions of pMES and cMES cell development<sup>36</sup>.

### Detection of emerging segmental units in the reconstructed stage-7 ectoderm

As described above, the stage-7\_nucleus dataset reconstructed the axial alignment of the ectoderm nuclei (Fig. 3). Next, we examined whether the stage-7 datasets could also reconstruct the repeated stripes of gene expression associated with segmentation in the plots. To address this question, we visualized the cell populations expressing *Pt-msx1*, *Pt-hh*, and *Pt-noto1* in the stage-7\_nucleus plot in comparison to the same-stage embryos stained for their transcripts (Fig. 5A, B). This comparison showed that the periodic stripe pattern in the embryo was recapitulated in the plot, where the complementary patterns of *Pt-hh* and *Pt-msx1*, as well as those of *Pt-noto1* and *Pt-msx1* were observed along the axis. Notably, patterns corresponding to the bi-splitting stripes in the head, tri-splitting stripes in the thorax, and stripes formed by oscillatory expression in the posterior region were observed in the UMAP plot. In contrast, the stage-7\_cell data did not exhibit any segmental patterns (Fig. 5C).

### Generation of genome-wide gene expression profiles along the AP axis in the ectoderm

The reconstruction of the axial alignment of the ectoderm using the stage-7\_nucleus dataset prompted us to predict gene expression profiles in a genome-wide and quantitative manner. To this end, we developed a method for extracting gene expression levels along the reconstructed AP axis (Fig. 5D–G; Supplementary Fig. 9A). In this method, a spline curve running along the axial alignment of the ectodermal nuclei was set using a contour map of the nucleus density (Fig. 5D, E), and the distance between each nucleus and the spline curve was calculated (Supplementary Data 4). The expression levels of each gene in the nuclei within a threshold distance (Fig. 5E) were plotted against the position along the reconstructed axis, and the plotted data were further smoothed (Fig. 5F, G) to obtain a one-dimensional expression profile for each gene in the genome (Fig. 5H, I; Supplementary Data 5). The expression profile comprised the values at 80 subdivisions of the reconstructed axis, starting from the anterior end of cluster 1 to the posterior end of cluster 10, which were defined by the expression of *Pt-six3-1* and *Pt-prd2* (Supplementary Fig. 9A). The distant range used for selecting nuclei along the spline curve was optimized so that the generated expression profiles reflected real gene expression patterns. The generated expression profiles of *Pt-hh*, *Pt-msx1*, and *Pt-noto1* represented the characteristic peak patterns observed in the head, thorax, and opisthosoma (Fig. 5H). Further comparisons also supported the fact that the obtained expression profiles reproduced the complex patterns of gene expression along the AP axis in detail (Fig. 5I; Supplementary Fig. 9B).



**Fig. 2 | Stage-7\_nucleus clusters.** **A** UMAP plot of stage-7\_nucleus RNA-seq data generated with a dimension of 1:50 and a resolution of 2.5. **B** Violin plots showing the number of genes (upper) and molecules (lower) detected per nucleus of the

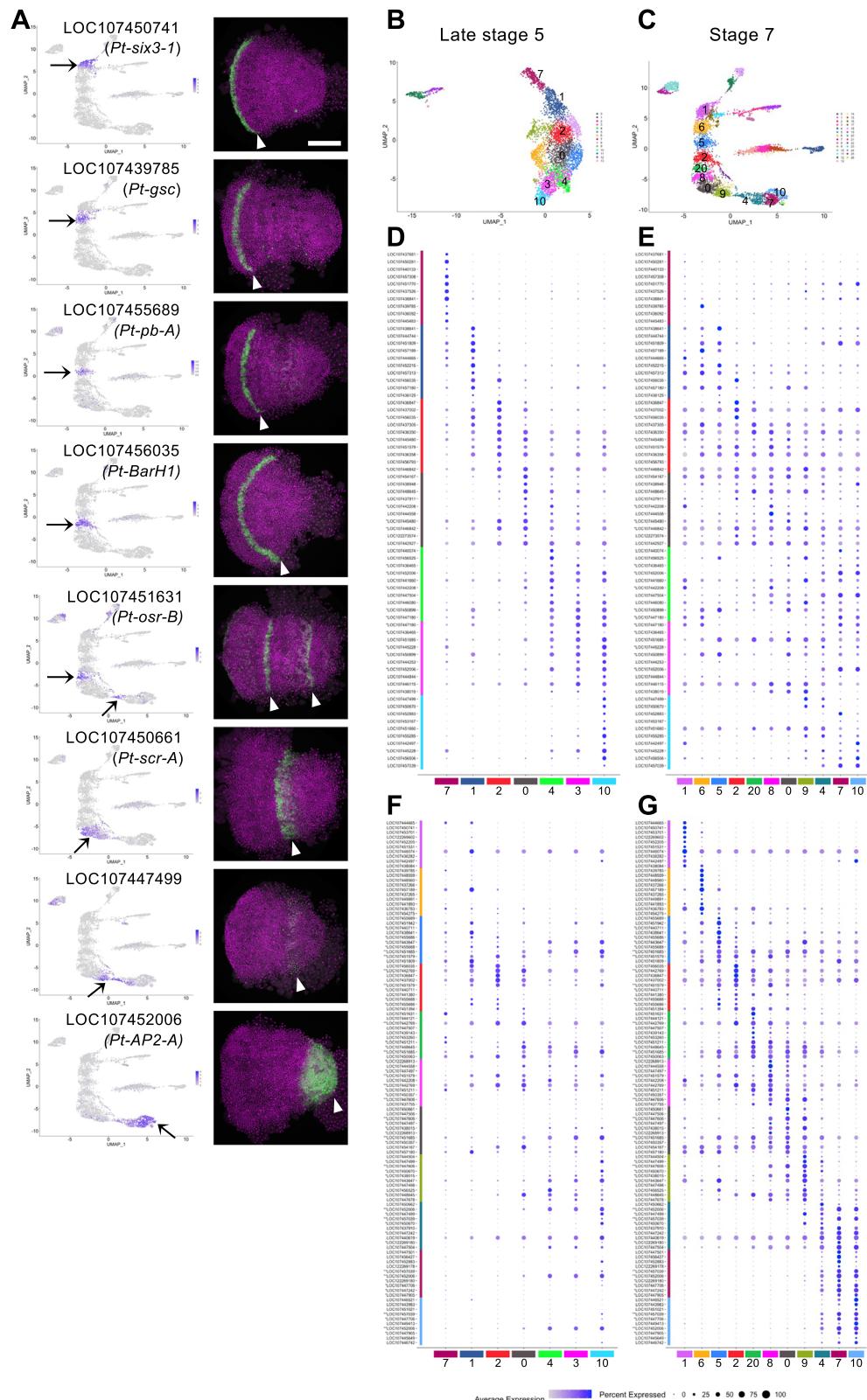
clusters in (A). **C** Heatmap showing expression levels (log-normalized) of markers in nuclei of each cluster. Top-10 markers of each cluster are shown, but those that have already appeared in earlier clusters are not shown.

**Table 2 | Marker genes used for the identification of cell types**

NCBI ID	Gene name	Late stage 5	Stage 7	Ref.
LOC107440133	<i>Pt-twist</i>	pMES	Mesoderm	36, 80
LOC107451427	<i>At_eW_012_A08</i>	Endoderm, CM cells	Endoderm, extraembryonic cells	36
LOC107456525	<i>Pt-Delta</i>	pEND, cMES	Stripes	36
LOC107451717	<i>Pt-ets4</i>	CM cells		65, 81
LOC107444265	<i>Pt-noggin-D</i>	Extraembryonic region		38
LOC107457189	<i>Pt-otd</i>	Periphery	Head	72
LOC107438948	<i>Pt-rhombo</i>	Intermediate region		38
LOC107438015	<i>Pt-gas1</i>	Center	Thorax	38
LOC107452006	<i>Pt-AP2-A</i>	Center	Opisthosoma	38
LOC107451809	<i>Pt-hh</i>	Periphery	Stripes	21, 36
LOC107447678	<i>Pt-msx1</i>	Center	Stripes	38
LOC107447988	<i>Pt-noto1</i>	Broad	Stripes	25
LOC107450741	<i>Pt-six3-1</i>		Anterior end	21
LOC107445228	<i>Pt-prd2</i>	Center, pMES	Stripes, posterior	38
LOC107452883	<i>Pt-aslH</i>	Posterior end	Stripes, posterior	38

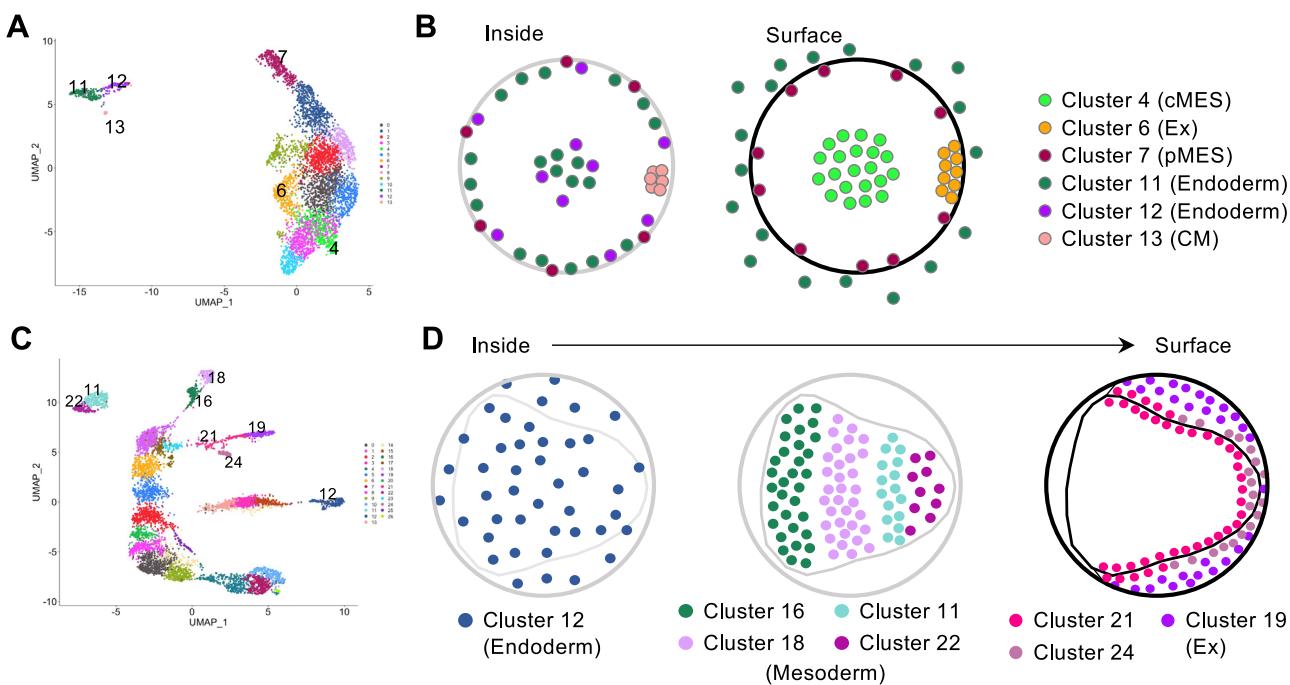
The same method was applied to the late stage-5\_nucleus dataset to obtain one-dimensional expression profiles along the radius of the germ disc, from the peripheral end of cluster 7 to the central end of cluster 10, which were defined by the expression of *Pt-twist* and *Pt-aslH* (Supplementary Fig. 9C, D). The late stage-5 expression profile also comprised values at 80 subdivisions, including the presumptive ectoderm and mesoderm. We confirmed that the predicted expression profiles of several selected genes recapitulated the actual expression patterns along the radius of the germ disc (Fig. 5); Supplementary Fig. 9E, Supplementary Data 6, 7). The genome-wide gene expression profiles predicted from the late stage-5\_nucleus and stage-7\_nucleus datasets are available in our database and in the figshare repository<sup>66,67</sup>.

**Genome-wide search for genes expressed in specific patterns**  
The quantitative expression profiles obtained allowed us to use a statistical method to perform a genome-wide search for genes that are expressed in specific patterns along the AP axis. We performed two rounds of hierarchical clustering of all genes expressed in the ectoderm supercluster using the mean and difference between the maximum and minimum values collected from the expression profiles (Fig. 6A). In the first round of clustering of 16,023 genes using the stage-7\_nucleus dataset, the genes were divided into 12 groups (Supplementary Fig. 10). From these, five groups were selected based on the following criteria: relatively small mean values and relatively large differences. In the second round, each group was further classified into two to five subgroups (Supplementary Data 8). Eleven



**Fig. 3 | Expression of stage-7 ectoderm cluster markers.** **A** Expression of stage-7 ectoderm cluster markers in the UMAP plots (left) and in stage-7 embryos detected using FISH (green, right). All embryos were counterstained with DAPI (purple). Arrows point to expression domains in the plot, and arrowheads to those in the embryo. Scale bar = 200  $\mu$ m. **B, C** UMAP plots of late stage-5 (B) and stage-7 (C) nucleus data. The clusters aligned along the reconstructed AP axis and examined in the following analysis are indicated by numbers. **D–G** Dot plots showing the

expression of top-10 markers of the late stage-5 ectoderm and mesoderm clusters 7, 1, 2, 0, 4, 3, and 10 (D, E) and those of the stage-7 ectoderm clusters 1, 6, 5, 2, 20, 8, 0, 9, 4, 7, and 10 (F, G) in the late stage-5 (D, F) and stage-7 (E, G) clusters. Average expression levels (log-normalized) and percent expressed in the nuclei in each cluster are shown by the colors and sizes of the circles. Asterisks indicate genes that appear multiple times in the list (\* twice, \*\* three times).



**Fig. 4 | Summary of mesoderm and endoderm clusters.** **A** Late stage-5 UMAP showing the position of the non-ectoderm clusters. **B** Schematic illustrations of the inside and surface layers of the late stage-5 germ disc showing cluster-11 and cluster-12 endoderm cells, cluster-7 pMES cells, cluster-4 cMES cells, cluster-13 CM cells, and cluster-6 extraembryonic cells. **C** Stage-7 UMAP showing the position of the

non-ectoderm clusters. **D** Schematic illustrations of the inside-to-surface layers of the stage-7 germ band showing cluster-12 endoderm cells, clusters-16, 18, 11, and 22 mesoderm cells, and clusters-21, 24, and 19 cells from the periphery of the embryo to the extraembryonic tissue.

subgroups were selected based on the same criteria (Fig. 6B). A total of 213 genes were selected (Supplementary Data 9). Although this simple clustering-based gene selection worked well, some genes known to be expressed with specific patterns along the AP axis in the stage-7 germ band [e.g., LOC107453167 (*Pt-kru*-1)] were not included in the 213 genes. These genes exhibited very low expression counts in the dataset.

Similar hierarchical clustering analyses were performed on the late stage-5 expression profiles of 15,061 genes, and ten subgroups containing 264 genes were selected (Supplementary Fig. 11, Supplementary Data 10, 11). The expression profiles of the 213 selected genes at stage 7 and 264 genes at late stage 5 showed a variety of patterns (Fig. 6B; Supplementary Fig. 11C).

#### Reconstruction of the AP pattern using the 213 selected genes

To examine the contribution of the selected genes to reconstructing the AP pattern, we prepared two submatrices from the normalized stage-7\_nucleus matrix. One submatrix was composed of 4102 ectoderm nuclei and the 213 genes identified in the above analysis (submatrix A), and the other was composed of the same nuclei and the full gene set, but excluding the 213 genes (submatrix B). Seurat analysis using submatrix A reconstructed the axial alignment of the nuclei, displaying the precise AP pattern with the repetitive stripe pattern (Fig. 6C; Supplementary Fig. 12A). However, the analysis using submatrix B failed to reconstruct the pattern (Fig. 6D; Supplementary Fig. 12B). These results indicate that the snRNA-seq count data of the 213 genes selected at stage 7 provided essential and sufficient information for mathematically reconstructing the AP pattern. Similarly, analysis of the late stage 5 dataset showed that the 264 selected genes contained rich information for reconstructing the AP pattern (Supplementary Fig. 11D).

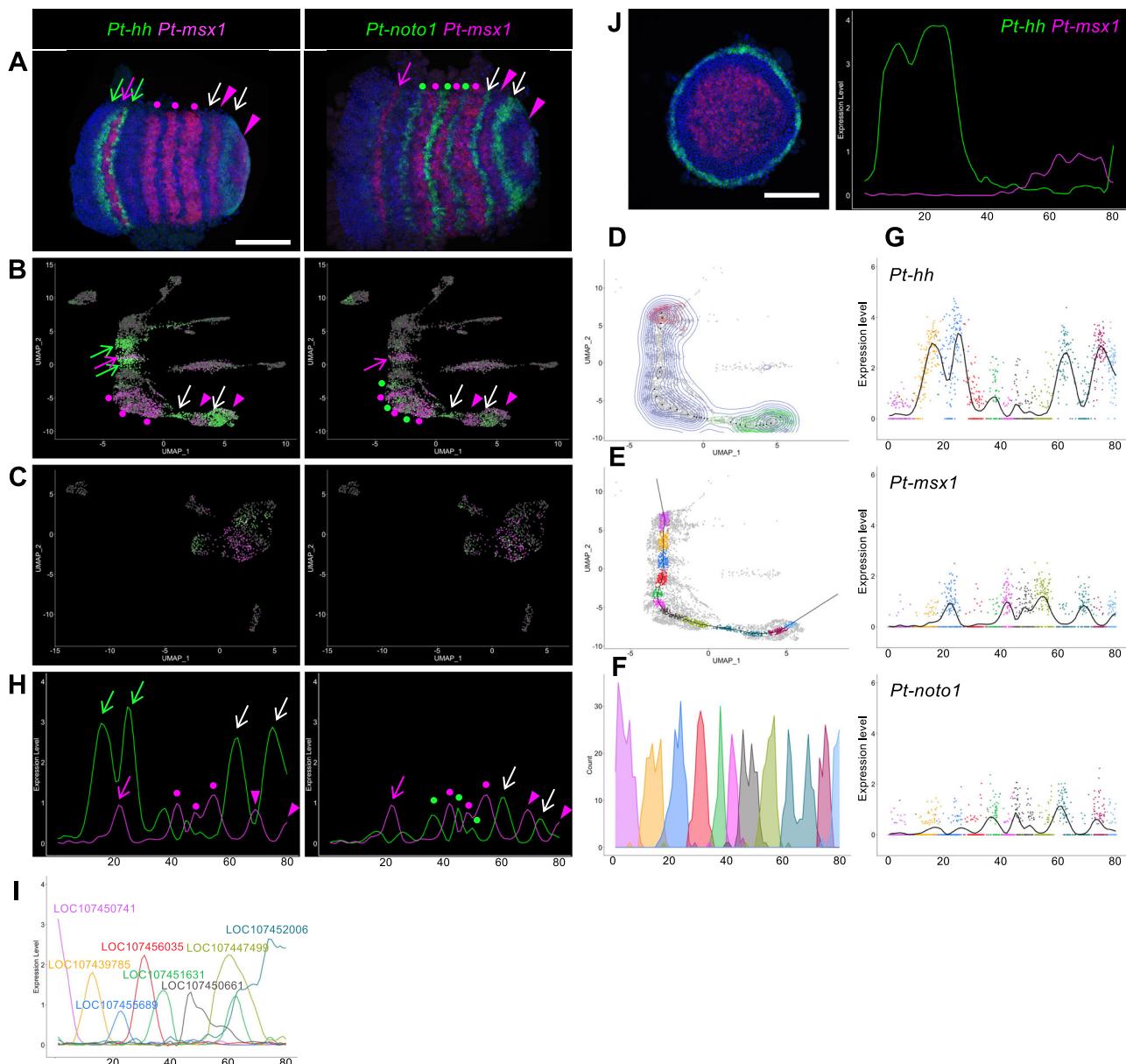
To investigate whether the structure of the recovered AP axis is dependent on the global or local contribution of each cluster, we modified the stage-7\_nucleus full matrix by randomly exchanging gene expression values among cells of a certain cluster and performed Seurat analysis. When the 213 genes were subjected to random exchange, the ectoderm alignment became discontinuous, specifically at the modified cluster, with the nuclei

closely gathered (Fig. 6E, F; Supplementary Fig. 13A). This reconstruction failure did not occur when using other 213 genes for random exchange (Fig. 6G; Supplementary Fig. 13B). We further changed the levels of the exchange. This experiment showed that the degree of discontinuity was dependent on the percentage of nuclei in a cluster subjected to exchange (Supplementary Fig. 13C). These results indicate that each small region possesses polarity, which contributes to the generation of the global AP pattern in the entire ectoderm at stage 7, and that the expression of the 213 genes mediates the mathematical reconstruction of the AP pattern.

#### Characterization of the AP axis using the changes in the expression profile

To quantitatively capture the features of the reconstructed AP pattern, we obtained the first-derivative values of the expression profiles of the 264 selected genes for late stage 5 (Supplementary Fig. 11C) and 213 genes for stage 7 (Fig. 6B) as a function of position (from 1 to 80) along the reconstructed AP axis (Supplementary Data 12, 13). The values of the first derivative of a graph correspond to the rate of change, which directly dictates the form of the graph. We postulated that the derivative values being zero and being local maxima/minima reflected the peaks of gene expression and the boundaries of the gene expression domains, respectively. At each AP position, the derivative values were sorted, and the top-50 positive and negative values were visualized using a heatmap (Fig. 7A, B). This revealed specific regions in which many genes exhibited sharp expression boundaries. There were two such regions in the late stage-5 heatmap (Fig. 7A), one near the anterior end within the anterior terminal cluster 7, including presumptive pMES cells, and the other in the posterior terminal cluster 10, corresponding to the region where the stripe formation for the posterior segments started<sup>62</sup>. In contrast, the stage-7 heatmap showed a wider and more complex distribution of regions, where many genes exhibited sharp expression boundaries (Fig. 7B), indicating the progression of AP pattern development.

Next, using the first-derivative values of the gene expression profiles, we calculated the correlation coefficients between the 80 subdivisions of the



**Fig. 5 | Detection of stripe pattern and generation of expression profile.**

A–C Expression of *Pt-hh* (left, green), *Pt-noto1* (right, green), and *Pt-msx1* (A red; B, C magenta) in the stage-7 embryos (A) and in the stage-7\_nucleus (B) and stage-7\_cell (C) UMAP plots. *Pt-hh* (green arrows) and *Pt-msx1* (magenta arrows) stripes in the head, *Pt-msx1* (magenta dots) and *Pt-noto1* stripes (green dots) in the thorax, and *Pt-hh* and *Pt-noto1* stripes in L4 and O1 segments (white arrows) and *Pt-msx1* stripes in O1 and O2 segments (magenta arrowheads) are indicated (A). Domains corresponding to these stripes are marked similarly (B). D–G Process for the generation of the expression profile. Contours based on the densities of ectoderm nuclei (blue) and of those expressing *Pt-six3-1* (red) and *Pt-prd2* (green), and points

plotted along the ridge (D). A formed spline curve (E) and distribution of nuclei along the linear axis (F). Plots of expression levels of *Pt-hh*, *Pt-msx1*, and *Pt-noto1* in the colored nuclei against the position along the axial line, overlaid with the calculated smoothing line (G). The color code in (E–G) corresponds to the cluster colors (Fig. 2A). H Expression profiles of *Pt-hh* and *Pt-msx1* (left) and *Pt-noto1* and *Pt-msx1* (right). The colors, arrows, and dots are the same as in (A). I Expression profiles of AP marker genes (Fig. 3A). J Expression of *Pt-hh* (green) and *Pt-msx1* (red) in a late stage-5 embryo (left) and the generated expression profile of these genes (right). Embryos (A, J) were counterstained with DAPI (blue), and scale bars = 200  $\mu$ m.

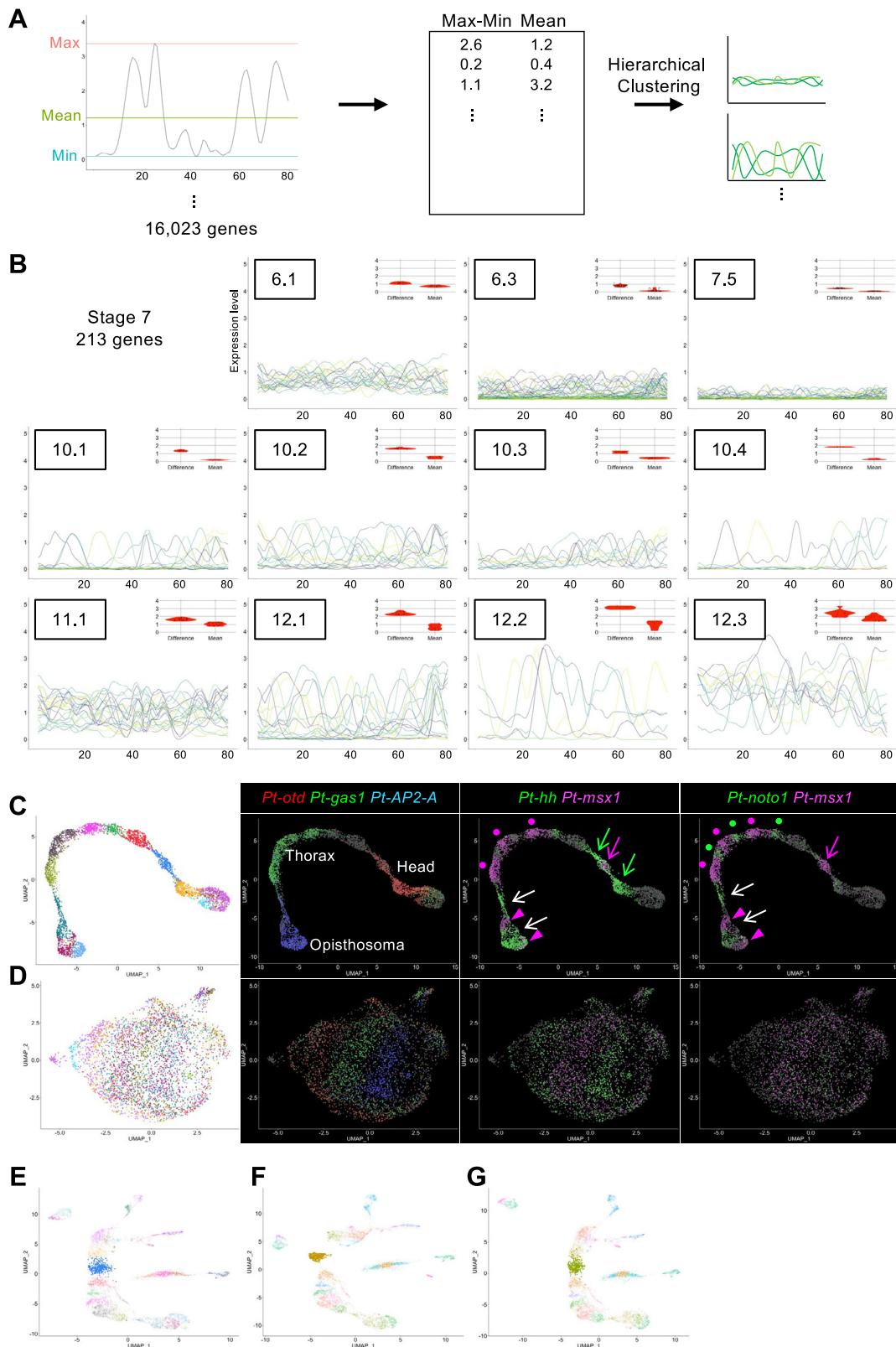
reconstructed AP axis to determine the correlated subdivisions, where a significant number of genes showed similar changes in expression along the AP axis. The resulting 2-D heatmap for the late stage-5 data showed an ambiguous pattern; in contrast, the 2-D heatmap for the stage-7 data exhibited a repetitive pattern of positively and negatively correlated regions (Fig. 7C–F). This repetitive nature was conspicuous in the middle and posterior parts of the reconstructed AP axis, matching the forming segmental units.

We further performed the same analysis using the stage-7 data after separating the 213 selected genes into two mutually exclusive subsets: subset a, composed of 105 genes that were also included in the 264 late stage-5

genes, and subset b, composed of the other 108 genes. The patterns obtained with these two subsets were clearly different: subset a yielded a repetitive pattern in the middle part of the reconstructed AP axis (Fig. 7G), and subset b in the posterior part (Fig. 7H). These results indicate that the genes expressed from before late stage 5 and those expressed after late stage 5 differentially served as repetitive pattern elements in the middle and posterior parts of the embryo.

#### Clustering of genes based on their local expression profiles

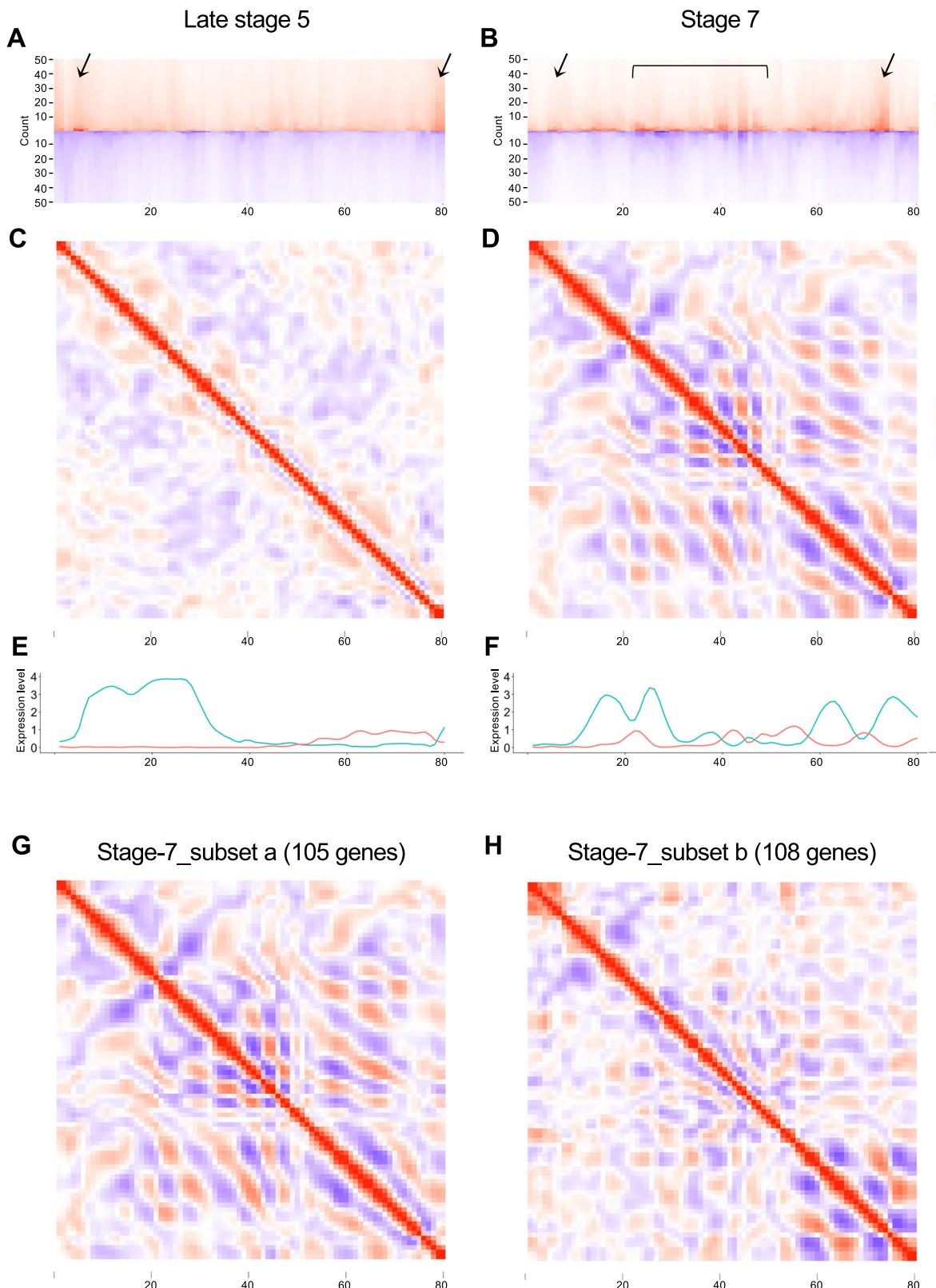
To identify genes with similar expression profiles within specific regions of the reconstructed stage-7 AP axis, we used the first-derivative values to



**Fig. 6 | Genome-wide search for genes with specific expression patterns.**

A Illustration of the search protocol. B Expression profiles of the 213 selected genes are visualized separately for each subgroup. Insets show violin plots of the difference and mean values of expression profiles of genes. C, D UMAP plots generated with submatrices A (C) and B (D). Expression patterns of body region markers and those

of *Pt-hh*, *Pt-msx1*, and *Pt-noto1* are also shown. Arrows and dots point to the expression corresponding to the stripes in Fig. 5A. E–G Randomization of the expression of the 213 selected genes in cluster 5. Single-nucleus UMAP plots showing nuclei of cluster 5 without randomization (E, blue) and with randomization of the 213 selected genes (F, brown) and of the other 213 genes (G, green).

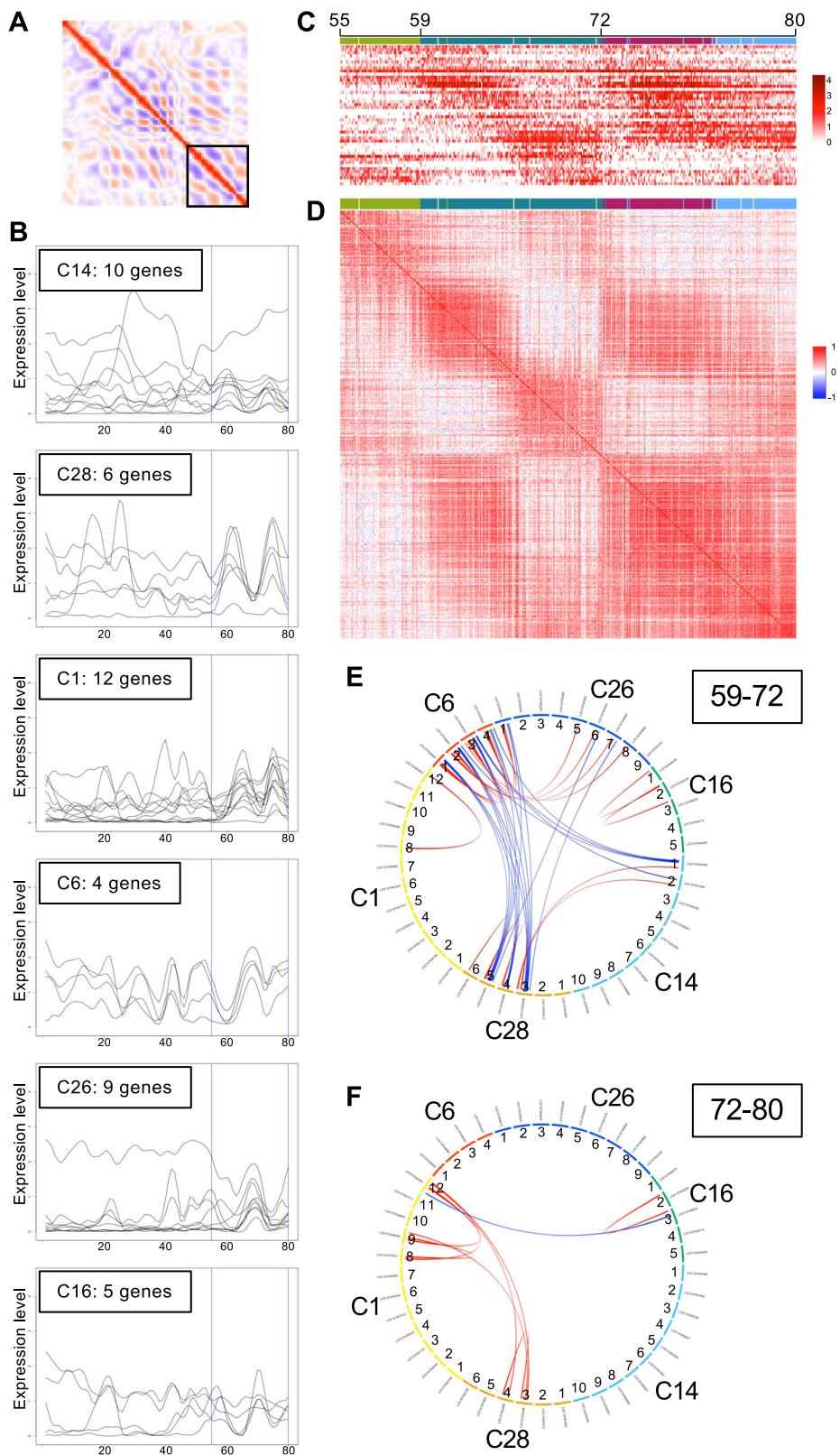


**Fig. 7 | Characterization of regions using first-derivative values.** A–F The first-derivative values of the 264 selected genes at late stage 5 and of 213 genes at stage 7 were sorted at each position, and the top 50 highest and lowest values are shown in red and blue, respectively (A, B). Arrows and parentheses indicate regions where high and low values are remarkable. 2-D plots of the correlation coefficients between

the 80 positions, calculated using the first-derivative values of the selected 264 and 213 genes (C, D). Expression profiles of *Pt-hh* (blue) and *Pt-msx1* (red) at late stage 5 (E) and stage 7 (F). G, H 2-D plots of the correlation coefficients between the 80 positions, calculated using the stage 7 first-derivative values of 105 (subset a) and 108 (subset b) genes.

**Fig. 8 | Grouping of genes in the posterior oscillatory region using first-derivative values.**

A Region C at positions 55–80 in the 2-D plot of the correlation coefficients. B Expression profiles of the 46 genes in the six groups. C Heatmap showing the expression levels of the 46 genes in the single nuclei in this region. Only the colored nuclei in Fig. 5E are shown. D 2-D plot of the correlation coefficients between the nuclei shown in (C), calculated using the expression levels of the 46 genes. The color bars at the top (C, D) show the cluster colors in Fig. 2. E, F Gene pairs with high correlation coefficients. Gene pairs with  $r > 0.5$  and  $r < -0.5$  are connected by red and blue lines, respectively. Correlation coefficients were calculated using the expression levels in the nuclei at positions 59–72 (E) and 72–80 (F).



group the 213 selected genes. We set three overlapping regions, positions 1–33 (region A), 28–58 (region B), and 55–80 (region C), which approximately corresponded to the head, thorax, and opisthosoma in the stage-7 embryos, respectively. For each region, the correlation coefficients were calculated between every pair of genes using the first-derivative values, which were subjected to hierarchical clustering analysis to generate

27, 33, and 29 groups of genes for regions A, B, and C, respectively (Supplementary Figs. 14–16, Supplementary Data 14). Many of these groups contained genes with clearly recognizable and similar peak patterns, some of which showed segmental periodicity intervals. For example, group 23 from the analysis of region A (abbreviated as group A23) contained *Pthh* and several other genes with similar bi-splitting peaks, and group A18

contained *Pt-msx1* and several other genes exhibiting similar peak positions between the *Pt-hh* bi-splitting peaks (Supplementary Fig. 14). Groups B3, B5, B10, and B29 contained genes with three or four peaks in region B (Supplementary Fig. 15). The peak positions were similar within each group but slightly different between the groups. Similarly, in region C, groups C1, C6, C14, C16, C26, and C28 contained genes with single or double peaks at distinct positions (Fig. 8A, B; Supplementary Fig. 16), which are candidates for the clock-like oscillatory genes. Among the groups mentioned here, we noted that some shared two or three genes; groups A23 and B29 shared LOC107444253 (*Hes-1-A*), LOC107447988 (*Pt-noto1*), and LOC107451809 (*Pt-hh*); groups A23 and C14 shared LOC107447988 (*Pt-noto1*) and LOC107451464 (*Pt-hairy*); groups B3 and C1 shared LOC107456088 (*Pt-mirr4*), LOC107457313 (*Pt-sox2A*), and LOC107451770; groups B5 and C6 shared LOC107436691, LOC107439303 (*hemicentin-2*), and LOC107453597 (*Pt-fgf8*); groups B5 and C26 shared LOC107444558 (*pax-5*) and LOC107447678 (*Pt-msx1*); groups B29 and C28 shared LOC107451809 (*Pt-hh*), LOC107452888 (*Pt-winged eye-B*), and LOC107455759 (*meteorin-like*). These detected shares of genes may provide clues for investigating the whole mechanisms underlying regional variations in pattern formation. These results demonstrate the utility of the first-derivative values of the expression profiles for classifying genes according to their expression profile patterns.

### Analysis of cell states associated with posterior stripe formation using the predicted clock-like genes

Oscillatory gene expression sequentially generates new segmental units in the posterior region of the *Parasteatoda* embryo, with each unit forming within approximately 5 h<sup>25</sup>. The posterior unit is less developed than the anterior unit. The expression profiles of *Pt-hh*, *Pt-msx1*, and *Pt-noto1* (Fig. 5) and positional correlation analyses (Fig. 7) indicated the presence of two segmental units within region C (positions 55–80) of the reconstructed stage-7 AP axis. Gene clustering analysis (Supplementary Fig. 16) identified six groups containing 46 genes with single or double peaks in this region (Fig. 8B). Using these 46 genes, we analyzed the individual cell states in the posterior region to uncover gene-regulatory interactions. Herein, genes are referred to as, for example, ‘C14 genes’ for genes in group C14 and ‘C14-1 gene’ for the first gene in group C14.

First, a heatmap was generated, which showed the expression levels of the 46 genes in individual nuclei in region C, which were aligned according to their positions along the reconstructed axis (Fig. 8C). This heatmap revealed heterogeneity even among the nuclei at similar positions. Nonetheless, we observed two repeated units in the aligned nuclei. The repeated units were more clearly detected in a 2-D heatmap of the correlation coefficients between the nuclei, based on the expression levels of the 46 genes (Fig. 8D); the anterior unit was approximately at positions 59–72, and the posterior unit was approximately at positions 72–80. Next, within each unit, the correlation coefficients between the 46 genes were calculated using the expression levels in the single nuclei, and gene pairs with positive and negative correlations ( $r > 0.5$  and  $r < -0.5$ ) were visualized (Fig. 8E, F). While the anterior unit (positions 59–72) exhibited negative correlations between the C6 and C28 genes and positive correlations within the C6 genes (Fig. 8E), these correlations, and most other correlations detected in the anterior unit, were not detected in the posterior unit (positions 72–80) (Fig. 8F). Instead, the posterior unit showed only one negative correlation between C1-12 and C16-3 and positive correlations among C1-8, C1-9, C1-12, C28-3, and C28-4 genes. These findings indicate that dynamic changes in gene-regulatory interactions occur during stripe formation in the posterior region. Notably, the C16-3 and C1-12 genes, corresponding to *Pt-Delta* and *Pt-eve*, oscillate in different phases in the posterior terminal region of the spider embryo<sup>19,36–38</sup>.

More comprehensive gene-to-gene correlation analyses using all the 213 genes across various regions of the reconstructed stage-7 AP axis (Supplementary Fig. 17, Supplementary Data 15) revealed distinct correlation patterns in different regions, suggesting region-specific genetic regulation during AP axis patterning in *Parasteatoda* embryos.

### Discussion

This study demonstrated that snRNA-seq data derived from stage-7 *Parasteatoda* embryos successfully reconstructed the alignment of cells along the AP axis in a UMAP plot. A faithful recapitulation of the stripe expression patterns was detected in this aligned representation. Reconstruction of an axial pattern using only data derived from scRNA-seq or snRNA-seq has not been reported in other animals. Our key observations indicate a polarized state of the ectoderm cells. Smaller clusters were aligned along the reconstructed AP axis with higher resolution parameters in the Seurat analysis (Fig. 11). The expression of the marker genes of each ectoderm cluster was also detected in neighboring clusters at lower levels (Fig. 3G). These results emphasize the gradual transition in gene expression along this axis. Randomized gene expression within the selected ectoderm cluster disrupted the axial array at the positions of the randomized clusters (Fig. 6F; Supplementary Fig. 13A). Collectively, these observations indicate that ectoderm cells are organized along the AP axis with distinct gene expression, establishing a polarized state that underlies the formation of the stripe pattern.

The successful reconstruction of AP polarity and recapitulation of the stripe pattern in the UMAP highlight a unique aspect of the *Parasteatoda* embryo. In contrast, applying scRNA-seq and snRNA-seq to *Drosophila* embryos effectively revealed differentiating cell states, such as mesoderm, endoderm, and nervous system cells<sup>54–56</sup>. However, cell states corresponding to the ectoderm stripes or (para)segments are detected only in the anterior and posterior regions, and ectoderm cells in the trunk region are primarily separated into odd- and even-numbered segments<sup>55</sup>. Reconstructing the *Drosophila* ectoderm pattern from single-cell data requires integration of information from stained embryos<sup>55,58,59</sup>. This difference in the detection of stripe patterns stems from several potential factors. The use of snRNA-seq in *Parasteatoda* and scRNA-seq in *Drosophila*<sup>55,56</sup> may have contributed to the observed differences. Sample collection methods may also have affected the results; *Drosophila* embryos were collected in 1.5- to 4-h time windows covering various stages<sup>54,56</sup>, while *Parasteatoda* sibling embryos were collected at essentially the same developmental stage. Furthermore, *Drosophila* embryogenesis involves simultaneous AP and dorsal–ventral (DV) patterning, leading to complex cell-state specifications, whereas *Parasteatoda* exhibits a preceding progression of AP patterning relative to DV patterning, simplifying the cell-state analysis along the AP axis<sup>68</sup>. These underscore the potential of the *Parasteatoda* embryo as an alternative model for dissecting the AP patterning.

Our analysis of stage-7 single-cell data revealed ectoderm superclusters corresponding to the head, thorax, and opisthosoma regions but failed to capture the precise polarity or stripe pattern. One major drawback of our scRNA-seq procedure, when using stage-7 spider embryos as the source, came from the low recovery rates of cells during sample preparation and library construction, although cell recovery was comparable between scRNA-seq and snRNA-seq when using late stage-5 embryos<sup>62</sup>. In general, snRNA-seq may detect fewer genes than scRNA-seq if the scale of the data is comparable<sup>69</sup>. Our snRNA-seq procedure, however, can have remarkably high recovery rates of nuclei from a small number of spider embryos, even at stage 7 (Table 1), while detecting more genes than scRNA-seq as a whole. The differences between scRNA-seq and snRNA-seq may have other aspects. The discrepancy in detecting the complex stripe pattern may reflect differences in sensitivity to temporal changes in transcription between the two techniques. Nascent pre-mRNAs captured in single-nucleus data may provide a more dynamic view of transcriptional changes than mature mRNAs captured in single-cell data. This hypothesis is supported by the observation that single-nucleus data obtained from late stage-5 sibling embryos clearly detected the cell states emerging from the germ disc center for posterior segmentation, which was not apparent in the single-cell data<sup>62</sup>. However, other factors may also contribute to these differences, including variations in the number of cells and nuclei, sample preparation procedures, and slight differences in developmental stage between the embryos used for the library construction.

Our analysis revealed several novel cell states. At least two endoderm cell states were detected at late stage 5 (Fig. 4A, B; Supplementary Fig. 4),

suggesting heterogeneity within the endoderm cell population beyond the previously described pEND and cEND cells<sup>36</sup>. These differences likely reflect the degree of differentiation. Despite originating from different regions (the center and periphery of the germ disc), stage-7 endoderm cells exhibited a relatively uniform cell state (Fig. 4C, D; Supplementary Fig. 8). In contrast, four distinct mesoderm clusters were identified at stage 7, reflecting differences along the AP axis. Further investigation is needed to determine whether mesoderm cells exhibit further AP polarization in later stages and to understand the potential interactions between the mesoderm and overlying ectoderm in shaping their respective cell fates. The classification of CM cells has been a subject of debate<sup>62,70,71</sup>. Our data showed a close proximity between CM cells and endoderm cells (Fig. 1C), indicating that CM cells may represent a specialized subpopulation of the endoderm.

We established a novel method to generate quantitative gene expression profiles along the reconstructed AP axis in a UMAP plot (Fig. 5; Supplementary Fig. 9). These profiles accurately reflected the gene expression patterns along the AP axis in the *Parasteatoda* embryo, including distinct phases of the segmental stripes. This approach enabled genome-wide, unbiased identification of genes exhibiting specific AP expression patterns. We identified 213 genes from stage-7\_nucleus data and 264 genes from late stage-5\_nucleus data that could be used to reconstruct the AP pattern, including the segmental stripe pattern at stage 7 (Fig. 6; Supplementary Fig. 11). Although this does not definitively prove that these genes directly establish the AP pattern during embryogenesis, it provides crucial clues for identifying the underlying molecular networks. By analyzing the first derivatives of the expression profiles, we identified gene groups with similar peak patterns (Supplementary Figs. 14–16), which are candidate genes involved in stripe formation. Furthermore, we examined gene expression in the posterior region at single-cell resolution to identify gene pairs that tend to be co-expressed or expressed in a mutually exclusive manner (Fig. 8E, F, Supplementary Fig. 17). This analysis provides valuable data for investigating potential gene interactions. These analytical approaches can be applied to embryos at different developmental stages. As the *Parasteatoda* spider lays hundreds of eggs simultaneously, we can analyze serial pools of embryos collected according to time-based staging. Embryos subjected to gene knockdown experiments can be analyzed similarly. These advantages of using spider embryos facilitate a further investigation of the dynamic cell-state regulation of the stripe-forming processes. The high-resolution, quantitative single-nucleus data generated in this study provide a robust foundation for a deeper understanding of the diverse mechanisms underlying the arthropod segmented body axis and offer a reliable connection to the mathematical modeling of cell-based patterning processes.

This study demonstrates that quantitative data representing key characters of an animal phylum can be obtained based on the genome of an organism without relying on information from established model organisms. This finding opens new avenues for exploring the origins of animal diversity.

## Methods

### Spider

Animal experiments were performed according to a protocol reviewed and approved by the Institutional Animal Care and Use Committee of JT Bio-History Research Hall (2020-5). We used laboratory stocks of the spider *Parasteatoda tepidariorum* (synonym, *Achaearanea tepidariorum*), maintained at 25 °C under 16-h light and 8-h dark photoperiods. The developmental stages of the embryos have been described previously<sup>72</sup>. Stage-7 embryos were collected 13 h after late stage 5, when the cumulus reached the rim of the germ disc. For single-cell and single-nucleus isolation and embryo fixation, eggs derived from one egg sac were used for each experiment; 20–120 eggs were dechorionated, and approximately half of the embryos that displayed a typical morphological appearance were selected. To ensure the reliability of the data, the development of 8–16 sibling embryos was carefully monitored in each experiment to confirm their normal development.

### Isolation of the cell and the nucleus

Spider embryos were dechorionated using commercial bleach for 4 min, and then rinsed with distilled water. Embryos were then transferred into CGBS-CMF [Chan and Gehring buffered saline<sup>73</sup>, Ca- and Mg-free: 55 mM NaCl, 40 mM KCl, 10 mM Tricine, pH 6.95] solution containing 0.01% (for cell) or 0.5% (for nucleus) bovine serum albumin (BSA: Nacalai Tesque, 01278-44). The vitelline membrane was manually removed with forceps. For single-cell isolation, devitellinized embryos were incubated in CGBS-CMF containing 1 mg/mL elastase solution (Worthington) at room temperature for 5 min. The cell suspension was then filtered through a 70-µm strainer (pluriStrainer-Mini, 43-10070) followed by two passes through a 10-µm strainer (pluriStrainer-Mini, 43-10010) with brief centrifugation. Cells were collected in a DNA LoBind tube (Eppendorf, 022431021). The elastase reaction was terminated by adding BSA to a final concentration of 1%.

Single-nucleus isolation was performed as previously described<sup>62</sup>. In brief, devitellinized embryos were resuspended in 400 µL homogenization buffer [250 mM sucrose, 10 mM Tris (pH 8.0), 25 mM KCl, 5 mM MgCl<sub>2</sub>, 0.1% Triton-X100, 0.1 mM DTT, 100 U/mL RNase inhibitor (SUPERaseIn RNase Inhibitor: Invitrogen, AM2694), and protease inhibitor at 1:100 dilution (Nacalai Tesque, 25955-11)]<sup>74</sup> and homogenized using a loose plastic pestle (Fisher scientific, 12-141-368) with 40 strokes. The homogenate was centrifuged, and the precipitate was resuspended in CGBS-CMF containing 0.5% BSA and 100 U/mL RNase inhibitor. The suspension was filtered through a 40-µm strainer (pluriStrainer-Mini, 43-10040) and washed with the same solution. Isolated single-cell and single-nucleus samples were stored with five volumes of CellCover (Anacite Laboratories, 800-050) at 4 °C until further use.

Fifteen and twenty embryos from different egg sacs were used for the single-cell and single-nucleus isolation, respectively. Concurrent with the cell and nucleus isolation, sibling embryos were fixed for subsequent staining.

### Construction and sequencing of libraries

scRNA-seq and snRNA-seq libraries were constructed using a BD Rhapsody WTA Amplification Kit (BD Biosciences, 665915) following the manufacturer's protocol with modifications. For scRNA-seq construction, the stored sample was centrifuged, resuspended in a 0.5× CGBS-CMF containing 0.01% BSA, and loaded onto a cartridge pre-rinsed with 0.5× CGBS-CMF. For snRNA-seq library construction, RNase inhibitor was added to the sample buffer at a final concentration of 100 U/mL, Proteinase K was added to the lysis buffer at a final concentration of 0.5 mg/mL, and the lysis step was extended to 5 min. Sequencing was performed using the Illumina HiSeqX Sequencing System with 150-bp paired-end reads. We obtained 435,917,554 and 485,625,106 raw reads for scRNA-seq and snRNA-seq, respectively. The data were deposited in the GEO repository (GSE287446).

### Generation of gene expression matrices

RNA-seq reads were processed using the BD Rhapsody WTA pipeline on the Seven Bridges Genomics Cloud platform<sup>75</sup> with default parameters, as described previously<sup>62</sup>. Low-quality reads were then filtered. Cells/nuclei and RNA molecules were identified based on the cell barcode and unique molecular index (UMI) information contained within the R1 reads. The R2 reads were aligned against the *Parasteatoda tepidariorum* genome<sup>45</sup>, GCF\_000365465.3\_Ptep\_3.0, using the STAR index created by STAR-2.5.2b<sup>76</sup>. For snRNA-seq data, the annotation file accompanying the genome information was modified to include both exonic and intronic reads for read counting. The alignment results were integrated with the cell barcode and UMI information to generate an expression matrix for each sample. These expression matrices were deposited in the GEO repository associated with the RNA-seq data.

### Clustering analysis with Seurat

Data were analyzed using the Seurat package (v.4.1.1)<sup>77,78</sup>. To filter out low-quality cells and nuclei and potential doublets, we selected cells and nuclei

based on the number of unique genes and total RNA molecules. Cells with more than 1500 unique genes and fewer than 20,000 total RNA molecules were selected for stage-7\_cell data. Similarly, nuclei with more than 2000 unique genes and fewer than 20,000 total RNA molecules were selected for stage-7\_nucleus data. The late stage-5\_nucleus data<sup>62</sup> were reprocessed, selecting nuclei with more than 1500 unique genes and fewer than 15,000 total RNA molecules (Supplementary Fig. 2). During the filtering, we removed 681 nuclei from the late stage-5\_nucleus data, 152 cells from the stage-7\_cell data, and 2182 nuclei from the stage-7\_nucleus data. This resulted in the final cell/nuclei counts presented in Table 1.

Following the instructions provided with the Seurat package, we performed log-normalization, identified 2000 variable features, scaled the data, and conducted principal component (PC) analysis to reduce the dimensionality of the data to 50. The processed data were then subjected to FindNeighbors and FindClusters functions, followed by nonlinear dimension reduction using UMAP (Uniform manifold approximation and projection). The data were visualized in a 2-D space (Fig. 1C–E). Multiple parameter settings for the dimension and resolution were tested (Fig. 1I; Supplementary Figs. 3, 5). Additionally, we performed Seurat analyses on the pre-filtered dataset and observed that the excluded cells/nuclei predominantly localized to clusters exhibiting lower counts in the UMAP (Supplementary Fig. 2). This observation suggests that the filtering step had minimal effects, if any, on the downstream analyses.

### Random selection of nuclei from the expression matrix

To analyze the stage-7\_nucleus data with a sample size equivalent to that of the stage-7\_cell data (865 nuclei), 865 nuclei were randomly selected from the full stage-7\_nucleus matrix using the sample function in R. A submatrix was generated using these selected nuclei. Subsequently, Seurat clustering analysis was performed on this submatrix with the following parameters: dimension, 1:50; resolution, 2.5. This random selection and subsequent analysis were independently repeated 10 times (Supplementary Fig. 6).

### Identification of cluster marker genes

The marker genes for each cluster were identified using the FindMarkers function in Seurat. Genes expressed in more than 25% of nuclei in a cluster and less than 90% in other clusters with adjusted *p*-values less than 0.001 were listed as markers for the cluster (Supplementary Data 2, 3).

### Visualization of gene expression on the UMAP plot

Gene expression was visualized on the UMAP plot using the FeaturePlot function of the Seurat package. To visualize the expression of two or three genes simultaneously, feature plots were generated using a gray75 to black color scale. These plots were then converted into negative images, overlaid, and color-coded using RGB values in ImageJ (Fiji) software (ver. 2.9.0/1.53t). UMAP plots depicting the expression of all genes in the expression matrices can be accessed through our web database and have been deposited in Figshare<sup>66,67</sup>.

### Assignment of nucleus positions along the reconstructed axis

In the UMAP plot of the stage-7\_nucleus RNA-seq data generated with the clustering parameters, a dimension of 1:50 and a resolution of 0.5 (Fig. 1D), a contour encompassing ectoderm clusters 0, 1, 2, 4, and 5 was drawn based on the density of 4129 nuclei using the geom\_density\_2d function in the ggplot2 package (v.3.4.2). Contours were also drawn for nuclei expressing *Pt-six3-1* and *Pt-prd2*; the former was expressed at the most anterior part of the germ band, and the latter was expressed in oscillatory stripes in the posterior region (Supplementary Fig. 9A). These were used to determine the anterior and posterior ends. Using Python (v.3.8.3), several points were manually plotted along the ectoderm contour ridge and at the *Pt-six3-1* and *Pt-prd2* contour peaks. Two additional points were plotted on the extension lines from the anterior and posterior termini. These extensions were used to correctly calculate the distances from the nuclei to the reconstructed axis around the anterior and posterior ends, as described below. A spline curve

was fitted through these points using the xspline function of R (v.4.2.1), which represents the reconstructed AP axis (Fig. 5E).

To assign the position of plotted nuclei, the axial spline curve was evenly divided into 10,000 points using Numpy (v.1.22.4). The positions of these points along the curve were determined by summing the distances between the neighboring points from the anterior end. The nearest point on the curve was identified for each nucleus in the UMAP plot. The position of each nucleus was assigned as the position of its nearest point on the curve and the distance to the nearest point. Nuclei located more anterior or posterior than the ends of the spline curve were omitted.

A similar analysis was performed on the UMAP plot of the late stage-5\_nucleus data (dimension: 1:45; resolution: 0.5) (Supplementary Fig. 9D) using 4230 nuclei from ectoderm, mesoderm, and extraembryonic clusters 0, 1, 2, 3, 5, and 6. The anterior and posterior ends were defined based on the expression of *Pt-twist* and *Pt-aslH*. Nuclei located near the anterior end and not expressing *Pt-twist* were omitted. The positions of the nuclei are listed in Supplementary Data 4.1 and 6.1.

### Generation of gene expression profiles

Nuclei within 0.3 units of the axial curve were selected for generating quantitative gene expression profiles (shown in Fig. 5E and Supplementary Fig. 9D, colored nuclei), specifically 1387 nuclei at stage 7 and 1138 nuclei at late stage 5 (Supplementary Data 4.2, 6.2). To mitigate the potential impact of the limited number of nuclei near the anterior and posterior ends on the subsequent smoothing step, the data were mirrored: nuclei from the anterior half were mirrored at the anterior end, and those from the posterior half were mirrored at the posterior end, resulting in a doubled expression matrix. Loess smoothing was performed with the geom\_smooth function in ggplot2 (span = 0.05) using the nuclear positions in the doubled matrix and their normalized gene expression counts. This generated a smoothed line with 160 points for the expression profile of each gene. The mirrored data were then reduced to 80 points by selecting points 41–120 for downstream analyses. Gene expression profiles at late stage 5 and stage 7 are shown in Supplementary Data 7 and 5. The patterns of the expression profiles are available in our web database and the Figshare repository<sup>66,67</sup>.

### Search for genes with specific expression patterns along the AP axis

First, to exclude genes expressed in small numbers of sparsely distributed cells, which could be difficult to distinguish from genes expressed in a region-specific manner, we determined the minimum distance between nuclei expressing each gene along the reconstructed AP axis. Based on these minimum distances, Euclidean distances were calculated between genes, and hierarchical clustering was performed using ward.D2 method, resulting in two gene groups. A small group of 946 sparsely expressed genes was excluded from further analysis (Supplementary Data 16), and a larger group of 16,023 genes was retained for subsequent analyses.

For each of the 16,023 genes, two values were extracted from their expression profiles: the mean expression level and the difference between the maximum and minimum expression (Supplementary Data 8). The Euclidean distances between genes were calculated based on these two values, and hierarchical clustering was performed. Several groups were selected based on relatively low mean expression levels and large differences in expression. Subsequently, genes within each selected group were subjected to a second round of clustering analysis. This process resulted in the selection of 213 genes at stage 7 (Supplementary Data 9). Additionally, we conducted the second-round clustering for other unselected groups, which resulted in forming subgroups that did not match the selection criteria (Supplementary Data 8). A similar analysis was performed on the late stage-5 data. After excluding 1149 sparsely expressed genes (Supplementary Data 17), the remaining 15,061 genes were subjected to two rounds of clustering to select 264 genes (Supplementary Data 10, 11).

## Construction of UMAP with the selected genes

The normalized stage-7\_nucleus expression matrix containing 19,692 genes and 6239 nuclei was reduced to two smaller matrices: matrix A, containing the 213 selected genes and 4102 ectoderm nuclei, and matrix B, containing the remaining 19,479 genes and the same 4102 ectoderm nuclei. Similarly, two matrices were generated from the late stage-5\_nucleus data: one containing 264 selected genes and 4230 nuclei, and the other containing 18,835 remaining genes and 4230 nuclei. These reduced matrices were then used for subsequent Seurat analyses (Fig. 6C, D; Supplementary Figs. 11D, 12). To visualize the expression of the excluded genes (the 213 and 264 genes, respectively) in the UMAP, their expression values were extracted from the original (unreduced) matrix using the IDs of the corresponding nuclei.

## Randomization of the expression of genes in a cluster

To randomize the expression of the 213 genes within the selected ectoderm clusters in the stage-7\_nucleus count matrix, we followed the procedure described below. A set of random numbers, equal to the number of nuclei in the cluster, was generated using the sample function in R. This process was repeated to create 213 sets of random numbers corresponding to the 213 genes. Based on each set of random numbers, the expression values of a specific gene were randomly exchanged among nuclei within the selected cluster. In the first experiment, randomization was performed for each ectoderm cluster using the selected 213 genes (Fig. 6E, F; Supplementary Fig. 13A). In the second experiment, randomization was performed for cluster 5 using the other 213 genes randomly selected from among the highly variable genes identified in the Seurat analysis. This experiment was repeated three times (Fig. 6G; Supplementary Fig. 13B). In the third experiment, randomization was performed for cluster 5 using the original 213 genes. However, in this experiment, the expression values were exchanged among a subset of cluster-5 nuclei (10–90% of the cluster, selected independently for each gene). Randomized matrices were used for subsequent Seurat analyses (Supplementary Fig. 13C).

## Calculation of the first-derivative value and correlation analysis

First-derivative values were calculated for each of the 80 positions in the expression profiles of the 213 stage-7 and 264 late stage-5 genes predicted to exhibit specific expression patterns (Supplementary Data 12, 13). For positions 2–79, the derivative was calculated as the average difference in expression levels between the current position and its two neighboring positions. For the terminal positions (1 and 80), the derivative was calculated as the difference between the current position and its single neighboring position (positions 2 and 79, respectively). The top 50 highest and lowest first-derivative values at each position are shown in Fig. 7A, B.

Using the first-derivative values of all these genes at 80 positions, Pearson correlation coefficients were calculated between the 80 positions using the cor function in R. These correlation coefficients were visualized in a heatmap (Fig. 7C, D). Similarly, the correlation coefficients calculated with subsets of stage-7 genes are shown in Fig. 7G, H.

## Grouping of genes using the first-derivative values

To group genes with similar expression profile patterns within the three regions (region A, positions 1–33; B, 28–58; and C, 55–80), the following approach was used: the 213 stage-7 genes were filtered to exclude those with the maximum values of the expression profiles less than 0.2 within the target region. Pearson correlation coefficients were calculated between genes based on their first-derivative values within the target region. Euclidean distances were then calculated from the correlation coefficients, and hierarchical clustering was performed using the ward.D2 method to group the genes. This resulted in 27 groups in region A, 33 groups in region B, and 29 groups in region C (Supplementary Figs. 14–16).

## Expression in single nuclei and correlation analysis

From the normalized expression matrix of the stage-7\_nucleus data, we selected the colored nuclei depicted in Fig. 5E, along with the 213 genes predicted to exhibit specific patterns. Several submatrices were generated

using this selected matrix. Each submatrix included nuclei located within a specific target region and genes expressed in more than 10% of the nuclei within the target region. For each submatrix, the correlation coefficients between cells and between genes were calculated using the cor function in R. Gene pairs with relatively high correlation coefficients ( $r < -0.5$  and  $r > 0.5$ ) were visualized by connecting them with blue and red lines, respectively (Fig. 8E, F; Supplementary Fig. 17C).

## cDNA cloning

Full-length or partial cDNAs were obtained from our laboratory stocks of EST clones or isolated using PCR. The cDNAs were used to synthesize probes for in situ hybridization. The cDNA clones and the PCR primers used in this study are listed in Supplementary Data 18.

## Fluorescence in situ hybridization

Embryo fixation and FISH were performed as described previously<sup>79</sup>. DIG-, DNP-, and FITC-labeled probes (Table 2; Supplementary Data 18)<sup>80,81</sup> were used and detected with anti-DIG-POD (1:1000 dilution; Roche 11 207 733 910), anti-DNP-HRP (1:200 dilution, PerkinElmer FP1129), and anti-FITC-HRP (1:200 dilution, PerkinElmer NEF710) antibodies in combination with 5-(and-6)-carboxyfluorescein and DyLight680 tyramides. Samples were counterstained with 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich) to visualize DNA. The stained embryos were individually mounted on glass slides with spacers and observed using a TCS SPE confocal system (Leica). Images were processed using ImageJ (Fiji) software ver. 2.14.0/1.54 f. Z-sections were projected using Max projection.

## Statistics and reproducibility

For all statistical analyses, the methods used are addressed in the relevant sections. For sample collection, embryos with a typical morphological appearance were randomly selected. For single-cell and single-nucleus data analyses, cells/nuclei expressing genes/molecules within the range of numbers specified in the “Methods” section were retained.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The stage-7 scRNA-seq and snRNA-seq reads and processed data have been deposited in GEO under accession number GSE287446. The late stage-5 snRNA-seq reads and the original matrix are available at GEO under accession number GSM6069861 in the data series GSE201705<sup>62</sup>, and the reprocessed late stage-5 data are available in GSE287446. Gene expression patterns in the UMAP and patterns of the expression profiles have been deposited in figshare<sup>66,67</sup> and are also available on our website (<https://www.brh2.jp>). Material requests should be addressed to [yasuko@brh.co.jp](mailto:yasuko@brh.co.jp).

## Code availability

The scripts used for analyzing the data and producing graphs and figures are available at <https://github.com/Takanori-Akaiwa>.

Received: 1 March 2025; Accepted: 3 June 2025;

Published online: 11 June 2025

## References

1. Peel, A. D., Chipman, A. D. & Akam, M. Arthropod Segmentation: beyond the *Drosophila* paradigm. *Nat. Rev. Genet.* **6**, 905–916 (2005).
2. Clark, E., Peel, A. D. & Akam, M. Arthropod segmentation. *Development* **146**, dev170480 (2019).
3. Cooke, J. & Zeeman, E. C. A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *J. Theor. Biol.* **58**, 455–476 (1976).
4. Meinhardt, H. Hierarchical inductions of cell states: a model for segmentation in *Drosophila*. *J. Cell Sci. Suppl.* **4**, 357–381 (1986).

5. Lacalli, T. C. Modeling the *Drosophila* pair-rule pattern by reaction-diffusion: gap input and pattern control in a 4-morphogen system. *J. Theor. Biol.* **144**, 171–194 (1990).
6. Palmeirim, I., Henrique, D., Ish-Horowicz, D. & Pourquié, O. Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* **91**, 639–648 (1997).
7. Fujimoto, K., Ishihara, S. & Kaneko, K. Network evolution of body plans. *PLoS ONE* **3**, e2772 (2008).
8. Lauschke, V. M., Tsiairis, C. D., François, P. & Aulehla, A. Scaling of embryonic patterning based on phase-gradient encoding. *Nature* **493**, 101–105 (2013).
9. Cotterell, J., Robert-Moreno, A. & Sharpe, J. A local, self-organizing reaction-diffusion model can explain somite patterning in embryos. *Cell Syst.* **1**, 257–269 (2015).
10. Clark, E. Dynamic patterning by the *Drosophila* pair-rule network reconciles long-germ and short-germ segmentation. *PLoS Biol.* **15**, e2002439 (2017).
11. Diaz-Cuadros, M. et al. In vitro characterization of the human segmentation clock. *Nature* **580**, 113–118 (2020).
12. Matsuda, M. et al. Recapitulating the human segmentation clock with pluripotent stem cells. *Nature* **580**, 124–129 (2020).
13. Diaz-Cuadros, M., Pourquié, O. & El-Sherif, E. Patterning with clocks and genetic cascades: segmentation and regionalization of vertebrate versus insect body plans. *PLoS Genet.* **17**, e1009812 (2021).
14. Miao, Y. et al. Reconstruction and deconstruction of human somitogenesis in vitro. *Nature* **614**, 500–508 (2023).
15. Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).
16. Akam, M. The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* **101**, 1–22 (1987).
17. Johnston, D. S. & Nüsslein-Volhard, C. The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201–219 (1992).
18. Damen, W. G. M., Weller, M. & Tautz, D. Expression patterns of *hairy*, *even-skipped*, and *runt* in the spider *Cupiennius salei* imply that these genes were segmentation genes in a basal arthropod. *Proc. Natl. Acad. Sci. USA* **97**, 4515–4519 (2000).
19. Stollewerk, A., Schoppmeier, M. & Damen, W. G. M. Involvement of Notch and Delta genes in spider segmentation. *Nature* **423**, 863–865 (2003).
20. Choe, C. P., Miller, S. C. & Brown, S. J. A pair-rule gene circuit defines segments sequentially in the short-germ insect *Tribolium castaneum*. *Proc. Natl. Acad. Sci. USA* **103**, 6560–6564 (2006).
21. Kanayama, M. et al. Travelling and splitting of a wave of hedgehog expression involved in spider-head segmentation. *Nat. Commun.* **2**, 500 (2011).
22. Sarrazin, A. F., Peel, A. D. & Averof, M. A segmentation clock with two-segment periodicity in insects. *Science* **336**, 338–341 (2012).
23. Green, J. & Akam, M. Evolution of the pair rule gene network: Insights from a centipede. *Dev. Biol.* **382**, 235–245 (2013).
24. Rosenberg, M. I., Brent, A. E., Payne, F. & Desplan, C. Dual mode of embryonic development is highlighted by expression and function of *Nasonia* pair-rule genes. *eLife* **3**, e01440 (2014).
25. Hemmi, N., Akiyama-Oda, Y., Fujimoto, K. & Oda, H. A quantitative study of the diversity of stripe-forming processes in an arthropod cell-based field undergoing axis formation and growth. *Dev. Biol.* **437**, 84–104 (2018).
26. Thomas, G. W. C. et al. Gene content evolution in the arthropods. *Genome Biol.* **21**, 15 (2020).
27. Hakeem, M. S. et al. Screens in fly and beetle reveal vastly divergent gene sets required for developmental processes. *BMC Biol.* **20**, 38 (2022).
28. Aase-Remedios, M. E. Evolution of the spider homeobox gene repertoire by tandem and whole genome duplication. *Mol. Biol. Evol.* **40**, msad239 (2023).
29. Cheatile Jarvela, A. M. & Wexler, J. R. Advances in genome sequencing reveal changes in gene content that contribute to arthropod macroevolution. *Dev. Genes Evol.* **233**, 59–76 (2023).
30. Sharma, P. P. The impact of whole genome duplication on the evolution of the arachnids. *Integr. Comp. Biol.* **63**, 825–842 (2023).
31. Driever, W. & Nüsslein-Volhard, C. A gradient of bicoid protein in *Drosophila* embryos. *Cell* **54**, 83–93 (1988).
32. Driever, W., Thoma, G. & Nüsslein-Volhard, C. Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen. *Nature* **340**, 363–367 (1989).
33. Kanayama, M., Akiyama-Oda, Y. & Oda, H. Early embryonic development in the spider *Achaearanea tepidariorum*: Microinjection verifies that cellularization is complete before the blastoderm stage. *Arthropod Struct. Dev.* **39**, 436–445 (2010).
34. Pechmann, M. Formation of the germ-disc in spider embryos by a condensation-like mechanism. *Front. Zool.* **13**, 35 (2016).
35. Akiyama-Oda, Y. & Oda, H. Cell migration that orients the dorsoventral axis is coordinated with anteroposterior patterning mediated by Hedgehog signaling in the early spider embryo. *Development* **137**, 1263–1273 (2010).
36. Oda, H. et al. Progressive activation of Delta-Notch signaling from around the blastopore is required to set up a functional caudal lobe in the spider *Achaearanea tepidariorum*. *Development* **134**, 2195–2205 (2007).
37. Schönauer, A. et al. The Wnt and Delta-Notch signalling pathways interact to direct pair-rule gene expression via *caudal* during segment addition in the spider *Parasteatoda tepidariorum*. *Development* **143**, 2455–2463 (2016).
38. Akiyama-Oda, Y. & Oda, H. Hedgehog signaling controls segmentation dynamics and diversity via *msx1* in a spider embryo. *Sci. Adv.* **6**, eaba7261 (2020).
39. Small, S., Blair, A. & Levine, M. Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J.* **11**, 4047–4057 (1992).
40. Arnosti, D. N., Barolo, S., Levine, M. & Small, S. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* **122**, 205–214 (1996).
41. Schroeder, M. D., Greer, C. & Gaul, U. How to make stripes: deciphering the transition from non-periodic to periodic patterns in *Drosophila* segmentation. *Development* **138**, 3067–3078 (2011).
42. Pechmann, M., McGregor, A. P., Schwager, E. E., Feitosa, N. M. & Damen, W. G. M. Dynamic gene expression is required for anterior regionalization in a spider. *Proc. Natl. Acad. Sci. USA* **106**, 1468–1472 (2009).
43. Hilbrant, M., Damen, W. G. M. & McGregor, A. P. Evolutionary crossroads in developmental biology: the spider *Parasteatoda tepidariorum*. *Development* **139**, 2655–2662 (2012).
44. Posnien, N. et al. A comprehensive reference transcriptome resource for the common house spider *Parasteatoda tepidariorum*. *PLoS ONE* **9**, e104885 (2014).
45. Schwager, E. E. et al. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* **15**, 62 (2017).
46. Oda, H. & Akiyama-Oda, Y. The common house spider *Parasteatoda tepidariorum*. *EvoDevo* **11**, 6 (2020).
47. Fujiwara, M., Akiyama-Oda, Y. & Oda, H. Virtual spherical-shaped multicellular platform for simulating the morphogenetic processes of spider-like body axis formation. *Front. Cell Dev. Biol.* **10**, 932814 (2022).
48. Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780 (2018).
49. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).

50. Farrell, J. A. et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).

51. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).

52. Mittnenzweig, M. et al. A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell* **184**, 2825–2842.e22 (2021).

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

54. Calderon, D. et al. The continuum of *Drosophila* embryonic development at single-cell resolution. *Science* **377**, eabn5800 (2022).

55. Sakaguchi, S. et al. Single-cell transcriptome atlas of *Drosophila* gastrula 2.0. *Cell Rep.* **42**, 112707 (2023).

56. Sun, J., Zhang, C., Gao, F. & Stathopoulos, A. Single-cell transcriptomics illuminates regulatory steps driving anterior-posterior patterning of *Drosophila* embryonic mesoderm. *Cell Rep.* **42**, 113289 (2023).

57. Qiu, C. et al. A single-cell time-lapse of mouse prenatal development from gastrula to birth. *Nature* **626**, 1084–1093 (2024).

58. Karaikos, N. et al. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **358**, 194–199 (2017).

59. Nitzan, M., Karaikos, N., Friedman, N. & Rajewsky, N. Gene expression cartography. *Nature* **576**, 132–137 (2019).

60. Liu, C. et al. Spatiotemporal mapping of gene expression landscapes and developmental trajectories during zebrafish embryogenesis. *Dev. Cell* **57**, 1284–1298.e5 (2022).

61. Vahid, M. R. et al. High-resolution alignment of single-cell and spatial transcriptomes with CytoSPACE. *Nat. Biotechnol.* **41**, 1543–1548 (2023).

62. Akiyama-Oda, Y., Akaiwa, T. & Oda, H. Reconstruction of the global polarity of an early spider embryo by single-cell and single-nucleus transcriptome analysis. *Front. Cell Dev. Biol.* **10**, 933220 (2022).

63. Leite, D. J. et al. An atlas of spider development at single-cell resolution provides new insights into arthropod embryogenesis. *EvoDevo* **15**, 5 (2024).

64. Medina-Jiménez, B. I., Budd, G. E. & Janssen, R. Single-cell RNA sequencing of mid-to-late stage spider embryos: new insights into spider development. *BMC Genomics* **25**, 150 (2024).

65. Iwasaki-Yokozawa, S., Nanjo, R., Akiyama-Oda, Y. & Oda, H. Lineage-specific, fast-evolving GATA-like gene regulates zygotic gene activation to promote endoderm specification and pattern formation in the Theridiidae spider. *BMC Biol.* **20**, 223 (2022).

66. Akiyama-Oda, Y., Akaiwa, T. & Oda, H. Gene expression patterns from single-nucleus RNA-seq of late germ-disc stage embryos of the spider *Parasteatoda tepidariorum* [Data set]. figshare <https://doi.org/10.6084/m9.figshare.28208540> (2025).

67. Akiyama-Oda, Y., Akaiwa, T. & Oda, H. Gene expression patterns from single-nucleus RNA-seq of early germ-band stage embryos of the spider *Parasteatoda tepidariorum* [Data set]. figshare <https://doi.org/10.6084/m9.figshare.28208471> (2025).

68. Akiyama-Oda, Y. & Oda, H. Axis specification in the spider embryo: dpp is required for radial-to-axial symmetry transformation and sog for ventral patterning. *Development* **133**, 2347–2357 (2006).

69. Bakken, T. E. et al. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS ONE* **13**, e0209648 (2018).

70. Edgar, A., Bates, C., Larkin, K. & Black, S. Gastrulation occurs in multiple phases at two distinct sites in *Latrodectus* and *Cheiracanthium* spiders. *EvoDevo* **6**, 33 (2015).

71. Prpic, N.-M. & Pechmann, M. Extraembryonic tissue in chelicerates: a review and outlook. *Philos. Trans. R. Soc. B* **377**, 20210269 (2022).

72. Akiyama-Oda, Y. & Oda, H. Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells. *Development* **130**, 1735–1747 (2003).

73. Chan, L.-N. & Gehring, W. Determination of Blastoderm Cells in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **68**, 2217–2221 (1971).

74. McLaughlin, C. N. et al. Single-cell transcriptomes of developing and adult olfactory receptor neurons in *Drosophila*. *eLife* **10**, e63856 (2021).

75. Tzani, I., Hagan, R. & Clarke, C. Tracing production instability in a clonally-derived CHO cell line using single cell transcriptomics. <https://doi.org/10.1002/bit.27715>. *Biotechnol. Bloeng.* (2021).

76. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

77. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

78. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

79. Akiyama-Oda, Y. & Oda, H. Multi-color FISH facilitates analysis of cell-type diversification and developmental gene regulation in the *Parasteatoda* spider embryo. *Dev. Growth Differ.* **58**, 215–224 (2016).

80. Yamazaki, K., Akiyama-Oda, Y. & Oda, H. Expression patterns of a twist-related gene in embryos of the spider *Achaearanea tepidariorum* reveal divergent aspects of mesoderm development in the fly and spider. *Zool. Sci.* **22**, 177–185 (2005).

81. Pechmann, M., Benton, M. A., Kenny, N. J., Posnien, N. & Roth, S. A novel role for Ets4 in axis specification and cell migration in the spider *Parasteatoda tepidariorum*. *eLife* **6**, e27590 (2017).

## Acknowledgements

We thank A. Noda for technical assistance; Prof. Y. Takahashi, Prof. K. Sano, and Prof. T. Nakano for encouragement and discussion; and the members of JT Biohistory Research Hall and the members of the Multicellular System Area, PRESTO JST, for discussion. This work was supported by JST PRESTO (JPMJPR2041) to Y.A. and JSPS Kakenhi (20K06676, 24K01771) to Y.A., and the JT Biohistory Research Hall.

## Author contributions

Y.A. and H.O. conceived the project, which was registered as a research plan of the PRESTO and BRH. T.A. conducted the laboratory experiments with help from Y.A. and H.O., and T.A. and Y.A. analyzed the data with support from H.O. Y.A. wrote the manuscript with support from H.O.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-08335-x>.

**Correspondence** and requests for materials should be addressed to Yasuko Akiyama-Oda.

**Peer review information** *Communications Biology* thanks Ariel Chipman, Siegfried Roth, Gangcai Xie, and the other anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Aylin Bircan. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025