A single-cell 3D spatiotemporal multi-omics atlas from *Drosophila* embryogenesis to metamorphosis

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

[1]BGI Research, Hangzhou 310030, China

[2]Shenzhen Key Laboratory of Gene Regulation and Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen 518005, China

[3]Department of Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen 518005, China

[4]Department of Pharmacology, School of Medicine, Southern University of Science and Technology, Shenzhen 518005, China

[5]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

[6]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, Berlin 14195, Germany

[7]China National GeneBank, BGI Research, Shenzhen 518083, China.

[8]BGI Research, Shenzhen 518083, China

[9]Joint Laboratory of Guangdong-Hong Kong Universities for Vascular Homeostasis and Diseases, School of Medicine, Southern University of Science and Technology, Shenzhen 518005, Guangdong, China

[10]These authors contributed equally

[11]Lead contact

*Authors for correspondence

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, 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. By integrating these multi-dimensional data, we constructed cell state trajectories that uncover the detailed profiles of tissue development. With a focus on the central nervous system (CNS) and midgut, we dissected the spatiotemporal dynamics of gene regulatory networks, cell type diversity, and morphological changes from a multi-omics perspective. This extensive atlas provides an unprecedentedly rich resource and serves as a systematic platform for studying *Drosophila* development with integrated single-cell data at an ultra-high spatiotemporal resolution.

Key words: *Drosophila* development, single-cell multi-omics, spatial transcriptomics, Stereo-seq, Flysta3D

INTRODUCTION

The advances in single-cell multi-omics technologies have revolutionized our understanding of biological processes, revealing cell-specific functional heterogeneities that underlie the complex physiologies of development, aging, and diseases. To date, the functional profile of a single cell can be characterized across multiple dimensions, including its cell surface epitopes, transcriptome, epigenome, and proteome (reviewed in Ref[1]). The development of spatial multi-omics techniques further added spatial context to these dimensions of information (reviewed in Ref[2]), and progress has been made in integrating these multi-modal data to construct a panoramic profile of context-specific functions of single cells and their communications with one another (reviewed in Ref[3]).

Drosophila melanogaster has long been a fundamental model organism for genetics and developmental biology research. Recent single-cell multi-omics studies have highlighted the versatility of *Drosophila* in characterizing transcriptomic and epigenomic dynamics of individual cells during embryogenesis[4,5], tissue development[6–8], tissue regeneration[9], and systemic aging[10]. These studies generated rich resources for dissecting the multi-omics profiles of various tissues at single cell precision across developmental stages. Nevertheless, the spatial context of such single-cell omics data is crucial to understanding their biological relevance but is often lost during standard single-cell sequencing procedures.

Embryogenesis is an intricately regulated process that transforms a totipotent zygote into a fully formed embryo with functional organs. Over the past several decades, research into *Drosophila* embryogenesis has yielded invaluable insights into this meticulous process and many of its features that are conserved in mammals (reviewed in Ref[11]). Recently, a few studies have addressed *Drosophila* embryogenesis from the perspective of single-cell multi-omics[4,5], but focused only on a few developmental time points or are limited in genome coverage. Until recently, genome-wide spatial transcriptomic profiling of developing *Drosophila* was lacking due to the miniature sizes of *Drosophila* samples and resolution limit of spatial transcriptomic techniques. Previously, we utilized spatial enhanced resolution omics sequencing (Stereo-seq)[12], a sequencing- and patterned DNA nanoball (DNB) array-based spatial transcriptomic platform with high spatial resolution and sensitivity, to address this gap. Using Stereo-seq, we generated 3D spatiotemporal transcriptomic maps of *Drosophila* late-stage embryos and larvae and analyzed the development of tissues within their actual 3D spatial context[13].

Here, we expanded our previous spatiotemporal transcriptomic atlas of *Drosophila* to cover its developmental lifespan from embryo to pupa. Using Stereo-seq and *Spateo*, a computational pipeline designed to analyze single-cell multi-modal spatial transcriptomic data[14], we reconstructed 3D transcriptomes at single cell spatial resolution. We further complemented embryo 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* embryos that includes transcriptomic and epigenomic information within an ultra-high-resolution spatial context. The data in this single-cell spatiotemporal multi-omics atlas of *Drosophila* development are curated in our database, Flysta3D, for easy access.

Based on the unprecedentedly rich data resource, we established multi-omics cell state trajectories of tissue development. Along these trajectories, we systematically characterized the spatiotemporal dynamics of cell differentiation potential, signaling pathways, and transcription factor (TF) regulatory networks. Focusing on two widely studied *Drosophila* tissues, central nervous system (CNS) and midgut, we delved into their cell type diversification, gene regulatory networks, and morphological changes from a multi-omics perspective. Given that we have produced extensive multi-omics datasets for the embryonic stages, the major focus of this paper will be the analysis of embryogenesis from a multi-omics perspective. The scStereo-seq data for the larval and pupal stages are not discussed extensively here but will be accessible via Flysta3D database. Flysta3D hosts all the datasets generated in this study and provides interactive 3D visualization of gene expression patterns, TF regulatory networks, signaling pathway activities, etc. in these datasets. Our database can facilitate systematic research on *Drosophila* development with its comprehensive information and broad range of applications.

RESULTS

Single-cell 3D spatial transcriptomes of *Drosophila* from embryogenesis to metamorphosis

To construct multi-omics atlas of *Drosophila* development, we started off by expanding and enhancing the 3D spatial transcriptomes of *Drosophila* development based on our previous work. Developing embryos were collected at 0.5 to 2 h intervals throughout the ~24 h course of embryogenesis (hereafter termed based on computationally inferred developmental age, see below). Larva samples were collected at early or late time points for each of the three larval stages (hereafter termed L1 to L3 early/late). Pupa samples were collected at 12 h intervals starting from pupation (hereafter termed P12 to P72) **(Figure 1A and Table S1)**. Cryosection was performed for each sample to obtain their sagittal sections of 7 or 8 μm thickness, and all available sections of each sample were subjected to *in situ* mRNA capture, library preparation, and sequencing at the Stereo-seq platform.

The subcellular spatial resolution achieved by Stereo-seq (~500 nm distance between DNBs) was not fully utilized in our previous dataset due to the lack of cell location information. Here, to address this, we attempted to achieve single-cell spatial resolution by nucleus staining and imaging of each Stereo-seq chip before library preparation. Cell segmentation was then performed based on the location of each nucleus. After sequencing and mapping, 2D spatial gene expression matrices were aligned with segmented images. Each DNB was then assigned to a cell bin, allowing for more precise single-cell transcriptome analysis **(Figure S1A)**. We then combined cell bins from all sections of individual samples, performed unsupervised clustering based on both gene expression profiles and spatial locations **(Data S1)**, and manually annotated the clusters according to marker gene expression and spatial morphology **(Table S2)**. 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 **(Table S1)**.

With the 2D single-cell spatial transcriptomic datasets, we reconstructed the spatial transcriptomes in 3D leveraging point cloud-based modeling method in *Spateo* package, which were optimized for cell bins. This approach offered enhanced structural details compared to our previous 3D modeling results and allowed the alignment of 3D models from different time points for morphometric analysis (see below). The 3D modeling of cell bin spatial transcriptomic data effectively captured the anatomical morphology of tissues with finer details than our previous 3D models **(Figure 1B and Figure S1B)**. In our previous study on limited embryo and larva samples, we demonstrated that Stereo-seq data reproducibly captured spatial gene expression patterns that largely overlapped with established *in situ* databases[15,16], as well as those that were absent in these databases. Based on this more comprehensive spatiotemporal transcriptomic dataset, we further identified a list of 338 genes without reported spatial expression patterns in embryos and reconstructed their patterns in 3D

(Table S3). We selected 9 genes to validate their spatial expression patterns with fluorescence *in situ* hybridization (FISH) and found high consistency with scStereo-seq data in terms of spatial gene expression patterns and tissue enrichment **(Figure 1C and Figure S1C)**. These results further substantiated the power of Stereo-seq in recapitulating spatial gene expression profiles and guiding *in vivo* validation.

A single-cell spatiotemporal multi-omics atlas of *Drosophila* embryogenesis

Despite its ability to provide a better representation of single-cell spatial transcriptomes, scStereo-seq had a higher dropout rate compared to droplet-based scRNA-seq due to a reduced number of DNBs assigned to each cell bin. This limitation curtailed the ability of scStereo-seq to detect genes that express at a lower level. To overcome this drawback and to augment our single-cell 3D spatial transcriptomic data with deeper transcriptomic and epigenomic information, we collected samples at 2-hour intervals across embryogenesis and performed droplet-based scRNA-seq and scATAC-seq **(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)**. These quality control statistics in our scRNA-seq data were comparable to or better than previous *Drosophila* embryo scRNA-seq studies[4,5] **(Figure 1D)**. Additionally, we obtained 240,573 single-cell chromatin accessibility profiles with scATAC-seq, with a median of 11,772 fragments per cell **(Table S1)**. The number of fragments captured per cell and other quality control statistics in our scATAC-seq data were also comparable to or better than previous *Drosophila* scATAC-seq datasets[5,17] **(Figure 1E and Figure S1D-F)** and achieved high coverage of previously reported scATAC-seq datasets[5,17], DNase I hypersensitive sites (DHS)[18], annotated transcription start sites (TSS)[19], and known enhancer sites[20–22] **(Figure 1F)**.

With the aggregated scRNA-seq data collected across embryogenesis, we performed an initial round of coarse unsupervised clustering and generated 45 cell clusters in the uniform manifold approximation and projection (UMAP) plot **(Figure S1G)**. We annotated these clusters and classified annotations at three levels (cell type/tissue substructure - tissue - germ layer, e.g., gastric caecum - midgut - endoderm) based on marker gene expression **(Figure 1A and Table S2)**. Similarly, we performed coarse unsupervised clustering in aggregated scATAC-seq data, generating 40 distinct clusters in the UMAP plot **(Figure S1H)**. Each cluster was also annotated through inspection of marker genes **(Figure 1A and 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 1G)**.

Given the deep genome coverage of our data, we further profiled tissue cell type heterogeneity by subclustering and annotating tissue clusters from scRNA-seq and scATAC-seq data. The resolution of subclustering for each tissue was determined based on previously reported cell type complexities. Detailed cell types were annotated based on marker gene expression and literature search **(Table S4)**. Through subclustering of the scRNA-seq data, we were able to examine the specific composition of embryo tissue 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[23] and chordotonal organs[24,25] **(Figure S2)**. These subclustering results indicated that we were able to extensively characterize the major cell types in the embryo and identify several rare ones, such as adult midgut progenitors (AMPs)[26] and entero-endocrine cells (EEs)[27] in the midgut (see below). We also identified subclusters representing most of these detailed cell types in the scATAC-seq data **(Figure S3)**. While we annotated tissue subclusters to the best of our knowledge, there could still be instances where clusters were not assigned their optimal annotations. We annotated some of the ambiguous or unknown cell clusters with reference to cell types they resembled most based on marker genes, such as “neuron-like” **(Figure S2 and S3)**. Therefore, community efforts are welcome to help further specify the annotations of tissue cell types. To verify the well annotated subclusters we identified in both datasets, we compiled a list of common tissue substructure/cell type markers, which are identified in both datasets **(Table S5)** and validated the expression specificity of 3 previously unreported cell type markers using FISH **(Figure 2A)**.

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.

Developmental age-matched integration of multi-omics data

Integration of multi-omics data offers more comprehensive perspectives when searching for key molecular factors regulating tissue development. Our multi-omics data were obtained from multiple batches of samples using different techniques. Considering the rapid spatiotemporal gene expression changes during embryogenesis, it is crucial to confirm that the developmental ages of samples were matched before integrating multi-omics data. Age matching was also necessitated by the fact that mated female flies might retain embryos in their reproductive tract for some time between fertilization and egg laying ("egg retention")[28], leading to possible deviations of the actual developmental age from the sample collection windows in individual scStereo-seq samples.

To precisely stage our samples, we applied *RAPToR*, a predictive model inferring the developmental age of biological samples based on transcriptomic profiles[29], on both embryo scStereo-seq and scRNA-seq data. The developmental age of embryo scStereo-seq samples was determined by *RAPToR* inference with the entire embryo as a pseudo-bulk input **(Table S1)**. The ages inferred by *RAPToR* aligned well with the collection window for the majority of samples. However, in a few instances, there was a significant discrepancy, with *RAPToR* inferring an age notably older than that derived from the collection window, indicating the presence of female egg retention **(Table S1)**. We further validated the age of these samples by inspecting their nuclear staining morphologies and found better agreement with *RAPToR* inference than collection window. Consequently, these samples are denoted by their *RAPToR*-inferred developmental age rather than the actual sample collection time window (e.g., E15.75 refers to an embryo sample with the inferred developmental age of 15.75 h). The resulting set of 43 embryo scStereo-seq samples we collected comprehensively covered *Drosophila* embryogenesis **(Figure S4A)**. The *RAPToR*-inferred developmental age of single cells from scRNA-seq data showed overall good agreement with their actual sample collection window **(Figure 2B and Figure S4B)**, but with significant tissue-dependent variations within each stage **(Figure S4C)**, likely because *RAPToR*, a model trained on bulk RNA-seq data, lacks cell type specificity for scRNA-seq data. Similar variations were also observed in age inference of individual cell bins of scStereo-seq samples **(Figure S4D)**. To infer the developmental age of cells from scATAC-seq data, we employed a previously described neural network model[5]. The resulting cell developmental age was also largely consistent with the sample collection window **(Figure 2C and Figure S4E)**.

Subsequently, we selected cells in the scRNA-seq data with a *RAPToR* inferred developmental age difference of 1 hour for integration with scStereo-seq data using

NovoSpaRc[30]. The integrated data enabled the imputation of spatial gene expression patterns with higher genome coverage, yielding markedly reduced signal background, enhanced tissue enrichment, and improved spatial gene expression patterns that exhibited greater resemblance to FISH validation results **(Figure 1C and Figure S1C)**.

We acquired scRNA-seq and scATAC-seq data from several hundred embryos per sample batch. Sample developmental age matched collection window for the majority of embryos, and the substantial sample size mitigated the influence of female egg retention. This was reinforced by the consistency between the model-predicted age and the actual collection window in both datasets **(Figure 2B-C)**. Additionally, the developmental ages of scRNA-seq and scATAC-seq were inferred with different models and might not be comparable. Consequently, we directly used sample collection window to integrate the scRNA-seq and scATAC-seq data for downstream analysis.

Together, our three multi-omics datasets exhibited coherence during embryo stages and can be integrated in a developmental age-specific manner in downstream analyses.

Construction of multi-omics tissue development trajectories

To delve into the detailed dynamics of developing tissues, we aimed to chronologically organize the tissue cell types in scRNA-seq and scATAC-seq data into continuous tissue developmental trajectories. Upon examining the subclustered and annotated data, we noticed that certain developmentally transitional cell types were categorized into different tissue clusters between assays, possibly due to differences in assay techniques, genome coverage, or clustering resolution (e.g., "muscle primordium" was annotated in the "mesoderm" cluster of the scATAC-seq dataset but in the "muscle" cluster of the scRNA-seq dataset) **(Figure S2 and Figure S3)**. To resolve this issue, instead of focusing on individual tissues, we included all cells of the same germ layer for collective and continuous analysis.

We first integrated the scRNA-seq and scATAC-seq data by finding integrated anchors for label transfer, imputing gene expression matrix from peak matrix of scATAC-seq data, and co-embedding them in the same UMAP space **(Figure 2D)**. Subsequently, unsupervised clustering was performed on the integrated germ layer data **(Figure S4F)**. Notably, a substantial number of late-stage cells annotated as "muscle" and "epidermis" in the scATAC-seq data did not correspond to any cell clusters in the scRNA-seq data **(Figure 2D)**. This discrepancy likely stemmed from technical limitations in capturing late-stage muscle (possibly due to their syncytial characteristics) and epidermal cells with our droplet-based scRNA-seq procedure. This was supported by a significant decrease in the fraction of these two cell types in late-stage scRNA-seq data **(Figure 1G)**.

To construct tissue developmental trajectories within a germ layer, we applied *PhyloVelo*[31] to the integrated data to establish velocity vector fields for three germ layers and re-annotated cell clusters based on marker genes and their chronological order along the velocity trajectories **(Figure 2E and Table S4)**. Due to their complexities, the trends of cell type differentiation are better visualized in 3D UMAP plots **(Data S2)**. With these velocity vector fields, we delineated multi-omics tissue development trajectories for all three germ layers **(Figure 2F)**.

Multi-omics tissue development trajectories allowed continuous and systematic tracing of various aspects of tissue- and cell type-specific dynamics during embryo development. To assess the differentiation dynamics of single cells, we employed *CytoTRACE*[32], which leveraged the number of detectably expressed genes as a robust indicator of differentiation potential. *CytoTRACE* analysis revealed diverse trends in differentiation dynamics across tissues during organ specification and maturation **(Figure 3A-B and Figure S4G)**. In general, mesodermal and endodermal tissues exhibited a slower decrease in differentiation potential compared to ectodermal ones. As anticipated, gonad cells maintained a consistently high level of potential throughout embryogenesis. Notably, the nervous system displayed the most rapid decline in differentiation potential throughout development, indicating its relatively faster pace towards terminal differentiation **(Figure 3A)**. Within each tissue, different cell types also exhibited varying rates of reduction in differentiation potential during development **(Figure 3B and Figure S4G)**. Genes whose expression level showed the strongest positive correlation with *CytoTRACE* scores were enriched in cell differentiation-related and ribosome protein genes **(Figure S4H-I and Table S6)**. The latter has been previously reported as indicators of both differentiation potential (reviewed in Ref[33]) and aging[10]. Conversely, genes whose expression level most negatively correlated with *CytoTRACE* scores included specific markers of differentiated tissues (e.g., GABAergic neuron-specific marker *Rdl*[34] and hemocyte-specific marker *Ppn*[35]) **(Figure S4H)**.

To characterize the activities of signaling pathways along tissue trajectories, we utilized 7 signaling pathway gene sets from *FlyphoneDB*[36] and examined the expression dynamics of core pathway component genes across tissue developmental trajectories **(Figure S5A-B)**. Throughout the trajectories, we observed the up-regulation of multiple tissue-specific signaling pathways that are well-documented in the literature. The BMP signaling pathway, known for its integral role in ectoderm dorsal-ventral patterning (reviewed in Ref[37]), and in the regulation of neuromuscular junctions (NMJ, reviewed in Ref[38]), demonstrated the highest level of activity in early ectoderm and muscles. Meanwhile, the FGFR signaling pathway, which has been widely recognized for its role in trachea branching morphogenesis (reviewed in Ref[39]), showed maximum activity during the early stages of tracheal development. Lastly, the Hippo signaling pathway, well-established for its contribution to myogenesis (reviewed in Ref[40]), was most active in early muscle clusters **(Figure 3C)**. Thus, our multi-omics tissue developmental trajectories could serve as a systematic framework for exploring cell-cell communication networks.

Spatiotemporal cell type succession during tissue development

Next, we aimed to visualize the spatiotemporal dynamics of the identified cell types along the multi-omics tissue development trajectories. Using the marker genes associated with cell types as a reference, we applied the label transfer method from *Seurat* to annotate scStereo-seq cell bins with the cell types identified in the multi-omics tissue development trajectories. At the tissue level, the transferred labels demonstrated good agreement with manually annotated scStereo-seq cell bin clusters **(Figure 3D)**. Considering their relatively defined and regular morphology, we selected fat body and foregut/hindgut (both of ectodermal origin[41]) as models and aligned their cell types with embryo scStereo-seq samples. 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 3E)**. Additionally, the top marker genes of each label-transferred cell type in scStereo-seq data were consistent with their counterparts in the integrated scRNA-seq and scATAC-seq data **(Figure S5C)**. These results suggested a precise mapping of cell types to their spatial locations in scStereo-seq data.

Within tissues, at the cell type level, the succession of different stages of cell types can be traced through their proportional changes over development **(Figure 3F)**. The spatial distribution of each cell type can be quantified by neighborhood enrichment, where a higher score indicates a greater level of spatial clustering **(Figure 3G)**. We observed significantly higher neighborhood enrichment in foregut/hindgut cell types compared to those in the fat body. When mapped to their spatial locations, different stages of foregut/hindgut cell types formed more aggregated clusters, while fat body cell types were more scattered and mixed **(Figure 3H)**. These observations suggested that these two tissues employ different cell differentiation strategies. In the fat body, differentiating cells are dispersed across the entire tissue, resulting in the mixing of cell types at different stages. In the foregut/hindgut, spatially defined “differentiation hubs” exist to continuously give rise to new cells, while cells outside the hubs do not contribute much to differentiation and proliferation. Consequently, cell types at different developmental stages occupy more distinct spatial locations. This hypothesis was further supported by the spatial distribution of cell bin *CytoTRACE* scores of scStereo-seq data. Cells with higher differentiation potential were more spatially aggregated in foregut/hindgut than in fat body **(Figure 3I)**. It is established that fat body cells originate from precursors arranged in segments that extend throughout the entire tissue[42]. This arrangement could account for the widespread dispersion of differentiating cells we observed here. On the other hand, the role of spatially clustered potential foregut/hindgut differentiation hubs might be associated with previously identified niches of digestive tract stem cells, where two defined groups of stem cells give rise to the adult foregut and hindgut, respectively[43].

Therefore, through label transfer, we were able to map cell types along tissue development trajectories to their spatial locations in scStereo-seq samples, allowing us to track their spatiotemporal dynamics. In the following analyses, we extended this approach to more complex CNS and midgut cell types to uncover their dynamics during development.

Transcription factor regulatory networks along tissue development trajectories

Transcription factors (TFs) play a pivotal role in orchestrating the proper formation and growth of tissues. To unravel the regulatory networks governed by TFs during tissue development and differentiation, we scrutinized the top marker genes of each cell type and investigated the enrichment of TF binding motifs in their promoter/TSS regions in scATAC-seq data. Motif enrichment analysis unveiled the regulatory TFs guiding the differentiation paths from each germ layer **(Table S7)**, encompassing both well-established cell type-specific regulators as well as potentially novel and uncharacterized ones.

Throughout the developmental trajectories, we pinpointed multiple well-characterized TFs that exhibited stage- and tissue-specific regulatory functions. We then traced the temporal dynamics of their regulatory activities along tissue development trajectories. Exemplary findings include motif enrichment of *GATAe* in Malpighian tubules[44], *Rfx* in both PNS and CNS[45], and *sage* in the salivary gland[46] within the ectoderm. In the mesoderm, we identified motif enrichment of *Mef2* in somatic muscle[47], *bin* in visceral muscle[48], and *srp* in fat body[49] and hemocytes[50]. The endoderm displayed motif enrichment of *CrebB* in early endoderm formation[51], along with *fkh*, *GATAe*[52], and other GATA family TFs (reviewed in Ref[53]) regulating late-stage endoderm specification **(Figure S6A-B)**. Additionally, we uncovered several previously uncharacterized TFs with potential spatiotemporally specific functions during embryogenesis. Notably, *CG34367*, a TF featuring a Homeobox (Hox) domain, exhibited significant and specific motif enrichment in the early primordium of all three germ layers, suggesting a ubiquitous role in early developmental regulation. Mammalian orthologs of *CG34367*, SHOX/SHOX2, are implicated in early organogenesis and their mutations are associated with genetic disorders including Turner syndrome[54,55]. The TF *crp*, ubiquitously expressed in multiple tissues and known for specifying terminal cells in tracheal tubes[56], demonstrated potential regulatory functions in the mesodermal fat body and hemolymph, as indicated by our analysis. Moreover, we observed significant motif enrichment of *CG9727* and *CG12219* in nervous systems, *CG7368* in cardiac mesoderm, and *CG12236* and *CG4360* in early endoderm **(Figure 4A-B)**, indicating their specific functions in these tissues.

To further explore the spatial regulon activities of these TFs, we applied *SCENIC*[57] to the integrated scStereo-seq and scRNA-seq data, revealing that the spatial patterns of regulon activities for both known and uncharacterized TFs were consistent with the motif enrichment analysis in terms of tissue specificity **(Figure 4C)**. The spatial expression patterns of these less-characterized TFs were also probed by BDGP *in situ* database and all of them exhibited weak signal or ubiquitous expression patterns in stages of their inferred functions **(Figure S6C)**. The lack of staining can be explained by poor probe efficiency or low expression levels of these TFs, while ubiquitously expressed TFs could perform regulatory functions in a tissue-specific manner. The ambiguous *in situ* staining results underscored the advantages of our multi-omics data in facilitating the elucidation of tissue-specific TF functions.

Subsequently, we employed *Pando*[58] on the integrated scRNA-seq and scATAC-seq data to delve into the detailed regulons of identified TFs. Notably, visceral muscles exhibited segregation into two distinct groups in our multi-omics trajectories (dashed rectangles in **Figure 2E-F**, also see **Figure S6D and Data S2**). Upon scrutinizing the regulons of *bin* in these two groups, we observed that gene modules that *bin* regulated varied between them **(Figure 4D)**. In the visceral muscle 1 group, *bin* activity was positively correlated with expression levels of muscle assembly genes (*Zasp52*, *sals*, *Zasp66*, and *kon*). Conversely, in the visceral muscle 2 group, *bin* activity was negatively correlated with expression of genes with similar functions (*Prm*, *Zasp52*, *slow*, and *CAP*). Intriguingly, *bin* activity appeared to be partially opposite in regulating muscle structure assembly in these two groups. Supporting this result, the expression level of *Zasp52*, a core component of indirect flight muscles[59], was significantly lower in scRNA-seq cells from visceral muscle 2 late cluster than those from visceral muscle 1 late cluster **(Figure S6E)**. It is known that *bin* is a cell fate determinant of transformation between somatic and visceral muscle through the BMP signaling pathway[48,60]. The contrasting effects *bin* exerted on some target genes in different visceral muscle cell groups may reflect its fine-tuning functions among muscle lineages. We further visualized the gene regulatory networks (GRNs) in which *bin* participated in these two groups **(Figure 4E)**. Inspection of GRNs in two visceral muscle groups uncovered several shared known muscle co-regulators of *bin*, such as *Mad*, a BMP pathway regulating TF functional at NMJ[61], and *CHES-1-like*, also a BMP pathway regulator[62]. *bin* also co-regulated with different nervous system-related TFs in the two lineages, including *klu* and *pnr* in visceral muscle 1, and *slp2* and *FoxP* in visceral muscle 2. *klu* had reported functions in motor neurons[63] while *FoxP* is important for motor coordination[64]. The organization of these GRNs highlighted the coordinated and cell type-specific co-regulation between nervous and muscle systems.

To further characterize the fat body- and hemolymph-specific regulon activities of *crp* identified above, we visualized its GRN in early fat body and discovered that *crp* co-regulated with *srp* **(Figure 4E)**. Upon inspecting their regulons, we found that *srp* activity was negatively correlated with lipid metabolism pathway genes (*Apoltp*, *Aldh7A1*, *apolpp*, *Jheh1*, and *Echs1*), while *crp* acted in a contrasting fashion. Additionally, *srp* positively regulated amino acid metabolism genes *Mtap*, *Hn*, and *Gdh*, while *crp* positively regulated glutathione metabolism genes *GstE6*, *GstE7*, *GstE11*, and *GstE12* **(Figure 4D)**. Target genes in the regulons of *srp* and *crp* largely overlapped in fat body and plasmatocytes, and this overlap increased along developmental trajectories of both tissues **(Figure 4F)**. The regulons of *srp* we identified were consistent with its role in inducing fat cell formation starting from early fat body development[49,65] and *crp* is known to affect cell growth and tissue size control[56]. Our analysis suggested an increasingly coordinated role of *crp* and *srp* within the same GRN during fat body and plasmatocyte development.

In tracing tissue development trajectories, we successfully identified both previously reported and potential TFs, uncovering their tissue specificity and regulatory networks. It is worth noting that TFs and their binding motifs were linked based on *CIS-BP* database[66]. While the motifs we mentioned here were indeed enriched in specific cell types, it remained possible that their actual binding TFs differ from database inference, or there are additional unknown regulators that could bind these motifs. Algorithms like *Pando* used correlation between gene expression levels of TFs and their target genes to infer up- or down-regulation effects of these TFs, which could be susceptible to capture sensitivity of current single-cell sequencing techniques. Thus, additional experiment validation is required to elucidate these tissue-specific regulatory networks.

Multi-omics dissection of gene regulation during embryonic CNS development

The *Drosophila* nervous system serves as a prominent model for investigating neuron development and functions. Thus, we examined the development of CNS from a multi-omics perspective based on our data. The subclustering results of the CNS scRNA-seq data identified most major CNS cell types, including neuroblasts (marked by *mira* and *wor*), ganglion mother cells (GMC, marked by *tap*), neural progenitors (marked by *insb* and *nerfin-1*), glioblasts (marked by *gcm* and *repo*), and various types of glial cells **(Table S4)**. The UMAP plot of CNS cells provided an intuitive representation of differentiation paths of neurons and glia **(Figure 5A)**. Subclustering scATAC-seq data also identified most of these CNS cell types **(Figure 5B)**. Integration of scRNA-seq and scATAC-seq data allowed detailed annotation of various mature neuron cell types by generating more distinct cell type specific markers **(Figure 5C-D and Table S4)**. In light of the significantly higher complexity of mature neuron cell types, we chose a higher resolution for their clustering and annotation. Each mature neuron cell type was annotated based on expression of neurotransmitters **(Figure 5D)**. Within mature neuron groups expressing the same neurotransmitters, cell subtypes were distinguished by a list of largely uncharacterized marker genes **(Figure 5E and Table**

S4). The complex trends of CNS cell differentiation are better visualized in 3D UMAP plots **(Data S3)**.

Remarkably, mature neurons displayed a significantly more striking diversity in the UMAP plot derived from scATAC-seq data compared to that from scRNA-seq data **(Figure 5A-B and Data S3)**. This observation suggested the possibility that mature neurons appear similar in their transcriptomic profiles during late embryogenesis, but various types of neurons are under highly distinct epigenetic regulations, likely in preparation for more complex neural differentiation during larval stages. Notably, co-embedding of scRNA-seq and scATAC-seq CNS cells in the same UMAP plot showed that mature neurons in scATAC-seq data displayed an overall distribution shift from those in scRNA-seq data **(Figure 5C and S7A)**. This shift was not observed in non-mature neuron cell types **(Figure S7A)**. Similarly, we also noted a temporal mismatch in the distribution of mature neuron subtypes between scATAC-seq and scRNA-seq data **(Figure S7B)**. This further reflected the potential divergence between transcriptomic and epigenomic profiles among mature neurons.

Differentiation trajectories of mature neurons revealed by scATAC-seq

To dissect the epigenetic regulation of mature neuron and identify potential regulators of cell subtype differentiation, we explored the 3D UMAP plot of scATAC-seq data and focused on a cell subset, in which a cell cluster expressing sensory neuron markers (e.g., *ct*, *lov*, and *robo3*) appeared to differentiate into three mature neuron clusters: GABAergic (GABA) neurons 2 & 4 and tyraminergic (TA) neurons 1 **(Figure 5F and Data S3)**. We employed *STREAM*[67] to map the differentiation trajectories of these cell clusters, which were then projected onto the 3D UMAP space. This enabled us to identify the branching events within the differentiation trajectories **(Figure 5F)**. Leveraging these trajectories, we further subclustered sensory neurons into two distinct groups according to their chromatin accessibility and differentiation outcomes **(Figure 5G)**. Focusing on the top DA peaks between sensory 1 & 2, as well as those among GABA 2, GABA 4, and TA 1, we conducted a TF motif enrichment analysis. This revealed *Kr* as a principal regulator of this differentiation process, with significant motif activity contrast between the differentiation branches **(Figure 5G and Figure S7C)**. *Kr* is a well-established transcription repressor and temporal determinant of neuron fate[68,69]. The gene *Kr* itself exhibited significantly higher chromatin accessibility and expression level in GABA 2 compared to GABA 4/TA 1 **(Figure S7D)**, suggesting a more active regulatory activity in GABA 2. On the contrary, the binding motifs of *Kr* are significantly less enriched in GABA 2 and most of its potential target genes (nearest genes of *Kr* binding motifs) displayed reduced chromatin accessibility and expression level in GABA 2 compared to GABA 4/TA 1 **(Figure 5G)**. This observation supports a working model based on previous knowledge that *Kr* performs its transcription repressor functions through local quenching of transcription activators[70,71], likely through closing up the proximal chromatin. We carefully examined the genes that showed significant changes in both chromatin accessibility and expression level among the neuron subtypes. Pathway enrichment of these genes showed that these *Kr* target genes were functionally enriched in processes including axon guidance and glycosylation **(Figure S7E)**.

While most potential targets of *Kr* showed decreased chromatin accessibility, their expression changes varied across target genes. This variability could be due to the impact from transcriptional co-factors of *Kr*. To identify co-regulators that influenced the expression levels of genes repressed by *Kr*, we conducted a motif enrichment analysis within *Kr* peaks **(Figure S7F)**. This revealed several previously characterized neuron differentiation regulators, including *hb*, *grh*, and *opa*, through the comparison between GABA 2 and TA 1. **(Figure 5H)**. It is well established that the sequential activities of *hb*, *Kr*, and *grh* determine the temporal fate of several neuroblast lineages during differentiation (reviewed in Ref[72]). *opa* is previously reported as a regulator of *Kr* activity during early embryogenesis[73], as well as a regulator of temporal patterning of neural progenitors that acts in coordination with *grh*[74]. Our observations suggested that the synergy of these regulators persist in more differentiated neuron subtypes. Both *hb* and *grh* are known to function as either transcription activators or repressors[75–77]. In the differentiation process we investigated here, the motif activities of *hb* and *grh* were mostly in up-regulated *Kr* target genes, even in the presence of repressive effect of *Kr*. Conversely, the motif activities of *opa* were enriched in down-regulated *Kr* target genes **(Figure 5H)**. We then examined the peaks around the binding motifs of these co-regulators in the chromosomal regions of their mutual target genes using the scATAC-seq data. We observed a dramatic overall increase in chromatin accessibility along the differentiation track from sensory 2 to GABA 4/TA 1, compared to the subtle changes during the transition from sensory 1 to GABA 2. This seems to be a general phenomenon regardless of the expression change between GABA 2 and GABA 4/TA 1, suggesting diverse and complex regulatory consequences depending on the cooperating TFs and the targets. This was further supported by the long distance of *Kr*/co-regulator peaks from the TSS of the nearest genes, which could be a few kb in length and mostly downstream of gene targets. As examples, the peaks and chromosomal regions of two genes with the most significant expression level changes, *side-III* (potentially co-regulated by *Kr*, *hb*, and *grh*) and *fz* (potentially co-regulated by *Kr* and *opa*), are plotted to demonstrate their coordinated regulatory roles outside of promoter regions (**Figure S7G**).

Overall, we observed a high clustering resolution in scATAC-seq data when characterizing mature neuron subtypes, which was able to facilitate the discovery of transcription regulators and their co-factors that govern the refined developmental trajectories.

Mutual and diverse GRNs among CNS cell subtypes

In pursuit of potential regulators of the diverse neuron cell types, we conducted motif enrichment analysis and pinpointed TF regulators across various stages of neural development **(Table S7)**. Among these, previously reported TFs, such as *seq*, governing dendrite and axon outgrowth[78], exhibited the highest motif activity in neuron progenitors. Additionally, *klu*, known to specify the identity of a specific group of neuroblasts[79], displayed sustained activity in several types of mature neurons **(Figure S8A)**. Our analysis also revealed cell type-specific activity for several less characterized TFs, including *BEAF-32*[80] and above-mentioned potential nervous system-specific regulator *CG12219* **(Figure S8A)**. *Pando* visualization of their regulons showed that in neuroblasts, *BEAF-32* and *seq* co-regulated multiple cell cycle regulator genes (e.g., *PolE2*, *fzy*, *mad2*, and *Mcm2*) and neuroblast determinants (e.g., *mira* and *CycE*) **(Figure S8B-C)**. In mature neuron clusters, *CG12219* and *klu* displayed similar activity patterns across cell types **(Figure S8A)** and their regulons largely coincided in mature neuron cell types **(Figure 5I)**. For example, in dopaminergic/serotonergic neurons 2, *CG12219* and *klu* co-regulated the same group of signal transduction genes (e.g., *Pkc53E*, *Oct1R*, *Syngr*, *Syt*α, and *CG34393*) in the same GRN; In cholinergic neurons 3, *klu* and *CG12219* co-regulated glucose metabolism (e.g., *Pgi* and *Pgm1*) gene groups in the same GRN **(Figure S8B-C)**.

Our findings strongly supported the existence of cell subtype-specific regulons for the same TF, as well as the cooperative actions of different TFs that are finely tuned for neuronal differentiation. These regulatory mechanisms may play a role in orchestrating the precise development of neurons.

Refined spatiotemporal CNS cell subtypes during embryogenesis

We subsequently applied label transfer to project identified CNS cell types onto scStereo-seq samples **(Figure 5J)**. As expected, there was a discernible shift in cell count fraction from undifferentiated neuroblasts and GMCs to differentiated neuron and glia cell types from early to late-stage samples **(Figure S8D)**. Co-embedding embryo scStereo-seq data with scRNA-seq data in the same UMAP space demonstrated high coherence **(Figure S8E)**. Among the transferred cell types, as expected from their anatomical distribution, neuroblasts and glia cell types exhibited the highest level of spatial aggregation, whereas mature neurons were largely dispersed **(Figure S8F)**. Upon inspecting their spatial loci, the distribution of CNS cell types aligned well with their stratified anatomical structures in early-stage samples, with less differentiated cell types occupying the outer layers of the CNS and more differentiated ones in the inner layers **(Figure 5J)**. These findings supported the precision of the label transfer method in identifying CNS cell subtypes in scStereo-seq samples, thereby facilitating the exploration of neuron functions within their spatial context.

Gene expression dynamics during CNS morphometric changes

The *Drosophila* CNS undergoes profound morphological transformations throughout embryogenesis, influenced by intrinsic factors such as cell proliferation and differentiation, as well as external cues like inter-organ communication (reviewed in Refs[72,81,82]). Leveraging our 3D spatial transcriptomes generated with scStereo-seq, we delved into transcriptomic dynamics during morphogenesis by simultaneously tracking changes in tissue morphology and gene expression. Employing morphometric analysis from the *Spateo* package, we were able to align the 3D point-cloud models of two time points in spatial coordinates. Subsequently, we linked cell bins between the two samples based on spatial adjacency and transcriptomic similarity **(Figure S9A)**. This enabled the generation of 3D vectors, concurrently characterizing cell migration paths and transcriptomic changes over continuous developmental stages **(Figure 5K)**. Finally, we computed morphometric parameters describing cell migration paths and correlated them with transcriptomic changes.

We characterized the morphometric changes in the CNS across 3D models of seven scStereo-seq samples, spanning developmental ages from 7 to 18 h. These changes were represented by parameters such as the acceleration of cell migration (proportional to the distance cells migrated given the same migration time between two samples) **(Figure 5K and Movie S1)**, curvature (bending of cell migration paths) **(Figure S9B)**, curl (rotation of paths) **(Figure S9C)**, and torsion (curve twisting of paths) **(Figure S9D)**. Throughout CNS development, we observed a shift in regions with the highest acceleration from the posterior end of the ventral nerve cord (VNC) to the anterior end of the brain **(Figure 5K)**. 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 **(Figure 5K and Figure S9C)**. 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 S9B-C)**.

The morphometric analysis yielded a set of genes exhibiting spatiotemporal expression changes relevant to CNS morphometric dynamics **(Table S8)**. Gene ontology (GO) enrichment revealed that genes linked to CNS morphometric changes were highly enriched in cell fate specification and pattern formation **(Figure S9E and Table S8)**. Additionally, gene group enrichment analysis highlighted the significance of Hox family transcription factors, such as *Antp*, *Ubx*, *abd-A*, and *Abd-B*, consistent with their critical roles in specifying CNS patterns and segment identity[83]. The spatial expression patterns of these Hox family genes in our 3D CNS models aligned with BDGP *in situ* results **(Figure S9F)**. Notably, the expression levels of these Hox genes were mostly negatively correlated with acceleration scores across developmental stages **(Figure S9G)**, suggesting that their expression is associated with the inhibition of CNS cell migration. It is reported that Hox genes' roles include repressing neuroblast formation and entry into neuroblast quiescence in embryonic CNS (reviewed in Ref[84]). It is possible that Hox genes inhibit CNS cell migration through repression of neuroblast differentiation.

Associations were also observed between CNS morphometric scores and known CNS development regulators (e.g., *mira*, *tll*, and *toy*) as well as several uncharacterized factors **(Table S8)**. For example, the expression level of *CG42394* was negatively correlated with acceleration, while that of *lncRNA:CR30009* displayed a positive correlation **(Figure 5L)**. We validated the CNS-specific expression of these potential regulators with FISH **(Figure 5M)**. Notably, this list includes multiple long non-coding RNA (lncRNA) genes besides *lncRNA:CR30009*, which was previously reported to be enriched in glia and co-localize with the glia marker gene *repo*[85]. 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 S9H)**. These observations implied that the two lncRNA genes may influence CNS morphometric changes through the regulation of neuroblasts. Therefore, by conducting morphometric analysis of the CNS, we were able to identify both known and potential regulators of CNS cell migration.

Cell type and functional diversity of developing midgut

The *Drosophila* midgut, serving as the functional equivalent of the mammalian small intestine, fulfills versatile roles in food digestion, nutrient uptake, immunity, and endocrine regulation. The diverse functions of the midgut are carried out by distinct types of cells and the regions they form (reviewed in Refs[86,87]). Nevertheless, the timing of differentiation of these cell types remained elusive. Our prior investigations indicated that certain functional cell types began to emerge during late embryogenesis[13].

Here, we delved deeper into the diversity of midgut cell types using our multi-omics data. The clustering resolution of our scRNA-seq data was adequate for distinguishing various midgut cell types. Consequently, we concentrated on the scRNA-seq data, combined endoderm and midgut cell clusters, and conducted high-resolution subclustering and annotation **(Figure 6A)**. The UMAP plot portrayed a multitude of intestinal cell types throughout the developmental and differentiation stages of the midgut **(Figure 6A-B)**. These included endoderm (marked by Notch signaling pathway genes *E(spl)m4-BFM* and *Brd*), adult midgut progenitors (AMPs, marked by *esg*)[26], 6 types of entero-endocrine cells [EEs, marked by *pros* and distinguished by specific expression endocrine genes **(Figure 6C)**], and 6 types of enterocytes (ECs, marked and distinguished by digestive enzyme and metabolism-related genes)[88]. Cell clusters in transitional states between midgut primordium and functional ECs were denoted as "midgut chambers", with each cluster distinguished by its top markers.

Differentiation and molecular characteristics of embryonic midgut cell types

We further investigated the molecular markers of distinct midgut cell types. 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 early endoderm, AMP/EE progenitors, and AMPs clusters, consistent with previous reports[27,89]. Additionally, the Wnt signaling pathway was enriched in multiple EE clusters[90] **(Figure S10A)**. Interestingly, AMPs and several EE clusters showed high enrichment in pathways related to autophagy and apoptosis **(Figure S10A)**. Indeed, the expression of autophagy-related genes *Atg101* and *Atg9*[91], as well as apoptosis-related genes *chrb* and *scyl*[92] are specifically enriched in AMPs and EEs within midgut **(Figure S10B)**. It is possible that AMPs and EEs employ cell death-related mechanisms to maintain homeostasis during embryogenesis. To further characterize the functions of these diverse midgut cells, we used *Hotspot*[93] to identify 17 gene modules from midgut scRNA-seq data **(Figure 6D)**, each with distinct functional GO and cell type-specific enrichment **(Figure 6E)**. For example, module 1, enriched in neuropeptide signaling pathways, was concentrated in all 6 EE clusters, aligning with their role in sensing stimuli and secreting neural signals for physiological regulation; Module 6, enriched in Cytochrome P450 family enzymes, was concentrated in EC (*Acbp3*+), suggesting a significant role in metabolism; Module 14, functionally enriched in genes regulating nervous system development and the Notch signaling pathway, was concentrated in AMPs and EE (*Mip*+) **(Figure S10C)**.

It is established that during metamorphosis, larval midgut cells undergo apoptosis, and adult midgut cells arise from AMPs to reconstitute the adult midgut[89,94]. Interestingly, in the UMAP plot, the differentiation trajectory of endoderm cells branched early on towards adult cell types (AMP/EE) and larval ones, which implied that the fates of these cell types were predetermined upon their differentiation from the endoderm primordium **(Figure 6A)**. We observed that AMPs and EEs originated from the same cluster of cells in the UMAP plot, marked by the expression of *esg* and *pros*. This is in line with previous reports indicating that AMPs and EEs derive from the same group of midgut cells[27], which we denoted as “AMP/EE progenitors” in our data **(Figure 2E-F)**. To further track the kinetics of cell type emergence, we employed *Dynamo*[95] to illustrate the transcriptomic vector fields of midgut development. *Dynamo* analysis revealed that the kinetics of cell state changes supported the notion that AMPs and EEs derived from the same group of progenitors **(Figure 6F)**. As previously reported, the Notch signaling pathway extensively participated in this differentiation process[27,89], along with stem cell differentiation factors such as *esg*, *pros*, *ase*, and *Sox100B*[88]. Notably, we observed highly specific dynamics of the innate immune signaling gene *Sting*[96] in AMPs, suggesting its role in the specification of AMPs **(Figure 6G)**. *Dynamo* analysis also facilitated the tracing of EC formation from their precursors in midgut chambers **(Figure 6F)** and suggested cell type-specific markers for their specification **(Figure S10D)**.

In summary, we categorized and examined the variety of cell types, each with unique functions, present in the embryonic midgut. This allowed us to reveal the differentiation trajectories of AMPs and EEs, as well as identify potential regulatory processes that govern their development and maintenance.

Spatial distribution of midgut cell types from embryonic to pupal stages

Next, we sought to map the midgut cell types we identified to their spatial locations. Using the cell type marker genes from scRNA-seq data as a reference, we located their counterparts in scStereo-seq data through label transfer **(Figure 6H)**. Co-embedding scRNA-seq and label transferred scStereo-seq data in the same UMAP space demonstrated high coherence **(Figure S10E)**. The top marker genes of label transferred scStereo-seq cell bins were also consistent with scRNA-seq cells **(Figure S10F)**, indicating precise mapping of cell types to their spatial locations. In the label transferred scStereo-seq 3D models, we observed the dynamics of changes in cell fraction throughout development, 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 6I)**. Neighborhood enrichment analysis suggested that although most cell types were sparse in their spatial distribution, EC (*Jon99Cii*+) and EC (*Try29F*+) were more aggregated compared to other cell types **(Figure S10G)**. Indeed, these cell types and the expression patterns of their marker genes occupied distinct spatial loci consistently across embryo scStereo-seq samples **(Figure 6J)**. Therefore, mapping cell types to our scStereo-seq data enabled the tracing of embryonic midgut cell type distribution within their spatiotemporal context. This approach provides a comprehensive understanding of how different cell types are spatially and temporally organized during embryonic midgut development.

We subsequently endeavored to identify midgut cell types in the larva and pupa scStereo-seq samples. We performed subclustering for larval and pupal midgut cell bins and used label transfer results as a reference for annotation. Given the substantial developmental changes and technical differences, we opted to use label transferred and re-annotated embryo scStereo-seq cell bins as the reference for label transfer of larva and pupa scStereo-seq samples, rather than embryo scRNA-seq cells. Cell clusters that demonstrated low confidence in label transfer were annotated as new larval or pupal cell types based on top marker genes **(Figure S11A-C and Table S4)**.

Compared to embryos, larvae displayed a more diverse array of intestinal epithelial cell types over development **(Figure 6K and Figure S11D)**. Notably, different ECs were densely clustered along the anterior-posterior axis of the midgut, as observed in the 3D models of their distribution **(Figure 6K)**. To understand their roles, we carried out *Hotspot* gene module analysis on each larva scStereo-seq sample. To fully leverage our scStereo-seq data, we took spatial location of cell bins into consideration during identification of gene modules. Examining the functional enrichment of gene modules, we noticed that various cell types performed unique functions, some of which were similar to those in the embryonic midgut, while others were new and specific to the larval midgut. For instance, the gene module concentrated in EC (*CG7298*+) in L1 early implied functional enrichment of chitin formation, suggesting that cell type specialized midgut chitinization (reviewed in Ref[97]) commenced in the early larval stages; the gene module concentrated in EC (*CG13075*+) in L3 late indicated functional enrichment of apoptosis and pattern specification, possibly related with midgut remodeling during late larval stages **(Figure 6L and Figure S11E)**.

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 completely in L3 late sample **(Figure 6K and Figure S11D)**. 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[98]. Accompanying this process, we observed a marked and widespread shift in the gene expression profiles of various cell types within the midgut, characterized by a considerable increase in the expression of ribosomal and mitochondrial genes and their representation in marker genes **(Table S4)**.

We subsequently examined the morphology of the midgut and the marker genes within the pupal midgut subclusters and identified a notable cluster of midgut cells that, rather than expressing digestion-related genes, exhibited strong expression of ecdysone-responsive genes, such as *Eig71Ek*, *Eig71Ea*, and *Edg78E* **(Table S2)**. These cells, which we designated as "midgut outer" in manual annotation, formed a sheath-like structure that enveloped the rest of midgut cells, which we labeled as "midgut inner" **(Figure S11F and Data S1)**. These structures, which receded after the P24 stage and re-emerged at the P72 stage, bore a strong resemblance to the yellow body and its surrounding midgut epithelium into which the midgut delaminates during metamorphosis[26] (reviewed in Ref[99]).

In conclusion, the use of label transfer-assisted spatial mapping and annotation unveiled spatially restricted and cell type-specific functions of larval and pupal midgut cell types. The spatial transcriptomic data from our pupa scStereo-seq samples also offered valuable resources for studying the regulation of midgut morphogenesis.

Emergence and layout of embryonic midgut regions

The adult *Drosophila* midgut is conventionally divided into five regions (hereafter termed R1~R5) based on morphological constraints, with each region performing distinct functions[100,101]. While functional regionalization of the midgut has been extensively studied at larva and adult stages, the precise timing of subregion emergence during embryonic development remains to be elucidated. In our previous study utilizing 3D spatial transcriptomic models, we observed the emergence of subregions with distinct digestive functions during late embryogenesis[13].

Here, we further characterized the process of embryonic midgut regionalization. Referring to regional marker genes summarized in Ref[101], we first identified 6 gene modules from adult midgut regional marker genes with *Hotspot* **(Figure S12A)** and established their correlation with expression profiles of adult midgut regions. Among them, modules 4 and 6 display similar correlation with both R1 and R2, so we denoted the two regions they corresponded to R1/R2-like 1 & 2, respectively **(Figure S12B)**. 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 groups displayed distinct marker gene expression **(Figure S12C)** and increasing levels of spatial clustering over development, suggesting that they occupied distinct areas in the midgut **(Figure S12D)**. This allowed us to investigate the timing of their appearance and trace the spatial distribution of these regions. Upon inspecting the expression of gene modules in scStereo-seq samples, we noted that the expression of R1-like gene modules initiated at the very early stages of endoderm development and gradually declined over time. Modules corresponding to other regions began to actively express around 13 h of embryogenesis **(Figure 6M)**. 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 6N)**. This suggested that although midgut underwent lysis and reformation during metamorphosis, its regional organization was already patterned during embryogenesis.

To profile the biological functions each region undertook, we examined GO enrichment of marker genes for the identified embryonic midgut regions. The R1-like region is functionally enriched in protein metabolism; 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[102]; the R4-like region is functionally enriched in stimuli sensing, proteolysis, and nucleic acid metabolism; the R5-like region is functionally enriched in metal ion homeostasis **(Figure S12E)**. These functions aligned well with their counterparts in adult midgut regions[100,101]. We analyzed the cell type composition of each region and observed that over development, each region acquired its major cell types. For instance, R1/R2-like 2 mainly composed of EC (*Muc55B*+) and R3-like mainly composed of copper cells and EC (*Jon65Aiii*+) **(Figure S12F)**.

Together, our scStereo-seq data demonstrated distinctive cell compositions in the embryonic midgut regions, which determined the spatially localized sub-organ functions maintained through adulthood.

Morphometric regulators during embryonic midgut development

The *Drosophila* midgut experiences significant morphological transformations during development. It forms from the fusion of two separate rudiments at the anterior and posterior ends of the embryo, evolves into a closed chamber, and eventually establishes a highly convoluted tube-like morphology (reviewed in Ref[103]). We used *Spateo* to model the morphological changes during the fusion of the anterior and posterior midgut around stage 12 (~8 h of development) and the convolution of midgut tubes during late embryogenesis **(Figure 6O and Movie S2)**. Morphometric analysis of cell migration modeled the fusion of early midgut and the torsion of late midgut, which revealed an association between midgut cell acceleration and multiple previously reported morphogenesis regulators across stages, including the GATA family TF *grn*, which is known to regulate the process of midgut fusion[104], and Notch signaling pathway component *Kaz-m1*, which displays a highly restricted expression pattern at the fusion site of the midgut and has a potential regulatory role[105] **(Table S8)**. The expression levels of both factors were negatively correlated with acceleration scores, suggesting that they were associated with inhibition of midgut cell migration **(Figure 6P)**. Starting from E15.77, gastric caecum-specific marker genes *Acbp4* and *Pebp1* ranked top in genes associated with all aspects of morphometric changes **(Table S8)**, in line with the timing of gastric caecum extrusion and formation from the midgut chamber (~15 h of development)[41]. In addition, multiple Acyl-CoA binding protein (Acbp) family genes demonstrated high correlation with morphometric scores in late-stage midgut, which aligned with their known function of linking nutrient sensing and shaping tissue plasticity[106]. We also observed several potential regulators or effectors of midgut morphological changes, such as *CG32633*, which consistently displayed positive correlation with cell migration acceleration across samples **(Figure 6P)**. Thus, morphometric analysis provided clues for identifying potential regulators during the complex morphogenesis process of embryonic midgut.

DISCUSSION

After our initial proof-of-principle application of Stereo-seq on *Drosophila*, we present here a single-cell 3D spatiotemporal multi-omics atlas spanning developmental lifespan of *Drosophila* from embryogenesis to metamorphosis. The current study builds upon our previous work by enhancing the Stereo-seq spatial transcriptomic dataset in several ways. Firstly, the sample collection window was expanded to include development from embryo to pupa. While ISH databases like BDGP and Fly-FISH have extensively probed spatial gene expression patterns in embryonic stages, there are still missing genes in these databases. Additionally, similar systematic databases are notably absent for the larval and pupal stages. Our scStereo-seq data effectively encapsulated the spatial gene expression patterns, unveiling the spatiotemporal gene expression dynamics for a list of over 300 genes in embryos, previously uncharted in ISH databases. Our data also serves as a valuable asset for delving into the spatial gene expression patterns in larvae and pupae. Secondly, our previous work using merged bins of a predetermined number of DNBs as units of analysis (e.g., bin 20 × 20 recognized 400 DNBs as a single "cell") did not accurately capture the transcriptomes of individual cells. Here, 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 for embryo samples, 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.

The integration of multi-omics data has enriched our analysis, enabling a more nuanced portrayal of cell states. As an example, the transcriptomes of CNS mature neurons are remarkably uniform, as demonstrated by their intertwined distribution in the UMAP space of scRNA-seq data. However, the incorporation of scATAC-seq data, which shows a higher degree of heterogeneity among mature neurons, allowed us to identify detailed neuron subtypes and investigate the regulatory mechanisms driving their differentiation. Leveraging the high heterogeneity of scATAC-seq data, we were able to dissect the differentiation process of a group of neuron subtypes in detail and identified the TF *Kr* as a key regulator. The wealth of chromatin accessibility information allowed us to further uncover TFs that co-regulated gene expression with *Kr*. We also mapped cell clusters, derived from the integration of scRNA-seq and scATAC-seq data, to their spatial positions in the scStereo-seq data. This mapping enabled us to model single-cell transcriptomic and epigenomic profiles within tissue- and developmental stage-specific contexts. Therefore, this multi-omics data integration offered an unprecedented high-resolution spatiotemporal framework for analyzing cell state dynamics, such as TF regulons and signaling pathways.

There were also inconsistencies between data generated from different techniques. For example, late-stage epidermis and somatic muscle cells identified in the scATAC-seq data lacked corresponding scRNA-seq counterparts. We observed a similar lack of coherence in CNS mature neurons, which, in addition to missing scRNA-seq data, could also resulted from a temporal mismatch between chromatin accessibility and actual gene expression in these neurons. In certain instances, the chromatin regions of neuron subtype-specific genes were open, but gene expression was delayed. This discrepancy resulted in inaccurately imputed gene expression when integrating scATAC-seq and scRNA-seq data and mismatches in the co-embedded UMAP space. Similar phenomena have been observed in the mammalian nervous system, such as the process of epigenetic priming during normal or pathological development (reviewed in Refs[107,108]).

Compared to our previous study, the point cloud-based 3D modeling of developing tissues in this work provided significantly more detailed structural information that more accurately reflects organ anatomical features. Furthermore, by aligning 3D models between different time points with *Spateo*, we were able to simultaneously track cell migration paths and alterations in gene expression. This morphometric analysis provided a unique perspective, enabling the identification of potential regulators of cell migration and differentiation. It is recognized that the eventual shape and size of an organ can be influenced by physical interactions with neighboring organs and signaling molecules from distant organs during development[109–111]. In addition to the intra-organ morphometric analysis presented here, these models can also be used to investigate the impact of inter-organ physical or biochemical contact on local gene expression changes, which in turn affect the final boundaries of organs. With the establishment of a complete synapse-resolution connectome of the *Drosophila* larval brain[112], our 3D transcriptomes have the potential to be spatially aligned with these synapse connectivity maps. By integrating spatial transcriptome and connectome data, we can simultaneously pinpoint the spatial locations of known and yet-to-be-identified neurons and deconvolute their molecular nature, leading to a deeper understanding of their physiological functions.

The extensive datasets we generated here can be leveraged in many ways. *Drosophila* larva and pupa have provided excellent models for studying the course of post-organogenesis development and metamorphosis, yet single-cell profiling of tissues at these stages remained scant. These stages of samples in our scStereo-seq data can be readily integrated with existing larva scRNA-seq datasets[6,7,113] to complement them with spatial information or provide a spatial framework for future single-cell studies of larval or pupal tissues. The study of early stages of *Drosophila* pupal development has been challenging due to significant tissue lysis and reformation. Our pupa scStereo-seq data provided valuable insights for investigation of tissue-specific transcriptomic changes during metamorphosis.

Moreover, the unique organization of 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. The organism-wide 3D high-resolution features of our previous Stereo-seq datasets have already facilitated the development of several approaches for various purposes, including quantitative spatiotemporal modeling of single-cell spatial transcriptomic datasets[14], visualization and analysis of spatial omics data[114], construction of databases and optimization of their accessibility[115], alignment of 2D spatial transcriptomic sections for 3D modeling[116,117], and more. The new dataset features in this study can further assist in the development of bioinformatic approaches in many other aspects, 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. With the rapid development of spatial transcriptomic techniques and consequently the mounting number of datasets, these methods will serve as invaluable tools to facilitate our interpretation of multi-omics datasets.

In order to make our data more accessible, we have incorporated our datasets into the Spateo Viewer platform. 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, the Spateo Viewer provides userfriendly access to our 3D models. It enables interactive visualization of gene expression, activity of gene groups, and a variety of other customizable parameters within spatiotemporal contexts. 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 THIS STUDY

Our multi-modal analysis of the dataset revealed a wealth of information and demonstrated its potential for systematic spatiotemporal analysis of *Drosophila* development. However, there are still some areas that require improvement in future studies.

The scRNA-seq and scATAC-seq data in this study were obtained from separate samples. We aimed to align cells from the same tissues and same developmental stages between datasets for integrated multi-omics analysis through the control of collection window and inference of single cell developmental age. However, it is still possible that data used for integration were from different states of cells. Methods for simultaneous capture of transcriptomic and chromatin accessibility profiles from single cells have been developed lately[118,119], which may provide better integration results, especially addressing the temporal mismatch issue between scRNA-seq and scATAC-seq data. With future technical improvements, spatial information of cells may also be captured simultaneously to generate actual multi-omics profiles for each single cell.

This multi-omics atlas was generated exclusively from the genetic background strain *w1118*. However, investigating developmental regulation or disease mechanisms often involves genetic perturbations, such as knockdown/knockout of key regulator genes, or changes in environmental conditions, such as pathogen infection and drug treatment. Therefore, in the future, we plan to expand our study to include *Drosophila* models with various genetic mutations or subjected to different infection and/or treatment conditions to establish organism-wide single-cell multi-omics atlases. This approach will be particularly beneficial for studying complex physiological processes that involve multiple tissues in response to genetic perturbations, such as the progression of neurodegenerative diseases, gut-brain axis communication, and multi-organ metabolic diseases. The pipeline established in this study can serve as a basis for such investigations, enabling the generation of comprehensive datasets that incorporate genetic and environmental variability.

ACKNOWLEDGEMENTS

This work was supported by Shenzhen Key Laboratory of Gene Regulation and Systems Biology (Grant No. ZDSYS20200811144002008 to Y.H.), National Natural Science Foundation of China (Grant No. 32100684 to Q.H.), National Key R&D Program of China (Grant No. 2022YFC3400405 to X.X. and L.L.), National Key R&D Program of China (Grant No. 2022YFC3400300 to M.W.), Shenzhen Science and Technology Innovation Program (Grant No. KQTD20180411143432337, China) (to Y.H., W.C., and Q.H.), Guangdong Genomics Data Center (2021B1212100001 to T.Y. and J.C.). Part of 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 Center for Computational Science and Engineering and Core Research Facilities at Southern University of Science and Technology. We thank Dr. Xiaojie Qiu and Dr. Martin Vingron for helpful suggestions on data analysis and Dr. Mariana Wolfner for helpful comments on the manuscript. We thank Biorender for graphical resources used in **Figure 1A** and Graphical Abstract.

AUTHOR CONTRIBUTIONS

Q.H., L.L., X.X., and Y.H. conceived the idea; C.L., Y.G., W.C., L.L., X.X., and Y.H. supervised the work; Q.H. prepared the samples; M.W., Z.T., Y.H., N.Y., W.M., W.L., C.L., and X.L. prepared the sequencing library and performed sequencing; M.W., Q.H., Z.T., L.K., J.Y., R.X., Y.Z., Z.C., Z.J., K.O., X.W., Y.B., and M.L. performed computational analysis; Q.H., M.W., Z.T., L.K., and Y.H. interpreted the data; Q.H., Y.Z., T.Y, Y.W., and Z.Y. performed experimental validation; T.Y., and J.C. established the data portal and Flysta3D website; Q.H., M.W., W.C., L.L., X.X., and Y.H. acquired fundings; Q.H., M.W., Z.T., and Y.H. 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.

FIGURE LEGENDS

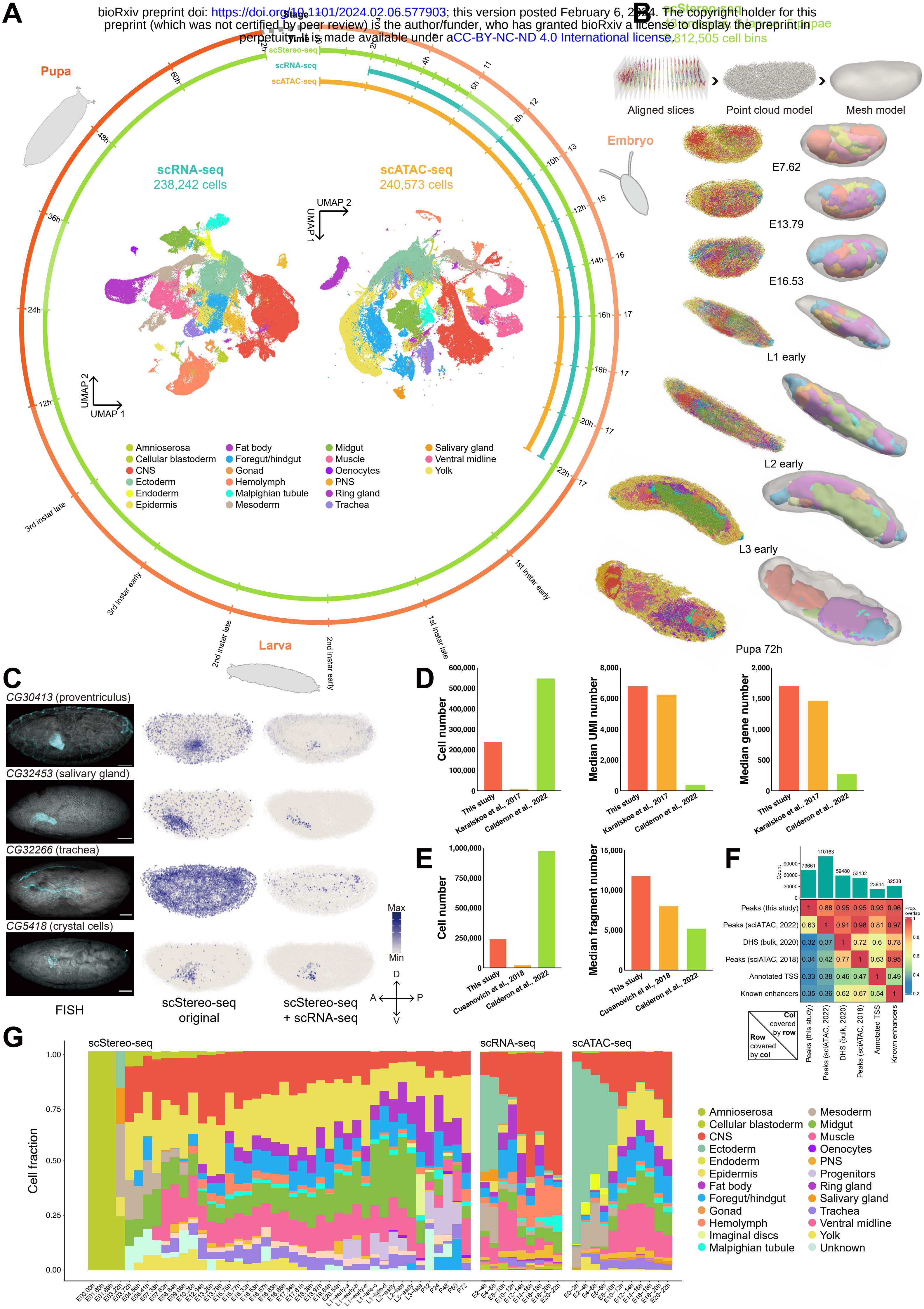
A
Stage
Time
scStereo-seq
scRNA-seq
scATAC-seq
Pupa
Embryo
Larva
2h
4h
6h
8h
10h
12h
14h
16h
18h
20h
22h
11
12
13
15
16
17
60h
48h
36h
24h
12h
3rd instar late
3rd instar early
2nd instar late
2nd instar early
1st instar late
1st instar early
scRNA-seq
238,242 cells
scATAC-seq
240,573 cells
UMAP 1
UMAP 2
Amnioserosa
Cellular blastoderm
CNS
Ectoderm
Endoderm
Epidermis
Fat body
Foregut/hindgut
Gonad
Hemolymph
Malpighian tubule
Mesoderm
Midgut
Muscle
Oenocytes
PNS
Ring gland
Trachea
Salivary gland
Ventral midline
Yolk
B
scStereo-seq
43 embryos, 8 larvae, 5 pupae
2,812,505 cell bins
Aligned slices
Point cloud model
Mesh model
E7.62
E13.79
E16.53
L1 early
L2 early
L3 early
Pupa 72h
C
CG30413 (proventriculus)
CG32453 (salivary gland)
CG32266 (trachea)
CG5418 (crystal cells)
FISH
scStereo-seq original
scStereo-seq + scRNA-seq
Max
Min
D
V
A
P
D
Cell number
600,000
500,000
400,000
300,000
200,000
100,000
0
This study
Karaiskos et al., 2017
Calderon et al., 2022
Median UMI number
8,000
6,000
4,000
2,000
0
Median gene number
2,000
1,500
1,000
500
0
E
Cell number
1,000,000
750,000
500,000
250,000
0
This study
Cusanovich et al., 2018
Calderon et al., 2022
Median fragment number
15,000
10,000
5,000
0
F
Count
90000
60000
30000
0
73661
110163
59480
53132
23844
32538
Prop. overlap
1
0.8
0.6
0.4
0.2
Peaks (this study) 1 0.88 0.95 0.95 0.93 0.96
Peaks (sciATAC, 2022) 0.63 1 0.91 0.98 0.81 0.97
DHS (bulk, 2020) 0.32 0.37 1 0.72 0.6 0.78
Peaks (sciATAC, 2018) 0.34 0.42 0.77 1 0.63 0.95
Annotated TSS 0.33 0.38 0.46 0.47 1 0.49
Known enhancers 0.35 0.36 0.62 0.67 0.54 1
Col covered by row
Row covered by col
Peaks (this study)
Peaks (sciATAC, 2022)
DHS (bulk, 2020)
Peaks (sciATAC, 2018)
Annotated TSS
Known enhancers
G
scStereo-seq
scRNA-seq
scATAC-seq
Cell fraction
1.00
0.75
0.50
0.25
0.00
E00.00h
E01.60h
E01.89h
E03.22h
E03.72h
E04.56h
E06.41h
E07.33h
E07.62h
E08.84h
E09.08h
E10.26h
E12.34h
E13.26h
E13.79h
E15.75h
E15.77h
E16.32h
E16.53h
E16.57h
E16.63h
E16.88h
E17.34h
E17.61h
E18.39h
E18.97h
E19.84h
E20.54h
L1-early-a
L1-early-b
L1-late-c
L1-late-d
L2-early
L2-late
L3-early
L3-late
P12
P24
P48
P60
P72
E2–4h
E4–6h
E8–10h
E10–12h
E12–14h
E14–16h
E16–18h
E18–20h
E20–22h
E0–2h
E2–4h
E4–6h
E6–8h
E10–12h
E12–14h
E14–16h
E16–18h
E18–20h
E20–22h
Amnioserosa
Cellular blastoderm
CNS
Ectoderm
Endoderm
Epidermis
Fat body
Foregut/hindgut
Gonad
Hemolymph
Imaginal discs
Malpighian tubule
Mesoderm
Midgut
Muscle
Oenocytes
PNS
Progenitors
Ring gland
Salivary gland
Trachea
Ventral midline
Yolk
Unknown

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

(A) Samples covered in this study. The outer rim indicates sample collection windows for three omics datasets, with each arc segment represents a collection window. Time points indicate hours after egg laying in embryos and hours after pupation in pupa. The inner panel shows UMAP plots of aggregated scRNA-seq and scATAC-seq data, color coded by tissue annotation. We did not obtain quality P36 scStereo-seq and E6-8h scRNA-seq data. CNS: central nervous system; PNS: peripheral nervous system. **(B)** 3D modeling of representative scStereo-seq samples using *Spateo*, showing point cloud (left) and mesh (right) models for the entire animal over developing stages. Models of epidermis, trachea, hemolymph, and muscle are not displayed in some samples for better visualization of internal organs. Tissue color codes are the same as **(A)**. Samples are not on the same scale. **(C)** FISH validation of representative genes from the list of genes without reported spatial expression patterns **(Table S3)**. For each gene, representative FISH images were obtained from stage 11-16 embryos from lateral or near-lateral view. Cyan: gene-specific RNA probes; grey: nuclei stained with DAPI. Arrowheads indicate structures with autofluorescence (e.g., trachea). Scale bars = 50 μm. All scStereo-seq samples are shown in lateral or near-lateral view. A-P: anterior-posterior; D-V: dorsal-ventral. Spatial expression patterns generated from original scStereo-seq or integrated scStereo-seq and scRNA-seq data are also from representative stage 13-17 embryos, projected along the Z-axis. See additional examples in **Figure S1C**. **(D)** Quality benchmark of scRNA-seq dataset in this study, showing cell number, median UMI number per cell, and median gene number per cell in datasets from this study and previous *Drosophila* embryo scRNA-seq studies. **(E)** Quality benchmark of scATAC-seq dataset in this study, showing cell number and median fragment number per cell from this study and previous *Drosophila* embryo scATAC-seq studies. **(F)** Heatmap showing proportion of scATAC-seq peaks in this study overlapping peaks in two previous *Drosophila* embryo scATAC-seq/scATAC-seq studies, bulk DHS peaks, and peaks in known TSSs and enhancers. **(G)** Bar plot showing cell type composition of data from scStereo-seq (some low-quality samples are filtered), scRNA-seq, and scATAC-seq over sample collection time. The y-axes are fraction of cell types annotated in each dataset. The x-axes are sample collection time points/windows (*RAPToR* inferred developmental age for embryo scStereo-seq data).

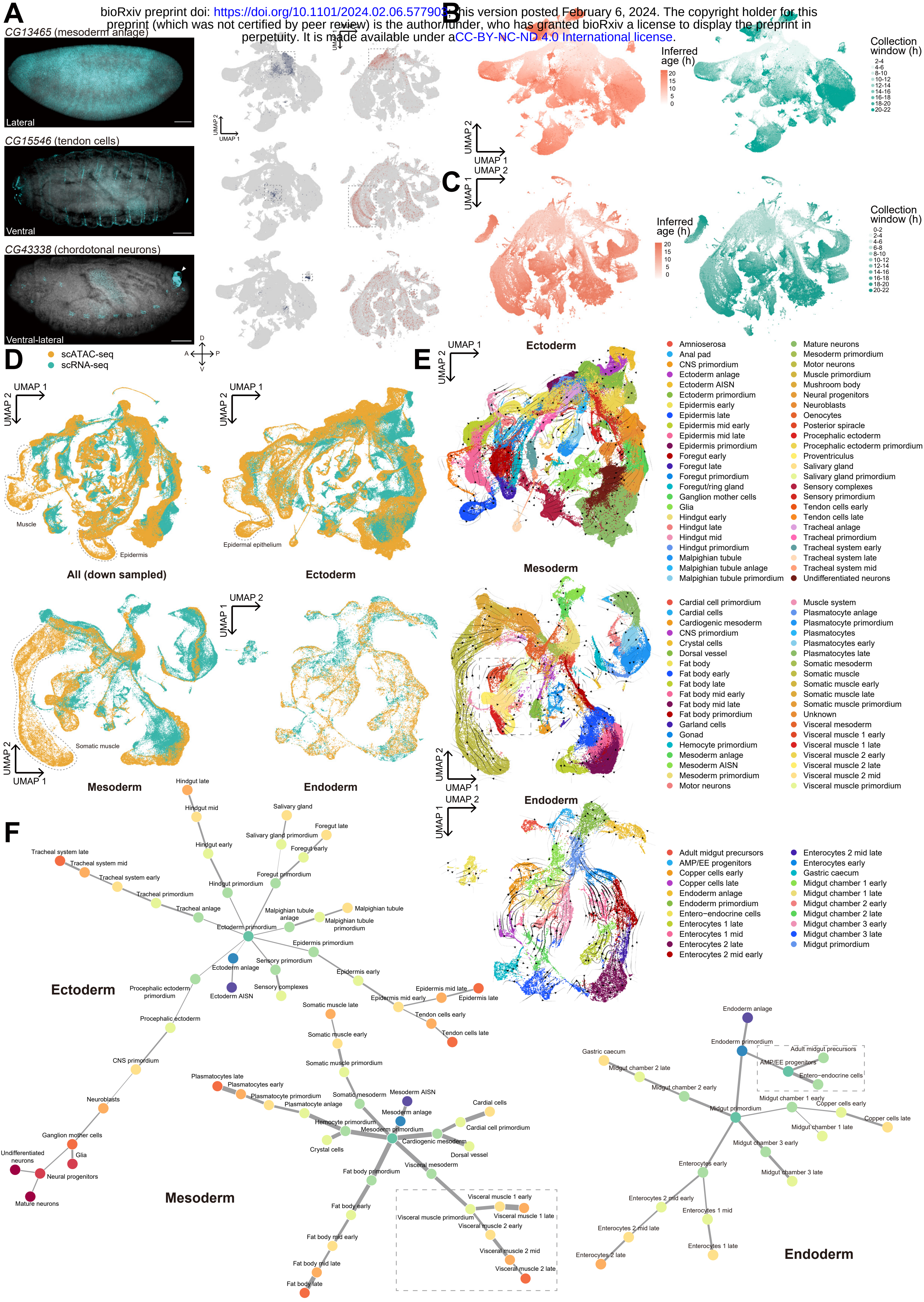
A
CG13465 (mesoderm anlage)
Lateral
CG15546 (tendon cells)
Ventral
CG43338 (chordotonal neurons)
Ventral-lateral
D
A
P
V
UMAP 2
UMAP 1
B
Inferred age (h)
20
15
10
5
0
Collection window (h)
2-4
4-6
8-10
10-12
12-14
14-16
16-18
18-20
20-22
C
Inferred age (h)
Collection window (h)
0-2
2-4
4-6
6-8
8-10
10-12
12-14
14-16
16-18
18-20
20-22
D
scATAC-seq
scRNA-seq
Muscle
Epidermis
All (down sampled)
Epidermal epithelium
Ectoderm
Somatic muscle
Mesoderm
Endoderm
E
Ectoderm
Amnioserosa
Anal pad
CNS primordium
Ectoderm anlage
Ectoderm AISN
Ectoderm primordium
Epidermis early
Epidermis late
Epidermis mid early
Epidermis mid late
Epidermis primordium
Foregut early
Foregut late
Foregut primordium
Foregut/ring gland
Ganglion mother cells
Glia
Hindgut early
Hindgut late
Hindgut mid
Hindgut primordium
Malpighian tubule
Malpighian tubule anlage
Malpighian tubule primordium
Mature neurons
Mesoderm primordium
Motor neurons
Muscle primordium
Mushroom body
Neural progenitors
Neuroblasts
Oenocytes
Posterior spiracle
Procephalic ectoderm
Procephalic ectoderm primordium
Proventriculus
Salivary gland
Salivary gland primordium
Sensory complexes
Sensory primordium
Tendon cells early
Tendon cells late
Tracheal anlage
Tracheal primordium
Tracheal system early
Tracheal system late
Tracheal system mid
Undifferentiated neurons
Mesoderm
Cardial cell primordium
Cardial cells
Cardiogenic mesoderm
CNS primordium
Crystal cells
Dorsal vessel
Fat body
Fat body early
Fat body late
Fat body mid early
Fat body mid late
Fat body primordium
Garland cells
Gonad
Hemocyte primordium
Mesoderm anlage
Mesoderm AISN
Mesoderm primordium
Motor neurons
Muscle system
Plasmatocyte anlage
Plasmatocyte primordium
Plasmatocytes
Plasmatocytes early
Plasmatocytes late
Somatic mesoderm
Somatic muscle
Somatic muscle early
Somatic muscle late
Somatic muscle primordium
Unknown
Visceral mesoderm
Visceral muscle 1 early
Visceral muscle 1 late
Visceral muscle 2 early
Visceral muscle 2 late
Visceral muscle 2 mid
Visceral muscle primordium
Endoderm
Adult midgut precursors
AMP/EE progenitors
Copper cells early
Copper cells late
Endoderm anlage
Endoderm primordium
Entero−endocrine cells
Enterocytes 1 late
Enterocytes 1 mid
Enterocytes 2 late
Enterocytes 2 mid early
Enterocytes 2 mid late
Enterocytes early
Gastric caecum
Midgut chamber 1 early
Midgut chamber 1 late
Midgut chamber 2 early
Midgut chamber 2 late
Midgut chamber 3 early
Midgut chamber 3 late
Midgut primordium
F
Ectoderm
Hindgut late
Hindgut mid
Salivary gland
Foregut late
Salivary gland primordium
Foregut early
Hindgut early
Tracheal system late
Tracheal system mid
Tracheal system early
Foregut primordium
Hindgut primordium
Tracheal primordium
Malpighian tubule
Tracheal anlage
Malpighian tubule anlage
Malpighian tubule primordium
Ectoderm primordium
Epidermis primordium
Ectoderm anlage
Sensory primordium
Epidermis early
Sensory complexes
Ectoderm AISN
Prochephalic ectoderm primordium
Epidermis mid late
Epidermis mid early
Epidermis late
Procephalic ectoderm
Tendon cells early
Tendon cells late
CNS primordium
Neuroblasts
Ganglion mother cells
Glia
Undifferentiated neurons
Neural progenitors
Mature neurons
Mesoderm
Somatic muscle late
Somatic muscle early
Somatic muscle primordium
Plasmatocytes late
Plasmatocytes early
Plasmatocyte primordium
Plasmatocyte anlage
Somatic mesoderm
Mesoderm AISN
Mesoderm anlage
Cardial cells
Hemocyte primordium
Mesoderm primordium
Cardial cell primordium
Cardiogenic mesoderm
Crystal cells
Dorsal vessel
Fat body primordium
Visceral mesoderm
Fat body early
Visceral muscle 1 early
Visceral muscle primordium
Visceral muscle 1 late
Visceral muscle 2 early
Fat body mid early
Visceral muscle 2 mid
Fat body mid late
Visceral muscle 2 late
Fat body late
Endoderm
Endoderm anlage
Endoderm primordium
Adult midgut precursors
Gastric caecum
AMP/EE progenitors
Midgut chamber 2 late
Entero-endocrine cells
Midgut chamber 2 early
Midgut chamber 1 early
Copper cells early
Midgut primordium
Copper cells late
Midgut chamber 1 late
Midgut chamber 3 early
Enterocytes early
Midgut chamber 3 late
Enterocytes 2 mid early
Enterocytes 1 mid
Enterocytes 2 mid late
Enterocytes 1 late
Enterocytes 2 late

Figure 2 Integration of scRNA-seq and scATAC-seq data for construction of tissue development trajectories.

(A) FISH validation of representative genes in the list of common tissue substructure/cell types **(Table S5)**. Left: representative FISH images of corresponding stages of gene expression enrichment, with sample viewpoints labeled. Cyan: gene-specific RNA probes; grey: nuclei stained with DAPI. Arrowheads indicate structures with autofluorescence (e.g., trachea). A-P: anterior-posterior; D-V: dorsal-ventral. Scale bars = 50 μm; Right: UMAP plots of marker gene expression specificity in aggregated scRNA-seq and scATAC-seq data. Cells with enriched marker gene expression/peak accessibility are highlighted in dashed rectangles. **(B)** UMAP plots of aggregated scRNA-seq data, color coded with *RAPToR* inferred developmental age (left) and actual sample collection window (right). **(C)** UMAP plots of aggregated scATAC-seq data, color coded with neural network model inferred developmental age (left) and actual sample collection window (right). **(D)** UMAP plots of co-embedded scRNA-seq and scATAC-seq data of all cells (down sampled) and three germ layers. Dashed lines mark cell clusters in scATAC-seq data that miss corresponding cells in scRNA-seq data, with their scATAC-seq annotations labeled. **(E)** Velocity fields of co-embedded UMAP plots of three germ layers in **(D)**, color coded with re-annotated cell types based on clustering of integrated data. Velocity trajectories point backward from chronologically older to younger cells. The dashed rectangle indicates visceral muscle groups discussed in the following analyses. AISN: anlage in statu nascendi. **(F)** Tissue development trajectories based on cluster phylogeny inferred from **(E)** for major tissues of three germ layers. Within each germ layer, widths of lines connecting subcluster annotations indicate gene expression similarities. Dashed rectangles indicate visceral muscle trajectories discussed in GRN analysis and AMP/EE trajectories discussed in midgut cell type identification.

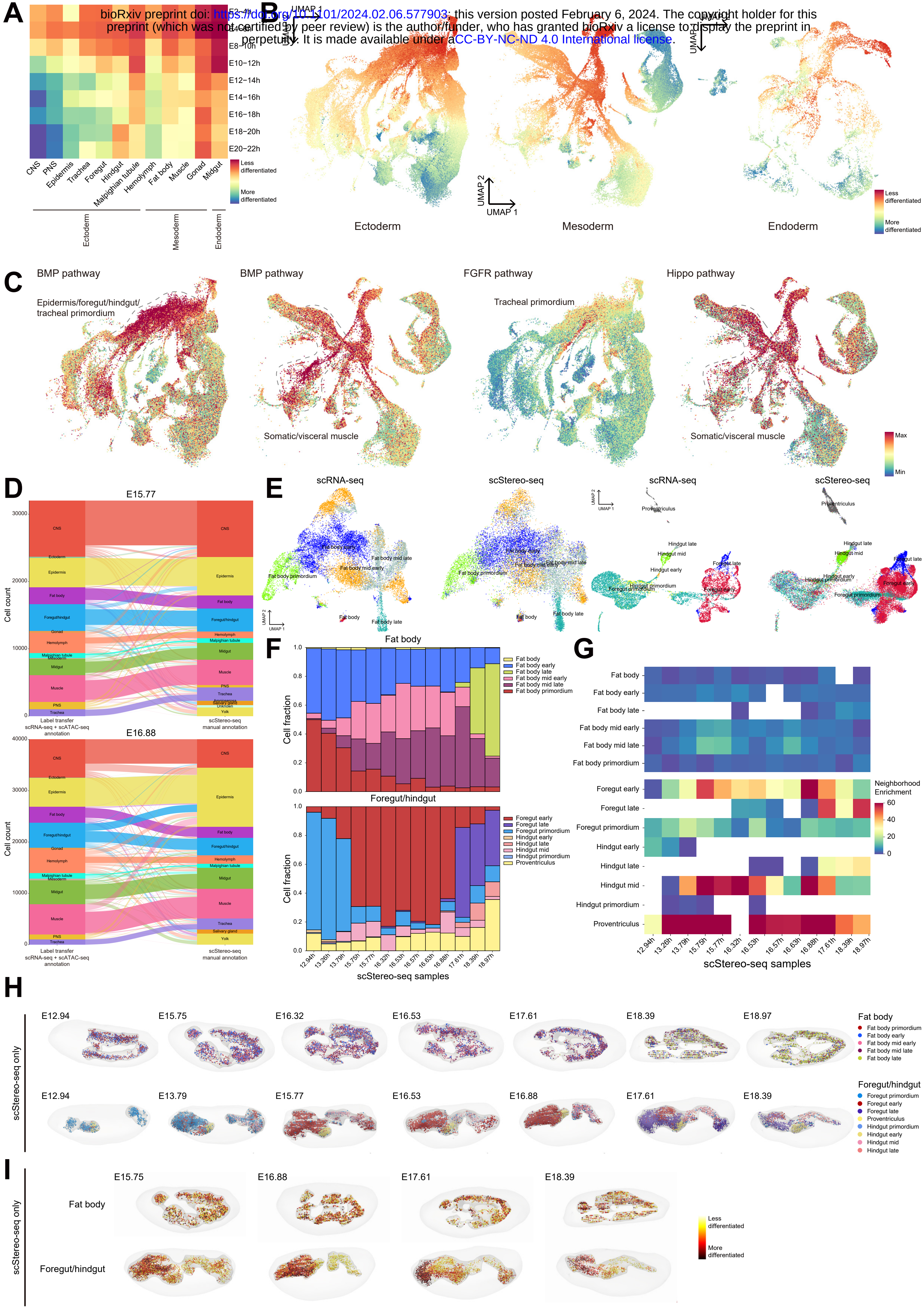
A
E2–4h
E4–6h
E8–10h
E10–12h
E12–14h
E14–16h
E16–18h
E18–20h
E20–22h
CNS
PNS
Epidermis
Trachea
Foregut
Hindgut
Malpighian tubule
Hemolymph
Fat body
Muscle
Gonad
Midgut
Ectoderm
Mesoderm
Endoderm
Less differentiated
More differentiated
B
UMAP 1
UMAP 2
Ectoderm
Mesoderm
Endoderm
Less differentiated
More differentiated
C
BMP pathway
Epidermis/foregut/hindgut/ tracheal primordium
BMP pathway
Somatic/visceral muscle
FGFR pathway
Tracheal primordium
Hippo pathway
Somatic/visceral muscle
Max
Min
D
E15.77
Cell count
CNS
Ectoderm
Epidermis
Fat body
Foregut/hindgut
Gonad
Hemolymph
Malpighian tubule
Mesoderm
Midgut
Muscle
PNS
Trachea
Amnioserosa
Salivary gland
Unknown
Yolk
Label transfer scRNA-seq + scATAC-seq annotation
scStereo-seq manual annotation
E16.88
E
scRNA-seq
scStereo-seq
scRNA-seq
scStereo-seq
Fat body primordium
Fat body early
Fat body mid early
Fat body mid late
Fat body late
Fat body
Proventriculus
Hindgut late
Hindgut mid
Hindgut early
Hindgut primordium
Foregut primordium
Foregut early
Foregut late
F
Fat body
Cell fraction
Fat body
Fat body early
Fat body late
Fat body mid early
Fat body mid late
Fat body primordium
Foregut/hindgut
Foregut early
Foregut late
Foregut primordium
Hindgut early
Hindgut late
Hindgut mid
Hindgut primordium
Proventriculus
12.94h
13.26h
13.79h
15.75h
15.77h
16.32h
16.53h
16.57h
16.63h
16.88h
17.61h
18.39h
18.97h
scStereo-seq samples
G
Fat body
Fat body early
Fat body late
Fat body mid early
Fat body mid late
Fat body primordium
Foregut early
Foregut late
Foregut primordium
Hindgut early
Hindgut late
Hindgut mid
Hindgut primordium
Proventriculus
Neighborhood Enrichment
60
40
20
0
scStereo-seq samples
H
scStereo-seq only
E12.94
E15.75
E16.32
E16.53
E17.61
E18.39
E18.97
Fat body
Fat body primordium
Fat body early
Fat body mid early
Fat body mid late
Fat body late
E12.94
E13.79
E15.77
E16.53
E16.88
E17.61
E18.39
Foregut/hindgut
Foregut primordium
Foregut early
Foregut late
Proventriculus
Hindgut primordium
Hindgut early
Hindgut mid
Hindgut late
I
scStereo-seq only
E15.75
E16.88
E17.61
E18.39
Fat body
Foregut/hindgut
Less differentiated
More differentiated

Figure 3 Spatiotemporal dynamics along multi-omics tissue developmental trajectories.

(A) Heatmap showing median tissue *CytoTRACE* scores based on scRNA-seq data along tissue development trajectories. *CytoTRACE* scores are scaled across all cells. **(B)** UMAP plots of scRNA-seq cells in the co-embedded UMAP space in **Figure 2E**, color coded with *CytoTRACE* scores. *CytoTRACE* scores are scaled within each germ layer. **(C)** Same as **(B)** but color coded with gene activity scores of core components of signaling pathways. Representative tissues enriched in signaling pathway activities are labeled. **(D)** Sankey plots showing agreement between scStereo-seq tissue manual annotations and transferred labels from integrated scRNA-seq and scATAC-seq data in representative scStereo-seq samples. **(E)** 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. **(F)** 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. **(G)** Heatmaps showing neighborhood enrichment scores of fat body and foregut/hindgut cell types across scStereo-seq samples. Blank cells indicate absence of label transferred cell types or lack of enrichment in corresponding samples. **(H)** 3D tissue models across representative embryo scStereo-seq samples, showing spatial distribution of label transferred cell types, mesh models for fat body or foregut/hindgut, and mesh models of the entire embryo. Due to high homology, some hindgut cells are annotated as foregut ones by label transfer. **(I)** Spatial distribution of cell bin *CytoTRACE* scores in representative fat body and foregut/hindgut models in **(H)**.

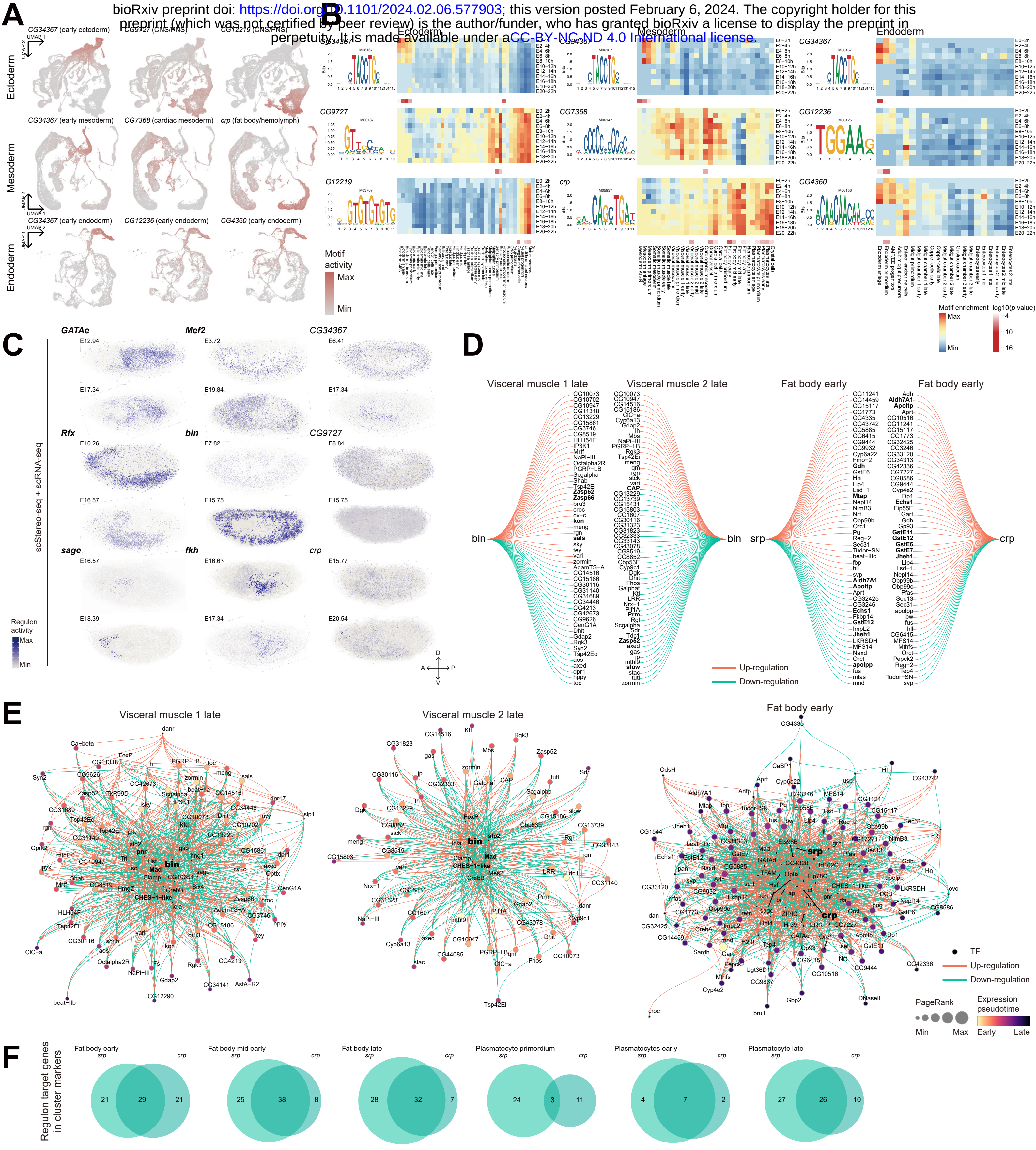
A
Ectoderm
Mesoderm
Endoderm
CG34367 (early ectoderm)
CG9727 (CNS/PNS)
CG12219 (CNS/PNS)
CG34367 (early mesoderm)
CG7368 (cardiac mesoderm)
crp (fat body/hemolymph)
CG34367 (early endoderm)
CG12236 (early endoderm)
CG4360 (early endoderm)
UMAP 1
UMAP 2
B
Ectoderm
Mesoderm
Endoderm
CG34367
CG9727
CG12219
CG7368
crp
CG12236
CG4360
M06167
M00187
M03707
M06147
M05937
M06125
M06156
Bits
E0–2h
E2–4h
E4–6h
E6–8h
E8–10h
E10–12h
E12–14h
E14–16h
E16–18h
E18–20h
E20–22h
Motif activity
Max
Min
Motif enrichment
log10(p value)
−4
−10
−16
C
scStereo-seq + scRNA-seq
GATAe
Mef2
CG34367
Rfx
bin
CG9727
sage
fkh
crp
E12.94
E17.34
E3.72
E19.84
E6.41
E17.34
E10.26
E16.57
E7.82
E15.75
E8.84
E15.75
E16.57
E18.39
E16.63
E17.34
E15.77
E20.54
Regulon activity
Max
Min
D
A
P
V
D
Visceral muscle 1 late
Visceral muscle 2 late
Fat body early
Fat body early
bin
bin
srp
crp
Up-regulation
Down-regulation
E
Visceral muscle 1 late
Visceral muscle 2 late
Fat body early
bin
srp
crp
TF
Up-regulation
Down-regulation
PageRank
Min
Max
Expression pseudotime
Early
Late
F
Regulon target genes in cluster markers
Fat body early
Fat body mid early
Fat body late
Plasmatocyte primordium
Plasmatocytes early
Plasmatocytes late
srp
crp
21
29
21
25
38
8
28
32
7
24
3
11
4
7
2
27
26
10

Figure 4 Transcription factor regulatory networks along multi-omics tissue development trajectories.

(A) The same UMAP plots as **Figure 2E** but only show scATAC-seq cells, color coded with motif activities of less-characterized TFs. **(B)** TF motif enrichment along tissue development trajectories, showing less-characterized TF genes in **(A)**, their binding motifs (left), motif enrichment 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. **(C)** Visualization of *SCENIC* regulon activity of some of TFs in **(A)** in representative samples from integrated scStereo-seq and scRNA-seq data, projected along the Z-axis. Previously reported tissue-specific TFs are in bold. All scStereo-seq samples are shown in lateral or near-lateral view. A-P: anterior-posterior; D-V: dorsal-ventral. **(D)** *Pando* identified regulons of TF *bin* in visceral muscle 1 late and visceral muscle 2 late, and those of TF *srp* and *crp* in fat body early. Genes in bold are discussed in detail in the main text. **(E)** *Pando* identified GRNs of TFs (highlighted in bold) and cell types in **(D)**. Other TFs in bold are discussed in detail in the main text. **(F)** Venn diagrams showing overlap between target genes in regulons of *srp* and *crp* along developmental trajectories of fat body and plasmatocytes.

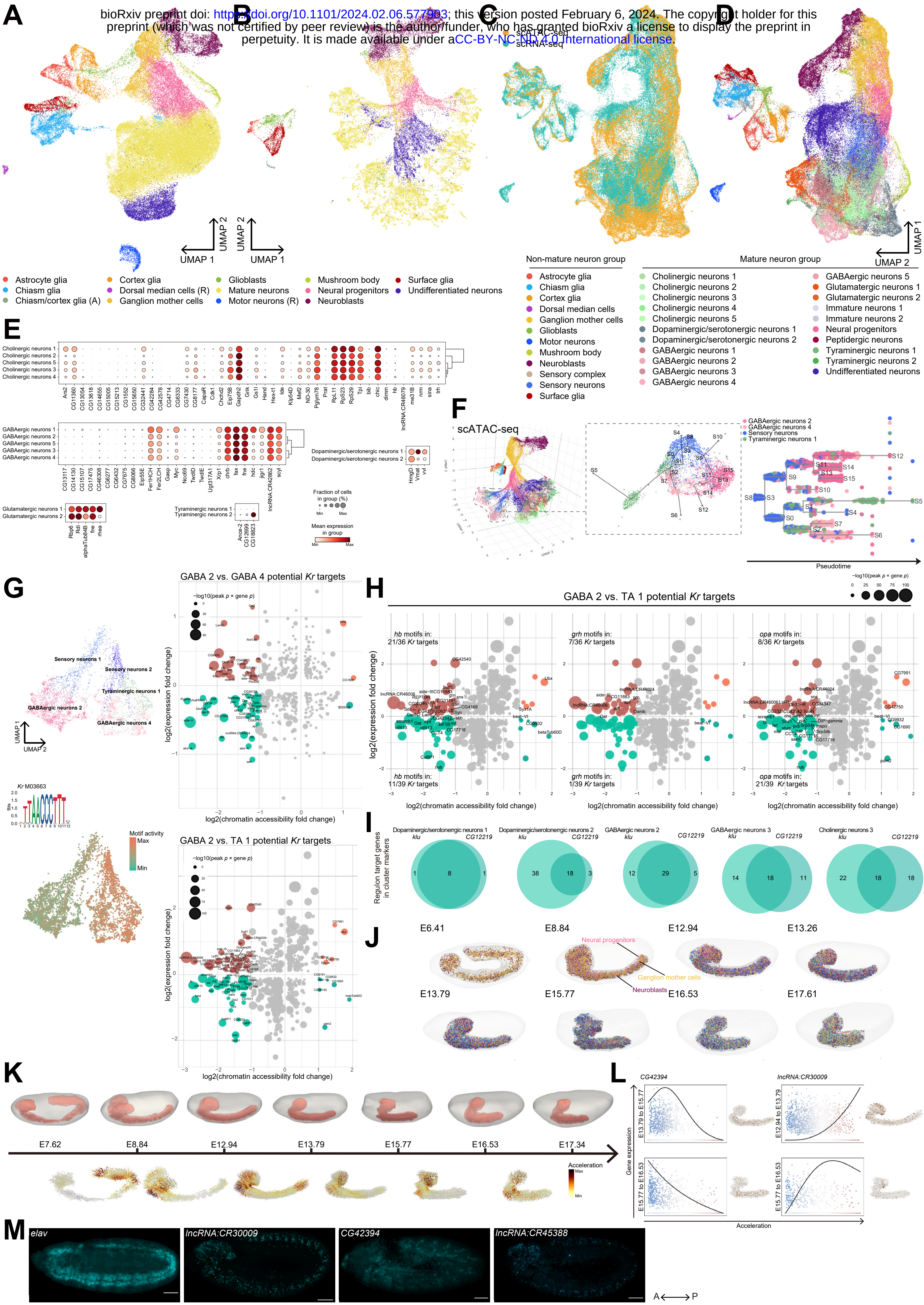
A
B
C
D
scATAC-seq
scRNA-seq
UMAP 1
UMAP 2
Astrocyte glia
Chiasm glia
Chiasm/cortex glia (A)
Cortex glia
Dorsal median cells (R)
Ganglion mother cells
Glioblasts
Mature neurons
Motor neurons (R)
Mushroom body
Neural progenitors
Neuroblasts
Surface glia
Undifferentiated neurons
Non-mature neuron group
Mature neuron group
Astrocyte glia
Chiasm glia
Cortex glia
Dorsal median cells
Ganglion mother cells
Glioblasts
Motor neurons
Mushroom body
Neuroblasts
Sensory complex
Sensory neurons
Surface glia
Cholinergic neurons 1
Cholinergic neurons 2
Cholinergic neurons 3
Cholinergic neurons 4
Cholinergic neurons 5
Dopaminergic/serotonergic neurons 1
Dopaminergic/serotonergic neurons 2
GABAergic neurons 1
GABAergic neurons 2
GABAergic neurons 3
GABAergic neurons 4
GABAergic neurons 5
Glutamatergic neurons 1
Glutamatergic neurons 2
Immature neurons 1
Immature neurons 2
Neural progenitors
Peptidergic neurons
Tyraminergic neurons 1
Tyraminergic neurons 2
Undifferentiated neurons
E
Fraction of cells in group (%)
Mean expression in group
Min
Max
F
scATAC-seq
S0 S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15
GABAergic neurons 2
GABAergic neurons 4
Sensory neurons
Tyraminergic neurons 1
Pseudotime
G
Sensory neurons 1
Sensory neurons 2
Tyraminergic neurons 1
GABAergic neurons 2
GABAergic neurons 4
Kr M03663
Motif activity
Max
Min
GABA 2 vs. GABA 4 potential Kr targets
GABA 2 vs. TA 1 potential Kr targets
−log10(peak p × gene p)
log2(expression fold change)
log2(chromatin accessibility fold change)
H
GABA 2 vs. TA 1 potential Kr targets
hb motifs in: 21/36 Kr targets
hb motifs in: 11/39 Kr targets
grh motifs in: 7/36 Kr targets
grh motifs in: 1/39 Kr targets
opa motifs in: 8/36 Kr targets
opa motifs in: 21/39 Kr targets
I
Regulon target genes in cluster markers
Dopaminergic/serotonergic neurons 1
Dopaminergic/serotonergic neurons 2
GABAergic neurons 2
GABAergic neurons 3
Cholinergic neurons 3
klu
CG12219
1 8 1
38 18 3
12 29 5
14 18 11
22 18 18
J
E6.41
E8.84
E12.94
E13.26
E13.79
E15.77
E16.53
E17.61
Neural progenitors
Ganglion mother cells
Neuroblasts
K
E7.62
E8.84
E12.94
E13.79
E15.77
E16.53
E17.34
Acceleration
Max
Min
L
CG42394
lncRNA:CR30009
E13.79 to E15.77
E12.94 to E13.79
E15.77 to E16.53
Gene expression
Acceleration
M
elav
lncRNA:CR30009
CG42394
lncRNA:CR45388
A
P

Figure 5 Gene regulation and morphometric dynamics in embryonic CNS.

(A) UMAP plot showing subclustering and annotation of CNS cells from scRNA-seq data. **(B)** Same as **(A)**, but for scATAC-seq data. Annotations with (R) or (A) indicate clusters identified only in scRNA-seq data or only in scATAC-seq data, respectively. **(C)** Co-embedding of CNS cells from scRNA-seq and scATAC-seq data in the same UMAP plot. **(D)** Same as **(C)**, but re-clustered and re-annotated. **(E)** Bubble plots showing expression level and enrichment of top marker genes of mature neuron cell types in **(D)**. **(F)** Left: 3D UMAP plot of scATAC-seq data (also see **Data S3**) and differentiation trajectories of selected cell clusters, with S0 through S15 denoting branching points of differentiation. S8 was set as the origin of differentiation; right: subway map plot showing differentiation trajectories and branching points of the same cell clusters. Each dot represents one cell from scATAC-seq data subset. **(G)** Upper left: UMAP plot showing subclustering of sensory neurons and their differentiation paths; lower left: the same UMAP plot color coded with *Kr* motif activity. *Kr* has 5 known motifs with highly similar sequence compositions. The composition and activity of representative motif M03663 are shown; right: scatter plot showing the genes associated with DA peaks and DE genes, comparing GABAergic neurons 2 with GABAergic neurons 4, and GABAergic neurons 2 with tyraminergic neurons 1. Nearest genes of *Kr* binding motifs are labeled. The size of each dot corresponds to the product of *p* values for DA peaks and DE genes. **(H)** The same scatter plots as **(G)**, comparing GABAergic neurons 2 with tyraminergic neurons 1 and labeled with nearest genes of binding motifs of *hb*, *grh*, and *opa*. **(I)** Venn diagrams showing overlap between target genes in regulons of *klu* and *CG12219* among representative mature neuron cell types. **(J)** 3D CNS models across representative embryo scStereo-seq samples, showing spatial distribution of cell types, mesh models of CNS, and mesh models of the entire embryo. Cell type color codes are the same as **(D)**. **(K)** 3D models of CNS, CNS cell migration trajectories, and acceleration scores across 7 scStereo-seq samples of developmental age between 7 and 18 h. **(L)** General linear model-based correlation between acceleration scores and expression levels of *CG42394* and *lncRNA:CR30009* in transitions between representative scStereo-seq samples. Spatial gene expression patterns in CNS 3D models are shown on the right of each plot. **(M)** FISH validation in stage 11-16 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. A-P: anterior-posterior; D-V: dorsal-ventral. Scale bars = 50 μm.

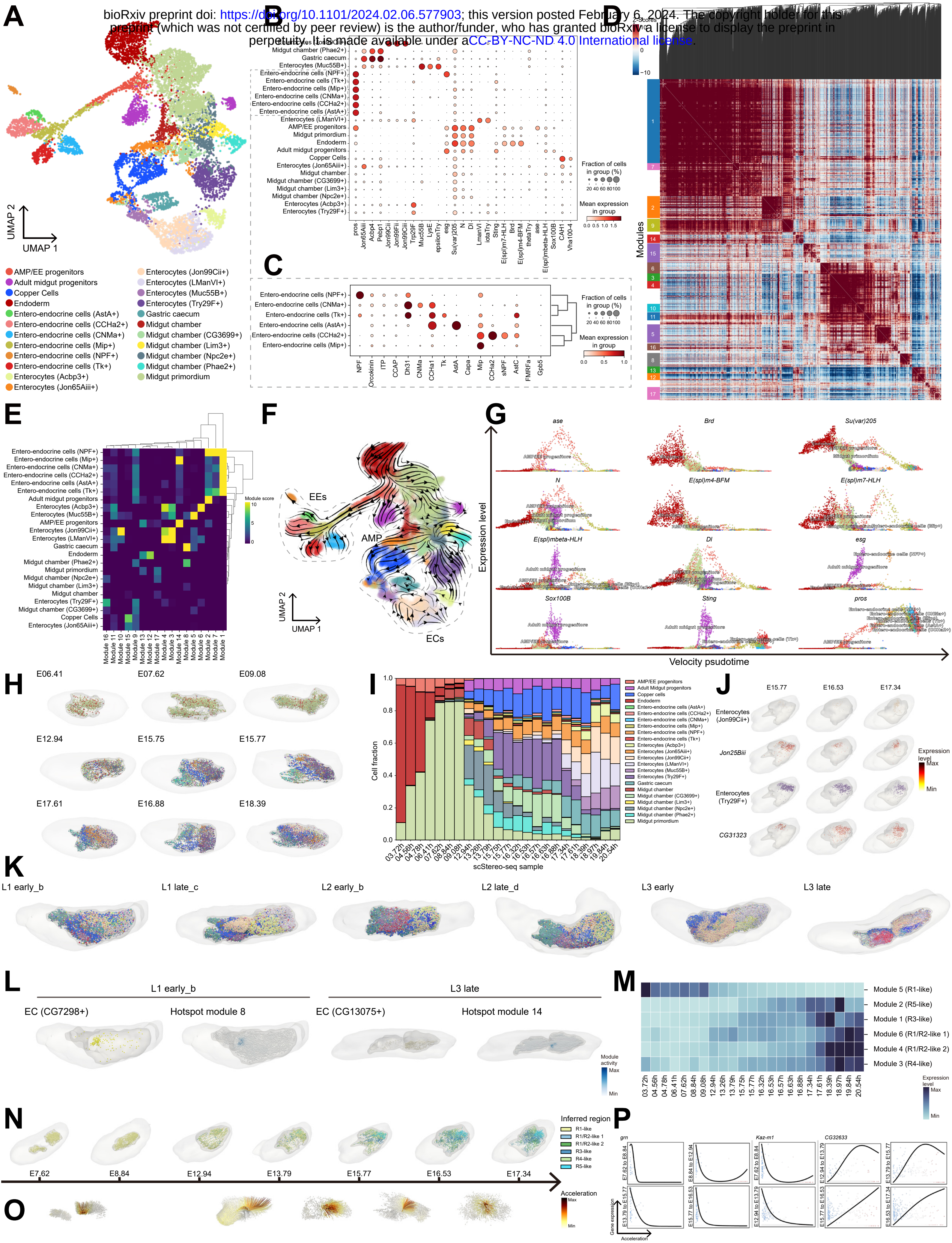
A
UMAP 2
UMAP 1
AMP/EE progenitors
Adult midgut progenitors
Copper Cells
Endoderm
Entero-endocrine cells (AstA+)
Entero-endocrine cells (CCHa2+)
Entero-endocrine cells (CNMa+)
Entero-endocrine cells (Mip+)
Entero-endocrine cells (NPF+)
Entero-endocrine cells (Tk+)
Enterocytes (Acbp3+)
Enterocytes (Jon65Aiii+)
Enterocytes (Jon99Cii+)
Enterocytes (LManVI+)
Enterocytes (Muc55B+)
Enterocytes (Try29F+)
Gastric caecum
Midgut chamber
Midgut chamber (CG3699+)
Midgut chamber (Lim3+)
Midgut chamber (Npc2e+)
Midgut chamber (Phae2+)
Midgut primordium
B
Enterocytes (Jon65Aiii+)
Midgut chamber (Phae2+)
Gastric caecum
Enterocytes (Muc55B+)
Entero-endocrine cells (NPF+)
Entero-endocrine cells (Tk+)
Entero-endocrine cells (Mip+)
Entero-endocrine cells (CNMa+)
Entero-endocrine cells (CCHa2+)
Entero-endocrine cells (AstA+)
Enterocytes (LManVI+)
AMP/EE progenitors
Midgut primordium
Endoderm
Adult midgut progenitors
Copper Cells
Enterocytes (Jon65Aiii+)
Midgut chamber
Midgut chamber (CG3699+)
Midgut chamber (Lim3+)
Midgut chamber (Npc2e+)
Enterocytes (Acbp3+)
Enterocytes (Try29F+)
pros
Jon65Aiii
Acbp4
Pebp1
Jon99Cii
Jon99Fii
Jon99Ciii
Trp29F
Muc55B
LysE
epsilonTry
esg
Su(var)205
N
Dl
LmanVI
iotaTry
Sting
E(spl)m7-HLH
Brd
E(spl)m4-BFM
thetaTry
ase
E(spl)mbeta-HLH
Sox100B
CAH1
Vha100-4
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0 1.5
C
Entero-endocrine cells (NPF+)
Entero-endocrine cells (CNMa+)
Entero-endocrine cells (Tk+)
Entero-endocrine cells (AstA+)
Entero-endocrine cells (CCHa2+)
Entero-endocrine cells (Mip+)
NPF
Orcokinin
ITP
CCAP
Dh31
CNMa
CCHa1
Tk
AstA
Capa
Mip
CCHa2
sNPF
AstC
FMRFa
Gpb5
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
D
Z-scores
10
0
−10
Modules
1
7
2
9
14
15
6
3
4
10
11
5
16
8
13
12
17
E
Entero-endocrine cells (NPF+)
Entero-endocrine cells (Mip+)
Entero-endocrine cells (CNMa+)
Entero-endocrine cells (CCHa2+)
Entero-endocrine cells (AstA+)
Entero-endocrine cells (Tk+)
Adult midgut progenitors
Enterocytes (Acbp3+)
Enterocytes (Muc55B+)
AMP/EE progenitors
Enterocytes (Jon99Cii+)
Enterocytes (LManVI+)
Gastric caecum
Endoderm
Midgut chamber (Phae2+)
Midgut primordium
Midgut chamber (Npc2e+)
Midgut chamber (Lim3+)
Midgut chamber
Enterocytes (Try29F+)
Midgut chamber (CG3699+)
Copper Cells
Enterocytes (Jon65Aiii+)
Module 16
Module 11
Module 10
Module 15
Module 9
Module 13
Module 12
Module 17
Module 4
Module 3
Module 14
Module 8
Module 5
Module 6
Module 2
Module 7
Module 1
Module score
10
5
0
F
EEs
AMP
ECs
UMAP 2
UMAP 1
G
Expression level
Velocity psudotime
ase
Brd
Su(var)205
N
E(spl)m4-BFM
E(spl)m7-HLH
E(spl)mbeta-HLH
Dl
esg
Sox100B
Sting
pros
AMP/EE progenitors
Adult midgut progenitors
Endoderm
Midgut primordium
Entero-endocrine cells (NPF+)
Entero-endocrine cells (Mip+)
Entero-endocrine cells (Tk+)
Entero-endocrine cells (CNMa+)
Entero-endocrine cells (AstA+)
Entero-endocrine cells (CCHa2+)
H
E06.41
E07.62
E09.08
E12.94
E15.75
E15.77
E17.61
E16.88
E18.39
I
Cell fraction
1.0
0.8
0.6
0.4
0.2
0.0
03.72h
04.56h
04.78h
06.41h
07.62h
08.84h
09.08h
12.94h
13.26h
13.79h
15.75h
15.77h
16.32h
16.53h
16.57h
16.63h
16.88h
17.34h
17.61h
18.39h
18.97h
19.84h
20.54h
scStereo-seq sample
AMP/EE progenitors
Adult Midgut progenitors
Copper cells
Endoderm
Entero-endocrine cells (AstA+)
Entero-endocrine cells (CCHa2+)
Entero-endocrine cells (CNMa+)
Entero-endocrine cells (Mip+)
Entero-endocrine cells (NPF+)
Entero-endocrine cells (Tk+)
Enterocytes (Acbp3+)
Enterocytes (Jon65Aiii+)
Enterocytes (Jon99Cii+)
Enterocytes (LManVI+)
Enterocytes (Muc55B+)
Enterocytes (Try29F+)
Gastric caecum
Midgut chamber
Midgut chamber (CG3699+)
Midgut chamber (Lim3+)
Midgut chamber (Npc2e+)
Midgut chamber (Phae2+)
Midgut primordium
J
E15.77
E16.53
E17.34
Enterocytes (Jon99Cii+)
Jon25Biii
Enterocytes (Try29F+)
CG31323
Expression level
Max
Min
K
L1 early_b
L1 late_c
L2 early_b
L2 late_d
L3 early
L3 late
L
L1 early_b
L3 late
EC (CG7298+)
Hotspot module 8
EC (CG13075+)
Hotspot module 14
Module activity
Max
Min
M
Module 5 (R1-like)
Module 2 (R5-like)
Module 1 (R3-like)
Module 6 (R1/R2-like 1)
Module 4 (R1/R2-like 2)
Module 3 (R4-like)
03.72h
04.56h
04.78h
06.41h
07.62h
08.84h
09.08h
12.94h
13.26h
13.79h
15.75h
15.77h
16.32h
16.53h
16.57h
16.63h
16.88h
17.34h
17.61h
18.39h
18.97h
19.84h
20.54h
Expression level
Max
Min
N
Inferred region
R1-like
R1/R2-like 1
R1/R2-like 2
R3-like
R4-like
R5-like
E7.62
E8.84
E12.94
E13.79
E15.77
E16.53
E17.34
O
Acceleration
Max
Min
P
grn
Kaz-m1
CG32633
E7.62 to E8.84
E8.84 to E12.94
E7.62 to E8.84
E12.94 to E13.79
E13.79 to E15.77
E13.79 to E15.77
E15.77 to E16.53
E12.94 to E13.79
E15.77 to E16.53
E16.53 to E17.34
Gene expression
Acceleration

Figure 6 Cell type diversity and functional regionalization in midgut.

(A) UMAP plot showing subclustering and annotation of endoderm and midgut cells from scRNA-seq data, derived from *Dynamo* analysis. **(B)** Bubble plot showing expression level and enrichment of top marker genes of cell types in **(A)**. **(C)** Same as **(B)** but within entero-endocrine cells. **(D)** Heat map showing correlation of functional gene modules identified by *Hotspot* in scRNA-seq data. Each row and each column represent a module marker gene, and *Z*-score indicates their correlation. **(E)** Heat map showing enrichment and clustering of *Hotspot* identified gene modules from **(D)** in midgut cell types in scRNA-seq data. **(F)** RNA velocity flow projected in UMAP space in **(A)**. Cell type color codes are the same as **(A)**. Dashed lines mark clusters representing adult midgut progenitors (AMPs), entero-endocrine cells (EEs), and enterocytes (ECs) discussed in the main text. **(G)** Dot plots showing relationship between velocity derived pseudotime and expression levels of genes of interest during differentiation of AMPs and EEs. Each dot represents one cell from midgut scRNA-seq data. **(H)** 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 **(A)**. **(I)** Bar plot showing cell type composition of midgut in scStereo-seq samples. Cell types are label transferred from scRNA-seq data. **(J)** Same as **(H)** but showing spatial distribution of copper cells, EC (*Jon99Cii*+), and EC (*Try29F*+) and their cell type marker genes in representative scStereo-seq samples. **(K)** Same as **(H)** but for larva scStereo-seq samples. Cell type color codes are the same as **Figure S11D**. Samples are not on the same scale. **(L)** 3D midgut models of L1 early and L3 late scStereo-seq samples, showing spatial distribution of representative ECs and their corresponding functional gene modules. **(M)** Heat map showing expression level of region-related gene modules across scStereo-seq samples. **(N)** Same as **(H)** but showing spatial distribution of inferred “adult midgut” regions. **(O)** 3D midgut cell migration trajectories and acceleration scores across 7 scStereo-seq samples of developmental age between 7 and 18 h. Sample viewpoints are different from **(N)** for better visualization of trajectories. **(P)** General linear model-based correlation between acceleration scores and expression levels of *grn*, *Kaz-m1*, and *CG32633* in transitions between representative scStereo-seq samples.

METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for the resources and reagents may be directed to the corresponding author 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.

Data and code availability

Raw data generated by Stereo-seq, scRNA-seq, and scATAC-seq in this study and associated analysis protocols and software can be accessed in our online database, Flysta3D. All data were analyzed with standard programs and packages, as detailed in Method details. Processed matrices 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). All custom codes using open-source software to support this study are provided in a public GitHub repository. Any additional information required to re-analyze the data reported in this study is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Fly strain maintenance

All Stereo-seq, scRNA-seq, and scATAC-seq samples were from *Drosophila* strain *w1118*. Flies were maintained on cornmeal-sucrose-agar media in a 25 ℃ incubator on a 12 h/12 h light/dark cycle.

Fly sample preparation

Samples were prepared and embedded for cryosection and Stereo-seq as previously described[13]. Unless otherwise specified, the samples were sectioned along the left-right axis to represent sagittal planes.

For scRNA-seq, single cells were isolated and fixed following protocols described in Ref[120] and stored at -20 ℃ until further use.

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

METHOD DETAILS

See method details in Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Figures, Tables, Movies, and Data can be found in the Supplemental Information.

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. 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. 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. 10.1093/bib/bbab454.

4. 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. 10.1126/science.aan3235.

5. 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. 10.1126/science.abn5800.

6. Brunet Avalos, C., Maier, G.L., Bruggmann, R., and Sprecher, S.G. Single cell transcriptome atlas of the Drosophila larval brain. eLife *8*, e50354. 10.7554/eLife.50354.

7. 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. 10.1038/s41592-019-0492-x.

8. Li, H., Janssens, J., De Waegeneer, M., Kolluru, S.S., Davie, K., Gardeux, V., Saelens, W., David, F.P.A., Brbić, M., Spanier, K., et al. (2022). Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. Science *375*, eabk2432. 10.1126/science.abk2432.

9. Worley, M.I., Everetts, N.J., Yasutomi, R., Chang, R.J., Saretha, S., Yosef, N., and Hariharan, I.K. (2022). Ets21C sustains a pro-regenerative transcriptional program in blastema cells of Drosophila imaginal discs. Curr. Biol. *32*, 3350-3364.e6. 10.1016/j.cub.2022.06.040.

10. Lu, T.-C., Brbić, M., Park, Y.-J., Jackson, T., Chen, J., Kolluru, S.S., Qi, Y., Katheder, N.S., Cai, X.T., Lee, S., et al. (2023). Aging Fly Cell Atlas identifies exhaustive aging features at cellular resolution. Science *380*, eadg0934. 10.1126/science.adg0934.

11. Mahowald, A.P., and Hardy, P.A. (1985). Genetics of Drosophila Embryogenesis. Annu. Rev. Genet. *19*, 149–177. 10.1146/annurev.ge.19.120185.001053.

12. 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.e21. 10.1016/j.cell.2022.04.003.

13. 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.e4. 10.1016/j.devcel.2022.04.006.

14. Qiu, X., Zhu, D.Y., Yao, J., Jing, Z., Zuo, L., Wang, M., Min, K.H. (Joseph), Pan, H., Wang, S., Liao, S., et al. (2022). Spateo: multidimensional spatiotemporal modeling of single-cell spatial transcriptomics. Preprint at bioRxiv, 10.1101/2022.12.07.519417 10.1101/2022.12.07.519417.

15. 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. 10.1186/gb-2002-3-12-research0088.

16. 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. 10.1016/j.cell.2007.08.003.

17. 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. 10.1038/nature25981.

18. 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.e9. 10.1016/j.devcel.2020.10.009.

19. 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. 10.1093/nar/gkac958.

20. 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. 10.1093/nar/gky957.

21. 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. 10.1038/nature13395.

22. 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. 10.1093/nar/gkab1032.

23. 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. 10.1126/science.1072387.

24. 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. 10.1186/s13630-015-0018-9.

25. 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*. 10.3389/fgene.2019.00024.

26. Micchelli, C.A., Sudmeier, L., Perrimon, N., Tang, S., and Beehler-Evans, R. (2011). Identification of adult midgut precursors in Drosophila. Gene Expr. Patterns GEP *11*, 12–21. 10.1016/j.gep.2010.08.005.

27. 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.

28. Horváth, B., and Kalinka, A.T. (2018). The genetics of egg retention and fertilization success in Drosophila: One step closer to understanding the transition from facultative to obligate viviparity. Evolution *72*, 318–336. 10.1111/evo.13411.

29. Bulteau, R., and Francesconi, M. (2022). Real age prediction from the transcriptome with RAPToR. Nat. Methods *19*, 969–975. 10.1038/s41592-022-01540-0.

30. 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. 10.1038/s41596-021-00573-7.

31. Wang, K., Hou, L., Wang, X., Zhai, X., Lu, Z., Zi, Z., Zhai, W., He, X., Curtis, C., Zhou, D., et al. (2023). PhyloVelo enhances transcriptomic velocity field mapping using monotonically expressed genes. Nat. Biotechnol. 10.1038/s41587-023-01887-5.

32. 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. 10.1126/science.aax0249.

33. Wang, R., and Amoyel, M. (2022). mRNA Translation Is Dynamically Regulated to Instruct Stem Cell Fate. Front. Mol. Biosci. *9*.

34. Liu, X., Krause, W.C., and Davis, R.L. (2007). GABAA Receptor RDL Inhibits Drosophila Olfactory Associative Learning. Neuron *56*, 1090–1102. 10.1016/j.neuron.2007.10.036.

35. Kramerova, I.A., Kawaguchi, N., Fessler, L.I., Nelson, R.E., Chen, Y., Kramerov, A.A., Kusche-Gullberg, M., Kramer, J.M., Ackley, B.D., Sieron, A.L., et al. (2000). Papilin in development; a pericellular protein with a homology to the

ADAMTS metalloproteinases. Dev. Camb. Engl. *127*, 5475–5485. 10.1242/dev.127.24.5475.

36. 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. 10.1093/genetics/iyab235.

37. Bier, E., and De Robertis, E.M. (2015). BMP gradients: A paradigm for morphogen-mediated developmental patterning. Science *348*, aaa5838. 10.1126/science.aaa5838.

38. 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. 10.1016/j.conb.2010.08.014.

39. 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. 10.1146/annurev.cellbio.19.031403.160043.

40. 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–re4. 10.1126/scisignal.2005096.

41. Campos-Ortega, J.A., and Hartenstein, V. (1997). The Embryonic Development of Drosophila melanogaster (Springer) 10.1007/978-3-662-22489-2.

42. 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. 10.1242/dev.120.9.2489.

43. 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. 10.4161/cc.10.7.14830.

44. 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. 10.1242/dev.178087.

45. 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. 10.1242/dev.00148.

46. 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. 10.1242/dev.092924.

47. 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.

10.1016/j.ydbio.2003.12.004.

48. 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. 10.1101/gad.917101.

49. 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. 10.1242/dev.128.7.1193.

50. 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. 10.1093/emboj/cdf545.

51. Eresh, S., Riese, J., Jackson, D.B., Bohmann, D., and Bienz, M. (1997). A CREB-binding site as a target for decapentaplegic signalling during Drosophila endoderm induction. EMBO J. *16*, 2014–2022. 10.1093/emboj/16.8.2014.

52. 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. 10.1016/j.ydbio.2004.11.021.

53. Stainier, D.Y.R. (2002). A glimpse into the molecular entrails of endoderm formation. Genes Dev. *16*, 893–907. 10.1101/gad.974902.

54. Blaschke, R.J., Monaghan, A.P., Schiller, S., Schechinger, B., Rao, E., Padilla-Nash, H., Ried, T., and Rappold, G.A. (1998). SHOT, a SHOX-related homeobox gene, is implicated in craniofacial, brain, heart, and limb development. Proc. Natl. Acad. Sci. U. S. A. *95*, 2406–2411. 10.1073/pnas.95.5.2406.

55. Rao, E., Weiss, B., Fukami, M., Rump, A., Niesler, B., Mertz, A., Muroya, K., Binder, G., Kirsch, S., Winkelmann, M., et al. (1997). Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. Nat. Genet. *16*, 54–63. 10.1038/ng0597-54.

56. 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. 10.1186/s12861-015-0069-6.

57. 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. 10.1038/nmeth.4463.

58. 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. (2022). Inferring and perturbing cell fate regulomes in human brain organoids. Nature, 1–8. 10.1038/s41586-022-05279-8.

59. Liao, K.A., González-Morales, N., and Schöck, F. (2016). Zasp52, a Core Z-disc Protein in Drosophila Indirect Flight Muscles, Interacts with α-Actinin via an

Extended PDZ Domain. PLoS Genet. *12*, e1006400. 10.1371/journal.pgen.1006400.

60. Popichenko, D., Sellin, J., Bartkuhn, M., and Paululat, A. (2007). Hand is a direct target of the forkhead transcription factor Biniou during Drosophilavisceral mesoderm differentiation. BMC Dev. Biol. *7*, 49. 10.1186/1471-213X-7-49.

61. Sulkowski, M., Kim, Y.-J., and Serpe, M. (2014). Postsynaptic glutamate receptors regulate local BMP signaling at the Drosophila neuromuscular junction. Development *141*, 436–447. 10.1242/dev.097758.

62. Yu, J., Liu, Y., Lan, X., Wu, H., Wen, Y., Zhou, Z., Hu, Z., Sha, J., Guo, X., and Tong, C. (2016). CHES-1-like , the ortholog of a non-obstructive azoospermia-associated gene, blocks germline stem cell differentiation by upregulating Dpp expression in Drosophila testis. Oncotarget *7*, 42303–42313. 10.18632/oncotarget.9789.

63. McDonald, J.A., Fujioka, M., Odden, J.P., Jaynes, J.B., and Doe, C.Q. (2003). Specification of motoneuron fate in Drosophila: Integration of positive and negative transcription factor inputs by a minimal eve enhancer. J. Neurobiol. *57*, 193–203. 10.1002/neu.10264.

64. Lawton, K.J., Wassmer, T.L., and Deitcher, D.L. (2014). Conserved role of Drosophila melanogaster FoxP in motor coordination and courtship song. Behav. Brain Res. *268*, 213–221. 10.1016/j.bbr.2014.04.009.

65. 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. 10.1242/dev.200815.

66. 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. 10.1016/j.cell.2014.08.009.

67. Chen, H., Albergante, L., Hsu, J.Y., Lareau, C.A., Lo Bosco, G., Guan, J., Zhou, S., Gorban, A.N., Bauer, D.E., Aryee, M.J., et al. (2019). Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat. Commun. *10*, 1903. 10.1038/s41467-019-09670-4.

68. Kao, C.-F., Yu, H.-H., He, Y., Kao, J.-C., and Lee, T. (2012). Hierarchical Deployment of Factors Regulating Temporal Fate in a Diverse Neuronal Lineage of the Drosophila Central Brain. Neuron *73*, 677–684. 10.1016/j.neuron.2011.12.018.

69. Isshiki, T., Pearson, B., Holbrook, S., and Doe, C.Q. (2001). Drosophila neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. Cell *106*, 511–521. 10.1016/s0092-8674(01)00465-2.

70. Nibu, Y., Senger, K., and Levine, M. (2003). CtBP-independent repression in the Drosophila embryo. Mol. Cell. Biol. *23*, 3990–3999. 10.1128/MCB.23.11.3990-3999.2003.

71. Shimell, M.J., Peterson, A.J., Burr, J., Simon, J.A., and O'Connor, M.B. (2000). Functional analysis of repressor binding sites in the iab-2 regulatory region of the abdominal-A homeotic gene. Dev. Biol. *218*, 38–52. 10.1006/dbio.1999.9576.

72. Doe, C.Q. (2017). Temporal Patterning in the Drosophila CNS. Annu. Rev. Cell Dev. Biol. *33*, 219–240. 10.1146/annurev-cellbio-111315-125210.

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

74. Abdusselamoglu, M.D., Eroglu, E., Burkard, T.R., and Knoblich, J.A. The transcription factor odd-paired regulates temporal identity in transit-amplifying neural progenitors via an incoherent feed-forward loop. eLife *8*, e46566. 10.7554/eLife.46566.

75. Huang, J.D., Dubnicoff, T., Liaw, G.J., Bai, Y., Valentine, S.A., Shirokawa, J.M., Lengyel, J.A., and Courey, A.J. (1995). Binding sites for transcription factor NTF-1/Elf-1 contribute to the ventral repression of decapentaplegic. Genes Dev. *9*, 3177–3189. 10.1101/gad.9.24.3177.

76. Attardi, L.D., and Tjian, R. (1993). Drosophila tissue-specific transcription factor NTF-1 contains a novel isoleucine-rich activation motif. Genes Dev. *7*, 1341–1353. 10.1101/gad.7.7b.1341.

77. Tran, K.D., Miller, M.R., and Doe, C.Q. (2010). Recombineering Hunchback identifies two conserved domains required to maintain neuroblast competence and specify early-born neuronal identity. Dev. Camb. Engl. *137*, 1421–1430. 10.1242/dev.048678.

78. Gunnar, E., Bivik, C., Starkenberg, A., and Thor, S. (2016). *sequoia* Controls the type I>0 daughter proliferation switch in the developing *Drosophila* nervous system. Development, dev.139998. 10.1242/dev.139998.

79. Xiao, Q., Komori, H., and Lee, C.-Y. (2012). klumpfuss distinguishes stem cells from progenitor cells during asymmetric neuroblast division. Dev. Camb. Engl. *139*, 2670–2680. 10.1242/dev.081687.

80. Avva, S.V.S.P., and Hart, C.M. (2016). Characterization of the Drosophila BEAF-32A and BEAF-32B Insulator Proteins. PloS One *11*, e0162906. 10.1371/journal.pone.0162906.

81. Marqués, G. (2005). Morphogens and synaptogenesis in Drosophila. J. Neurobiol. *64*, 417–434. 10.1002/neu.20165.

82. Corty, M.M., and Coutinho-Budd, J. (2023). Drosophila glia take shape to sculpt the nervous system. Curr. Opin. Neurobiol. *79*, 102689. 10.1016/j.conb.2023.102689.

83. Hirth, F., Hartmann, B., and Reichert, H. (1998). Homeotic gene action in embryonic brain development of Drosophila. Development *125*, 1579–1589. 10.1242/dev.125.9.1579.

84. Estacio-Gómez, A., and Díaz-Benjumea, F.J. (2014). Roles of Hox genes in the patterning of the central nervous system of Drosophila. Fly (Austin) *8*, 26–32. 10.4161/fly.27424.

85. 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. 10.1242/dev.175265.

86. 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. 10.1016/j.cois.2021.02.008.

87. Lemaitre, B., and Miguel-Aliaga, I. (2013). The Digestive Tract of Drosophila melanogaster. Annu. Rev. Genet. *47*, 377–404. 10.1146/annurev-genet-111212-133343.

88. 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. U. S. A. *117*, 1514–1523. 10.1073/pnas.1916820117.

89. 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. 10.1126/science.1181958.

90. Lin, G., Xu, N., and Xi, R. (2008). Paracrine Wingless signalling controls self-renewal of Drosophila intestinal stem cells. Nature *455*, 1119–1123. 10.1038/nature07329.

91. Xu, T., Nicolson, S., Denton, D., and Kumar, S. (2015). Distinct requirements of Autophagy-related genes in programmed cell death. Cell Death Differ. *22*, 1792–1802. 10.1038/cdd.2015.28.

92. Scuderi, A., Simin, K., Kazuko, S.G., Metherall, J.E., and Letsou, A. (2006). scylla and charybde, homologues of the human apoptotic gene RTP801, are required for head involution in Drosophila. Dev. Biol. *291*, 110–122. 10.1016/j.ydbio.2005.12.014.

93. DeTomaso, D., and Yosef, N. (2021). Hotspot identifies informative gene modules across modalities of single-cell genomics. Cell Syst. *12*, 446-456.e9. 10.1016/j.cels.2021.04.005.

94. Jiang, H., and Edgar, B.A. (2009). EGFR signaling regulates the proliferation of Drosophila adult midgut progenitors. Development *136*, 483–493. 10.1242/dev.026955.

95. Qiu, X., Zhang, Y., Martin-Rufino, J.D., Weng, C., Hosseinzadeh, S., Yang, D., Pogson, A.N., Hein, M.Y., Min, K.H. (Joseph), Wang, L., et al. (2022). Mapping transcriptomic vector fields of single cells. Cell *185*, 690-711.e45. 10.1016/j.cell.2021.12.045.

96. Martin, M., Hiroyasu, A., Guzman, R.M., Roberts, S.A., and Goodman, A.G. (2018). Analysis of Drosophila STING Reveals an Evolutionarily Conserved

Antimicrobial Function. Cell Rep. *23*, 3537-3550.e6. 10.1016/j.celrep.2018.05.029.

97. Moussian, B. (2013). The apical plasma membrane of chitin-synthesizing epithelia. Insect Sci. *20*, 139–146. 10.1111/j.1744-7917.2012.01549.x.

98. 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. CB *19*, 1741. 10.1016/j.cub.2009.08.042.

99. Hakim, R.S., Baldwin, K., and Smagghe, G. (2010). Regulation of midgut growth, development, and metamorphosis. Annu. Rev. Entomol. *55*, 593–608. 10.1146/annurev-ento-112408-085450.

100. Marianes, A., and Spradling, A.C. (2013). Physiological and stem cell compartmentalization within the Drosophila midgut. eLife. 10.7554/eLife.00886.

101. 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. 10.1016/j.celrep.2013.04.001.

102. 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. 10.1038/srep27242.

103. Miguel-Aliaga, I., Jasper, H., and Lemaitre, B. (2018). Anatomy and Physiology of the Digestive Tract of Drosophila melanogaster. Genetics *210*, 357–396. 10.1534/genetics.118.300224.

104. Brown, S., and Castelli-Gair Hombría, J. (2000). Drosophila grain encodes a GATA transcription factor required for cell rearrangement during morphogenesis. Dev. Camb. Engl. *127*, 4867–4876. 10.1242/dev.127.22.4867.

105. Wurmbach, E., Wech, I., and Preiss, A. (1999). The Enhancer of split complex of Drosophila melanogaster harbors three classes of Notch responsive genes. Mech. Dev. *80*, 171–180. 10.1016/S0925-4773(98)00212-3.

106. Li, X., and Karpac, J. (2023). A distinct Acyl-CoA binding protein (ACBP6) shapes tissue plasticity during nutrient adaptation in Drosophila. Nat. Commun. *14*, 7599. 10.1038/s41467-023-43362-4.

107. Burns, A.M., and Gräff, J. (2021). Cognitive epigenetic priming: leveraging histone acetylation for memory amelioration. Curr. Opin. Neurobiol. *67*, 75–84. 10.1016/j.conb.2020.08.011.

108. Ernst, C., and Jefri, M. (2021). Epigenetic priming in neurodevelopmental disorders. Trends Mol. Med. *27*, 1106–1114. 10.1016/j.molmed.2021.09.005.

109. Malita, A., and Rewitz, K. (2021). Interorgan communication in the control of metamorphosis. Curr. Opin. Insect Sci. *43*, 54–62. 10.1016/j.cois.2020.10.005.

110. Boulan, L., Andersen, D., Colombani, J., Boone, E., and Léopold, P. (2019). Inter-Organ Growth Coordination Is Mediated by the Xrp1-Dilp8 Axis in Drosophila. Dev. Cell *49*, 811-818.e4. 10.1016/j.devcel.2019.03.016.

111. Boulan, L., and Léopold, P. (2021). What determines organ size during development and regeneration? Development *148*, dev196063. 10.1242/dev.196063.

112. Winding, M., Pedigo, B.D., Barnes, C.L., Patsolic, H.G., Park, Y., Kazimiers, T., Fushiki, A., Andrade, I.V., Khandelwal, A., Valdes-Aleman, J., et al. (2023). The connectome of an insect brain. Science *379*, eadd9330. 10.1126/science.add9330.

113. 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. 10.7554/eLife.50822.

114. Yuan, Z., Pan, W., Zhao, X., Zhao, F., Xu, Z., Li, X., Zhao, Y., Zhang, M.Q., and Yao, J. (2023). SODB facilitates comprehensive exploration of spatial omics data. Nat. Methods *20*, 387–399. 10.1038/s41592-023-01773-7.

115. Zheng, Y., Chen, Y., Ding, X., Wong, K.H., and Cheung, E. (2023). Aquila: a spatial omics database and analysis platform. Nucleic Acids Res. *51*, D827–D834. 10.1093/nar/gkac874.

116. Liu, X., Zeira, R., and Raphael, B.J. (2023). Partial alignment of multislice spatially resolved transcriptomics data. Genome Res. *33*, 1124–1132. 10.1101/gr.277670.123.

117. Wang, G., Zhao, J., Yan, Y., Wang, Y., Wu, A.R., and Yang, C. (2023). Construction of a 3D whole organism spatial atlas by joint modelling of multiple slices with deep neural networks. Nat. Mach. Intell. *5*, 1200–1213. 10.1038/s42256-023-00734-1.

118. 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. *119*, e2206450119. 10.1073/pnas.2206450119.

119. 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. 10.7554/eLife.63632.

120. 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. 10.1186/s12915-017-0383-5.