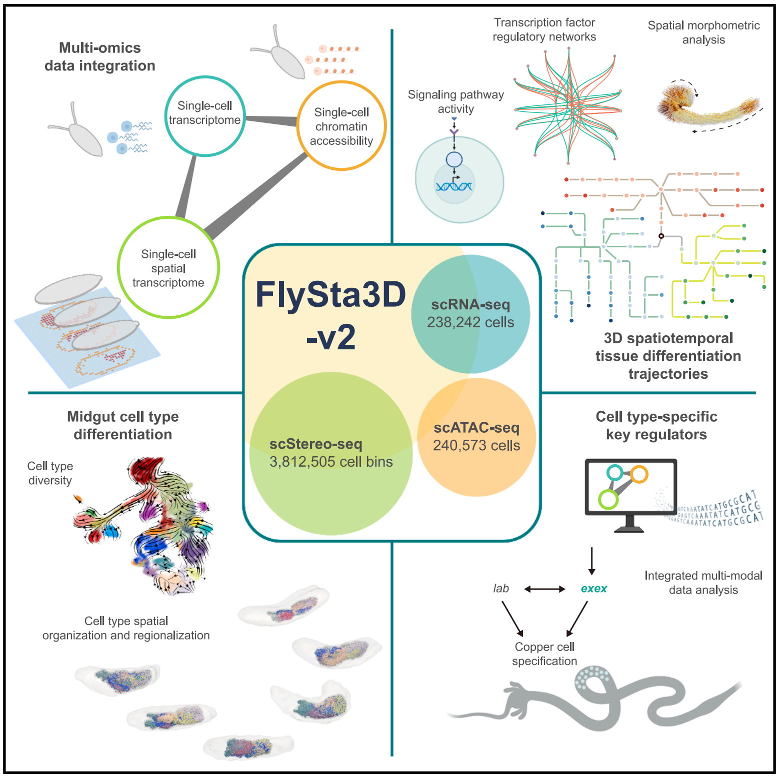
Cell

Resource

A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation

Graphical abstract



Highlights

- 3D single-cell spatiotemporal multi-omics profiling from *Drosophila* embryo to pupa
- Multimodal construction of tissue cell-type differentiation trajectories
- Spatiotemporal characterization of tissue development and regulatory networks
- Functional validation of cell-type-specific regulators in the developing midgut

Authors

Mingyue Wang, Qinan Hu,
Zhencheng Tu, ..., Longqi Liu, Xun Xu,
Yuhui Hu

Correspondence

huqn@sustech.edu.cn (Q.H.),
liulongqi@genomics.cn (L.L.),
xuxun@genomics.cn (X.X.),
huyh@sustech.edu.cn (Y.H.)

In brief

This work introduces Flysta3D-v2, a 3D spatiotemporal multi-omics database that spans *Drosophila* development from embryogenesis to metamorphosis. Through the integration of multimodal data, we detail the differentiation trajectories of developing *Drosophila* tissues within spatiotemporal contexts and elaborate the regulatory networks at the cell-type level. With a focus on the midgut, we identify cell-type-specific regulators using multi-omics data analysis and verify them through mutant studies.

Wang et al., 2025, Cell *188*, 1–20
August 21, 2025 © 2025 Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2025.05.047

CellPress

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

Cell

CellPress

Resource

A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation

Mingyue Wang,[1,2,3,4,15] Qinan Hu,[5,6,7,15,*] Zhencheng Tu,[1,8,15] Lingshi Kong,[5,6,7,15] Tengxiang Yu,[5,6,7] Zihan Jia,[1,8] Yuetian Wang,[5,6,7] Jiajun Yao,[9] Rong Xiang,[1,8] Zhan Chen,[1,8] Yan Zhao,[5,6,10] Yanfei Zhou,[5,6,7] Qing Ye,[1] Kang Ouyang,[1,8] Xianzhe Wang,[1] Yinqi Bai,[1] Zhenyu Yang,[5,6,7] Hanxiang Wang,[5,6,7] Yanru Wang,[5,6,7] Hanxiang Jiang,[5,6,7] Tao Yang,[11] Jing Chen,[11] Yunting Huang,[11] Ni Yin,[11] Wenyuan Mo,[11] Wenfu Liang,[11] Chang Liu,[1] Xiumei Lin,[1] Chuanyu Liu,[1,12] Ying Gu,[1,12] Wei Chen,[4,7] Longqi Liu,[1,2,3,*] Xun Xu,[1,2,*] and Yuhui Hu[5,6,7,13,14,16,*]

[1]BGI Research, Hangzhou 310030, China
[2]State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Hangzhou 310030, China
[3]School of Computer Science and Technology, Zhejiang University of Science and Technology, Hangzhou 310023, China
[4]Department of Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen 518055, China
[5]Department of Pharmacology, School of Medicine, Southern University of Science and Technology, Shenzhen 518055, China
[6]Joint Laboratory of Guangdong-Hong Kong Universities for Vascular Homeostasis and Diseases, School of Medicine, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China
[7]Shenzhen Key Laboratory of Gene Regulation and Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen 518055, China
[8]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[9]School of Ecology and Environment, Northwestern Polytechnical University, Xi'an 710072, China
[10]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, Berlin 14195, Germany
[11]China National GeneBank, BGI Research, Shenzhen 518083, China
[12]BGI Research, Shenzhen 518083, China
[13]Research Center for Chemical Biology and Omics Analysis, Guangming Advanced Research Institute, Southern University of Science and Technology, Shenzhen 518055, China
[14]SUSTech Homeostatic Medicine Institute, School of Medicine, Southern University of Science and Technology, Shenzhen 518055, China
[15]These authors contributed equally
[16]Lead contact
*Correspondence: huqn@sustech.edu.cn (Q.H.), liulongqi@genomics.cn (L.L.), xuxun@genomics.cn (X.X.), huyh@sustech.edu.cn (Y.H.)
https://doi.org/10.1016/j.cell.2025.05.047

SUMMARY

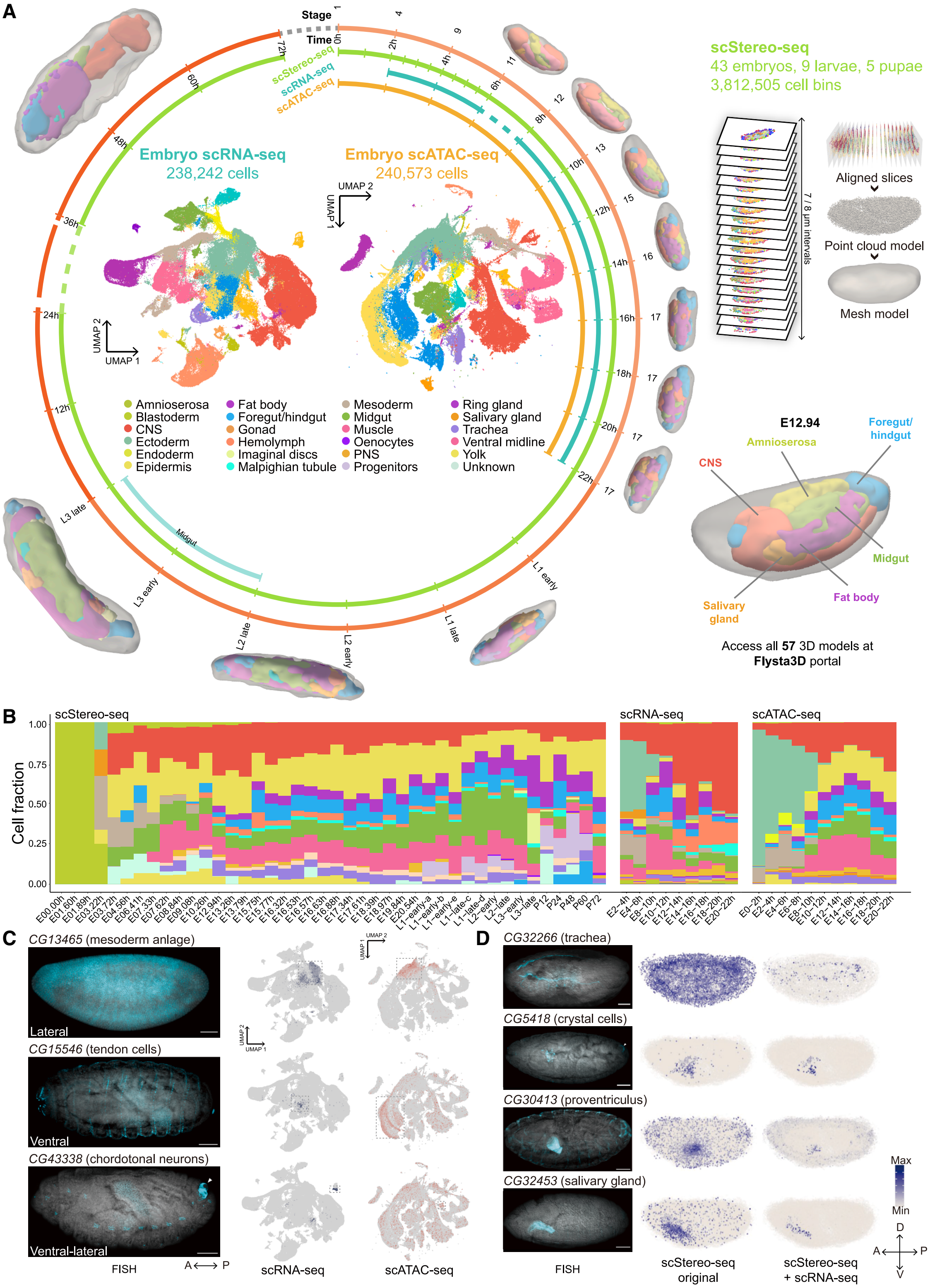
The development of a multicellular organism is a highly intricate process tightly regulated by numerous genes and pathways in both spatial and temporal manners. Here, we present Flysta3D-v2, a comprehensive multi-omics atlas of the model organism *Drosophila* spanning its developmental lifespan from embryo to pupa. Our datasets encompass 3D single-cell spatial transcriptomic, single-cell transcriptomic, and single-cell chromatin accessibility information. Through the integration of multimodal data, we generated developmentally continuous *in silico* 3D models of the entire organism. We further constructed tissue development trajectories that uncover the detailed profiles of cell-type differentiation. With a focus on the midgut, we identified transcription factors involved in midgut cell-type regulation and validated *exex* as a key regulator of copper cell development. This extensive atlas provides a rich resource and serves as a systematic platform for studying *Drosophila* development with integrated single-cell data at ultra-high spatiotemporal resolution.

INTRODUCTION

The development of a metazoan is an intricately regulated process that transforms a totipotent zygote into a fully formed organism with functional organs, necessitating the continuous coordinated regulation of molecular components within a three-dimensional space. The advances in single-cell multi-omics technologies have revolutionized our approaches to investigating organism development and cell heterogeneities at multiple dimensions.[1] The development of spatial multi-omics techniques further augments them with spatial contexts,[2] and progress has been made in integrating these multimodal data to construct panoramic profiles of context-specific functions of single cells and their communications with one another.[3]

Research into the model organism *Drosophila* has yielded invaluable insights into the meticulous processes of embryogenesis and organogenesis. Significantly, many of its features are conserved in mammals.[4] Recently, a few studies have addressed

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047



(legend on next page)

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

Drosophila development from the perspective of single-cell multi-omics.[5,6] The spatial contexts of gene expression at the single-cell level are crucial to understanding their biological relevance but are often lost during standard single-cell sequencing procedures. Previously, we utilized spatial enhanced resolution omics sequencing (Stereo-seq),[7] a sequencing- and patterned DNA nanoball (DNB) array-based spatial transcriptomic platform with high spatial resolution and sensitivity, to address this gap.[8]

Here, we expanded our previous spatiotemporal transcriptomic study of *Drosophila* to cover its developmental lifespan from embryo to pupa. Using Stereo-seq and *Spateo*, a computational pipeline designed to analyze single-cell multimodal data,[9] we reconstructed 3D transcriptomes at single-cell spatial resolution. We further complemented single-cell Stereo-seq (scStereo-seq) data with single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) data to create a multi-omics atlas of *Drosophila* embryogenesis. This atlas includes transcriptomic and epigenomic information within an ultra-high-resolution spatial context, which allowed us to establish multi-omics cell state trajectories with spatiotemporal dynamics of associated molecular regulatory networks.

To elucidate the developmental regulation of various cell types from a 3D spatiotemporal multi-omics perspective, we employed the midgut as a model for investigation. The *Drosophila* midgut fulfills versatile roles in food digestion, nutrient uptake, immunity, and endocrine regulation, carried out by distinct types of cells and the regions they form, such as enterocytes (ECs, responsible for nutrient absorption), entero-endocrine cells[10] (EEs, responsible for stimuli sensing and hormone secretion), and copper cells (associated with metal ion homeostasis and gastric acid secretion).[11,12] During metamorphosis, larval midgut cells undergo autophagy-dependent cell death,[13] and adult midgut cells reconstitute from a group of specifically designated stem cells, termed adult midgut progenitors (AMPs).[14,15] Nevertheless, the timing of differentiation and regulatory mechanisms has not been fully elucidated. Focusing on the developing midgut, we investigated their cell-type diversification, gene regulatory networks (GRNs), and regionalization from a multi-omics perspective. Through multimodal analyses, we identified multiple potential factors involved in cell-type-specific regulation during midgut development and validated a homeodomain (HD) transcription factor (TF) *exex* as a key regulator of copper cells using mutant analysis.

The data in this single-cell spatiotemporal multi-omics atlas of *Drosophila* development are curated in our online database, Flysta3D version 2.0 (Flysta3D-v2, https://db.cngb.org/stomics/flysta3d-v2/). This database will facilitate systematic research on *Drosophila* development with its comprehensive information and broad range of applications.

RESULTS

Reconstruction of *Drosophila* 3D spatial transcriptomes from embryogenesis to metamorphosis at single-cell resolution

To construct a multi-omics atlas of *Drosophila* development, we first expanded and enhanced the 3D spatial transcriptomes by collecting samples across developmental stages, including embryos (at 0.5 to 2 h intervals), larvae (early and late stages of L1 to L3), and pupae (P12 to P72) (STAR Methods; Figure 1A; Table S1).

To achieve single-cell spatial resolution, we performed nucleus staining and cell segmentation for each chip, allowing for more precise single-cell transcriptomic analysis (Figure S1A). Utilizing the Stereo-seq platform, we generated organism-wide single-cell spatial transcriptomes for 43 embryo, 9 larva, and 5 pupa samples throughout *Drosophila* development, with a total of 3,812,505 cell bins (Figure 1A; Table S1). We then combined cell bins from all sections of individual samples, performed unsupervised clustering based on both gene expression profiles and spatial locations (all clustering results are in Mendeley Data: https://doi.org/10.17632/4zf847bxcd.1), and manually annotated the clusters (Table S2). Integrating spatial locations and cell annotations with *Spateo*,[9] we achieved 3D spatial transcriptomes to reconstruct diverse tissues with fine anatomical morphology (Figure 1A). Based on this comprehensive spatiotemporal transcriptomic dataset, we generated a list of 338 genes without reported spatial expression patterns in established *in situ* databases[16,17] and reconstructed their patterns in 3D (Table S3). Consequently, we generated 3D spatial transcriptomes for *Drosophila* samples, covering the developmental period from embryo to pupa.

A single-cell spatiotemporal multi-omics atlas of *Drosophila* embryonic development

To augment our single-cell 3D spatial transcriptomic data with deeper transcriptomic and epigenomic information, we

Figure 1. A single-cell spatiotemporal multi-omics atlas of developing *Drosophila*

(A) Overall summary of this study. Time points in sample collection windows indicate hours after egg laying in embryos and hours after pupation in pupa. UMAP plots of aggregated scRNA-seq and scATAC-seq data for embryonic stages are color coded by tissue annotation. Quality P36 scStereo-seq and E6-8h scRNA-seq data were not obtained. In 3D modeling of representative scStereo-seq samples, models of epidermis, trachea, hemolymph, and muscle are not displayed in some samples for better visualization of internal organs. 3D models are not on the same scale. CNS, central nervous system; PNS, peripheral nervous system.

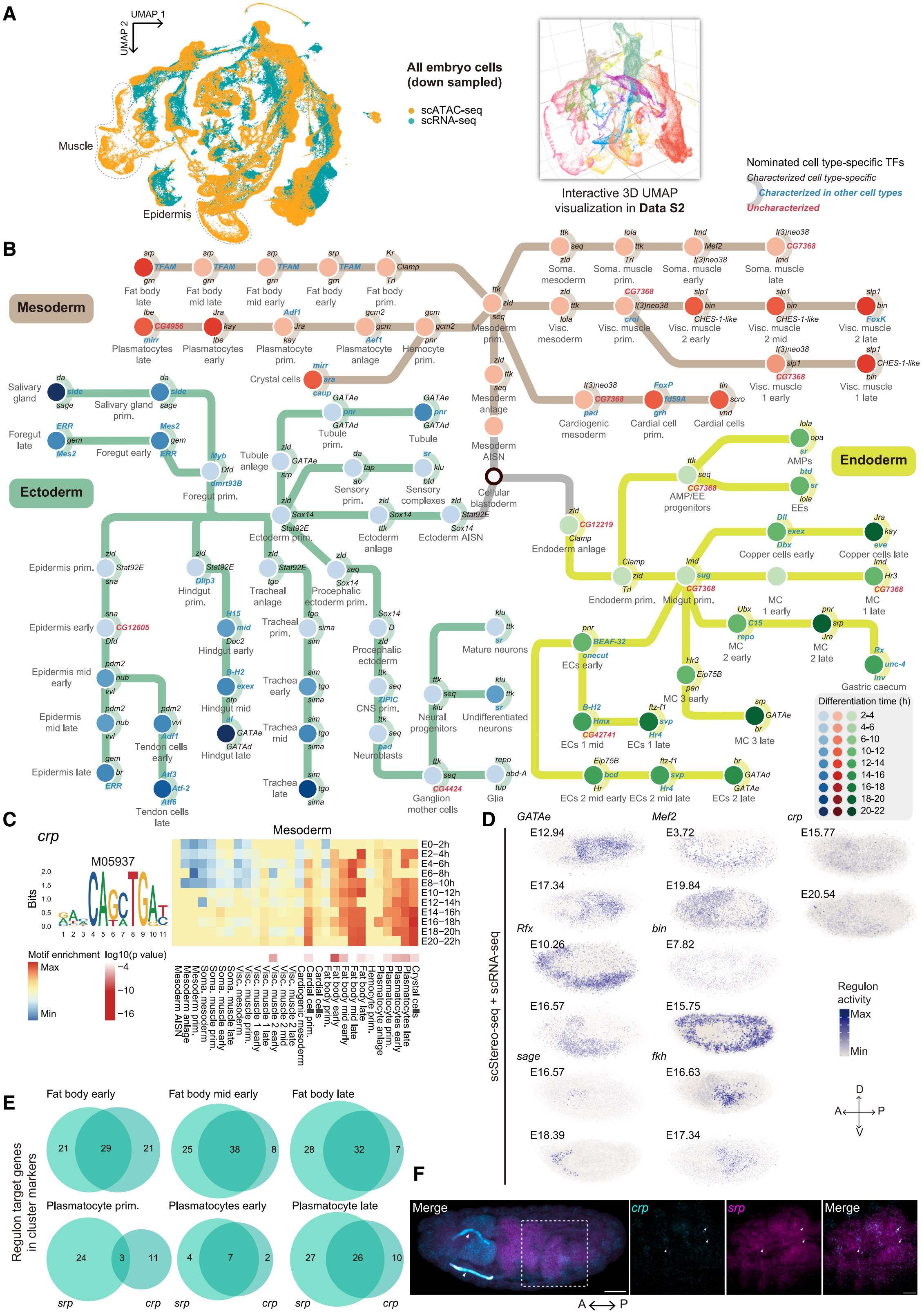
(B) Bar plot showing cell-type composition of scStereo-seq, scRNA-seq, and scATAC-seq data. Cell-type color codes are the same as those in (A).

(C) Validation of representative cell-type-specific marker genes in Table S5. Left: representative FISH images of corresponding stages of gene expression enrichment, with sample viewpoints labeled. Right: UMAP plots of marker gene expression specificity in aggregated scRNA-seq and scATAC-seq data. Cells with enriched marker gene expression or peak accessibility are highlighted in dashed rectangles.

(D) Evaluation of scStereo-seq/scRNA-seq integration. For each gene, representative FISH images were obtained from stage 13–17 embryos from lateral view. Spatial expression patterns generated from original scStereo-seq or integrated scStereo-seq/scRNA-seq data are projected along the z axis. Cyan: gene-specific RNA probes; gray: nuclei stained with DAPI. Arrowheads indicate structures with autofluorescence. All scale bars, 50 μm. A-P, anterior-posterior; D-V, dorsal-ventral.

Related to Figures S1 and S2.

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047



(legend on next page)

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

performed droplet-based scRNA-seq and scATAC-seq on embryos collected at 2-h intervals across embryogenesis (Figure 1A). Following quality control, we obtained 238,242 single-cell transcriptomes with scRNA-seq, with a median of 6,841 unique molecular identifiers (UMIs) and 1,707 genes per cell (Table S1). We also obtained 240,573 single-cell chromatin accessibility profiles with scATAC-seq, with a median of 11,772 fragments per cell (Table S1). The statistics of our datasets were comparable to or better than previous *Drosophila* embryo scRNA-seq and scATAC-seq datasets[5,6,18] (Figures S1B–S1D). Additionally, our scATAC-seq data achieved high coverage of previously reported scATAC-seq datasets,[6,18] DNase I hypersensitive sites (DHSs),[19] annotated transcription start sites (TSSs),[20] and known enhancer sites[21–23] (Figure S1E). For more comprehensive analyses at larval stages, we generated an additional control scRNA-seq dataset from the L3 midgut of *NP1-Gal4* > *UAS-mCherry-shRNA* (where *NP1-Gal4* is a midgut EC-specific driver; shRNA, short hairpin RNA). Following quality control measures, we obtained 17,988 single-cell transcriptomes, with a median of 8,846 UMIs and 1,795 genes per cell (Table S1).

With the aggregated scRNA-seq data collected across embryogenesis, we first performed coarse unsupervised clustering and generated 45 cell clusters in the uniform manifold approximation and projection (UMAP) plot (Figure S1F). We annotated these clusters and classified annotations at three levels (cell type-tissue-germ layer, e.g., gastric caecum-midgut-endoderm) (Figure 1A; Table S2). We also performed coarse unsupervised clustering in aggregated scATAC-seq data (Figure S1G). Similarly, we generated 40 distinctly annotated clusters in the UMAP plot (Figure 1A; Table S2). The data we collected achieved extensive coverage of major tissues, as reflected by the proportion of cells representing each tissue and their dynamics over developmental stages (Figure 1B).

We further profiled tissue cell-type heterogeneity by subclustering (Data S1; Table S4). In scRNA-seq data, we were able to extensively characterize the cell-type composition of embryonic tissues, including rare cell types. For example, the subclustering of the peripheral nervous system (PNS) cluster allowed for the distinct identification of neurons and glia from external sensory[24] and chordotonal organs[25,26] (File 1 in Data S1). We also identified subclusters representing most of these cell types in scATAC-seq data (File 2 in Data S1). To verify the subclusters we identified, we compiled a list of common cell-type markers identified in both datasets (Table S5) and validated the expression specificity of 3 previously unreported cell-type markers using fluorescence *in situ* hybridization (FISH) (Figure 1C).

In summary, we generated a compendium of scStereo-seq, scRNA-seq, and scATAC-seq datasets throughout *Drosophila* embryogenesis. The high granularity and temporal continuity of our multi-omics data opened the possibility of cell-type- and developmental-age-dependent integration of these multi-omics data.

Integration of multi-omics data and construction of tissue differentiation trajectories

To unify the sample ages and integrate multimodal data, we applied *RAPToR*[27] on embryo scStereo-seq and scRNA-seq data and a previously described neural network model[6] on scATAC-seq data. The resulting set of 43 embryo scStereo-seq samples we collected comprehensively covered *Drosophila* embryogenesis (Figure S2A). The inferred developmental age of single cells from scRNA-seq data (Figures S2B and S2C) and scATAC-seq data (Figures S2D and S2E) showed good agreement with their actual sample collection window. Thus, we used the sample collection window to integrate the scRNA-seq and scATAC-seq data for downstream analysis.

To integrate scStereo-seq and scRNA-seq data, we selected cells with an inferred developmental age difference of 1 h for integration using *NovoSpaRc*.[28] To evaluate this integration, we selected 9 previously uncharacterized tissue-specific genes (Table S3) and established the ground truth of their expression patterns with FISH. We observed that, compared with scStereo-seq data alone, the integrated data yielded markedly reduced signal background, enhanced tissue enrichment, and improved spatial patterns that exhibited greater resemblance to FISH results (Figures 1D and S2F).

To integrate scRNA-seq and scATAC-seq data, an expression matrix was imputed from scATAC-seq peak matrix and co-embedded with scRNA-seq expression matrix in the same UMAP space for clustering (Figures 2A, S3A, and S3B). Next, we aimed to organize all embryonic tissue cell states in each germ layer of scRNA-seq and scATAC-seq data into continuous developmental trajectories. We applied *PhyloVelo*[29] to the integrated data to establish velocity vector fields for three germ layers and re-annotated cell clusters based on marker genes (Files 1–2 in Data S1; see also Table S4) and their order along the velocity trajectories

Figure 2. Multi-omics tissue differentiation trajectories and key cell-type-specific TFs

(A) UMAP plots of co-embedded scRNA-seq and scATAC-seq data of aggregated and down-sampled embryo cells. Dashed lines mark cell clusters that are largely missing in one of the datasets, with their annotations labeled.

(B) Metro plot showing tissue differentiation trajectories based on cluster phylogeny inferred from Figure S3C for major tissues of three germ layers. The top 3 TFs of each cell type, ranked by significance of differential motif activity, are labeled along trajectories. Refer to Table S6 for a full list of TFs. AISN, anlage in statu nascendi; AMP, adult midgut progenitor; EE, entero-endocrine cells; EC, enterocyte; MC, midgut chamber; prim., primordium; soma., somatic; visc., visceral.

(C) TF motif enrichment of *crp* in mesoderm, showing its binding motifs (left), motif enrichment heatmap (upper right), and enrichment *p* value heatmap (lower right).

(D) *pySCENIC* visualization of regulon activity of TFs in representative samples from integrated scStereo-seq/scRNA-seq data, projected along the z axis. A-P, anterior-posterior; D-V, dorsal-ventral.

(E) Venn diagrams showing overlap between target genes in regulons of *srp* and *crp* along tissue differentiation trajectories of fat body and plasmatocytes.

(F) FISH validation of colocalization between *crp* and *srp* in stage 13–17 embryos. Left: representative confocal images of merged channels at 10× magnification. The dashed square indicates selected areas for zoom-in. Arrowheads indicate structures with autofluorescence. A-P, anterior-posterior. Scale bar, 50 μm. Right: confocal images of zoom-in areas at 63× magnification. Arrowheads indicate representative loci of colocalization. Scale bar, 20 μm.

Related to Figures S3 and S4.

(Figure S3C; see also File 3 in Data S1). With these velocity vectors, we were able to determine the differentiation trajectories of each tissue and arrange cell types in a chronological order (Figure 2B). Due to their complexities, the branches of cell-type differentiation are better visualized in the 3D UMAP space (Files 4–7 in Data S1).

Signaling pathways play critical roles in cell fate determination and are indispensable to tissue differentiation. To characterize the activities of signaling pathways during development, we utilized 7 major signaling pathway gene sets from *FlyphoneDB*[30] and examined the expression dynamics of core pathway component genes across tissues (Figure S3D). We observed multiple well-documented tissue-specific signaling pathways. The bone morphogenetic protein (BMP) signaling pathway, known for its integral role in ectoderm dorsal-ventral patterning[31] and in the regulation of neuromuscular junctions (NMJs),[32] demonstrated the highest level of activity in early ectoderm and muscles. Meanwhile, the fibroblast growth factor receptor (FGFR) signaling pathway, which has been recognized for its role in trachea branching morphogenesis,[33] showed maximum activity during the early stages of tracheal development. Lastly, the Hippo signaling pathway, well-established for its contribution to myogenesis,[34] was most active in early muscle clusters (Figure S3E). These results indicated that our datasets could serve as a systematic framework for exploring cell signaling networks.

Cell-type-specific transcription factor activities along tissue differentiation

To decipher the cell-type-specific regulatory activities directed by TFs along the tissue differentiation trajectories, we scrutinized our integrated scRNA-seq/scATAC-seq data. We first identified TF binding motifs demonstrating cell-type-specific activities in scATAC-seq data. We further filtered the candidate TF list with their target gene enrichment in cell-type marker genes of scRNA-seq data. Overlapping candidates in this integrated analysis represented the most active cell-type-specific TFs along the differentiation paths, including both well-established regulators and potential ones that are yet to be characterized (Figure 2B; Table S6).

Representative cell-type-specific TF activities include motifs of *GATAe* in Malpighian tubules,[35] *Rfx* in both PNS and CNS,[36] and *sage* in the salivary gland[37] within the ectoderm. In the mesoderm, we identified motif enrichment of *Mef2* in somatic muscle,[38] *bin* in visceral muscle,[39] and *srp* in fat body[40] and hemocytes.[41] The endoderm displayed motif enrichment of GATA family TFs *grn*, *fkh*, and *GATAe*[42,43] regulating late-stage endoderm specification (Figure S4A). We also uncovered several previously uncharacterized TFs with potential spatiotemporally specific functions during embryogenesis. The TF *crp*, ubiquitously expressed in multiple tissues and known for specifying terminal cells in tracheal tubes,[44] demonstrated potential regulatory functions in the mesodermal fat body and hemolymph (Figure 2C). Moreover, we observed significant motif enrichment of *CG9727* and *CG12219* in nervous systems, *CG7368* in cardiac mesoderm, and *CG12236* and *CG4360* in early endoderm, implying their specific functions in these tissues (Figure S4A).

To further explore the spatial regulon activities of these TFs, we applied *SCENIC*[45] to the integrated scStereo-seq/scRNA-seq data, revealing that the spatial patterns of TF regulon activities were consistent with the motif enrichment analysis in tissue specificity (Figure 2D). The spatial expression patterns of these less-characterized TFs were also probed by the Berkeley *Drosophila* Genome Project (BDGP) *in situ* database,[46] and all of them exhibited weak signal or ubiquitous expression patterns in stages of their inferred functions (Figure S4B). Thus, our multi-omics data provided additional evidence for elucidating the tissue-specific regulatory roles of these TFs.

Subsequently, we employed *Pando*[47] on the integrated scRNA-seq/scATAC-seq data to delve into the detailed regulons of identified TFs. In the fat body- and hemolymph-specific regulon activities, we discovered that *crp* and *srp* were in the same GRN (Figure S4C). Notably, we observed that target genes in the regulons of *srp* and *crp* largely overlapped in early fat body (Figure S4D), and this overlap increased along developmental trajectories of mesodermal tissues, including fat body and plasmatocytes (Figure 2E). The regulons of *srp* we identified were consistent with its role in inducing fat cell formation starting from early fat body development,[40,48] and *crp* is known to affect cell growth and tissue size control.[44] Our analysis implied an increasingly coordinated role of *crp* and *srp* within the same GRN during fat body and plasmatocyte development. Consistent with this, FISH results confirmed an overlap between cells expressing *crp* and those expressing *srp* in late-stage embryos (Figure 2F).

By tracing the tissue differentiation trajectories, we identified both previously reported and potential TFs during cell-type differentiation, revealing their tissue specificity and coordinated regulatory networks.

Investigating the origins and modes of tissue differentiation with spatiotemporal cell-type mapping

To further investigate the evolution dynamics of cell types along tissue differentiation trajectories within their 3D spatial context, we aimed to map them to their spatial locations. We applied the label transfer method from *Seurat*[49] to annotate scStereo-seq cell bins with the cell types identified in the multi-omics tissue differentiation trajectories. At the tissue level, the transferred labels demonstrated good agreement with manually annotated scStereo-seq cell bin clusters (Figure S5A).

We selected fat body (counterpart of mammalian liver) and foregut/hindgut (counterpart of mammalian stomach/large intestine, both of ectodermal origin[50]) as models and aligned their cell types with embryo scStereo-seq samples (Figures 3A and 3B). Within these tissues, the distribution of cell bins from label-transferred scStereo-seq and cells from scRNA-seq data exhibited a coherent pattern when plotted in the same UMAP space (Figure S5B). The prediction scores showed that each scStereo-seq cell bin was assigned an unambiguous label from scRNA-seq/scATAC-seq annotations (Figures S5C and S5D). Within each tissue, differentiation trends of various stages of cell types were traced through their proportional changes over development. These changes coincided with their actual developmental tendencies (Figure S5E). These results suggested a reliable mapping of cell types to their spatial locations in scStereo-seq data.

The mapping of tissue cell types at various developmental stages offers a unique opportunity to examine the distribution of differentiation origins within a spatiotemporal context.

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

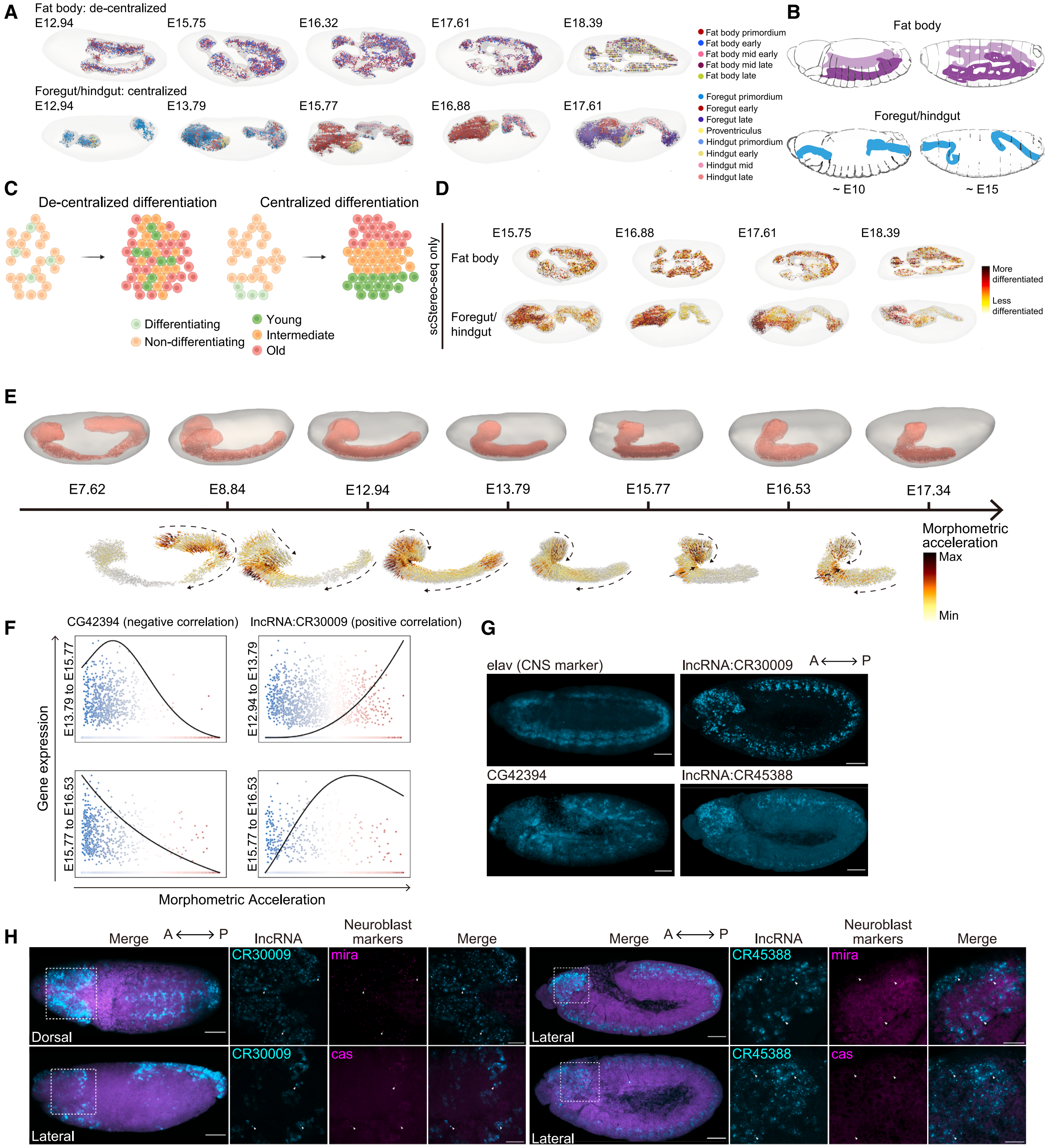


Figure 3. Spatiotemporal dynamics and morphometric changes of tissue cell types

(A) 3D tissue models across representative embryo scStereo-seq samples, showing spatial distribution of cell types, mesh models for fat body or foregut/hindgut, and mesh models of the entire embryo.

(B) Schematic representation of anatomical morphology of developing fat body and foregut/hindgut.

(C) Schematic representation of centralized and decentralized differentiation modes.

(D) Spatial distribution of cell bin *CytoTRACE* scores in representative fat body and foregut/hindgut models in (A).

(E) 3D models of CNS, CNS cell migration trajectories, and acceleration scores across 7 scStereo-seq samples of developmental age between 7 and 18 h. Dashed lines indicate directions of morphometric acceleration.

(legend continued on next page)

Specifically, tissues with centralized differentiation origins (such as stem cells) will generate spatially distinct cell types at different stages of development. By contrast, tissues with decentralized differentiation origins will produce a mixture of cell types at different stages (Figure 3C).

It is previously established that fat body cells originate from precursors arranged in repeated segments that extend throughout the entire tissue.[51] This indicates that the fat body differentiation is decentralized. To verify this, we quantified the levels of spatial aggregation of fat body cell types with neighborhood enrichment analysis, where a higher score indicates a greater level of spatial clustering. We observed a low level of neighborhood enrichment in the fat body (Figure S5F). Consistently, we noticed in 3D models that different stages of fat body cell types were scattered and mixed (Figure 3A). This was further supported by the spatial distribution of cell bin *CytoTRACE* scores of scStereo-seq data. *CytoTRACE* leverages the number of detectably expressed genes as a robust indicator of differentiation potential.[52] We observed that cells of various differentiation potentials were intermingled in fat body scStereo-seq data (Figure 3D).

On the other hand, the spatial distribution of the differentiation origin of embryonic foregut/hindgut has not been fully mapped. Previous studies identified niches of digestive tract stem cells for adult guts, where spatially defined groups of stem cells give rise to the adult foregut and hindgut, respectively.[53] It has also been documented that the embryonic hindgut derives from a narrow ring-shaped domain regulated by multiple signaling pathways, which control the outgrowth of hindgut segments.[54] The presence of gut stem cells and the distribution of differentiation origins remain unclear during organogenesis in embryos. Our data exhibited a significantly higher neighborhood enrichment in foregut/hindgut cell types (Figure S5F). Cell types of various developmental stages and differentiation potentials also demonstrated distinct spatial distributions (Figures 3A and 3D). These findings suggested that the embryonic foregut/hindgut differentiate in a centralized manner, providing evidence for the existence of clustered, rather than scattered, differentiation origins in the embryonic foregut and hindgut.

In summary, spatial mapping of tissue cell types offered an intuitive perspective for analyzing differentiation origins and identifying potential stem cell niches.

Transcriptomic dynamics during CNS morphometric changes

The 3D models created using *Spateo* facilitate tissue morphometric analysis, which links cell bins between samples and concurrently tracks cell migration paths and transcriptomic changes across developmental stages, offering a unique perspective for identifying potential regulators of cell migration and differentiation (Figure S6A).

Here, we utilized our CNS[55–57] scStereo-seq data as a model for morphometric analysis. The developing CNS demonstrated a stratified anatomy, where neuroblasts (neuronal stem cells) and ganglion mother cells (GMCs, progenitors of neurons and glia) were situated on the ventral side of the ventral nerve cord (VNC). Meanwhile, differentiated neurons and glia were located on the dorsal side.[58] This stratification was mirrored in the proportions of these cell types in our 3D CNS models (Figure S6B), which substantiated the spatial resolution of cell types in our scStereo-seq data.

We performed morphometric analysis in the CNS across 7 scStereo-seq samples, spanning developmental ages from 7 to 18 h (Video S1). Throughout CNS development, we observed a shift in regions with the highest acceleration from the posterior end of the VNC (before E8.84) to the anterior end of the brain (after E13.79) (Figures 3E and S6A). The decline in acceleration and curl scores in the VNC was likely linked to the completion of germ band retraction, indicating that the shortening of the VNC during early development primarily relied on the migration of posterior cells toward the anterior end. Conversely, the increase in acceleration and curl scores in the anterior brain region might reflect active cell organization in brain lobes during late embryogenesis (Figures 3E and S6C). As anticipated from CNS morphology, regions with the highest curvature and curl scores concentrated around the curved joint between the VNC and the brain (Figure S6C).

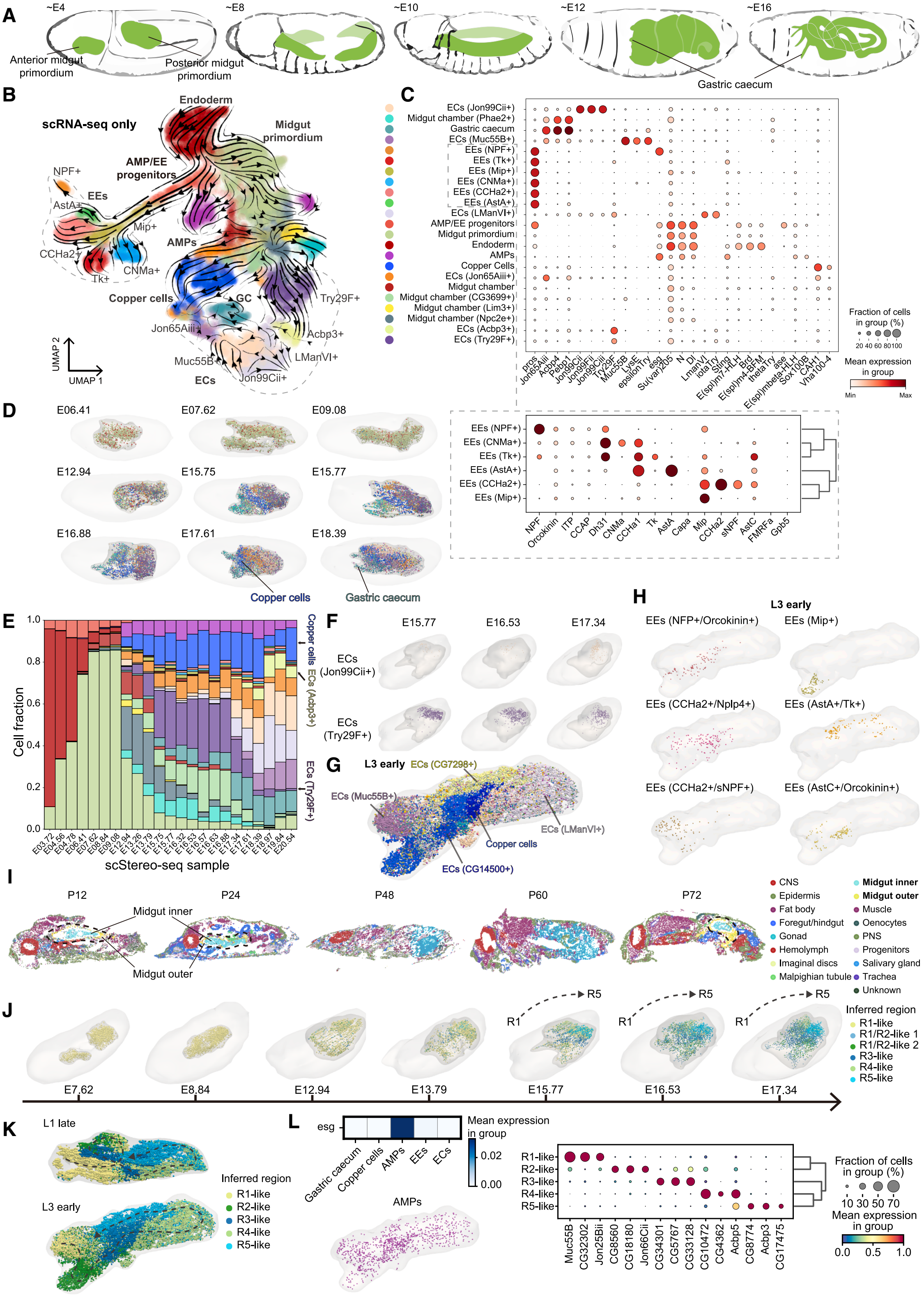
The morphometric analysis yielded a set of genes exhibiting spatiotemporal expression changes relevant to CNS morphometric dynamics (Table S7). Gene Ontology (GO) enrichment revealed that genes linked to CNS morphometric changes were highly enriched in cell fate specification and pattern formation. Notably, gene expression terms were more enriched at earlier stages, while signal transduction terms were more enriched at later stages (Figure S6D). These observations indicated that the process of CNS morphogenesis is closely linked to intrinsic cellular fate determination. This was further supported by contributions to CNS morphometric scores from known CNS development regulators (e.g., *mira*, *tll*, and *toy*). Additionally, we identified multiple uncharacterized factors. For example, the expression level of *CG42394*, which codes an uncharacterized transmembrane protein, was negatively correlated with acceleration, while that of *lncRNA:CR30009* displayed a positive correlation (Figure 3F). We validated the CNS-specific expression of these potential regulators with FISH and observed that *CG42394* exhibited a segmented expression pattern, while *lncRNA:CR30009* was expressed more ubiquitously in the CNS

(F) General linear model-based correlation between acceleration scores and expression levels of *CG42394* and *lncRNA:CR30009* in transitions between representative scStereo-seq samples.

(G) FISH validation in stage 13–17 embryos of gene candidates identified in CNS morphometric analysis. Representative images of pan-neuronal marker gene *elav* and candidate genes *CG42394*, *lncRNA:CR30009*, and *lncRNA:CR45388* are shown. All samples are shown in lateral view. Scale bars, 50 μm.

(H) FISH validation of colocalization between candidate lncRNA genes and neuroblast markers *mira* and *cas* in stage 13–17 embryos, with sample viewpoints labeled. Left: representative confocal images of merged channels at 10× magnification. The dashed square indicates selected areas for zoom-in. Scale bars, 50 μm. Right: confocal images of zoom-in areas at 63× magnification. Arrowheads indicate representative loci of colocalization. Scale bar, 20 μm. A-P, anterior-posterior; D-V, dorsal-ventral.

Related to Figures S5 and S6.



(legend on next page)

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

(Figure 3G). Notably, the list of potential morphometric regulators included multiple long non-coding RNA (lncRNA) genes besides *lncRNA:CR30009*, which was previously reported to be enriched in glia and colocalize with the glia marker gene *repo*.[59] Examining these lncRNA genes in our scRNA-seq data, we observed that the expression of *lncRNA:CR30009* and *lncRNA:CR45388* showed the highest correlation with neuroblast and glioblast marker genes (Figure S6E). We further confirmed their colocalization with neuroblast marker genes *mira* and *cas* using FISH (Figure 3H). These observations indicated that the two lncRNA genes may influence CNS cell migration through the regulation of neuroblasts.

Therefore, morphometric analysis on the CNS provided a unique angle to investigate genes associated with cell migration and identified both known and potential regulators of CNS cell migration.

Gene expression and spatial layout of embryonic midgut cell subclusters

The *Drosophila* midgut is a widely used model for investigating regulatory mechanisms of cell differentiation. The anatomical morphology and spatial distribution of cell types are highly dynamic during embryo midgut development[60] (Figure 4A). Among the various midgut cell types, EEs and ECs can be further classified into multiple subclasses based on their marker genes and physiological functions in adults.[61,62] However, their dynamics during embryonic and larval development remain unclear. Here, we first aimed to classify embryonic midgut cell types into subcategories and portray their differentiation dynamics with *Dynamo*.[63] Possibly due to technical limitations during sample collection, we identified very few EEs in scATAC-seq data (Figure S3A). Thus, we concentrated on the scRNA-seq data for high-resolution subclustering, annotation, and cell state lineage depiction (Figure 4B). This revealed 23 cell subclusters of the developing midgut and identified the presence and differentiation paths of AMPs (marked by *esg*),[64] EEs (marked by *pros* and distinguished by specific expression endocrine genes), and ECs (marked and distinguished by digestive enzyme and metabolism-related genes)[62] (Figure 4C).

The kinetics of cell state changes revealed by *Dynamo* implied that, in line with previous studies, AMPs and EEs derived from the same group of progenitors,[10] which we denoted as "AMP/EE progenitors" in our data. The differentiation paths of AMP/EE progenitors indicated their relatively independent origins compared with ECs. Pathway enrichment analysis of cluster marker genes revealed that embryonic midgut cells are regulated by distinct signaling pathways, reflecting their versatile functions. Of note, the Notch signaling pathway was enriched in AMP/EE clusters, consistent with previous reports.[10,15] Additionally, the Wnt signaling pathway was enriched in specific EE clusters[65] (Figure S7A). Interestingly, the AMP and EE clusters showed high enrichment in autophagy pathways (Figure S7A). Consistently, they exhibited high expression levels of autophagy-related genes (Figure S7B). It is possible that AMPs and EEs employ autophagy-related mechanisms to maintain midgut homeostasis during embryogenesis.

It is known that some *Drosophila* midgut cell types occupy distinct spatial locations to carry out their functions. For instance, copper cells are exclusively located in a small segment in the middle of the midgut.[66] To explore the spatial patterns of midgut cell subclusters, we mapped the subclusters identified above to their respective spatial locations through label transfer from scRNA-seq to embryo scStereo-seq data (Figure 4D). As expected from their anatomy, gastric caeca and copper cells exhibited high spatial clustering (Figure 4D). Co-embedding scRNA-seq and label-transferred scStereo-seq data in the same UMAP space demonstrated high coherence (Figure S7C). The prediction scores showed that each scStereo-seq cell bin can be assigned an unambiguous label from scRNA-seq (Figure S7D). These observations supported accurate mapping of cell types to their spatial locations.

In the label-transferred scStereo-seq 3D models, we observed the dynamics in cell-type fraction throughout embryogenesis, reflecting the different timings of emergence of these cell types. For example, EC (*Try29F+*) appeared around 13 h of development, while EC (*Acbp3+*) did not form until around 17 h (Figure 4E). Neighborhood enrichment analysis suggested that all EC subclusters exhibited a higher aggregated spatial

Figure 4. Cell-type diversity and functional regionalization in developing midgut

(A) Schematic representation of anatomical morphology of developing midgut during embryogenesis.

(B) UMAP plots showing subclustering, annotation, and RNA velocity flow of endoderm cells from scRNA-seq data, derived from *Dynamo* analysis. Velocity trajectories point in the direction of differentiation. AMPs, adult midgut progenitors; ECs, enterocytes; EEs, entero-endocrine cells; GC, gastric caecum. Cell-type color codes are in (C).

(C) Bubble plot showing expression levels and enrichment of top marker genes of cell types in (B). Marker genes of EEs are further classified.

(D) 3D midgut models across representative embryo scStereo-seq samples, showing spatial distribution of cell types, mesh models of midgut, and mesh models of the entire embryo. Cell-type color codes are the same as (B).

(E) Bar plot showing cell-type composition of midgut in scStereo-seq samples. Cell types are label transferred from scRNA-seq data. Cell-type color codes are the same as (B).

(F) Same as (D) but showing spatial distribution of EC (*Jon99Cii+*) and EC (*Try29F+*) in representative scStereo-seq samples.

(G) 3D models of L3 early midgut, showing annotated cell types with distinct spatial patterns.

(H) 3D models of L3 early midgut, showing spatial distribution of identified EE subclusters.

(I) Clustering and annotation results of 2D spatial transcriptomes of representative pupa scStereo-seq sample sections. Clusters annotated as "midgut inner" and "midgut outer" are highlighted and labeled. Samples are not on the same scale.

(J) 3D embryo and midgut models across representative embryo scStereo-seq samples, showing spatial distribution of inferred midgut regions.

(K) Same as (J) but for larva scStereo-seq samples, showing only midgut models, and dashed lines indicate inferred midgut backbones.

(L) Left: same as (H) but for AMPs. Heatmap shows the expression enrichment of *esg*; right: bubble plot showing expression level and enrichment of top marker genes of AMPs in identified midgut regions in the L3 early scStereo-seq sample.

Related to Figures S7, S8, and S9.

distribution after 13 h, indicating that mature EC functions are related to their specific spatial aggregation (Figure S8A). Across embryo scStereo-seq samples, we observed that the ECs marked by *Jon99Cii* and *Try29F* occupied distinct spatial locations, suggesting that these ECs responsible for producing digestive enzymes tend to take up more posterior locations of the midgut tract (Figure 4F). We further examined the spatial distribution of marker genes of cell subclusters with high neighborhood enrichment scores in the BDGP *in situ* database. The observation of distinct distribution patterns underscores the spatial specificity of these cell subclusters (Figure S8B).

In summary, we categorized and examined the variety of cell subclusters present in the embryonic midgut. Our analysis unveiled EC subclusters that occupy distinct spatial locations and traced their distribution over development.

Gene expression and spatial layout of larval and pupal midgut cell subclusters

We subsequently endeavored to identify midgut cell types in the larva and pupa scStereo-seq samples through subclustering (Figures S8C–S8E; Table S4). Compared with embryos, larval midgut grew significantly in size and displayed a more diverse array of intestinal cell types over development (Figures 4G and S8C). Notably, different ECs were densely clustered along the anterior-posterior axis of the midgut, as observed in the 3D models of their distribution (Figure S8F). In the adult midgut, it has been established that EEs can be classified into several subclusters, each expressing a unique combination of peptide hormones and occupying distinct regions along the midgut length.[61] Similarly, in our larval scStereo-seq data, EEs could be classified into subclusters expressing various peptide hormones (Figure S8D). These EE subclusters also displayed a distinct spatial distribution, akin to that observed in the adult midgut (Figure 4H). These observations indicated that spatial patterning of EEs might also occur in larvae similar to adults.

During the L3 stage, substantial changes occurred among midgut cell types. The anterior gastric caecum and the posterior EC (*Acbp3+*) in the midgut contracted and decreased in number in L3 early sample, eventually vanishing in the L3 late sample (Figures S8F and S8G). This suggested that significant remodeling and reorganization of the midgut takes place during the L3 stage, which coincided with the previously established timing of midgut cell death in this region before metamorphosis.[13] Consistently, the L3 late midgut showed higher expression of cell death-related genes compared with other larval stages, including those involved in autophagy and lysosomal pathways (e.g., *Atg8*, *CathD*, and *Cp1*)[13] (Figure S8H). Notably, these genes showed higher expression levels in AMPs and EEs (Figure S8H), consistent with previous observations in embryo scRNA-seq data. This suggests the critical functions of AMPs and EEs in the regulation of autophagy throughout embryonic and larval midgut, consistent with the significant roles of autophagy pathways during metamorphosis.[13]

The spatial transcriptomic data from our pupa scStereo-seq samples offered valuable resources for studying the regulation of midgut metamorphosis. Examining pupa scStereo-seq samples, we observed that the pupal midgut delaminated into two layers, which we annotated as "midgut inner" and "midgut outer" in manual annotation (Figure 4I; also refer to clustering results in Mendeley Data: https://doi.org/10.17632/4zf847bxcd.1). These structures receded after stage P24 and re-emerged at stage P72. Previous studies have found that during metamorphosis, part of AMPs locate to an epithelial layer surrounding the larval midgut, which degenerates into a structure termed yellow body.[67,68] The midgut structure observed in pupal scStereo-seq samples could reflect such delamination. Examining the expression profiles of these two layers, we observed that the midgut inner cluster was enriched in metallothionein and ferritin family genes, while the midgut outer cluster expressed multiple antimicrobial pathway genes (e.g., *DptA* and *Drsl2*) (Figure S8I). These results provided clues for investigation of gene regulation during midgut metamorphosis.

In conclusion, the use of label transfer-assisted spatial mapping and annotation unveiled spatially distinct larval and pupal midgut cell types and their dynamics over development.

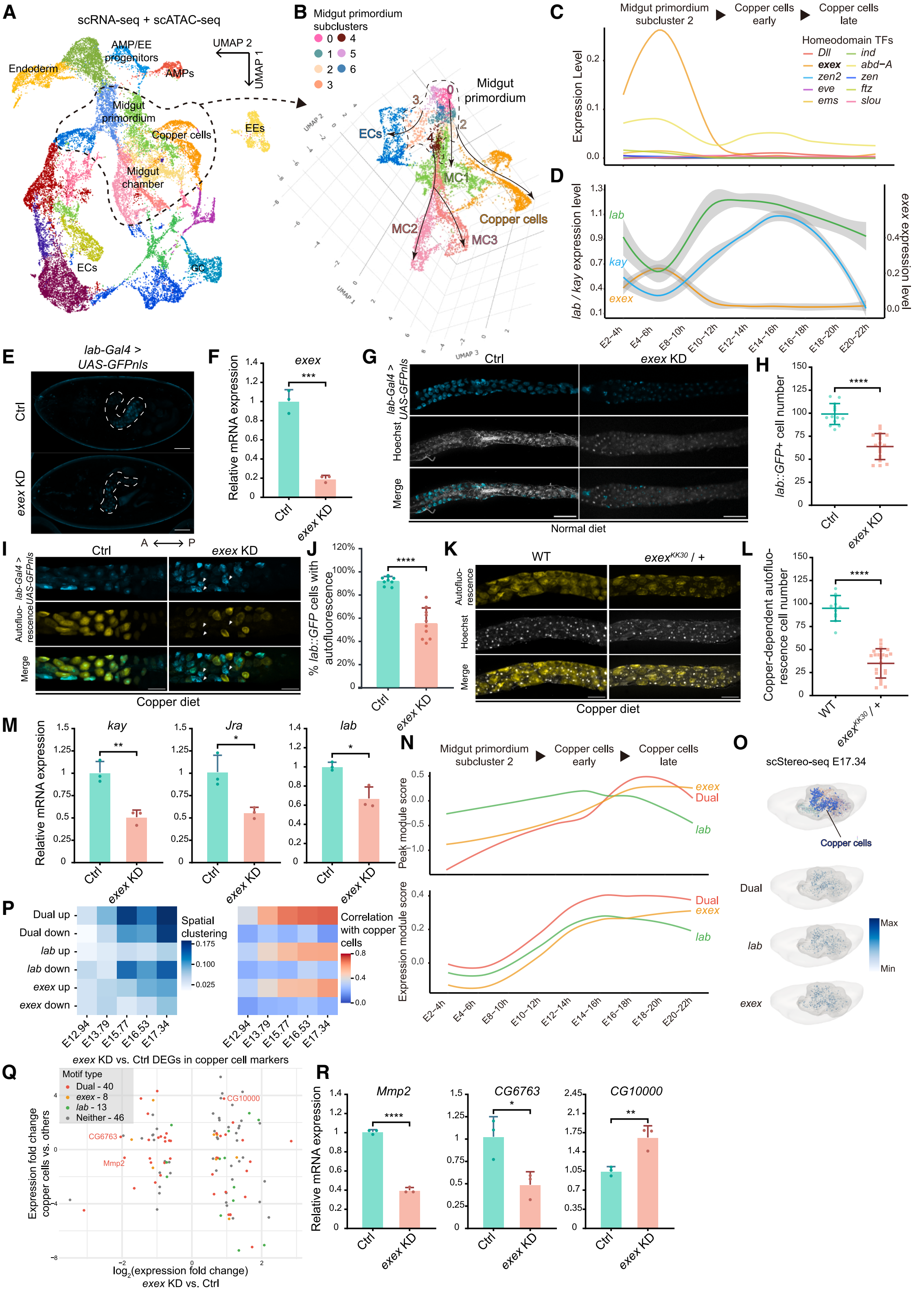
Emergence and spatial distribution of midgut functional regions

The adult *Drosophila* midgut is conventionally divided into five regions (hereafter termed R1–R5) based on morphological constraints, with each region performing distinct roles.[69,70] Our prior investigations indicated that certain functional midgut regions begin to emerge during late embryogenesis.[8]

Here, we further characterized the process of embryonic midgut regionalization. Referring to regional marker genes summarized in Buchon et al.,[70] we first identified 6 gene modules from adult midgut marker genes with *hotspot*[71] (Figure S9A) and established their correlation with expression profiles of adult midgut regions (Figure S9B). With these regional markers as references, we identified cell groups exhibiting transcriptomic similarity to adult R1 to R5 in scStereo-seq midgut cell bins, which we termed R1-like to R5-like. Each regional cell group displayed increasing levels of spatial clustering over development (Figure S9C), suggesting that they gradually occupied distinct areas in the midgut during embryogenesis. Upon inspecting the expression of gene modules in scStereo-seq samples, we noted that modules corresponding to most regions began to actively express around 13 h of embryogenesis (Figure S9D). Simultaneously, the spatial distribution of regions started to crystallize around the same time point, mirroring the spatial order as observed in the adult midgut (R1 to R5 from anterior to posterior) (Figure 4J). This suggested that the late-stage embryonic midgut exhibited regional compartmentalization akin to its adult counterpart. We examined gene GO enrichment of marker genes of the identified embryonic midgut regions (Figure S9E). For instance, the R1/R2-like regions are functionally enriched in fatty acid metabolism; the R3-like region is functionally enriched in ion transport and pH regulation, consistent with the acidic nature of this region[72]; and the R5-like region is functionally enriched in metal ion homeostasis (Figure S9F). These functions aligned well with their counterparts in adult midgut regions.[69,70]

We conducted profiling of larval midgut regions in a similar manner. Successfully, we identified modules corresponding to regions R1 through R5, which were arranged sequentially along the anterior-posterior axis of the larval midgut (Figure 4K). Notably, the identified larval EE subclusters also exhibited distinct

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047



(legend on next page)

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

regional enrichment akin to their adult counterparts, with each subcluster predominantly found in one specific region (Figure S9G). Previous studies have reported that adult intestinal stem cells (ISCs) also display a regional specificity, with ISCs in each region responsible for the differentiation and regeneration of cell types specific to their corresponding regions.[69] In our scStereo-seq data, we noted that while AMPs (marked by the expression of *esg*) were distributed throughout the entire midgut without spatial clustering (Figures 4L, S9G, and S9H), AMPs located in each identified region exhibited distinct expression profiles (Figure 4L). These findings suggested that the regional specification of midgut stem cells occurs as early as the larval stages.

Together, our scStereo-seq data demonstrated distinctive regions and their gene expression profiles in the embryonic and larval midgut, which dictated the spatially localized sub-organ functions similar to adults.

Identification of *exex* as a copper cell regulator through multimodal analyses

In pursuit of previously uncharacterized cell-type-specific regulators during midgut development, we traced the endoderm trajectories of integrated scRNA-seq/scATAC-seq data (Figures 5A and S3C). We first performed differential motif analysis within each lineage to identify TFs that exhibited cell lineage-specific activities. In the early midgut, we identified multiple motifs corresponding to previously reported TFs regulating AMP and EE differentiation, including those of *ttk*[73] and *lola*.[62,74] During late midgut differentiation, we also identified known TFs responsible for midgut development regulation, such as *fkh*[75] and *cad*[76] (Figure S10A).

Examining the midgut lineages, we noticed that the copper cells derived early from midgut primordium into a highly specialized lineage (Figures 5A and S3C). We observed that differentiated copper cells emerged around 10 h into embryogenesis (Figure 4E). This was consistent with the time point inferred from the BDGP *in situ* database, in which a copper cell marker, *Vha100-4*, appears after stage 13 (∼10 h) of embryogenesis (Figure S10B). To further trace the copper cell lineage back to its origin in the midgut primordium, we performed subclustering of the midgut primordium cluster, yielding 7 cell groups, each destined for various midgut cell types along differentiation trajectories (Figure 5B; Table S2; Data S1). Among these, subcluster 2 led to the development of copper cells, where we observed both specific transcript expression and elevated motif activity of *lab* (Figure S10C). These findings aligned with previous findings that the homeobox (Hox) TF *lab* is a well-established specific determinant of copper cells in the *Drosophila* midgut, being both necessary and sufficient for copper cell specification and maintenance.[77]

To further explore the regulatory factors driving the differentiation of copper cells, we examined their differentially active TF binding motifs and compared them to sibling lineages using scATAC-seq data. Differential motif analysis revealed that multiple HD TF motifs (characterized by their signature TAAT motif sequences[78]) were significantly enriched in copper cells compared with other midgut chamber lineages (Figure S10D). The evolutionarily conserved Hox TFs, a subset of HD TFs, are each responsible for regulating multiple body patterning and cell-type

Figure 5. *exex* is a copper cell regulator during midgut development

(A) UMAP plot showing subclustering and annotation of endoderm and midgut cells from integrated scRNA-seq/scATAC-seq data. AMPs, adult midgut progenitors; EEs, entero-endocrine cells; GC, gastric caecum; ECs, enterocytes.

(B) 3D UMAP plot showing subclustering of midgut primordium and cell types each subcluster differentiates into. MC, midgut chamber.

(C) Line plot showing expression levels of HD TFs corresponding to top copper cell differential motifs in Figure S10D in the copper cell lineage, deduced from scRNA-seq data.

(D) Same as (C) but for *exex*, *kay*, and *lab*, showing intervals predicted with a span of 0.75.

(E) Confocal images of representative control (*lab-Gal4 > UAS-mCherry-shRNA; UAS-GFPnls*, $n = 5$) and *exex* KD (*lab-Gal4 > UAS-exex-shRNA; UAS-GFPnls*, $n = 6$) stage 13–17 embryos. Copper cell regions are marked by dashed lines. A-P, anterior-posterior. Scale bars, 50 μm.

(F) Bar plots showing qPCR quantification of mRNA expression levels of *exex* and *lab* in copper cell regions. Bars represent mean relative value of gene expression, with error bars indicating standard deviation. Each data point represents one biological replicate.

(G) Confocal images of representative control ($n = 15$) and *exex* KD ($n = 23$) unfixed midgut copper cell regions from L3 raised in normal diet. Scale bars, 200 μm.

(H) Line plot showing the number of *lab::GFP*-positive cells in control and *exex* KD L3 midgut samples. Lines represent mean ± standard deviation.

(I) Same as (G) but from L3 treated in media supplemented with 500 μM $CuSO_4$. Arrowheads indicate representative *lab::GFP*-positive cells without copper-dependent autofluorescence in *exex* KD samples. Scale bars, 100 μm.

(J) Bar plot showing the percentage of *lab::GFP*-positive cells emitting copper-dependent autofluorescence in control and *exex* KD L3 midgut samples, with error bars indicating standard deviation.

(K) Same as (I) but from wild-type (*Oregon-R-P2*) and heterozygous $exex^{KK30}$ L3. Scale bars, 100 μm.

(L) Line plot showing the number of copper-dependent autofluorescence cells in wild-type and heterozygous $exex^{KK30}$ L3 midgut samples. Lines represent mean ± standard deviation.

(M) Same as (F) but for *kay*, *Jra*, and *lab*.

(N) Line plots showing peak module score (from scATAC-seq data) and expression module score (from scRNA-seq data) of up-regulated copper cell marker genes potentially regulated by peaks containing *exex*, *lab*, or dual motifs.

(O) Representative 3D midgut models showing the spatial activity patterns of genes with dual and *lab* or *exex* single motifs in scStereo-seq. Spatial patterns of copper cells are shown as a reference.

(P) Heatmaps showing Moran's I index (indicator of spatial clustering) and spatial correlation scores with copper cells for up- or down-regulated copper cell markers with dual and *lab* or *exex* single motifs.

(Q) Scatter plot showing expression fold change of copper cell marker genes in DEGs between *exex* KD and control in bulk RNA-seq data.

(R) Same as (M) but for *Mmp2*, *CG6763*, and *CG10000*. Unpaired t tests were used for all statistical analysis. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$. Related to Figures S10 and S11.

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

specification processes. It is believed that their pleiotropy is achieved through interaction with other cell-type-specific regulators.[78] Thus, there could be previously uncharacterized HD TFs involved in copper cell specification.

To assess this hypothesis, we examined the expression level of HD TFs corresponding to top copper cell differential motifs along the lineage. Among them, we noted that *exex* exhibited a significantly high level of expression at the origin of the lineage, which subsequently decreased after copper cell differentiation (Figures 5C and 5D). Moreover, multiple *exex* motifs were highly enriched in differential motifs of both early and late stages of copper cell development (Figure S10A). First identified as a regulator of axon guidance in the nervous system,[79] *exex* was also found to specify the fate of an EE subset in the midgut.[61] Indeed, we observed an expression enrichment of *exex* in an EE subset in our scRNA-seq data (Figure S10E). During embryogenesis, *exex* is expressed in both CNS and midgut[80] (Figure S10B). However, its function in copper cell differentiation was not previously investigated.

To examine the potential involvement of *exex* in copper cell regulation, we investigated its expression level and motif activity in the copper cell lineage. Compared with *lab*, the transcript expression of *exex* was present in but not restricted to midgut primordium subcluster 2 and decreased to an undetectable level in differentiated copper cells. However, *exex* motifs retained specific high activities throughout the lineage (Figures S10C and S10E). These expression profiles and regulatory characteristics of *exex* persisted into larval stages. Analysis of the L3 midgut scRNA-seq data revealed that although *exex* expression is more widespread compared with *lab* (Figure S10F), its target genes specifically retained copper cell activities, similar to those of *lab* (Figure S10G). These observations further implied the involvement of *exex* in copper cell development from embryonic to larval stages.

exex is required for copper cell specification and homeostasis through regulation of *kay* and *lab* expression

To validate the function of *exex*, we performed copper-cell-specific knockdown (KD) using *lab-Gal4* to drive the expression of both *UAS-exex-shRNA* and a *UAS-GFPnls* reporter. Examining late-stage embryos, we observed that *exex* KD embryos showed weaker and scarcer *lab::GFP* signals in the copper cell region compared with control embryos (Figure 5E). Consistently, we observed that RNAi KD resulted in a substantial decrease in *exex* expression and number of *lab::GFP*-positive cells in the L3 midgut (Figures 5F–5H and S11A). The remaining *lab::GFP*-positive cells exhibited noticeably smaller sizes and aberrant morphology compared with the typical large, flattened, and spherical copper cell features observed in the control (Figure 5G).

We then examined if physiological functions of copper cells were perturbed upon *exex* KD. Copper cells characteristically emit an orange autofluorescence when excited by ultraviolet light following dietary copper treatment.[77] This is thought to result from the interaction between metallothionein proteins and accumulated copper ions in copper cells.[81] In *exex* KD L3 treated with a copper diet, we observed a significant decrease in both the number of *lab::GFP*-positive cells and the proportion of *lab::GFP*-positive cells that also exhibited orange fluorescence (Figures 5I and 5J). This indicated a loss of copper ion homeostasis in copper cells upon *exex* deficiency. We further investigated copper cell phenotypes in copper-fed L3 carrying $exex^{KK30}$, an *exex* loss-of-function allele.[80] This strain is homozygous lethal in our hands. In heterozygous L3, we noted a significant loss of cells with copper-dependent autofluorescence and altered morphology (Figures 5K and 5L). Together, these findings confirmed the essential role of *exex* in copper cell development.

We subsequently explored the potential mechanisms by which *exex* regulates copper cell specification. We noted that motif activities of *kay* and *Jra* ranked top in copper cells (Figure S10D). These two TFs are known activators of *lab* expression in the midgut.[82,83] Given the temporal asynchrony between the expression level of *exex* and those of *kay* and *lab* (Figure 5D), it is plausible that *exex* activity precedes the other two in primordial copper cells. The temporal order was further supported by the identification of *exex* motif sequences at the TSS region of *kay* (Figure S11B). Additionally, the expression level of *kay* was significantly reduced upon *exex* KD. Although *exex* motif was not found in the TSS region of *Jra*, we still noted a similar level of expression decrease. Consistent with these findings, we observed a moderate yet significant reduction in *lab* expression (Figure 5M).

Collectively, these observations underscored the pivotal role of *exex* in regulating both copper cell development and copper homeostasis within the *Drosophila* midgut through modulating the expression levels of *kay* and *lab*.

Potential co-regulation between *exex* and *lab* during copper cell development

Although the expression levels of *exex* and *lab* displayed a temporal sequence, their binding motifs remained consistently active within the copper cell lineage. Consequently, we explored whether *exex* also exerts coordinated regulatory functions with *lab*. To investigate this hypothesis, we inspected all the peaks containing *lab* or *exex* motifs and found that peaks bearing motifs of both TFs (dual motifs) showed greater overlap with copper cell marker genes compared with those bearing only *lab* or *exex* motifs (single motifs) (Figure S11C; Table S8). We observed that compared with single motif peaks, dual motif peaks were closer to the TSS of copper cell marker genes (Figure S11D), potentially exerting stronger transcription regulation. In line with this observation, copper cell marker genes regulated by dual motifs demonstrated higher chromatin accessibility and expression levels compared with those regulated solely by *lab* motifs (Figure 5N).

Additionally, we visualized the spatial expression patterns of up- or down-regulated marker genes with either dual or single motifs in copper cells using scStereo-seq data. We observed that the up-regulated genes with dual motifs displayed an increasing trend of spatial aggregation over development, while the genes with single motifs did not exhibit spatial clustering patterns (Figures 5O and 5P). Compared with the genes with single motifs, those with dual motifs showed progressively increasing and stronger spatial correlation with copper cells over time (Figure 5P). These findings suggested that genes regulated by dual motifs exhibit a growing specificity to copper cells over development.

To further detect global changes resulting from *exex* copper-cell-specific KD, we conducted bulk RNA-seq on copper

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

cell-enriched segments of L3 midgut (Data S2). Upon examining the transcripts per million (TPM) of *exex*, we confirmed its reduction in KD samples (Figure S11E). Investigating the 415 differentially expressed genes (DEGs) between *exex* KD and control, we observed that 107 (25.8%) of them overlapped with copper cell markers. Among these genes, we noticed a higher number of dysregulated genes with dual motifs compared with those with single motifs (Figure 5Q). Inspecting representative dual motif genes, including metalloendopeptidase genes *Mmp2* and *CG6763*, as well as an uncharacterized glycosyltransferase gene *CG10000*, we noted significant expression dysregulation upon *exex* KD (Figure 5Q), which was confirmed using qPCR (Figure 5R). We further observed the presence of dual motifs in the copper cell differentially accessible (DA) peaks of these genes (Figure S11F). These observations implied that, in addition to regulating *lab* expression, the function of *exex* might also encompass serving as a transcriptional co-regulator with *lab* during copper cell development.

In summary, through *in vivo* functional validation, we confirmed the critical roles of *exex* in copper cell developmental regulation. These results underscored the utility of our multi-omics data in identifying regulatory factors at the cell-type level within developing tissues.

DISCUSSION

A comprehensive high-resolution multi-omics atlas of *Drosophila* development

After our initial proof-of-principle application of Stereo-seq on *Drosophila*, we present here a single-cell 3D spatiotemporal multi-omics atlas spanning the developmental lifespan of *Drosophila* from embryogenesis to metamorphosis. This study enhanced Stereo-seq spatial transcriptomic dataset in several ways. Firstly, the sample collection window was expanded to include development from embryo to pupa. Since systematic databases are notably absent for the larval and pupal stages, our data serve as a valuable asset for investigating the spatial gene expression patterns in larvae and pupae. Secondly, we incorporated imaging data from nucleus staining with Stereo-seq to enable cell segmentation and established single-cell spatial transcriptomes. Finally, we integrated droplet-based scRNA-seq and scATAC-seq data with scStereo-seq data, which improved genome coverage and incorporated epigenomic information. The plethora of multi-omics data generated in this study provided many unique angles for dissecting the molecular underpinnings of various aspects of tissue development, as we have shown in this study.

While we annotated tissue subclusters to the best of our knowledge, there could still be instances where clusters were not assigned their optimal annotations. Therefore, community efforts are welcome to help further specify the annotations of tissue cell types.

Identification of cell-type-specific regulatory factors through integrative multimodal data analysis

Our investigation of midgut cell-type regulators demonstrated the power of our high-resolution multi-omics data in supporting detailed delineation of tissue development at the cell-type level. Utilizing a combination of analysis tools to identify TFs that perform cell-type-specific functions, we identified both previously reported and potential regulators. Nevertheless, some of our analyses did not detect well-established, cell-type-specific TFs, such as *grn* in differentiating midgut[84] and *lab* in copper cells.[77] This could be attributed to the paucity of their downstream targets. These TFs, despite being the primary determinants for certain cell types and appearing in cell-type marker genes (Figure S11G), might regulate only a few target genes, resulting in a low ranking in enrichment-based analyses. Another explanation is the divergence between a TF's transcript expression and its motif activities. For instance, *lab* motif activity was not restricted in copper cells, but *lab* transcript expression was highly exclusive to copper cells, leading to specific activation of its target genes in copper cells (Figure S11H).

The identification of the copper cell regulator *exex* in this study was facilitated by a multimodal analysis approach. Solely relying on transcript expression data, *exex* remained elusive due to its low expression levels in copper cells within our scRNA-seq data. HD family TFs share highly similar binding motifs with the consensus TAAT sequence, which complicates the identification of *lab* co-regulators based only on motif information from scATAC-seq data. However, through comprehensive multimodal analyses, we successfully revealed the importance of *exex* in copper cells and its regulatory coordination with *lab*. These findings underscored the importance of the integration of multiple analysis methods.

Potential exploitations of integrated multi-omics datasets

The extensive datasets we generated here can be leveraged in many ways. We used the midgut as an example to demonstrate how our scStereo-seq data can provide a spatial framework for single-cell studies of larval or pupal tissues, such as through integration with existing larva scRNA-seq datasets.[85–87] The study of *Drosophila* early pupal development has been challenging due to significant tissue lysis and reformation. Our pupa scStereo-seq data provided valuable insights for the investigation of tissue-specific transcriptomic changes during metamorphosis.

Moreover, our datasets can serve as a source of inspiration for the development of multiple types of bioinformatic algorithms and methods and can serve as benchmarking resources for such algorithms, such as cell segmentation of spatial transcriptomic data, integration of multi-omics data, spatial mapping of cell types, machine-learning-based cell-type and age prediction, and cell lineage tracing, among others.

To make our data more accessible, we have incorporated our datasets into the Spateo Viewer platform.[9] This platform is a versatile and scalable web application specifically designed for the exploration of spatial transcriptomics data. Accessible through our online data portal, Flysta3D-v2, the Spateo Viewer provides user-friendly access to our 3D models with interactive visualization of gene expression and a variety of other customizable parameters within spatiotemporal contexts. A detailed tutorial of Spateo Viewer can be found in the software repository[88] and https://db.cngb.org/stomics/help. We believe that our comprehensive multi-omics database will serve as a catalyst for systematic research into *Drosophila* development, facilitating a deeper understanding of organism-wide spatiotemporal dynamics.

Limitations of the study

Our multimodal analysis demonstrated its potential for systematic investigation of *Drosophila* development at the tissue cell-type level. However, there are still certain technical limitations of our datasets.

First, the single-cell resolution of Stereo-seq data was achieved through imaging-based cell segmentation that relies on nuclei staining images, which may not always be accurate. Potential issues include the same cell being recognized as two cells in separate sections or overlapping cells not being properly defined. These technical issues may be addressed with robust cell membrane staining instead of nuclei staining for better delineation of cell boundaries. On the computational side, improved algorithms utilizing machine learning can enhance cell recognition across sections.[89,90]

Second, the scRNA-seq and scATAC-seq data in this study were obtained from separate experiments. These technical and batch variations resulted in differences in genome coverage and cell composition across datasets. Despite our efforts in age-matched integration, it is still possible that data used for integration were from different states of cells, confounding the analysis results. To overcome this issue, methods for simultaneous capture of transcriptomic and chromatin accessibility profiles from single cells have been developed lately[91–93] and can be applied in future studies.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Yuhui Hu (huyh@sustech.edu.cn).

Materials availability

All materials used for Stereo-seq, MGI C4 scRNA-seq, and MGI C4 scATAC-seq are commercially available. All unique/stable reagents generated in this study are available from the lead contact upon request.

Data and code availability

Raw data generated by Stereo-seq, scRNA-seq, and scATAC-seq in this study and associated protocols can be accessed in our online database, Flysta3D-v2 (https://db.cngb.org/stomics/flysta3d-v2/). All data were analyzed with standard programs and packages, as detailed in STAR Methods. Processed matrices of single-cell datasets can be accessed through Mendeley Data (https://doi.org/10.17632/tvvjfr3c6j.1, https://doi.org/10.17632/29695x8txs.1, and https://doi.org/10.17632/4zf847bxcd.1). Raw data generated by bulk RNA-seq of copper cells can be accessed through Mendeley Data (https://doi.org/10.17632/pvft366s4d.1 and https://doi.org/10.17632/wjjhnfzy9n.1). Due to the file size limit, the scATAC-seq matrix was divided into two parts in a tissue-balanced manner. All custom codes using open-source software to support this study are provided in a public GitHub repository (https://github.com/Flysta3D/multi_omics_atlas). Any additional information required to reanalyze the data reported in this study is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the Shenzhen Key Laboratory of Gene Regulation and Systems Biology (grant no. ZDSYS20200811144002008 to Y. Hu), the Shenzhen Medical Research Fund (grant no. D2401016 to Y. Hu and L.K.), the National Natural Science Foundation of China (grant no. 32100684 to Q. H.), the Guangdong Basic and Applied Basic Research Foundation (grant no. 2024A1515012343 to Q.H.), the National Key R&D Program of China (grant no. 2022YFC3400405 to X.X. and L.L.), the National Key R&D Program of China (grant no. 2022YFC3400300 to M.W.), the Shenzhen Science and Technology Innovation Program (grant no. KQTD20180411143432337 to Y. Hu, W. C., and Q.H.), and the Guangdong Genomics Data Center (grant no. 2021B1212100001 to T. Yang and J.C.). Part of the data analysis work was performed on the STOmic Cloud (https://cloud.stomics.tech). Library sequencing was performed in China National GeneBank. This work was supported by the Center for Computational Science and Engineering and Core Research Facilities at Southern University of Science and Technology. This work was supported by the Key Laboratory of Spatial Omics of Zhejiang Province. We thank Dr. Xiaojie Qiu and Dr. Martin Vingron for helpful suggestions on data analysis, Dr. Yuxuan Lyu for providing transgenic fly strains, and Dr. Mariana Wolfner for helpful comments on the manuscript. We thank BioRender for the graphical resources used in figure panels.

AUTHOR CONTRIBUTIONS

M.W., Q.H., L.L., X.X., and Y. Hu conceived the idea. Q.H., Y.G., W.C., L.L., X. X., and Y. Hu supervised the work. Q.H. prepared the samples. M.W., Z.T., Y. Huang, W.L., N.Y., W.M., Chang Liu, Chuanyu Liu, and X.L. prepared the sequencing library and performed sequencing. M.W., Q.H., Z.T., L.K., J.Y., R. X., Y. Zhao, Z.C., Z.J., Q.Y., K.O., X.W., and Y.B. performed computational analysis. Q.H., M.W., Z.T., L.K., and Y. Hu interpreted the data. Q.H., T. Yu, Y. Zhou, Yuetian Wang, H.W., Yanru Wang, H.J., and Z.Y. performed experimental validation. T. Yang and J.C. established the data portal and Flysta3D-v2 website. X.X., Q.H., M.W., W.C., L.L., and Y. Hu acquired funding. Q.H., M.W., Z.T., and Y. Hu wrote the manuscript. All authors reviewed and edited the manuscript.

DECLARATION OF INTERESTS

The chip, procedure, and application of Stereo-seq are covered in pending patents. Employees of BGI have stock holdings in BGI.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT in order to polish the language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Fly strain maintenance
  - Fly sample preparation
- METHOD DETAILS
  - Stereo-seq library preparation and sequencing
  - scRNA-seq library construction and sequencing
  - scATAC-seq library construction and sequencing
  - scStereo-seq data processing
  - scRNA-seq data processing
  - scATAC-seq data processing
  - Developmental age inference
  - Spatial TF enrichment analysis
  - Integration of scRNA-seq and scStereo-seq data
  - Integration of scRNA-seq and scATAC-seq Data
  - Generation of vector fields with *PhyloVelo*
  - Construction of tissue differentiation trajectory

  - ○ Cell differentiation potential analysis
  - ○ Signal pathway activity analysis
  - ○ TF motif enrichment for GRN analysis
  - ○ Nomination of cell type-specific TFs
  - ○ Differential motif analysis
  - ○ GRN analysis with *Pando*
  - ○ Spatial stratification of CNS cell types
  - ○ CNS morphometric analysis
  - ○ Sub-clustering and re-annotation of endoderm scRNA-seq data
  - ○ Midgut marker gene enrichment
  - ○ Gene module identification in midgut cell types
  - ○ RNA velocity field analysis
  - ○ Gene module analysis in midgut regions
  - ○ Fluorescence *in situ* hybridization
  - ○ Generation and assaying of transgenic flies
  - ○ RT-qPCR
  - ○ Bulk RNA-seq library prep and analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2025.05.047.

Received: February 6, 2024
Revised: January 28, 2025
Accepted: May 30, 2025

REFERENCES

1. Baysoy, A., Bai, Z., Satija, R., and Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. Nat. Rev. Mol. Cell Biol. *24*, 695–713. https://doi.org/10.1038/s41580-023-00615-w.
2. Vandereyken, K., Sifrim, A., Thienpont, B., and Voet, T. (2023). Methods and applications for single-cell and spatial multi-omics. Nat. Rev. Genet. *24*, 494–515. https://doi.org/10.1038/s41576-023-00580-2.
3. Kang, M., Ko, E., and Mersha, T.B. (2022). A roadmap for multi-omics data integration using deep learning. Brief. Bioinform. *23*, bbab454. https://doi.org/10.1093/bib/bbab454.
4. Mahowald, A.P., and Hardy, P.A. (1985). Genetics of Drosophila Embryogenesis. Annu. Rev. Genet. *19*, 149–177. https://doi.org/10.1146/annurev.ge.19.120185.001053.
5. Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The Drosophila embryo at single-cell transcriptome resolution. Science *358*, 194–199. https://doi.org/10.1126/science.aan3235.
6. Calderon, D., Blecher-Gonen, R., Huang, X., Secchia, S., Kentro, J., Daza, R.M., Martin, B., Dulja, A., Schaub, C., Trapnell, C., et al. (2022). The continuum of Drosophila embryonic development at single-cell resolution. Science *377*, eabn5800. https://doi.org/10.1126/science.abn5800.
7. Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., et al. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell *185*, 1777–1792. https://doi.org/10.1016/j.cell.2022.04.003.
8. Wang, M., Hu, Q., Lv, T., Wang, Y., Lan, Q., Xiang, R., Tu, Z., Wei, Y., Han, K., Shi, C., et al. (2022). High-resolution 3D spatiotemporal transcriptomic maps of developing Drosophila embryos and larvae. Dev. Cell *57*, 1271–1283. https://doi.org/10.1016/j.devcel.2022.04.006.
9. Qiu, X., Zhu, D.Y., Lu, Y., Yao, J., Jing, Z., Min, K.H., Cheng, M., Pan, H., Zuo, L., King, S., et al. (2024). Spatiotemporal modeling of molecular holograms. Cell *187*, 7351–7373. https://doi.org/10.1016/j.cell.2024.10.011.
10. Takashima, S., Adams, K.L., Ortiz, P.A., Ying, C.T., Moridzadeh, R., Younossi-Hartenstein, A., and Hartenstein, V. (2011). Development of the Drosophila entero-endocrine lineage and its specification by the Notch signaling pathway. Dev. Biol. *353*, 161–172. https://doi.org/10.1016/j.ydbio.2011.01.039.
11. Hung, R.-J., Li, J.S.S., Liu, Y., and Perrimon, N. (2021). Defining cell types and lineage in the Drosophila midgut using single cell transcriptomics. Curr. Opin. Insect Sci. *47*, 12–17. https://doi.org/10.1016/j.cois.2021.02.008.
12. Lemaitre, B., and Miguel-Aliaga, I. (2013). The Digestive Tract of Drosophila melanogaster. Annu. Rev. Genet. *47*, 377–404. https://doi.org/10.1146/annurev-genet-111212-133343.
13. Denton, D., Shravage, B., Simin, R., Mills, K., Berry, D.L., Baehrecke, E.H., and Kumar, S. (2009). Autophagy, not apoptosis, is essential for midgut cell death in Drosophila. Curr. Biol. *19*, 1741–1746. https://doi.org/10.1016/j.cub.2009.08.042.
14. Jiang, H., and Edgar, B.A. (2009). EGFR signaling regulates the proliferation of Drosophila adult midgut progenitors. Development *136*, 483–493. https://doi.org/10.1242/dev.026955.
15. Mathur, D., Bost, A., Driver, I., and Ohlstein, B. (2010). A Transient Niche Regulates the Specification of Drosophila Intestinal Stem Cells. Science *327*, 210–213. https://doi.org/10.1126/science.1181958.
16. Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E., et al. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol. *3*, RESEARCH0088. https://doi.org/10.1186/gb-2002-3-12-research0088.
17. Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell *131*, 174–187. https://doi.org/10.1016/j.cell.2007.08.003.
18. Cusanovich, D.A., Reddington, J.P., Garfield, D.A., Daza, R.M., Aghamirzaie, D., Marco-Ferreres, R., Pliner, H.A., Christiansen, L., Qiu, X., Steemers, F.J., et al. (2018). The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature *555*, 538–542. https://doi.org/10.1038/nature25981.
19. Reddington, J.P., Garfield, D.A., Sigalova, O.M., Karabacak Calviello, A., Marco-Ferreres, R., Girardot, C., Viales, R.R., Degner, J.F., Ohler, U., and Furlong, E.E.M. (2020). Lineage-Resolved Enhancer and Promoter Usage during a Time Course of Embryogenesis. Dev. Cell *55*, 648–664. https://doi.org/10.1016/j.devcel.2020.10.009.
20. Martin, F.J., Amode, M.R., Aneja, A., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., et al. (2023). Ensembl 2023. Nucleic Acids Res. *51*, D933–D941. https://doi.org/10.1093/nar/gkac958.
21. Rivera, J., Keränen, S.V.E., Gallo, S.M., and Halfon, M.S. (2019). REDfly: the transcriptional regulatory element database for Drosophila. Nucleic Acids Res. *47*, D828–D834. https://doi.org/10.1093/nar/gky957.
22. Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature *512*, 91–95. https://doi.org/10.1038/nature13395.
23. Gao, T., Zheng, Z., Pan, Y., Zhu, C., Wei, F., Yuan, J., Sun, R., Fang, S., Wang, N., Zhou, Y., et al. (2022). scEnhancer: a single-cell enhancer resource with annotation across hundreds of tissue/cell types in three species. Nucleic Acids Res. *50*, D371–D379. https://doi.org/10.1093/nar/gkab1032.
24. Moore, A.W., Jan, L.Y., and Jan, Y.N. (2002). hamlet, a Binary Genetic Switch Between Single- and Multiple- Dendrite Neuron Morphology. Science *297*, 1355–1358. https://doi.org/10.1126/science.1072387.
25. Styczynska-Soczka, K., and Jarman, A.P. (2015). The Drosophila homologue of Rootletin is required for mechanosensory function and ciliary

rootlet formation in chordotonal sensory neurons. Cilia *4*, 9. https://doi.org/10.1186/s13630-015-0018-9.

26. zur Lage, P., Newton, F.G., and Jarman, A.P. (2019). Survey of the Ciliary Motility Machinery of Drosophila Sperm and Ciliated Mechanosensory Neurons Reveals Unexpected Cell-Type Specific Variations: A Model for Motile Ciliopathies. Front. Genet. *10*, 24. https://doi.org/10.3389/fgene.2019.00024.
27. Bulteau, R., and Francesconi, M. (2022). Real age prediction from the transcriptome with RAPToR. Nat. Methods *19*, 969–975. https://doi.org/10.1038/s41592-022-01540-0.
28. Moriel, N., Senel, E., Friedman, N., Rajewsky, N., Karaiskos, N., and Nitzan, M. (2021). NovoSpaRc: flexible spatial reconstruction of single-cell gene expression with optimal transport. Nat. Protoc. *16*, 4177–4200. https://doi.org/10.1038/s41596-021-00573-7.
29. Wang, K., Hou, L., Wang, X., Zhai, X., Lu, Z., Zi, Z., Zhai, W., He, X., Curtis, C., Zhou, D., et al. (2024). PhyloVelo enhances transcriptomic velocity field mapping using monotonically expressed genes. Nat. Biotechnol. *42*, 778–789. https://doi.org/10.1038/s41587-023-01887-5.
30. Liu, Y., Li, J.S.S., Rodiger, J., Comjean, A., Attrill, H., Antonazzo, G., Brown, N.H., Hu, Y., and Perrimon, N. (2022). FlyPhoneDB: an integrated web-based resource for cell–cell communication prediction in Drosophila. Genetics *220*, iyab235. https://doi.org/10.1093/genetics/iyab235.
31. Bier, E., and De Robertis, E.M. (2015). BMP gradients: A paradigm for morphogen-mediated developmental patterning. Science *348*, aaa5838. https://doi.org/10.1126/science.aaa5838.
32. Bayat, V., Jaiswal, M., and Bellen, H.J. (2011). The BMP signaling pathway at the Drosophila neuromuscular junction and its links to neurodegenerative diseases. Curr. Opin. Neurobiol. *21*, 182–188. https://doi.org/10.1016/j.conb.2010.08.014.
33. Ghabrial, A., Luschnig, S., Metzstein, M.M., and Krasnow, M.A. (2003). Branching Morphogenesis of the Drosophila Tracheal System. Annu. Rev. Cell Dev. Biol. *19*, 623–647. https://doi.org/10.1146/annurev.cellbio.19.031403.160043.
34. Wackerhage, H., Del Re, D.P., Judson, R.N., Sudol, M., and Sadoshima, J. (2014). The Hippo signal transduction network in skeletal and cardiac muscle. Sci. Signal. *7*, re4. https://doi.org/10.1126/scisignal.2005096.
35. Martínez-Corrales, G., Cabrero, P., Dow, J.A.T., Terhzaz, S., and Davies, S.-A. (2019). Novel roles for GATAe in growth, maintenance and proliferation of cell populations in the Drosophila renal tubule. Development *146*, dev178087. https://doi.org/10.1242/dev.178087.
36. Dubruille, R., Laurençon, A., Vandaele, C., Shishido, E., Coulon-Bublex, M., Swoboda, P., Couble, P., Kernan, M., and Durand, B. (2002). Drosophila Regulatory factor X is necessary for ciliated sensory neuron differentiation. Development *129*, 5487–5498. https://doi.org/10.1242/dev.00148.
37. Fox, R.M., Vaishnavi, A., Maruyama, R., and Andrew, D.J. (2013). Organ-specific gene expression: the bHLH protein Sage provides tissue specificity to Drosophila FoxA. Development *140*, 2160–2171. https://doi.org/10.1242/dev.092924.
38. Cripps, R.M., Lovato, T.L., and Olson, E.N. (2004). Positive autoregulation of the Myocyte enhancer factor-2 myogenic control gene during somatic muscle development in Drosophila. Dev. Biol. *267*, 536–547. https://doi.org/10.1016/j.ydbio.2003.12.004.
39. Zaffran, S., Küchler, A., Lee, H.H., and Frasch, M. (2001). biniou (FoxF), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in Drosophila. Genes Dev. *15*, 2900–2915. https://doi.org/10.1101/gad.917101.
40. Hayes, S.A., Miller, J.M., and Hoshizaki, D.K. (2001). serpent, a GATA-like transcription factor gene, induces fat-cell development in Drosophila melanogaster. Development *128*, 1193–1200. https://doi.org/10.1242/dev.128.7.1193.
41. Waltzer, L., Bataillé, L., Peyrefitte, S., and Haenlin, M. (2002). Two isoforms of Serpent containing either one or two GATA zinc fingers have different roles in Drosophila haematopoiesis. EMBO J. *21*, 5477–5486. https://doi.org/10.1093/emboj/cdf545.
42. Okumura, T., Matsumoto, A., Tanimura, T., and Murakami, R. (2005). An endoderm-specific GATA factor gene, dGATAe, is required for the terminal differentiation of the Drosophila endoderm. Dev. Biol. *278*, 576–586. https://doi.org/10.1016/j.ydbio.2004.11.021.
43. Stainier, D.Y.R. (2002). A glimpse into the molecular entrails of endoderm formation. Genes Dev. *16*, 893–907. https://doi.org/10.1101/gad.974902.
44. Wong, M.M.-K., Liu, M.-F., and Chiu, S.K. (2015). Cropped, Drosophila transcription factor AP-4, controls tracheal terminal branching and cell growth. BMC Dev. Biol. *15*, 20. https://doi.org/10.1186/s12861-015-0069-6.
45. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.
46. Hammonds, A.S., Bristow, C.A., Fisher, W.W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E., and Celniker, S.E. (2013). Spatial expression of transcription factors in Drosophila embryonic organ development. Genome Biol. *14*, R140. https://doi.org/10.1186/gb-2013-14-12-r140.
47. Fleck, J.S., Jansen, S.M.J., Wollny, D., Zenk, F., Seimiya, M., Jain, A., Okamoto, R., Santel, M., He, Z., Camp, J.G., et al. (2023). Inferring and perturbing cell fate regulomes in human brain organoids. Nature *621*, 365–372. https://doi.org/10.1038/s41586-022-05279-8.
48. Tsuyama, T., Hayashi, Y., Komai, H., Shimono, K., and Uemura, T. (2023). Dynamic de novo adipose tissue development during metamorphosis in Drosophila melanogaster. Development *150*, dev200815. https://doi.org/10.1242/dev.200815.
49. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., et al. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat. Biotechnol. *42*, 293–304. https://doi.org/10.1038/s41587-023-01767-y.
50. Campos-Ortega, J.A., and Hartenstein, V. (1997). The Embryonic Development of Drosophila melanogaster (Springer Berlin Heidelberg) https://doi.org/10.1007/978-3-662-22489-2.
51. Hoshizaki, D.K., Blackburn, T., Price, C., Ghosh, M., Miles, K., Ragucci, M., and Sweis, R. (1994). Embryonic fat-cell lineage in Drosophila melanogaster. Development *120*, 2489–2499. https://doi.org/10.1242/dev.120.9.2489.
52. Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science *367*, 405–411. https://doi.org/10.1126/science.aax0249.
53. Singh, S.R., Zeng, X., Zheng, Z., and Hou, S.X. (2011). The adult Drosophila gastric and stomach organs are maintained by a multipotent stem cell pool at the foregut/midgut junction in the cardia (proventriculus). Cell Cycle *10*, 1109–1120. https://doi.org/10.4161/cc.10.7.14830.
54. Takashima, S., and Murakami, R. (2001). Regulation of pattern formation in the *Drosophila* hindgut by *wg*, *hh*, *dpp*, and *en*. Mech. Dev. *101*, 79–90. https://doi.org/10.1016/S0925-4773(00)00555-4.
55. Marqués, G. (2005). Morphogens and synaptogenesis in Drosophila. J. Neurobiol. *64*, 417–434. https://doi.org/10.1002/neu.20165.
56. Corty, M.M., and Coutinho-Budd, J. (2023). Drosophila glia take shape to sculpt the nervous system. Curr. Opin. Neurobiol. *79*, 102689. https://doi.org/10.1016/j.conb.2023.102689.
57. Doe, C.Q. (2017). Temporal Patterning in the Drosophila CNS. Annu. Rev. Cell Dev. Biol. *33*, 219–240. https://doi.org/10.1146/annurev-cellbio-111315-125210.
58. Crews, S.T. (2019). Drosophila Embryonic CNS Development: Neurogenesis, Gliogenesis, Cell Fate, and Differentiation. Genetics *213*, 1111–1144. https://doi.org/10.1534/genetics.119.300974.

59. McCorkindale, A.L., Wahle, P., Werner, S., Jungreis, I., Menzel, P., Shukla, C.J., Abreu, R.L.P., Irizarry, R.A., Meyer, I.M., Kellis, M., et al. (2019). A gene expression atlas of embryonic neurogenesis in Drosophila reveals complex spatiotemporal regulation of lncRNAs. Development *146*, dev175265. https://doi.org/10.1242/dev.175265.
60. Miguel-Aliaga, I., Jasper, H., and Lemaitre, B. (2018). Anatomy and Physiology of the Digestive Tract of Drosophila melanogaster. Genetics *210*, 357–396. https://doi.org/10.1534/genetics.118.300224.
61. Guo, X., Yin, C., Yang, F., Zhang, Y., Huang, H., Wang, J., Deng, B., Cai, T., Rao, Y., and Xi, R. (2019). The Cellular Diversity and Transcription Factor Code of *Drosophila* Enteroendocrine Cells. Cell Rep. *29*, 4172–4185. https://doi.org/10.1016/j.celrep.2019.11.048.
62. Hung, R.-J., Hu, Y., Kirchner, R., Liu, Y., Xu, C., Comjean, A., Tattikota, S. G., Li, F., Song, W., Ho Sui, S., et al. (2020). A cell atlas of the adult Drosophila midgut. Proc. Natl. Acad. Sci. USA *117*, 1514–1523. https://doi.org/10.1073/pnas.1916820117.
63. Qiu, X., Zhang, Y., Martin-Rufino, J.D., Weng, C., Hosseinzadeh, S., Yang, D., Pogson, A.N., Hein, M.Y., Hoi Joseph Min, K., Wang, L., et al. (2022). Mapping transcriptomic vector fields of single cells. Cell *185*, 690–711. https://doi.org/10.1016/j.cell.2021.12.045.
64. Micchelli, C.A., Sudmeier, L., Perrimon, N., Tang, S., and Beehler-Evans, R. (2011). Identification of adult midgut precursors in Drosophila. Gene Expr. Patterns *11*, 12–21. https://doi.org/10.1016/j.gep.2010.08.005.
65. Lin, G., Xu, N., and Xi, R. (2008). Paracrine Wingless signalling controls self-renewal of Drosophila intestinal stem cells. Nature *455*, 1119–1123. https://doi.org/10.1038/nature07329.
66. Dubreuil, R.R. (2004). Copper cells and stomach acid secretion in the *Drosophila* midgut. Int. J. Biochem. Cell Biol. *36*, 745–752. https://doi.org/10.1016/j.biocel.2003.07.004.
67. Hakim, R.S., Baldwin, K., and Smagghe, G. (2010). Regulation of midgut growth, development, and metamorphosis. Annu. Rev. Entomol. *55*, 593–608. https://doi.org/10.1146/annurev-ento-112408-085450.
68. Takashima, S., Younossi-Hartenstein, A., Ortiz, P.A., and Hartenstein, V. (2011). A novel tissue in an established model system: the Drosophila pupal midgut. Dev. Genes Evol. *221*, 69–81. https://doi.org/10.1007/s00427-011-0360-x.
69. Marianes, A., and Spradling, A.C. (2013). Physiological and stem cell compartmentalization within the Drosophila midgut. eLife *2*, e00886. https://doi.org/10.7554/eLife.00886.
70. Buchon, N., Osman, D., David, F.P.A., Fang, H.Y., Boquete, J.-P., Deplancke, B., and Lemaitre, B. (2013). Morphological and molecular characterization of adult midgut compartmentalization in Drosophila. Cell Rep. *3*, 1725–1738. https://doi.org/10.1016/j.celrep.2013.04.001.
71. DeTomaso, D., and Yosef, N. (2021). Hotspot identifies informative gene modules across modalities of single-cell genomics. Cell Syst. *12*, 446–456. https://doi.org/10.1016/j.cels.2021.04.005.
72. Overend, G., Luo, Y., Henderson, L., Douglas, A.E., Davies, S.A., and Dow, J.A.T. (2016). Molecular mechanism and functional significance of acid generation in the Drosophila midgut. Sci. Rep. *6*, 27242. https://doi.org/10.1038/srep27242.
73. Wang, C., Guo, X., Dou, K., Chen, H., and Xi, R. (2015). Ttk69 acts as a master repressor of enteroendocrine cell specification in Drosophila intestinal stem cell lineages. Development *142*, 3321–3331. https://doi.org/10.1242/dev.123208.
74. Hao, X., Wang, S., Lu, Y., Yu, W., Li, P., Jiang, D., Guo, T., Li, M., Li, J., Xu, J., et al. (2020). Lola regulates Drosophila adult midgut homeostasis via non-canonical hippo signaling. eLife *9*, e47542. https://doi.org/10.7554/eLife.47542.
75. Kusch, T., and Reuter, R. (1999). Functions for Drosophila brachyenteron and forkhead in mesoderm specification and cell signalling. Development *126*, 3991–4003. https://doi.org/10.1242/dev.126.18.3991.
76. Ryu, J.-H., Kim, S.-H., Lee, H.-Y., Bai, J.Y., Nam, Y.-D., Bae, J.-W., Lee, D.G., Shin, S.C., Ha, E.-M., and Lee, W.-J. (2008). Innate Immune Homeostasis by the Homeobox Gene Caudal and Commensal-Gut Mutualism in Drosophila. Science *319*, 777–782. https://doi.org/10.1126/science.1149357.
77. Hoppler, S., and Bienz, M. (1994). Specification of a single cell type by a Drosophila homeotic gene. Cell *76*, 689–702. https://doi.org/10.1016/0092-8674(94)90508-8.
78. Svingen, T., and Tonissen, K.F. (2006). Hox transcription factors and their elusive mammalian gene targets. Heredity *97*, 88–96. https://doi.org/10.1038/sj.hdy.6800847.
79. Odden, J.P., Holbrook, S., and Doe, C.Q. (2002). Drosophila HB9 is expressed in a subset of motoneurons and interneurons, where it regulates gene expression and axon pathfinding. J. Neurosci. *22*, 9143–9149. https://doi.org/10.1523/JNEUROSCI.22-21-09143.2002.
80. Broihier, H.T., and Skeath, J.B. (2002). Drosophila Homeodomain Protein dHb9 Directs Neuronal Fate via Crossrepressive and Cell-Nonautonomous Mechanisms. Neuron *35*, 39–50. https://doi.org/10.1016/S0896-6273(02)00743-2.
81. McNulty, M., Puljung, M., Jefford, G., and Dubreuil, R.R. (2001). Evidence that a copper-metallothionein complex is responsible for fluorescence in acid-secreting cells of the Drosophila stomach. Cell Tissue Res. *304*, 383–389. https://doi.org/10.1007/s004410100371.
82. Riese, J., Tremml, G., and Bienz, M. (1997). D-Fos, a target gene of Decapentaplegic signalling with a critical role during Drosophila endoderm induction. Development *124*, 3353–3361. https://doi.org/10.1242/dev.124.17.3353.
83. Casas-Tinto, S., Gomez-Velazquez, M., Granadino, B., and Fernandez-Funez, P. (2008). FoxK mediates TGF-β signalling during midgut differentiation in flies. J. Cell Biol. *183*, 1049–1060. https://doi.org/10.1083/jcb.200808149.
84. Brown, S., and Castelli-Gair Hombría, J. (2000). Drosophila grain encodes a GATA transcription factor required for cell rearrangement during morphogenesis. Development *127*, 4867–4876. https://doi.org/10.1242/dev.127.22.4867.
85. Brunet Avalos, C., Maier, G.L., Bruggmann, R., and Sprecher, S.G. (2019). Single cell transcriptome atlas of the Drosophila larval brain. eLife *8*, e50354. https://doi.org/10.7554/eLife.50354.
86. Bageritz, J., Willnow, P., Valentini, E., Leible, S., Boutros, M., and Teleman, A.A. (2019). Gene expression atlas of a developing tissue by single cell expression correlation analysis. Nat. Methods *16*, 750–756. https://doi.org/10.1038/s41592-019-0492-x.
87. Kurmangaliyev, Y.Z., Yoo, J., LoCascio, S.A., and Zipursky, S.L. (2019). Modular transcriptional programs separately define axon and dendrite connectivity. eLife *8*, e50822. https://doi.org/10.7554/eLife.50822.
88. Yao, J., et al. (2024). Spateo-viewer. https://github.com/aristoteleo/spateo-viewer.
89. Li, Y., Lac, L., Liu, Q., and Hu, P. (2024). ST-CellSeg: Cell segmentation for imaging-based spatial transcriptomics using multi-scale manifold learning. PLOS Comput. Biol. *20*, e1012254. https://doi.org/10.1371/journal.pcbi.1012254.
90. Chen, H., Li, D., and Bar-Joseph, Z. (2023). SCS: cell segmentation for high-resolution spatial transcriptomics. Nat. Methods *20*, 1237–1243. https://doi.org/10.1038/s41592-023-01939-3.
91. Wu, H., Li, X., Jian, F., Yisimayi, A., Zheng, Y., Tan, L., Xing, D., and Xie, X. S. (2022). Highly sensitive single-cell chromatin accessibility assay and transcriptome coassay with METATAC. Proc. Natl. Acad. Sci. USA *119*, e2206450119. https://doi.org/10.1073/pnas.2206450119.
92. Swanson, E., Lord, C., Reading, J., Heubeck, A.T., Genge, P.C., Thomson, Z., Weiss, M.D., Li, X.-J., Savage, A.K., Green, R.R., et al. (2021). Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. eLife *10*, e63632. https://doi.org/10.7554/eLife.63632.
93. Deng, Q., Cai, P., Luo, Y., Zhang, Z., Ma, W., Huang, Z., Chen, X., Hao, S., Ma, W., Xu, J., et al. (2024). Deciphering gene regulatory programs in

mouse embryonic skin through single-cell multiomics analysis. Preprint at bioRxiv. https://doi.org/10.1101/2024.10.11.617797.

94. Lareau, C.A., Duarte, F.M., Chew, J.G., Kartha, V.K., Burkett, Z.D., Kohlway, A.S., Pokholok, D., Aryee, M.J., Steemers, F.J., Lebofsky, R., et al. (2019). Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. Nat. Biotechnol. *37*, 916–924. https://doi.org/10.1038/s41587-019-0147-6.

95. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics Oxf. Engl. *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

96. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinform. Oxf. Engl. *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

97. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods *18*, 100–106. https://doi.org/10.1038/s41592-020-01018-x.

98. Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, 1431–1443. https://doi.org/10.1016/j.cell.2014.08.009.

99. Dai, M., Pei, X., and Wang, X.-J. (2022). Accurate and fast cell marker gene identification with COSG. Brief. Bioinform. *23*, bbab579. https://doi.org/10.1093/bib/bbab579.

100. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. *8*, 329–337. https://doi.org/10.1016/j.cels.2019.03.003.

101. Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. Media *3*, 361–362. https://doi.org/10.1609/icwsm.v3i1.13937.

102. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

103. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

104. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

105. Hu, Y., Comjean, A., Attrill, H., Antonazzo, G., Thurmond, J., Chen, W., Li, F., Chao, T., Mohr, S.E., Brown, N.H., et al. (2023). PANGEA: a new gene set enrichment tool for Drosophila and common research organisms. Nucleic Acids Res. *51*, W419–W426. https://doi.org/10.1093/nar/gkad331.

106. Shi, Q., Liu, S., Kristiansen, K., and Liu, L. (2022). The FASTQ+ format and PISA. Bioinformatics Oxf. Engl. *38*, 4639–4642. https://doi.org/10.1093/bioinformatics/btac562.

107. Kruchten, N., Seier, A., and Parmer, C. (2024). Plotly PY: An interactive, open-source, and browser-based graphing library for Python (Version 5.24.1). https://doi.org/10.5281/zenodo.14503524.

108. Kibirige, H., et al. (2025). has2k1/plotnine: v0.14.5. https://doi.org/10.5281/zenodo.1325308.

109. Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., et al. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat. Protoc. *15*, 2247–2276. https://doi.org/10.1038/s41596-020-0336-2.

110. Olsen, L.R. (2023). rearrr. https://github.com/LudvigOlsen/rearrr.

111. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

112. Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nat. Methods *15*, 1053–1058. https://doi.org/10.1038/s41592-018-0229-2.

113. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat. Methods *18*, 1333–1341. https://doi.org/10.1038/s41592-021-01282-5.

114. Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq Reads with STAR. Curr. Protoc. Bioinformatics. *51*, 11.14.1–11.14.19. https://doi.org/10.1002/0471250953.bi1114s51.

115. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.

116. Chen, H., and Boutros, P.. VennDiagram: Generate High-Resolution Venn and Euler Plots. https://doi.org/10.32614/CRAN.package.VennDiagram.

117. Hu, Y., Sopko, R., Foos, M., Kelley, C., Flockhart, I., Ammeux, N., Wang, X., Perkins, L., Perrimon, N., and Mohr, S.E. (2013). FlyPrimerBank: an online database for Drosophila melanogaster gene expression analysis and knockdown evaluation of RNAi reagents. G3 (Bethesda) *3*, 1607–1616. https://doi.org/10.1534/g3.113.007021.

118. Alles, J., Karaiskos, N., Praktiknjo, S.D., Grosswendt, S., Wahle, P., Ruffault, P.-L., Ayoub, S., Schreyer, L., Boltengagen, A., Birchmeier, C., et al. (2017). Cell fixation and preservation for droplet-based single-cell transcriptomics. BMC Biol. *15*, 44. https://doi.org/10.1186/s12915-017-0383-5.

119. Liu, C., Wu, T., Fan, F., Liu, Y., Wu, L., Junkin, M., Wang, Z., Yu, Y., Wang, W., Wei, W., et al. (2019). A portable and cost-effective microfluidic system for massively parallel single-cell transcriptome profiling. Preprint at bioRxiv. https://doi.org/10.1101/818450.

120. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. *5*, 1752–1779. https://doi.org/10.1214/11-AOAS466.

121. Kibirige, H. (2025).

122. Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A.C., Kuemmerle, L. B., Rybakov, S., Ibarra, I.L., Holmberg, O., Virshup, I., et al. (2022). Squidpy: a scalable framework for spatial omics analysis. Nat. Methods *19*, 171–178. https://doi.org/10.1038/s41592-021-01358-2.

123. Kruchten, N., Seier, A., and Parmer, C. (2024) An Interactive, Open-Source, and Browser-Based Graphing Library for Python, [Version 5.24.1]. https://doi.org/10.5281/zenodo.14503524.

124. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J.Mach. Learn. Res. *12*, 2825–2830.

125. Hu, Y. (2005). Efficient, High-Quality Force-Directed Graph Drawing. Math. J. *10*, 37–71.

126. Chen, H. (2022) VennDiagram: Generate High-Resolution Venn and Euler Plots, [Version 1.7.3].

127. Waskom, M.L. (2021). seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021. https://doi.org/10.21105/joss.03021.

128. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

129. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics Oxf. Engl. *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-Digoxigenin antibody	Abcam	ab420; RRID: AB_304362
Rabbit anti-Biotin antibody	Abcam	ab53494; RRID: AB_867860
Alexa Fluor 488 Tyramide SuperBoost Kit, goat anti-mouse IgG	Invitrogen	B40941
Alexa Fluor 594 Tyramide SuperBoost Kit, goat anti-rabbit IgG	Invitrogen	B40925
Critical commercial assays		
Tissue-Teck OCT	Sakura	4583
20X Saline-sodium citrate (SSC)	Thermo	AM9770
Cornmeal-sucrose-agar *Drosophila* media	Hopebio	HB8590
Drosophila incubator	Laifu	PGX-280A-3H
Cryostat	Leica	CM1950
Halobarcon oil 700	Macklin	H909293
Hoechst 33342	MCE	HY-15559
m-cresol purple	Yuanye	S19181
Fluorescence stereo microscope	Soptop	SZX12
Confocal microscope	Zeiss	LSM 980 NLO
40 μm cell strainer	NEST	258369
70 μm cell strainer	NEST	258368
DNBelab C4 Pocket Single-Cell Lab Package	MGI	940-000506-00; 940-000507-00; 940-000508-00; 940-000509-00; 940-0005010-00
Stereo-seq STOmics Gene Expression Package	MGI	101KT114
VAHTS Universal V8 RNA-seq Library Prep Kit for Illumina	Vazyme	NR605
DIG RNA Labeling Mix	Roche	11277073910
Biotin RNA Labeling Mix	Roche	11685597910
T7 RNA polymerase	Fermentas	EP0111
RNA Isolator Total RNA Extraction Reagent	Vazyme	R401
HiScript II Q Select RT SuperMix for qPCR	Vazyme	R233
ChamQ SYBR qPCR Mastermix	Vazyme	Q311
Deposited data		
Sequencing data generated by Stereo-seq, scRNA-seq, and scATAC-seq	This study	https://db.cngb.org/stomics/flysta3d-v2/
Processed matrices for scStereo-seq and scRNA-seq data	This study	https://doi.org/10.17632/tvvjfr3c6j.1
Processed matrices for embryo scATAC-seq data (part 1)	This study	https://doi.org/10.17632/29695x8txs.1
Processed matrices for embryo scATAC-seq data (part 2)	This study	https://doi.org/10.17632/4zf847bxcd.1
Raw sequencing reads of bulk RNA-seq data of *lab-Gal4 > UAS-mCherry-shRNA* copper cells	This study	https://doi.org/10.17632/pvft366s4d.1
Raw sequencing reads of bulk RNA-seq data of *lab-Gal4 > UAS-exex-shRNA* copper cells	This study	https://doi.org/10.17632/wjjhnfzy9n.1
Experimental models: Organisms/strains		
Drosophila melanogaster w1118	Tsinghua Fly Center	THJ0265

(*Continued on next page*)

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Drosophila melanogaster lab-Gal4	Bloomington Drosophila Stock Center	BDSC43652
Drosophila melanogaster w; NP1-Gal4	Gift from Dr. Yuxuan Lyu	N/A
Drosophila melanogaster w1118; UAS-GFPnls	Bloomington Drosophila Stock Center	BDSC4776
Drosophila melanogaster UAS-mCherry-shRNA	Bloomington Drosophila Stock Center	BDSC35785
Drosophila melanogaster UAS-exex-shRNA	Tsinghua Fly Center	TH04061.N
Drosophila melanogaster Oregon-R-P2	Bloomington Drosophila Stock Center	BDSC2376
Drosophila melanogaster exex^{KK30}	Bloomington Drosophila Stock Center	BDSC9930
Oligonucleotides		
Primers for generating FISH probes and qPCR primers	This study	Table S9
Software and algorithms		
bap2	Lareau et al.[94]	https://github.com/buenrostrolab/bap
BEDTools	Quinlan and Hall[95]	http://code.google.com/p/bedtools
BWA	Li and Durbin[96]	http://maq.sourceforge.net
Cellpose	Stringer et al.[97]	http://www.github.com/mouseland/cellpose
CIS-BP	Weirauch et al.[98]	http://cisbp.ccbr.utoronto.ca/
COSG	Dai et al.[99]	https://github.com/genecell/COSG
CytoTRACE	Gulati et al.[52]	https://cytotrace.stanford.edu/
DoubletFinder	Mcginnis et al.[100]	https://github.com/chris-mcginnis-ucsf/DoubletFinder
Dynamo	Qiu et al.[63]	https://github.com/aristoteleo/dynamo-release
FlyPhoneDB	Liu et al.[30]	https://www.flyrnai.org/tools/fly_phone/web/
Gephi	Bastian et al.[101]	http://gephi.org
Harmony	Korsunsky et al.[102]	https://github.com/immunogenomics/harmony
HOMER	Heinz et al.[103]	http://homer.ucsd.edu/homer/download.html
Hotspot	DeTomaso and Yosef[71]	https://github.com/YosefLab/Hotspot
MACS2	Zhang et al.[104]	https://pypi.python.org/pypi/MACS2/
NovoSpaRc	Moriel et al.[28]	https://pypi.org/project/novosparc
Pando	Fleck et al.[47]	https://github.com/quadbiolab/Pando
PANGEA	Hu et al.[105]	https://www.flyrnai.org/tools/pangea/web/home/7227
Phylovelo	Wang et al.[29]	https://github.com/kunwang34/PhyloVelo
PISA	Shi et al.[106]	https://github.com/shiquan/PISA
Plotly	Kruchten et al.[107]	https://github.com/plotly/plotly.py
Plotnine	Kibirige et al.[108]	https://github.com/has2k1/plotnine
pySCENIC	Van de Sande et al.[109]	https://github.com/aertslab/pySCENIC
RAPToR	Bulteau and Francesconi[27]	https://github.com/LBMC/RAPToR
rearrr	Olsen[110]	https://github.com/LudvigOlsen/rearrr
scanpy	Wolf et al.[111]	https://scanpy.readthedocs.io/en/stable/
scVI	Lopez et al.[112]	https://github.com/YosefLab/scVI
Seurat	Hao et al.[49]	https://github.com/satijalab/seurat/
Signac	Stuart et al.[113]	https://github.com/timoast/signac/
Spateo	Qiu et al.[9]	https://github.com/aristoteleo/spateo-release
squidpy	Palla et al., 2021	https://squidpy.readthedocs.io/en/stable/
STAR	Dobin et al.[114]	https://github.com/alexdobin/STAR
Velocyto	La Manno et al.[115]	http://velocyto.org/velocyto.py
VennDiagram	Chen and Boutros[116]	https://cran.r-project.org/web/packages/VennDiagram/index.html

(*Continued on next page*)

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FlyPrimerBank	Hu et al.[117]	https://www.flyrnai.org/cgi-bin/DRSC_primerbank.pl
Other		
Custom codes generated using open-source software	This study	https://github.com/Flysta3D/multi_omics_atlas

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Fly strain maintenance

All Stereo-seq, scRNA-seq, and scATAC-seq samples were obtained from *Drosophila* strain *w1118*. Flies were maintained on cornmeal-sucrose-agar media in a 25 °C incubator on a 12 h/12 h light/dark cycle. Detailed strain information can be found in the key resources table.

Fly sample preparation

Samples were prepared and embedded for cryosection and Stereo-seq as previously described.[8] Unless otherwise specified, embryo samples were sectioned along the left-right axis to represent sagittal planes.

For scRNA-seq, single cells from embryos were isolated and fixed following protocols described in Alles et al.[118] and stored at -20 °C until further use. Single cells from midgut samples of wandering third instar larvae were isolated as described in Hung et al.[62] and fixed in the same manner until further use.

For scATAC-seq, embryos were collected from a population cage and transferred to a 70 μm cell strainer, dechorionated in commercial bleach for 3 min, rinsed with ddH2O, and dried on a Kimwipe. Samples were then snap-frozen in liquid nitrogen and stored at -80 °C until further use.

METHOD DETAILS

Stereo-seq library preparation and sequencing

Tissue processing and imaging

Stereo-seq library was prepared following the protocol in Chen et al.[7] with modification. Unless otherwise mentioned, buffer recipes and primer sequences used in this study were the same as Ref[7]. Briefly, embedded samples were balanced at -20 °C for 30 min and sectioned using a Leica CM1950 cryostat at a thickness of 7 or 8 μm. All available resulting slices were collected for sequencing. To minimize batch effects, six embryo/L1 slices, or four L2 slices, or two L3/pupa slices were mounted on one 1 cm × 1 cm Stereo-seq chip simultaneously. The mounted chips were then incubated at 37 °C for 3 min and subsequently fixed in cold methanol (Sigma, 34860) at -20 °C for 30 min. After fixation, chips were treated with staining solution consisted of 0.1 × SSC (Ambion, AM9770), 1/200 nucleic acid dye (Invitrogen, Q10212), and 2 U/μl RNase inhibitor (NEB, M0314L). Sections were incubated in staining solution for 3 min and then rinsed with 0.1 × saline sodium citrate (SSC) buffer supplemented with 2 U/μl RNase inhibitor. Sections were then imaged using a Motic Custom PA53 FS6 microscope (10 × objective) before RNA capture.

RNA capture and in situ *reverse transcription*

Tissue slices, mounted on Stereo-seq chips, were subjected to treatment with 0.1% pepsin (Sigma, P7000) in a 0.01 M HCl solution, followed by incubation at 37 °C for 6 min for permeabilization. Permeabilized slices were washed twice using 0.1 × SSC buffer supplemented with 0.05 U/ml RNase inhibitor. After permeabilization, RNAs were released from the tissue and captured by DNA nanoballs (DNBs) on Stereo-seq chips. Chips were incubated in reverse transcription mixture at 42 °C overnight. After reverse transcription, chips were washed twice with 0.1 × SSC buffer and subsequently incubated in tissue removal buffer at 37 °C for 30 min. cDNAs were collected by treating the chip with Exonuclease I (NEB, M0293L) for 1 h at 37 °C. Residual cDNAs were collected through a final rinse of the chip with 0.1× SSC buffer. The flowthrough, combined with the Exonuclease I treated product, was purified using 0.8 × VAHTS DNA clean beads (Vazyme, N411-03).

Library construction and sequencing

Purified cDNAs were amplified using the KAPA HiFi Hotstart Ready Mix (Roche, KK2602) with 0.8 μM cDNA-PCR primer. PCR amplification was performed with the following steps: pre-heating at 95 °C for 5 min; 15 amplification cycles with 98 °C for 20 s, 58 °C for 20 s, and 72 °C for 3 min; final incubation at 72 °C for 5 min. Amplified cDNAs were purified using 1 × VAHTS DNA clean beads. For library construction, 20 ng cDNAs were fragmented using in-house Tn5 transposase at 55 °C for 10 min. 0.02% SDS was added to terminate the reaction. Fragmented cDNAs were then subjected to amplification using the KAPA HiFi Hotstart Ready Mix with the addition of 3 μL each 10 μM Stereo-seq-Library-F primer and Stereo-seq-Library-R primer. PCR amplification was performed with the following steps: pre-heating at 95°C for 5 min; 13 amplification cycles with 98 °C for 20 s, 58 °C for 20 s, and 72 °C for

30 s; final incubation at 72 °C for 5 min. Amplified PCR products were purified using VAHTS DNA clean beads. Construction of sequencing libraries and sequencing with MGI DNBSEQ-T10 sequencer were performed following manufacturer's protocols.

scRNA-seq library construction and sequencing

Sequencing libraries were prepared using the DNBelab C Series High-throughput Single-Cell RNA Library Preparation Kit (MGI, 94000004700) following manufacturer's protocol. Briefly, fixed single cell suspensions, preserved in methanol, were balanced at 4 °C for 5 min and centrifuged at 2,000 g at 4°C for 5 min. Cell pellets were washed twice with wash buffer [PBS with 1% RNAse inhibitor and 0.04% bovine serum albumin (BSA)] before resuspension in 100 μL wash buffer. Cells were counted and 20,000 cells were aliquoted for droplet generation, emulsion breakage, beads collection, reverse transcription, second strand synthesis, cDNA amplification, and droplet index product amplification to generate barcoded libraries. Sequencing libraries were quantified using Qubit ssDNA Assay Kit (Invitrogen, Q10212) and subsequently sequenced with MGI DNBSEQ-T10 sequencer.

scATAC-seq library construction and sequencing

Sequencing libraries were prepared using the DNBelab C Series Single-Cell ATAC Library Prep set (MGI, 1000021878) following the manufacturer's protocol. Briefly, single nuclei were extracted by grinding flash-frozen embryos in the lysis buffer using a 2 mL homogenizer (Sigma-Aldrich, D8938). Nuclei were washed twice with wash buffer and resuspended in 100 μL wash buffer. Nuclei were counted before 50,000 nuclei were aliquoted and subjected to in-house Tn5 transposase treatment. Treated single-nucleus suspension was then converted to barcoded scATAC-seq libraries through droplet encapsulation, pre-amplification, emulsion breakage, beads collection, DNA amplification, and purification. Indexed sequencing libraries were constructed following the manufacturer's protocol and quantified using Qubit ssDNA Assay kit. Sequencing libraries were subsequently sequenced with MGI DNBSEQ-T10 sequencer.

scStereo-seq data processing

Cell segmentation

Manual registration of the DNB image with the nucleic acid staining image was performed as described in Chen et al.[7]. Nucleus identification and cell segmentation was performed on the aligned images using *Cellpose*.[97] Additionally, *spateo.cs.expand_labels* function from *Spateo*[9] was applied to augment each segmented nucleus with an additional 10 DNBs, approximating the actual size of *Drosophila* cells. Furthermore, *spateo.io.read_bgi* function was employed to perform cell segmentation on the aligned bin1 matrix. Within each segmented cell, UMI counts from all DNBs corresponding to the segmentation were aggregated, preserving the counts on a per-gene basis. These aggregated counts were then summed to generate a gene/cell matrix for downstream analysis. To accurately determine the centroid of each cell, *rearrr* package[110] was used to facilitate the identification of the central position within the segmented cell.

Section alignment

spateo.align.morpho_align function was used to perform registration on all slices of each sample, which involved aligning and co-registering all slices within a sample to ensure accurate spatial alignment throughout the sample volume.

Data quality control

Cell bins were first filtered to retain those with mitochondrial gene content ≤ 10%. Cell bins were further filtered with *filter_cells* function from the *spateo.pp.filter* module (*min_area=20*, *min_expr_genes=20*). Gene filtering was applied with *filter_genes* function from the same module (*min_cells=3*, *min_counts*=1). *spateo.tl.pearson_residuals* function was then used to normalize data from all slices from the same sample. Normalized data were combined to create an integrated matrix for each sample.

Dimension reduction and clustering

dynamo.tl.compute_neighbors function was first used to compute a neighbor graph with parameters *n_neighbors=6* and *n_pca_components=50*. Dimensionality reduction was performed using *spateo.tl.pca_spateo* function with parameter *n_pca_components=50*, along with the *spateo.tl.reduceDimension* function from the *dyn.tl* module with parameter *n_components=2*. Cluster assignment was performed using *spateo.tl.scc* function. Marker genes for each cluster were determined using *scanpy.tl.rank_genes_groups* function from *SCANPY*[111] with *t-test* method. Based on marker genes and spatial morphology, clusters were manually curated and annotated into cell types, tissues, and germ layers.

3D model reconstruction and alignment

Spateo was employed to reconstruct 3D models of spatial transcriptomic data. Specifically, 3D point-cloud models of scStereo-seq samples were generated using *spateo.tdr.construct_pc* function with default parameters. To align spatial coordinates for continuous morphometric analysis, the scStereo-seq sample of the first time point was set as the reference by transforming the coordinates of the sample through *spateo.tdr.rotate_model* function, such that the centroid of the embryo was located at origin. The anterior-posterior axis and dorsal-ventral axis corresponded to the x and y axis in the coordinate system, respectively. *spateo.tl.model_align* function was used with default parameters to align all other sample models to the same coordinate system. Finally, 3D mesh models of samples were created through *spateo.tdr.construct_surface* function with parameters adjusted for each sample. 3D models of individual tissues were generated with the same methods as those for the entire embryo.

Gauging scStereo-seq capture efficiency

The precise cell count has not been definitively established for all embryonic tissues, thereby limiting our ability to accurately gauge the percentage of cells captured by scStereo-seq. Nevertheless, it is understood that following the 14th cell cycle of the syncytial embryo (~2 h into development), there are around 6,000 nuclei, with the majority forming individual cells and the remainder residing in the yolk.[50] This aligns with the cell numbers observed in our early scStereo-seq samples. For instance, the E2.95 sample captured 5,564 cells, and the E3.25 sample captured 6,175 cells. These results implied that there was no significant cell loss during scStereo-seq sample preparation at these stages. Owing to handling mishaps, certain scStereo-seq samples, particularly those from pupal stages, which were structurally amorphous, did not have all cryosection slices preserved (Table S1). To ensure the integrity of downstream analyses, we used samples with complete slice capture whenever possible.

scRNA-seq data processing

Construction of gene/cell matrix

Raw reads generated by DNBelab C4-based scRNA-seq were processed as previously described.[119] Briefly, reads were pre-processed with *PISA*[106] and aligned to the *Drosophila melanogaster* genome (dm6) using *STAR*.[114] Gene expression levels were quantified with *PISA* to generate gene/cell matrix, and genes present in < 3 cells were filtered.

Data quality control

Cells were filtered by the following criteria: feature number > 500 and mitochondrial gene content ≤ 10% for embryo data; feature number > 500, mitochondrial gene content < 25% and UMI between 1000 and 80000 for L3 midgut data. *DoubletFinder*[100] was used to identify and remove doublets, with a doublet anticipation of 3%. Batch effects were corrected with *scVI*[112] for embryo data.

Dimension reduction, clustering, and annotation

Seurat[49] was used for data normalization and dimension reduction. Louvain method was used for cell clustering. Based on marker genes, clusters were manually curated and annotated into cell types, tissues, and germ layers. For tissue subclustering, annotated tissues were divided into their respective matrices and subjected to *RunUMAP* function, followed by re-clustering at different resolutions (see Table S4) to identify detailed cell subclusters. Based on marker genes, subclusters were manually curated and annotated into cell subclusters. For all plots, hex color codes of tissues can be found in Table S10.

scATAC-seq data processing

Construction of peak/cell matrix

Raw reads generated by DNBelab C4-based scATAC-seq were pre-processed with *PISA*. Retained reads were then aligned to the *Drosophila melanogaster* genome (dm6) using *BWA*[96] with default parameters to generate BAM files. Subsequently, fragment files from the same scATAC-seq library were created using *bap2*.[94] Processed fragments were used for peak calling with *MACS2*[104] for each sample, including only autosome peaks. To minimize batch effects, samples within the same time window were matched pairwise and peak integration was performed with *IDR*.[120] IDR integrated peaks were then merged using *BEDTools*.[95] The resulting merged peak list per time window was used to construct the peak/cell matrix for each time window. Peaks present in < 1% cells were filtered. Further analysis was carried out using *Signac*.[113]

Data quality control

Cells were filtered by the following criteria: peak region fragment number > 4000 and < 40000; reads percentage in peaks ≥ 15%; blacklist ratio < 0.05; TSS enrichment score > ~1.7 (adjusted for each time window); reads per peak (reads per cell divided by peak number per cell) > 1.5 - 2.1 (adjusted for each time window); nucleosome signal < 4. Doublets were removed through a modified *scrublet* algorithm.[6] After calculation of doublet scores, cells above the 95th percentile were retained. To assess the quality of our data, peaks from our data were juxtaposed with annotated TSS sites (2 kb upstream and 200 bp downstream of TSS),[20] peaks from previous *Drosophila* embryo scATAC-seq datasets,[18] curated sets of known embryonic enhancers,[21–23] and bulk DHS peaks.[19] For each pair of datasets, the percentage of elements in one set that overlapped with elements in the other set was calculated, permitting a minimum overlap of 1 bp, and reciprocally.

Dimension reduction and clustering

Dimension reduction and clustering were performed following pipelines described in Ref[6]. Briefly, peak/cell matrix was normalized using latent semantic indexing with log-scaled term frequency. The 2nd to the 50th principal components were retained after running *RunPCA* function from *Seurat* and L2-normalization was applied to these components. To minimize batch effects, *Harmony*[102] was applied on the PCA matrix prior to running *RunUMAP* function from *Seurat* (*min.dist = 0.3*) for clustering. Louvain clustering algorithm was used and cluster resolution for each time window was set to 0.3. Two rounds of dimension reduction and clustering were performed for each time window. In both rounds, clusters with a cell count < 5% of total cell count and clusters dominated by cells from a single batch (> 80%) were discarded. After the first round of clustering, remaining clusters were used in cluster-specific peak re-calling to generate a new peak/cell matrix, including only autosome peaks. For each round of clustering, batch effects were removed with *Harmony*.[102] The resulting clusters were then re-clustered and selected for the final whole embedding.

Integrative UMAP embedding

To create an integrative peak list encompassing samples from all time windows, peak profiles from all 11 time windows after the peak re-calling were merged using *BEDTools*. With this peak list, 11 chromatin accessibility objects were created, containing retained cells after two rounds of clustering. These objects were then merged to generate a single chromatin accessibility profile. Dimension reduction and clustering were performed as described above with some modifications, including removal of peak representation filter. Clusters with < 100 cells, instead of clusters with a cell count < 5% of total cell count, were excluded. Additionally, *Runharmony* function from *Harmony* was not run on the integrative matrix to preserve differences between time windows, which may reflect epigenetic changes during development.

Tissue and cell type annotation

To annotate cell types at the tissue level, a gene activity matrix was created using *GeneActivity* function from *Signac*, which computed the counts in both gene body and promoter region (2 kb upstream of TSS) to generate a gene/cell matrix. Marker genes for each cluster were identified using the *FindAllMarkers* function from *Seurat* with the parameters *min.pct = 0.1* and *logfc.threshold = 0.1*. Based on marker genes, clusters were manually curated and annotated into cell types, tissues, and germ layers (Table S4). Tissue subclustering was performed as described for scRNA-seq data.

Cell composition discrepancies between datasets

We noticed that the proportion of cells representing some tissues differed among the three datasets (Figure 1B), which was attributable to different technical procedures during sample preparation stage. For instance, the proportion of epidermis was significantly greater in scStereo-seq compared to the other two datasets. During the single-cell isolation phase for scRNA-seq and scATAC-seq, the chorion (the outermost layer of the eggshell with dorsal appendages at the anterior end) was removed from embryos as it contributes significant impurities to the cell suspension. Conversely, for the scStereo-seq samples, the chorion was left intact to facilitate the handling of embryos and the determination of orientation during embedding. Epidermal cells derived from or attached to the chorion were thus retained in scStereo-seq dataset but largely lost in the other two. There was also a significant decrease in the fraction of epidermis and muscle cell types in scRNA-seq data compared to scATAC-seq (Figure 1B). This discrepancy stemmed from technical limitations in capturing muscle (likely due to their syncytial characteristics) and epidermal cells with our single cell suspension generation procedure. scATAC-seq, on the other hand, fractioned cells for single nuclei during library preparation and better captured these two cell types. Additionally, due to drastic morphological differences between larval and embryonic tissues, cells representing each tissue also varied between them.

Developmental age inference

Necessity of age inference and matching

Considering the rapid spatiotemporal gene expression changes during embryogenesis, it is crucial to confirm that the developmental ages of embryo samples were matched before integrating multi-omics data. Age matching involving scStereo-seq samples was also necessitated by instances where mated female flies might retain embryos in their reproductive tract for some time between fertilization and egg laying (“egg retention”)26, leading to possible deviations of the actual developmental age from the sample collection windows in individual scStereo-seq samples. scRNA-seq and scATAC-seq data were acquired from several hundred embryos per sample batch. Sample developmental age matched collection window for most embryos, and the substantial sample size mitigated the influence of sporadic female egg retention.

Age inference of scRNA-seq and scStereo-seq data

Inference of developmental age for each cell in scRNA-seq and scStereo-seq data was performed by *RAPToR Drosophila* reference[27]. For scRNA-seq data, normalized gene expression matrices of single cells were used as input; for scStereo-seq data, pseudo bulk expression matrices of entire embryo samples were used as input. Overlapping genes between our data and reference data were used for age inference. *ae* function was used to generate age predictions from input. The inferred age of scStereo-seq data aligned well with the collection window for most samples and matched better with nuclear staining morphologies.

Age inference of scATAC-seq data

Inference of developmental age for each cell in scATAC-seq data was performed following a neural network-based algorithm described in Calderon et al.[6]. The peak/cell matrix described above was used to train a new model from our own dataset. In brief, peak/cell matrix was divided into 11 partitions and 10 of them were used for model training, reserving 1 partition for testing. The midpoint of each time window was considered as the developmental age. The constructed model was then used to infer the developmental age of all cells.

Spatial TF enrichment analysis

To analyze spatial changes in TF regulon activities, enrichment scores for all 815 known *Drosophila melanogaster* TFs were calculated using *pySCENIC*[109] on imputed scStereo-seq matrix. Results were visualized with *Plotnine*[121] by overlaying registered sections. Gene expression patterns were mapped along the anterior-posterior (*x*) and dorsal-ventral (*y*) axes, with transparency inversely proportional to TF regulon activity levels.

Integration of scRNA-seq and scStereo-seq data

Gene expression imputation

NovoSpaRc[28] was used for gene expression imputation in scStereo-seq data based on scRNA-seq data. Single cells within 1 h difference of inferred age in scRNA-seq data were used for imputation of each scStereo-seq sample. Gene expression matrices from scRNA-seq data were used as the primary input for *NovoSpaRc* and scStereo-seq gene expression matrices were used as reference atlas. Optimal transport of cells to their spatial locations was computed with the following parameters: *alpha_linear = 0.8, epsilon = 5e-3*.

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

Label transfer of scRNA-seq annotation

After dimensionality reduction, top 20 dimensions of scRNA-seq data were used for label transfer. Canonical correlation analysis (CCA) was performed using *FindTransferAnchors* function from *Seurat* to identify corresponding anchors between scRNA-seq and scStereo-seq data. *TransferData* function was used to transfer annotations from scRNA-seq data to scStereo-seq data. *MapQuery* function was used to integrate embeddings of scRNA-seq and scStereo-seq data and project them in the same UMAP space for visualization. Given the substantial developmental changes and technical differences, for larval midgut samples (except L3 late), label transferred and re-annotated embryo scStereo-seq midgut cell bins, instead of scRNA-seq midgut cells, were used as the reference for label transfer. Cell clusters with low transfer confidence scores were manually re-annotated. Label transfer was performed on L3 late midgut with similar methods, with label transferred and re-annotated L3 early midgut as the reference.

Neighborhood enrichment

To calculate spatial enrichment scores of each cell type within each scStereo-seq sample and generate heatmap and bar plot visualization, *squidpy.pl.nhood_enrichment* function from *Squidpy*[122] was used with the mode parameter *zscore*.

Integration of scRNA-seq and scATAC-seq Data

CCA data integration

scRNA-seq data and scATAC-seq data of three germ layers (ectoderm, mesoderm, and endoderm) were integrated for analysis. The gene activity matrix for the scATAC-seq data was calculated using *GeneActivity* function from *Signac*. Both scRNA-seq expression matrix and scATAC-seq gene activity matrix were normalized using *NormalizeData* function from *Seurat*. To mitigate biases introduced by different sequencing techniques, CCA was performed using *FindTransferAnchors* function from *Seurat*. CCA was conducted with scATAC-seq gene activity matrix as the query and scRNA-seq expression matrix as the reference to identify transfer anchors. Leveraging these transfer anchors, *TransferData* function from *Seurat* was employed to obtain the imputed scATAC-seq matrix (*refdata = ‘scRNA data matrix‘, weight.reduction = ‘scATAC pca.l2 matrix‘, dim = 2:50*). Subsequently, imputed scATAC-seq matrix was merged with scRNA-seq data matrix. Dimension reduction was performed to co-embed them into the same UMAP space. After running *RunPCA* function, the 1st to the 50th principal components were used to run *RunUMAP* function to obtain the co-embedded UMAP matrix. 3D UMAP plots were generated with *plotly*,[123] and the third dimension of UMAP were calculated by *RunUMAP* function (*n.components = 3L*).

Annotation of integrated data

Cell type marker genes were generated using *FindAllMarkers* function from *Seurat* with the thresholds minimum cell fraction = 0.1 and logFC > 0.1. The top 50 markers were selected for manual annotation. Top 200 marker genes generated by the *COSG*[99] were also examined for additional reference, where the minimum cell fraction was also set to 0.1. Based on marker genes identified by both methods, clusters were manually curated and annotated (Table S4). For all plots, hex color codes of tissue cell types can be found in Table S10.

Generation of vector fields with *PhyloVelo*

Phylovelo[29] was used to generate velocity fields and infer trajectories using monotonically expressed genes. First, duplicate genes were filtered using *drop_duplicate_genes* function. Log-normalization was then performed to filter genes with < 10 read counts. Subsequently, velocities were inferred using the normalized data and the top 5% of genes with the highest Spearman correlations. *k=15* was set for the k-nearest-neighbor (kNN) graph for the final velocity embedding.

Construction of tissue differentiation trajectory

Based on cell differentiation Alta, Silhouette indexes of adjacent cell types were calculated to evaluate the divergence time between adjacent cell types using *scikit-learn*.[124] The adjacent cell types and their corresponding Silhouette indexes were used as nodes and edges, respectively, to construct an undirected graph. The layout of graph was arranged by using the Yifan Hu proportional algorithm[125] and rendered in *Gephi*.[101]

Cell differentiation potential analysis

CytoTRACE[52] was utilized to estimate the differentiation potential of single cells with default parameters. A sample size of 3,000 were used. *plotCytoTRACE* and *plotCytoGenes* functions were used to visualize the results.

Signal pathway activity analysis

Gene components of signaling pathways were obtained from *FlyphoneDB*.[30] scRNA-seq data were filtered by the following criteria: cells with $\geq$ 200 genes, and genes present in $\geq$ 3 cells. Filtered data were then subjected to preliminary normalization using *scanpy.pp.normalize_per_cell* function, followed by log transformation with *scanpy.pp.log1p* function. *scanpy.pp.scale* function was applied to perform further data standardization. *scanpy.tl.score_genes* function was used to assess the expression scores of major genes from each signaling pathway in individual cells. *sns.clustermap* function was used to compute the mean expression scores of each pathway across cell types and collection windows and generate heatmaps.

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

TF motif enrichment for GRN analysis

Overlapping genes between top 200 markers from *Seurat* and top 200 markers from *COSG* in each cell type in integrated scRNA-seq/scATAC-seq data were selected as potential TF target genes. Genomic regions of potential target genes were overlapped with differentially accessible peaks (DA peaks, identified by *FindAllMarker* function of *Seurat*, using test "*LR*" and parameters *minimum cell fraction = 0.1* and *logFC > 0.1*) specific to each cell type. Only peaks annotated as promoter-TSS peaks by the *annotatePeak.pl* program from *Homer*[103] were used for motif enrichment. Motifs from the *CIS-BP* database[98] were enriched in the selected peak regions using the *FindMotifs* function from *Signac*. A *p*-value cutoff of $< 10^{-4}$ was used to filter significant motifs. Activity scores were then generated for these significant motifs. After applying *AddMotifs* function in *Signac*, motif activity score matrix was created using *RunChromVar* function from *Signac*. Venn plots of target genes overlapped between two TFs were created with *VennDiagram* package.[126]

Nomination of cell type-specific TFs

Cell type-specific marker genes within each germ layer were identified with *FindAllMarker* function of *Seurat* (*minimum cell fraction = 0.1*, *logFC > 0.1*, *p.adjust < 0.05*). Top 200 up and down-regulated markers were selected. Next, genomic regions 2kb up- or downstream of TSS regions of these genes were overlapped with with peak regions (fragment number > 0) and overlapping peaks were used for motif enrichment. Motifs from the *CIS-BP* database were enriched in the selected peak regions using the *FindMotifs* function from *Signac*. The enriched motifs were combined with differential motif results for TF nomination.

Differential motif analysis

Differential motif activities between each cell type or each lineage were calculated by *FindAllMarkers* function of *Seurat*, setting the average difference calculation method as *mean.fxn = rowMeans*.

GRN analysis with *Pando*

Pando[47] was used to generate GRN in desired cell groups. For each cell type, motifs retained after enrichment analysis were used as input for the motif position weight matrix (PWM) and overlapping genes between top 200 markers from *Seurat* and top 200 markers from *COSG* in each cell type were used as target genes input. To infer GRNs, TF correlation threshold was set to be ≥ 0.1 and *Signac* method was used for the peak-to-gene assignment. To identify TF-target gene modules, *p*-value threshold was set to 0.05. The number of variables in the model and *R-square* threshold were adjusted based on the total number of modules found (*nvar = 10*, *rsq = 0.2*, or *nvar = 1*, *rsq = 0.05*). After module inference, modules with *p*-values > 0.05 or *p*-values $> 10^{-4}$ were discarded, depending on the total number of modules identified. GRN and TF graphs were then visualized.

Spatial stratification of CNS cell types

To examine the layered distribution of CNS cell types, 3D VNC models were first stratified with *spateo.dd.digitize* function along the D-V axis. The digital layers were further divided into equal bins with *pd.cut* function. Cell type distribution and proportion within each digital bin was calculated.

CNS morphometric analysis

Morphometric trajectories between two scStereo-seq samples were modeled with *spateo.tdr.cell_directions* from *Spateo* with default parameters. Each cell from the earlier time point was assigned to their most likely counterpart from the later time point to generate morphometric vector fields with *SparseVFC* algorithm through *spateo.tdr.morphofiled* function. Subsequently, geometry quantities were calculated with *spateo.tdr.morphofield_acceleration*, *spateo.tdr.morphofield_curvature*, *spateo.tdr.morphofield_curl*, and *spateo.tdr.morphofield_torsion* functions. A general linear model (GLM) regression was performed with *spateo.tl.glm_degs* function to identify genes with highest correlation with morphometric property changes. Gene set enrichment was then performed with *PANGEA*.[105]

Sub-clustering and re-annotation of endoderm scRNA-seq data

Endoderm cells were isolated from the scRNA-seq *Seurat* object using *subset* function. Cell clusters with germ layer annotation of endoderm were extracted, and the cluster annotated as "hindgut anlage" was excluded. The dimensionality of the data was reduced first by PCA (30 components) on the top 3,000 most variable genes, and then further by UMAP (*dims = 1:30*). Cell clusters were identified first with the *Seurat FindNeighbors* function (*dims = 1:30*), and then *FindClusters* function with a resolution of 1.4. This process resulted in a total of 37 subclusters. Marker genes for each cluster were identified using *Seurat FindAllMarkers* function (*only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.5*). Clusters were annotated based on these markers. Clusters annotated as "visceral muscle", "neurons", "plasmatocytes", and "proventriculus" were then excluded (Table S4).

Midgut marker gene enrichment

Marker genes of each cell subcluster were generated using *Findmarker* function of *Seurat* package with parameters *logFC > 0.5* and *pValue < 0.05*. Top 100 markers were used. To screen for more cell type-specific markers, *COSG* package was also used to generate marker genes of each cell type. Common marker genes identified with both methods were kept for the following analysis. Gene ontology and pathway enrichment were identified with *PANGEA*.

Gene module identification in midgut cell types

Hotspot[71] was utilized to identify co-expressed gene modules in endoderm cells. A weighted kNN graph with 30 neighbors was computed and genes were ranked in descending order based on their *Z* scores. The top 2,000 genes were selected and genes with an auto-correlation FDR > 0.05 were discarded. Gene modules were generated through agglomerative clustering, with the minimum number of genes per module set at 40. 17 modules were identified, and 376 genes were not assigned to a module. *Hotspot* eigengene module scores were determined by *calculate_module_scores* function. Mean scores of each module were calculated based on cell types and visualized using *clustermap* function in *seaborn*.[127]

RNA velocity field analysis

Based on re-annotated endoderm cells in scRNA-seq data, *Velocyto* command line interface[115] was used to calculate spliced and unspliced transcripts of captured genes. *Dynamo*[63] was used to model differentiation dynamics of cell trajectories. Raw count matrices of spliced and unspliced transcripts were processed with *recipe_monocle* function in *Dynamo*. The top 2,000 highly variable genes were used as feature genes for dimension reduction using UMAP and default parameters. Kinetic parameters were calculated with *dynamo.tl.dynamics* function. Velocity vector flows were projected to 2D UMAP space and visualized with *dynamo.tl.cell_velocities* function. Continuous vector fields were generated in the UMAP space using *vf.VectorField* function. Velocity pseudotime of single cells were determined with *ddhodge* function. Expression patterns of genes of interest were plotted against single-cell pseudotime. Using the vector fields, speed, acceleration, and curvature scores for single cells were calculated with *vf.speed*, *vf.acceleration*, and *vf.curvature* functions, respectively.

Gene module analysis in midgut regions

A set of 1,500 regional marker genes from Ref[70] were used for midgut region analysis. These markers were overlapped with genes captured in scStereo-seq data and 1,452 of them were retained to create a kNN graph using *hs.create_knn_graph* function (*weighted_graph=True, n_neighbors=100*) from *Hotspot*. A local correlation matrix was constructed to assess the correlation among cells within each midgut region. The top 50 marker genes of each region were selected for local correlation calculation with *hs.local_correlation_z* function. Gene modules were filtered to retain the ones that contained $\geq$ 8 genes and had a false discovery rate (FDR) < 0.05. The similarity between gene modules and midgut regions were evaluated by similarity scores, which calculated the ratio of region-specific genes in each module. Identified midgut regions were filtered to retain the ones that contained $\geq$ 20 cells. Due to significant developmental gaps during larval stages, midgut regions were inferred independently for each sample with the same workflow.

Fluorescence *in situ* hybridization

FISH was performed following protocols described in Fly-FISH database[17] with the following modification: Alexa Fluor 488 Tyramide SuperBoost Kit, goat anti-mouse (Invitrogen, B40922) or Alexa Fluor 594 Tyramide SuperBoost Kit, streptavidin (Invitrogen, B40935) was used in the signal developmental step after primary antibody incubation following vendor protocols. Primers to generate RNA probes from embryo cDNA are listed in Table S9. Samples were imaged with a Zeiss LSM 980 NLO confocal microscope with a 10 × objective. Imaging results were processed and analyzed with ImageJ.[128]

Generation and assaying of transgenic flies

For copper-cell-specific knockdown and morphology observation, *lab-Gal4; UAS-GFPnls* strain was crossed with *UAS-mCherry-shRNA* or *UAS-exex-shRNA* strain. When indicated, larvae were transferred to culture media containing 500 μM $CuSO_4$ for 24 h before imaging. Dissected L3 midgut samples (from gastric caecum to the branching point of Malpighian tubules) were incubated in PBS containing 5 μg/mL Hoechst 33342 for 10 min and rinsed with PBS before being mounted in halocarbon oil 700. For $exex^{KK30}$ larvae, the remaining carcass was used for DNA extraction and genotyping to determine homozygosity. Samples were imaged with a Zeiss LSM 980 NLO confocal microscope with a 5 × or 10 × objective. For sample examination at lower magnification, dissected L3 midgut samples were imaged with a stereo microscope. Imaging data was processed with ImageJ and the number of GFP positive cells were manually counted. Image capture and processing parameters were set identically between control and experimental groups.

RT-qPCR

For evaluation of knockdown efficiency in copper cells, copper cell segments from L3 midgut were excised and subjected to RNA extraction. RNA extraction was performed with RNA Isolator (Vazyme R401) following vendor's protocol. For each sample, 1 μg total RNA was subjected to reverse transcription using HiScript II Q Select RT Supermix (Vazyme R233) following vendors' protocol. cDNA products were diluted 10 × and used as qPCR template. qPCR was performed with ChamQ SYBR qPCR Mastermix (Vazyme Q311) following vendor's protocol. qPCR primer pairs used in this study were designed by FlyPrimerBank.[117] Primer sequences are listed in Table S9. For each primer pair, qPCR of serial dilution of cDNA samples was performed to generate a standard curve. Gene expression levels were quantified based on Ct values (normalized to internal control *αTub84B*) and standard curves.

Please cite this article in press as: Wang et al., A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation, Cell (2025), https://doi.org/10.1016/j.cell.2025.05.047

Bulk RNA-seq library prep and analysis

Extracted RNAs were subjected to RNA-seq library preparation using VAHTS Universal V8 RNA-seq Library Prep Kit for Illumina (NR605). PE150 next generation sequencing was performed at HaploX Shenzhen. The raw data were aligned to the reference genome *dm6* by *STAR* using a parameter *peOverlapNbasesMin 5* to avoid high ratio of unmapped short reads. *featureCounts*[129] was used to calculate gene region read counts. TPM (transcripts per million) method was used to normalize read counts.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details of experiments, including statistical tests, significance definition, and statistical methods, can be found in the figure legends.

Supplemental figures

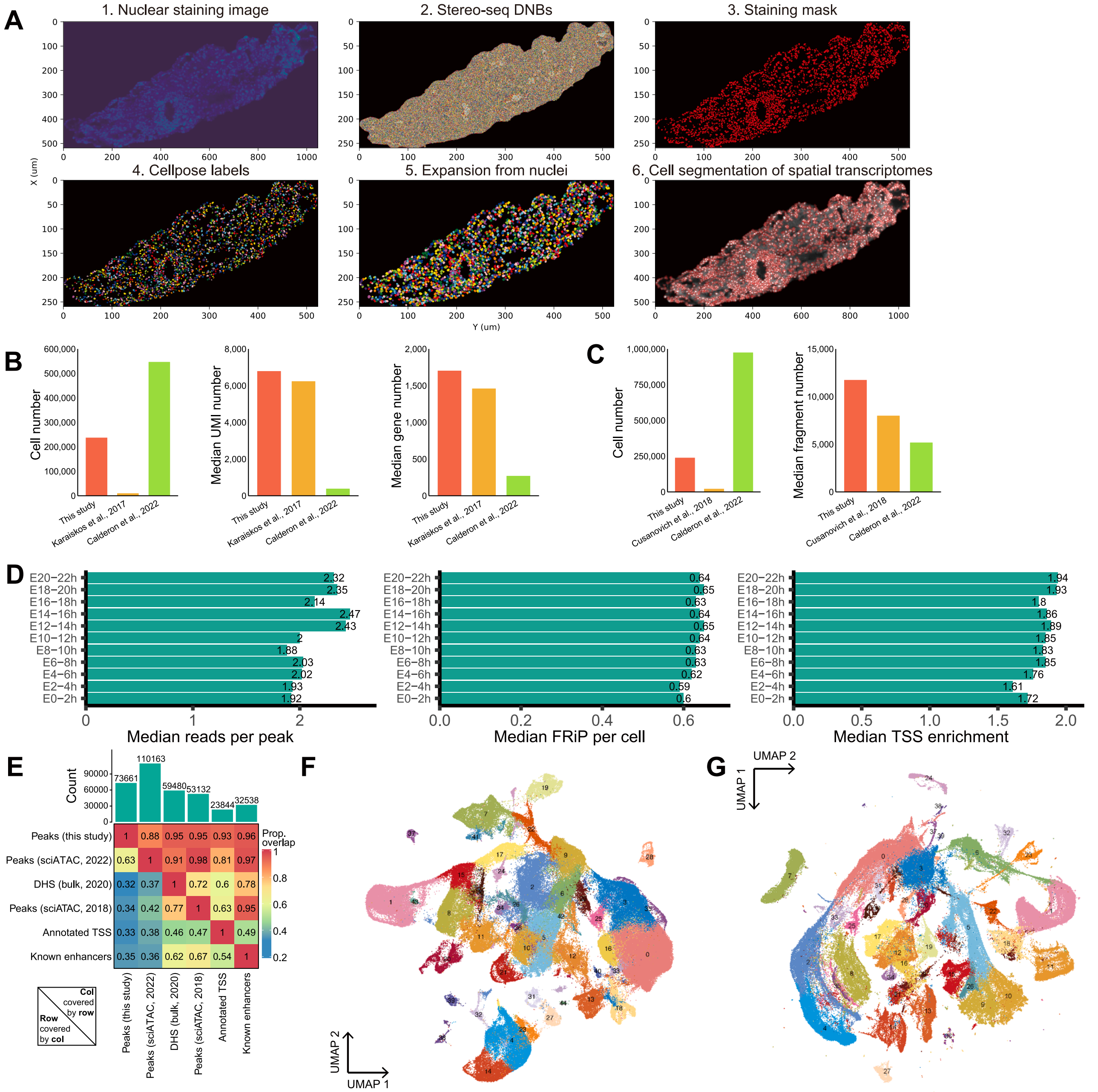


Figure S1. Data quality control and pre-processing of multi-omics datasets, related to Figure 1

(A) Cell segmentation workflow of a representative L1 early section.

(B and C) Quality benchmark of embryo (B) scRNA-seq and (C) scATAC-seq datasets in this study, showing quality statistics of this and previous *Drosophila* embryo scRNA-seq or scATAC-seq studies.

(D) Additional quality control statistics for scATAC-seq data, with bar plots showing median reads per peak, median fraction of reads in peaks (FRiP) per cell, and median TSS enrichment scores across sample batches.

(E) Heatmap showing proportion of embryo scATAC-seq peaks in this study overlapping peaks in two previous *Drosophila* embryo scATAC-seq studies, bulk DHS peaks, and peaks in known TSSs and enhancers.

(F and G) UMAP plots of aggregated (F) scRNA-seq and (G) scATAC-seq data after coarse unsupervised clustering, colored and labeled by cluster number.

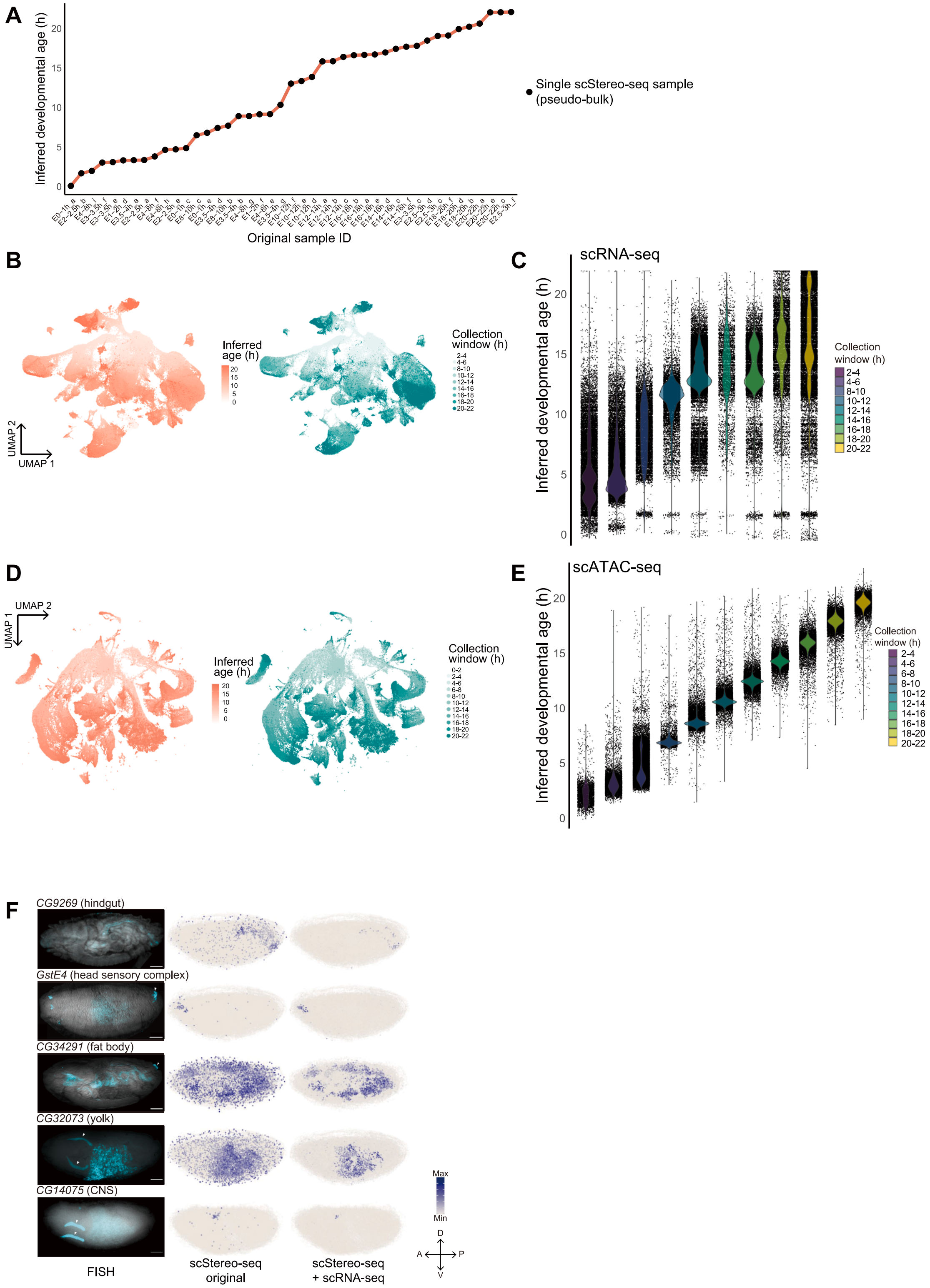


Figure S2. Developmental-age-matched integration of multi-omics data, related to Figure 1

(A) Line plot showing *RAPToR*-inferred developmental age of embryo scStereo-seq samples.

(B) UMAP plot of aggregated embryo scRNA-seq data, color coded with inferred developmental age or actual sample collection window.

(legend continued on next page)

(C) Violin plot showing *RAPToR*-inferred developmental age of cells from scRNA-seq data.
(D) Same as (B) but for scATAC-seq data.
(E) Violin plot showing neural network model-inferred developmental age of scATAC-seq samples.
(F) Additional examples for Figure 1D.

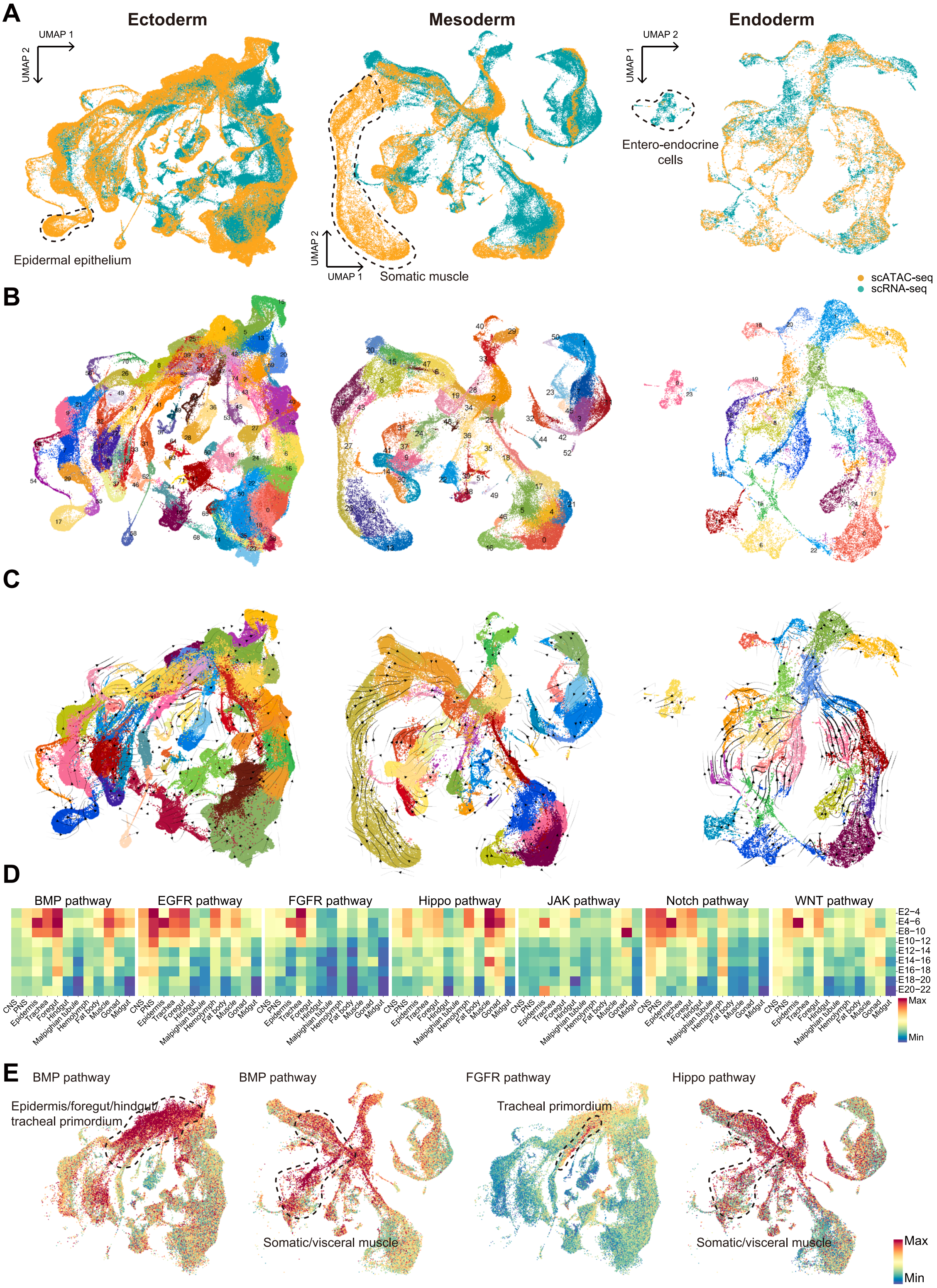


Figure S3. Construction of multi-omics tissue differentiation trajectories, related to Figure 2

(A) UMAP plots of co-embedded scRNA-seq and scATAC-seq data of three germ layers. Dashed lines mark cell clusters that are largely missing in one of the datasets, with their annotations labeled.

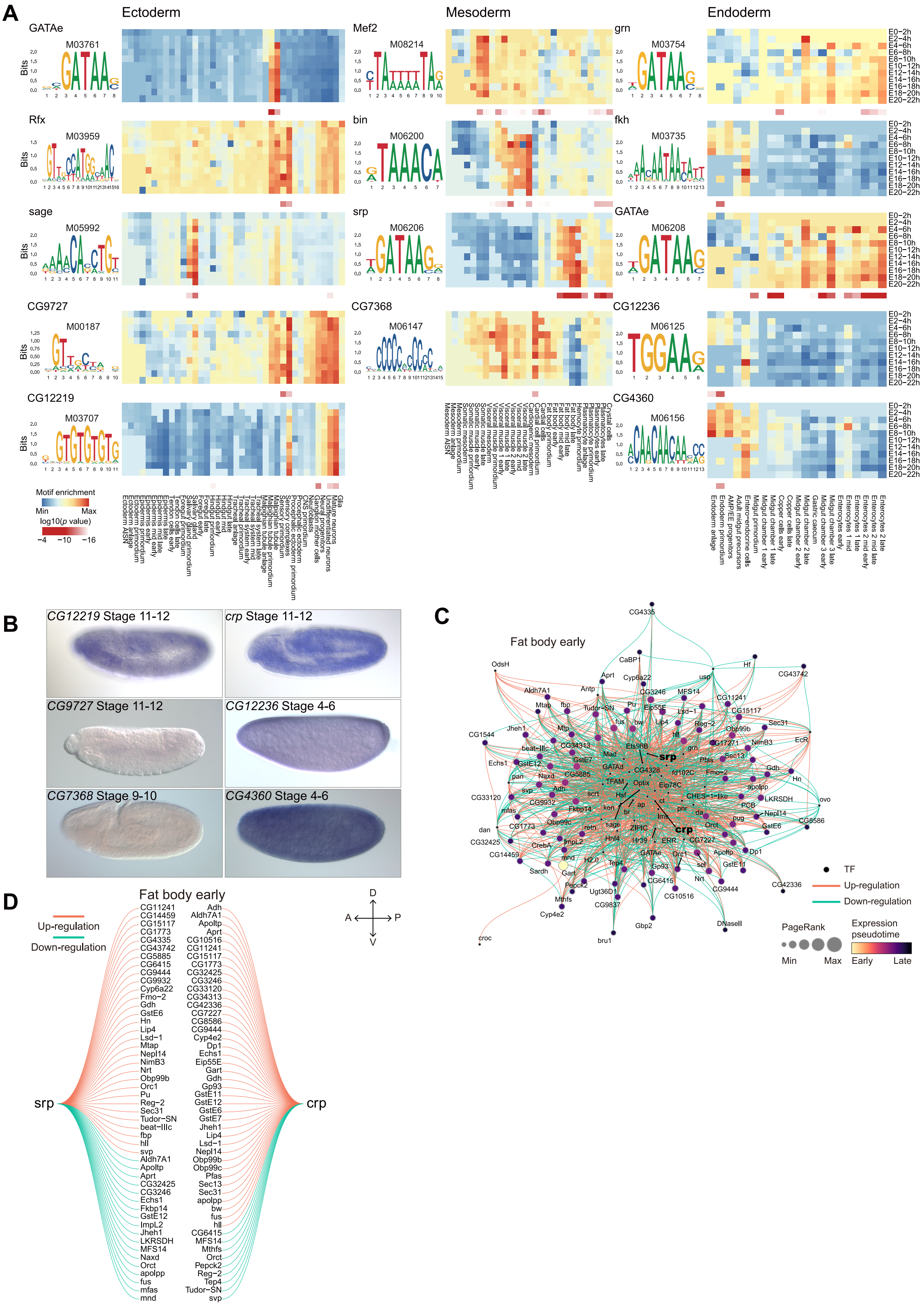
(legend continued on next page)

(B) Co-embedded UMAP plots of integrated scRNA-seq/scATAC-seq data of three germ layers after unsupervised clustering, color coded and labeled by cluster number.

(C) Velocity fields of co-embedded UMAP plots of three germ layers in (B), inferred with *PhyloVelo* and color coded with re-annotated cell types based on clustering of integrated scRNA-seq/scATAC-seq data. Velocity trajectories point backward from chronologically older to younger cells. See detailed annotations in File 3 of Data S1.

(D) Heatmap showing median gene activity scores of core components of signaling pathways across tissues, based on scRNA-seq data.

(E) UMAP plots of scRNA-seq cells in the co-embedded UMAP space, color coded with gene activity scores of core components of signaling pathways. Representative tissues enriched in signaling pathway activities are labeled.



(legend on next page)

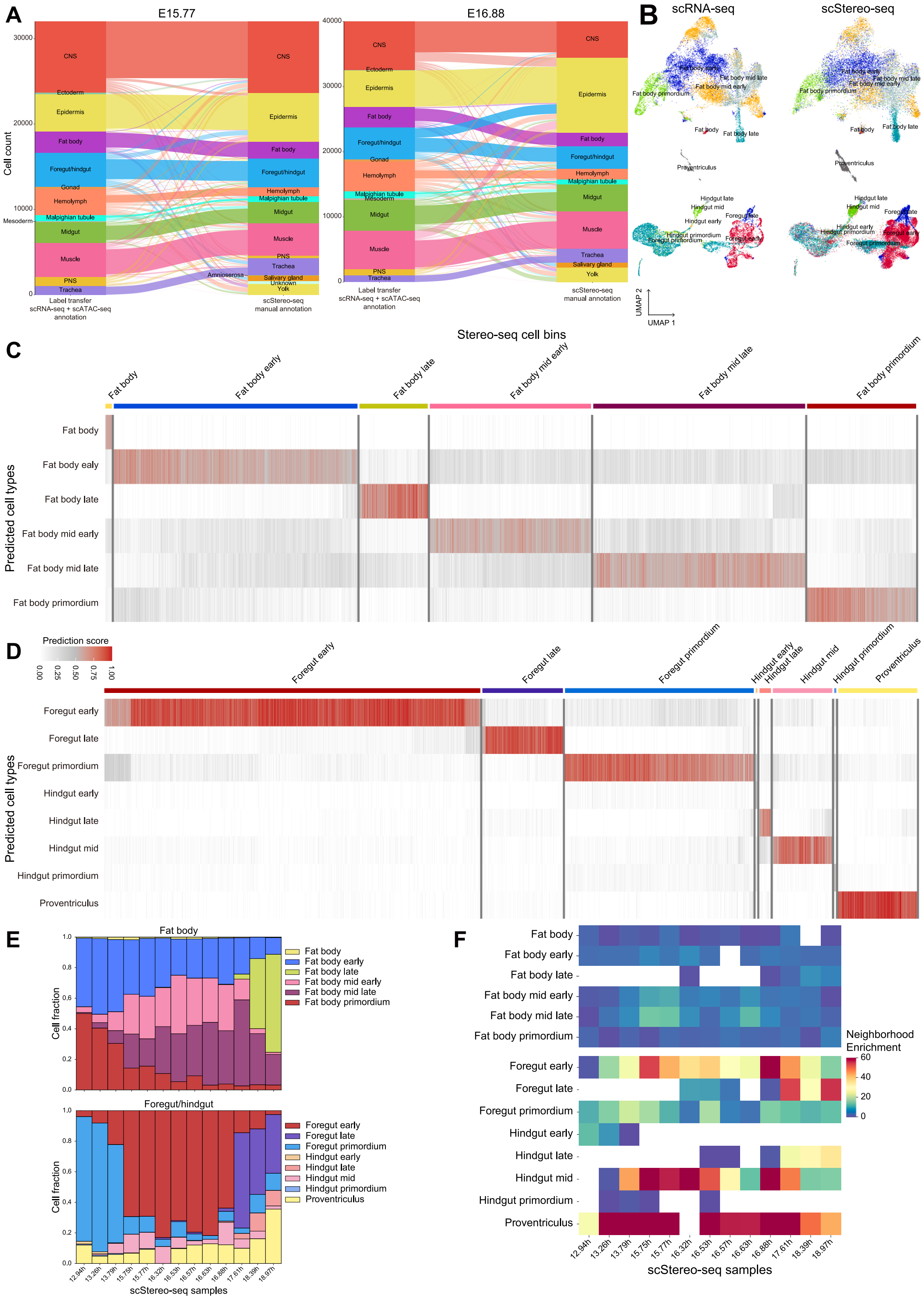
Figure S4. Transcription factor regulatory networks along tissue differentiation trajectories, related to Figure 2

(A) TF motif enrichment along tissue differentiation trajectories, showing TF binding motifs (left), motif activity heatmap (upper right), and enrichment *p* value heatmap (lower right) across tissue types and developmental stages in cells from three germ layers in scATAC-seq data. Previously characterized TFs are in bold.

(B) BDGP *in situ* patterns of less-characterized TFs in (A). A-P, anterior-posterior; D-V, dorsal-ventral.

(C) *Pando* identified GRNs of TFs *srp* and *crp* in fat body early cell cluster.

(D) *Pando* identified regulons of TFs *srp* and *crp* in fat body early cluster.



(legend on next page)

Figure S5. Spatiotemporal cell-type dynamics along tissue development trajectories, related to Figure 3

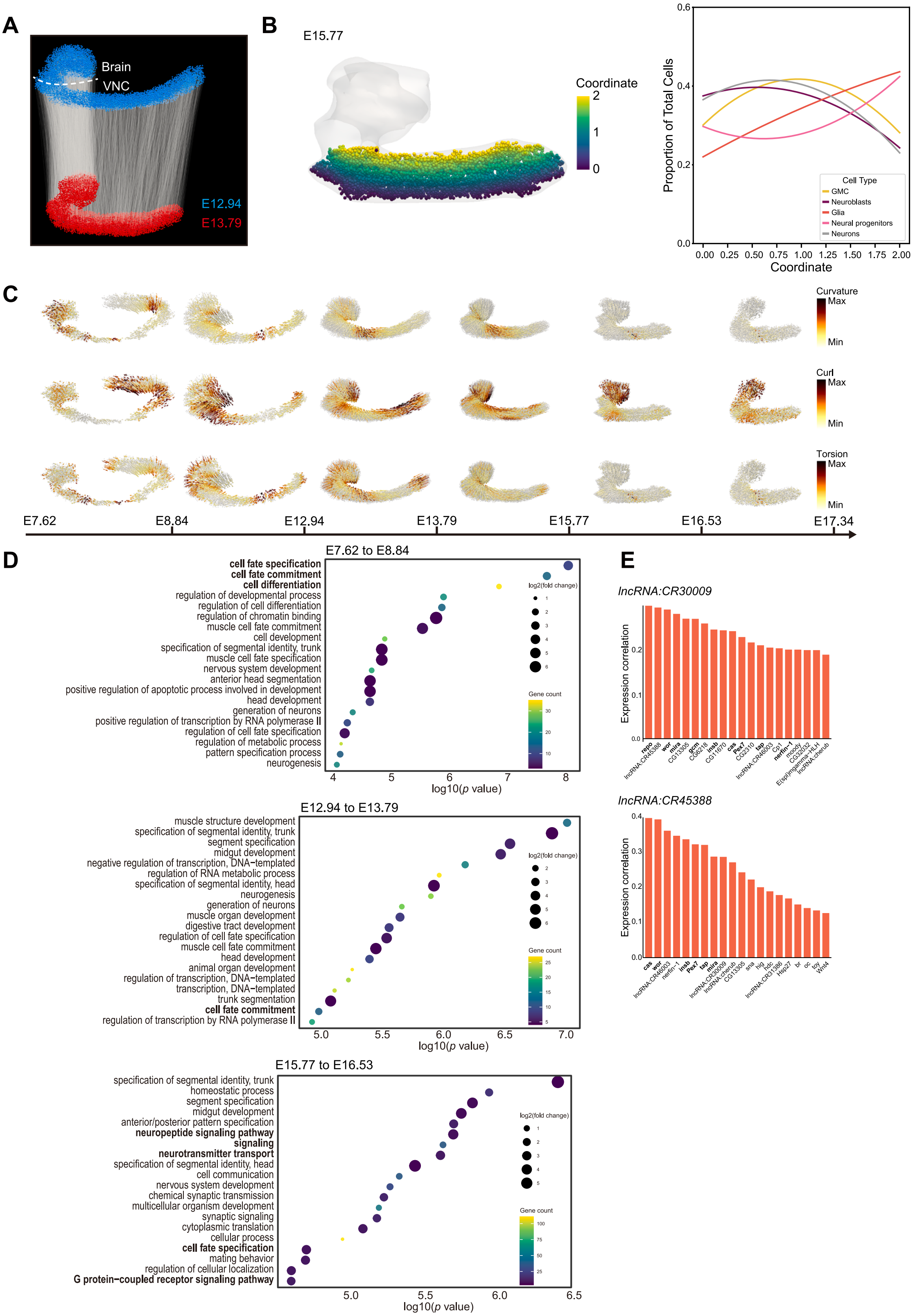
(A) Sankey plots showing agreement between scStereo-seq manual tissue annotations and transferred labels from integrated scRNA-seq/scATAC-seq data in representative scStereo-seq samples.

(B) Co-embedding of fat body and foregut/hindgut cells from scRNA-seq and scStereo-seq (pooled samples) data in the same UMAP plots, labeled with original scRNA-seq annotations or transferred annotations.

(C and D) Heatmap showing prediction scores of scStereo-seq cell bins from (C) fat body and (D) foregut/hindgut, predicted with label-transferred cell types from scRNA-seq data.

(E) Bar plots showing cell-type composition of fat body and foregut/hindgut in scStereo-seq samples. Cell types are label transferred from scRNA-seq data.

(F) Heatmaps showing neighborhood enrichment scores of fat body and foregut/hindgut cell types across scStereo-seq samples. Blank cells indicate the absence of label-transferred cell types or lack of enrichment in corresponding samples.



(legend on next page)

Figure S6. Gene expression dynamics during CNS morphometric changes, related to Figure 3

(A) Scheme of 3D model aligning and cell bin connection in CNS morphometric analysis. The boundary between brain and VNC is shown.

(B) Proportion of CNS cell types distributed along the dorsal-ventral axis of the VNC in a representative scStereo-seq 3D model. GMC, ganglion mother cells.

(C) Visualization of curvature, curl, and torsion scores of morphometric changes in CNS across the 7 scStereo-seq samples in Figure 3E.

(D) Bubble plots showing GO enrichment of genes associated with changes in four CNS morphometric scores among representative scStereo-seq samples. Pathways in bold are discussed in the main text.

(E) Bar plots showing genes that display the highest expression correlation with *lncRNA:CR30009* and *lncRNA:CR45388* in scRNA-seq data. Gene names in bold indicate neuroblast or glioblast markers.

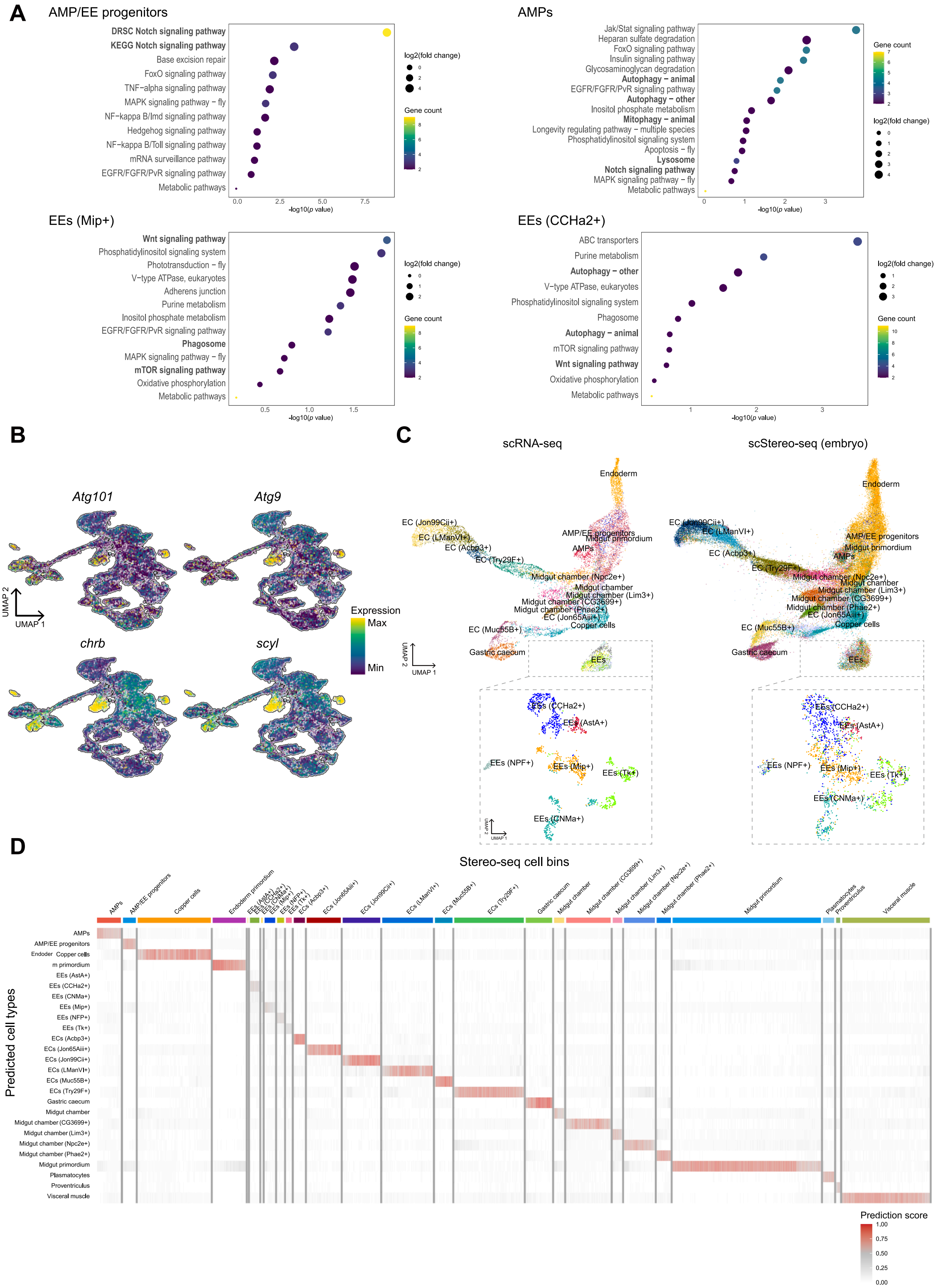


Figure S7. Diversity of embryonic midgut cell types and functions, related to Figure 4

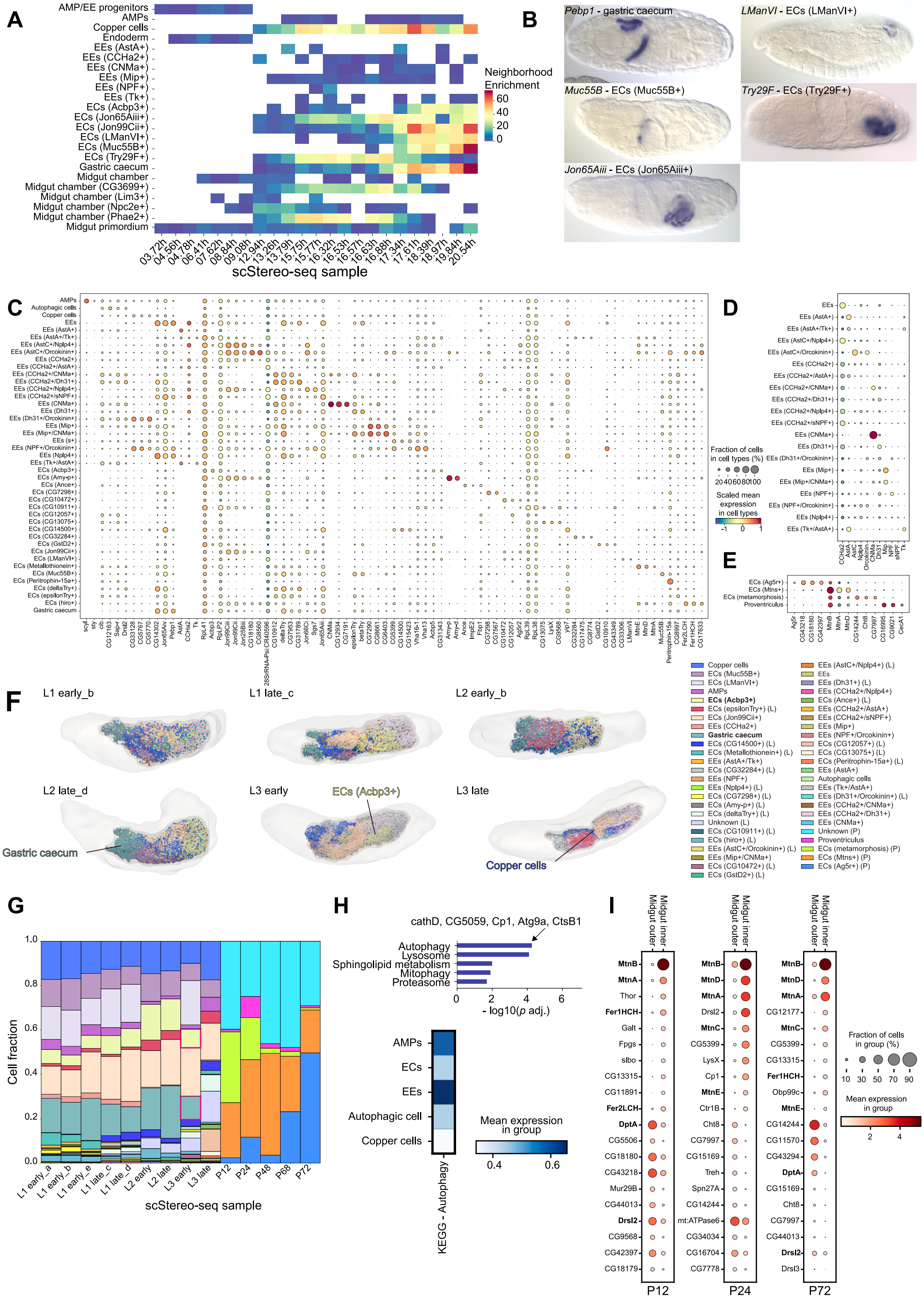
(A) Bubble plots showing Kyoto Encyclopedia of Genes and Genomes (KEGG) and *Drosophila* RNAi Screening Center (DRSC) PathON pathway enrichment of midgut cell-type marker genes from representative clusters. Pathways in bold are discussed in the main text.

(legend continued on next page)

(B) UMAP plots of embryonic midgut scRNA-seq data showing expression levels of cell death-related genes *Atg101*, *Atg*9, *chrb*, and *scyl*.

(C) Co-embedding of midgut (upper) and EEs (lower) from scRNA-seq and scStereo-seq data in the same UMAP plots. Original scRNA-seq annotations or transferred scStereo-seq annotations are labeled.

(D) Heatmap showing prediction scores of scStereo-seq midgut cell bins, predicted with label-transferred cell types from scRNA-seq data.



(legend on next page)

Figure S8. Spatial distribution of midgut cell types, related to Figure 4

(A) Heatmap showing neighborhood enrichment scores of cell types across scStereo-seq samples. Blank cells indicate the absence of label-transferred cell types or lack of enrichment in corresponding samples.

(B) BDGP *in situ* patterns of marker genes of midgut cell types with high neighborhood enrichment scores in scStereo-seq data.

(C–E) Bubble plots showing expression level and enrichment of top marker genes of (C) all larval midgut, (D) larval entero-endocrine cells, and (E) pupal midgut cell types in scStereo-seq data. AMPs, adult midgut progenitors; ECs, enterocytes; EEs, entero-endocrine cells.

(F) 3D midgut models across representative larva scStereo-seq samples, showing spatial distribution of cell types, mesh models of midgut, and mesh models of the entire embryo.

(G) Bar plot showing cell-type composition of larval and pupal midgut in scStereo-seq samples. Annotations with (L) or (P) indicate clusters identified only in larva or pupa samples, respectively. Cell types diminishing in the L3 late sample are labeled in bold and highlighted in L3 early. Cell-type color codes are the same as (F).

(H) Upper: bar plot showing KEGG pathway enrichment of up-regulated genes in the L3 late sample compared with other larval midgut samples; lower: heatmap showing mean expression levels of KEGG autophagy-related genes in L3 late midgut cell types.

(I) Bubble plots showing expression and enrichment differences between midgut inner and midgut outer clusters in representative pupa scStereo-seq samples. Genes discussed in the main text are in bold.

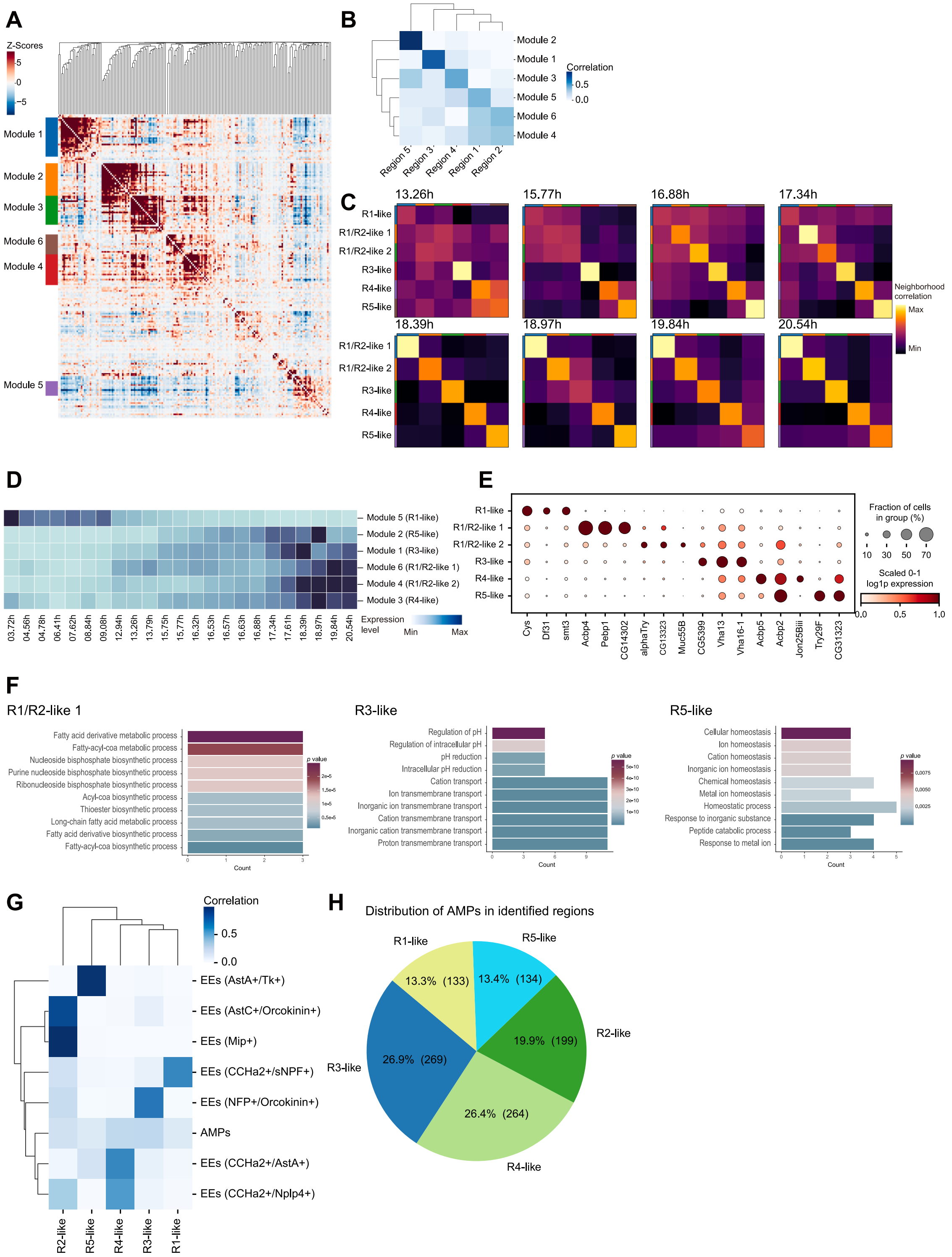


Figure S9. Functional regionalization of embryonic and larval midgut, related to Figure 4

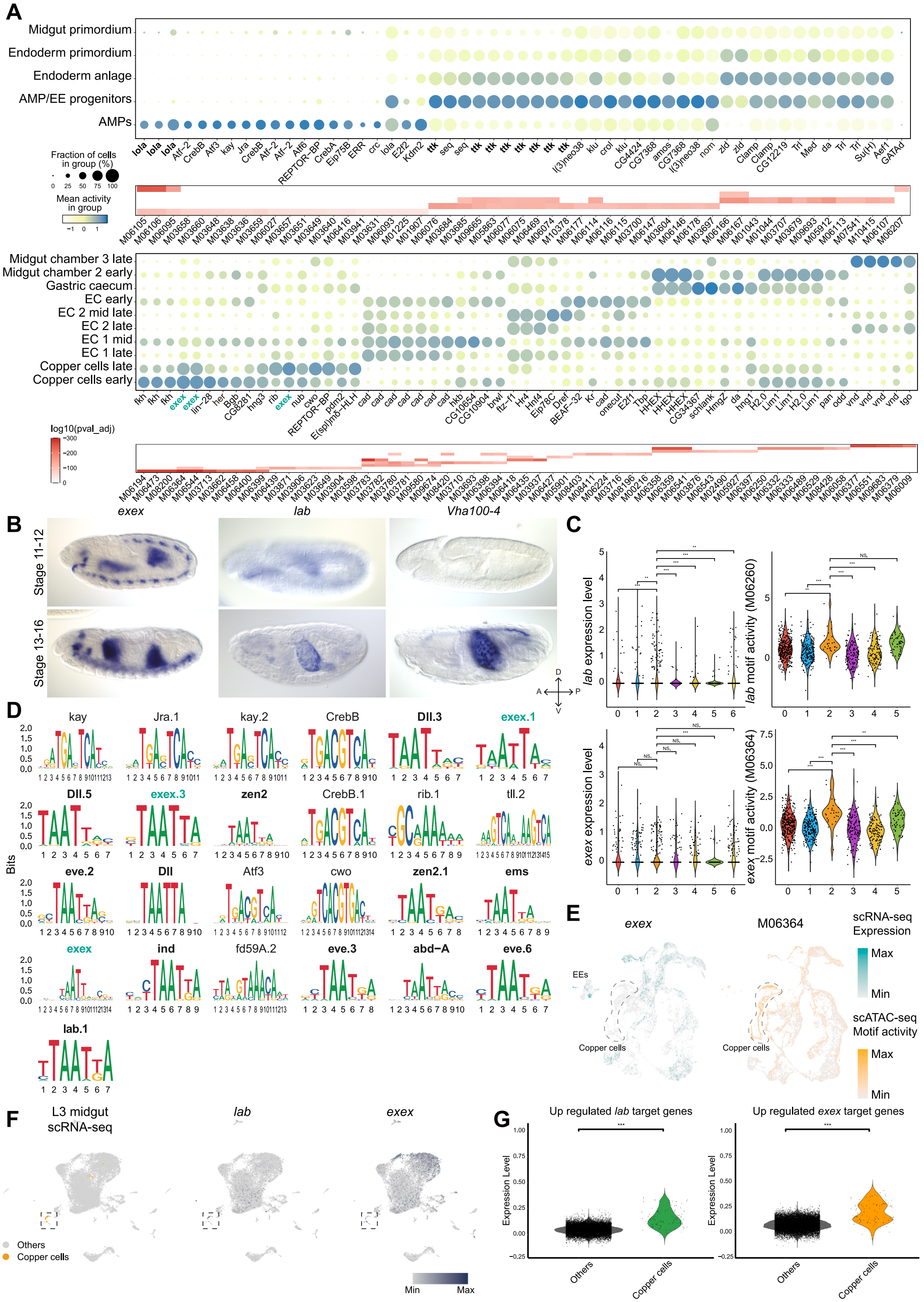
(A) Heatmap showing correlation of functional gene modules identified by *Hotspot* from adult midgut region marker genes in scStereo-seq data. Each row and each column represent a module marker gene, and the *Z* score indicates their correlation.

(B) Heatmap showing correlation between *Hotspot* identified gene modules and adult midgut region markers.

(C) Heatmaps showing neighborhood correlation of midgut regions across scStereo-seq samples.

(legend continued on next page)

(D) Heatmap showing expression level of region-related gene modules across scStereo-seq samples.
(E) Bubble plot showing expression level and enrichment of top marker genes of identified midgut regions.
(F) Bar plots showing GO enrichment of marker genes in representative regions identified in embryonic midgut.
(G) Heatmap showing enrichment of AMPs and EE subclusters in identified midgut regions in the L3 early scStereo-seq sample.
(H) Pie chart showing percentage of AMPs distributed in identified midgut regions in the L3 early scStereo-seq sample.



(legend on next page)

Figure S10. Identification of midgut copper-cell-specific regulators with multi-omics data analysis, related to Figure 5

(A) Bubble plots of top differential motif activity between lineages (upper, showing TF gene names) and heatmaps of enrichment p value (lower, showing corresponding motif codes) across early and late midgut cell types from Figure 5A. Cell-type labels are identical in y axes between upper and lower panels.

(B) BDGP *in situ* patterns of *exex*, *lab*, and copper cell marker gene *Vha100-4* in stage 11–12 and stage 13–16 embryos. A-P, anterior-posterior; D-V, dorsal-ventral.

(C) Violin plots showing transcript expression and representative motif activity of *lab* and *exex* in midgut primordium subclusters. Subcluster 6 did not include cells from scATAC-seq. Unpaired t tests were used for statistical comparison between subcluster 2 and other subclusters. NS, not significant; $^{**}p < 0.01$, $^{***}p < 0.001$.

(D) Differentially accessible TF binding motifs between copper cell lineage and other midgut chamber lineages. *lab* motifs are not enriched among top differential motifs, and their representative motif is shown in the lower left corner. Homeodomain-containing TFs are in bold.

(E) The same UMAP plots as Figure 5A but showing cells from scRNA-seq or scATAC-seq separately, color coded with gene expression levels of *exex* and activities of its representative binding motifs, respectively.

(F) Left: UMAP plot of L3 midgut scRNA-seq data from *NP1-Gal4* > *UAS-mCherry-shRNA* sample. The cell cluster representing copper cells is highlighted; right: same as the left panel but color coded with expression level of *lab* and *exex*.

(G) Violin plots showing expression level of *lab* and *exex* target gene modules in L3 midgut scRNA-seq data. Unpaired t tests were used for statistical analysis. $^{***}p < 0.001$.

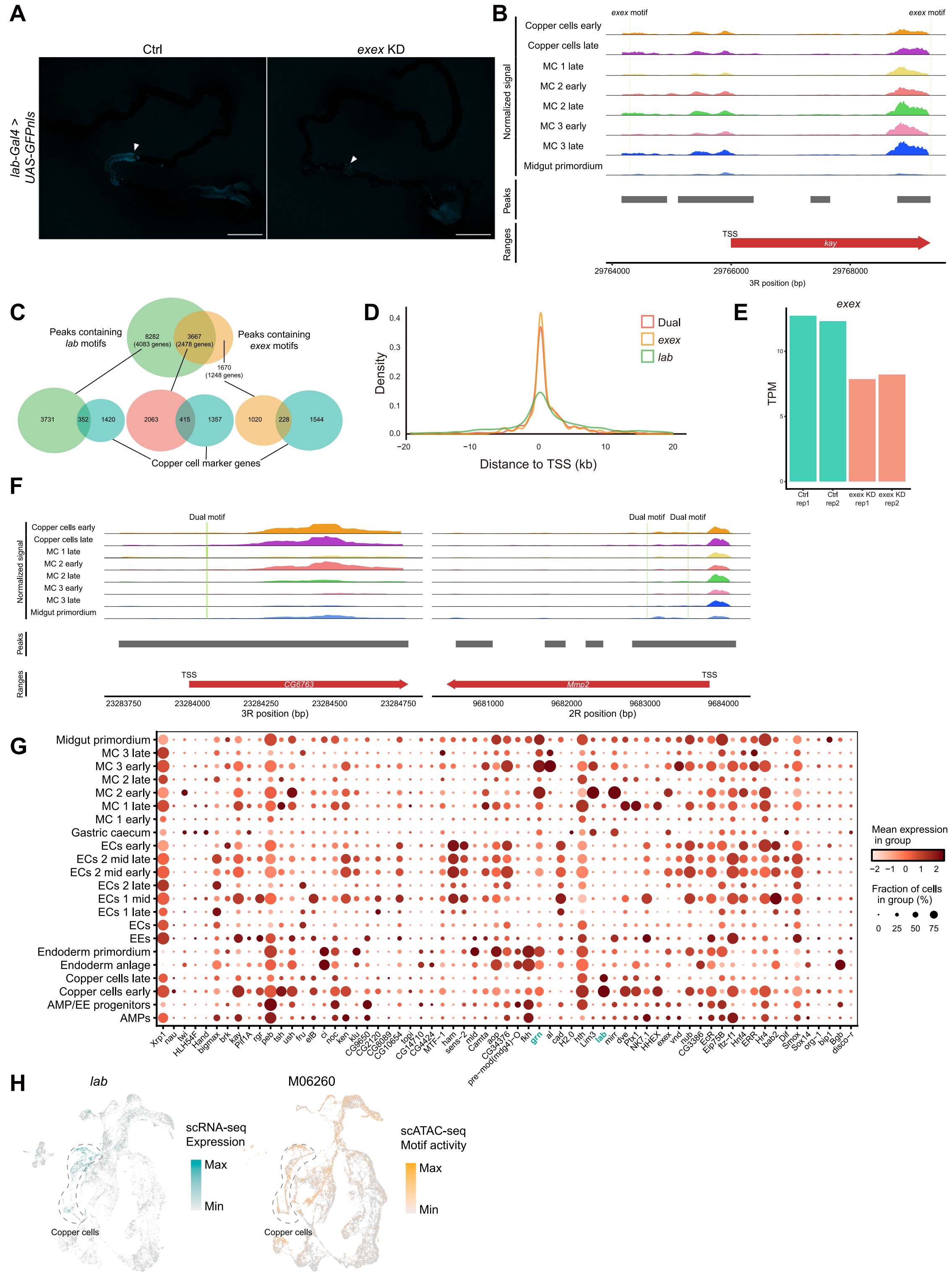


Figure S11. Identification of exex as a copper-cell-specific regulator with multi-omics data analysis, related to Figure 5

(A) Representative fluorescence stereoscope images of the entire midgut samples from control and *exex* copper-cell-specific RNAi KD L3. Arrowheads indicate copper cell regions. Scale bars, 1 mm.

(B) chromosomal regions around TSS of *kay*, showing detected peaks in midgut cell types from scATAC-seq data and binding motifs of *exex*.

(legend continued on next page)

(C) Upper: Venn diagram showing overlap between peaks containing *lab* and *exex* motifs. Number of nearest genes of the peaks are labeled. Lower: Venn diagrams showing overlap between genes in the upper panel and copper cell marker genes.

(D) Line plot showing distances between TSS of copper cell marker genes and peaks containing *exex*, *lab*, or dual motifs in scATAC-seq data.

(E) Bar plots showing TPM of *exex* in control and *exex* KD samples from bulk RNA-seq data.

(F) Same as (B) but for *Mmp2* and *CG6763*.

(G) Bubble plot showing expression level and enrichment of top TF marker genes of midgut cell types.

(H) The same UMAP plots as Figure 5A but showing cells from scRNA-seq or scATAC-seq separately, color coded with gene expression levels of *lab* and activities of its representative binding motifs, respectively.