Zebrahub-Multiome: Uncovering Gene Regulatory Network Dynamics During Zebrafish Embryogenesis

Yang Joon Kim[1*], Shruthi Vijay Kumar[1], Benjamin Iovino[1], Alejandro Granados[1], Sarah Ancheta[1], Xiang Zhao[1], Kyle Awayan[1], Amanda Seng[1], Michael Borja[1], Sheryl Paul[1], Honey Mekonen[1], Ritwicq Arjyal[1], Angela Detweiler[1], Yasin Şenbabaoğlu[1], Rafael Gómez-Sjöberg[1], Norma Neff[1], Merlin Lange[1*], Loïc A. Royer[1*]

[1]*CZ Biohub, San Francisco, USA.*

During embryonic development, gene regulatory networks (GRNs) drive molecular differentiation of cell types. However, the temporal dynamics of these networks remain poorly understood. Here, we present Zebrahub-Multiome, a comprehensive, time-resolved atlas of zebrafish embryogenesis, integrating single-cell chromatin accessibility (scATAC-seq) and gene expression (scRNA-seq) from over 94,000 cells sampled across six key developmental stages (10 to 24 hours post-fertilization). Our analysis reveals early-stage GRNs shared across multiple lineages, followed by the emergence of lineage-specific regulatory programs during later stages. We also observe a shift in transcription factor (TF) influence from broad, multi-lineage roles in early development to more specialized, cell-type-specific functions as development progresses. Using *in silico* genetic perturbations, we highlight the dynamic role of TFs in driving cell fate decisions, emphasizing the gradual specialization of regulatory circuits. All data and analyses are made accessible through an interactive web portal, enabling users to explore zebrafish gene regulatory dynamics across time and cell types. This resource provides a foundation for hypothesis generation and deeper insights into vertebrate development.

Introduction

The intricate process of embryonic development is orchestrated by a complex interplay of gene regulatory networks (GRNs), in which transcription factors (TFs) interact with sets of *cis*-regulatory elements to control the transcription of target genes. Understanding the dynamic roles of TFs across developmental stages and cell types is crucial for deciphering the regulatory logic governing embryogenesis[1,2]. Over the past decade, advancements in single-cell genomic technologies have enabled remarkable progress in developmental biology through systematic and *in-toto* measurement of transcriptomes over time at the whole-embryo scale[3–5]. These technologies have successfully mapped the emergence of cell types in various model organisms, including worm[6], fly[7], zebrafish[4,8–10], and mouse[11,12].

More recently, multi-omics approaches combining genome-wide chromatin accessibility (scATAC-seq) with single-cell RNA sequencing (scRNA-seq) have deepened our understanding of gene regulation during cell fate specification[13–15]. These technical advances have enabled more accurate inference of gene regulatory networks by incorporating both epigenomic and transcriptomic information[16–20].

Although multi-omic technologies have been employed to investigate the dynamics of gene regulatory networks (GRNs), several limitations persist. First, most GRN studies have focused on cell-type-specific analyses[14,17], often overlooking the temporal dynamics of GRNs within individual cell types. Second, a prior study[21] attempted to integrate both cell-type and temporal resolution in their analysis, but relied on bulk RNA-seq and ATAC-seq, techniques better suited for homogeneous cell populations. These bulk approaches fail to capture the cellular heterogeneity present in developing organisms, where numerous cell types coexist. Furthermore, their work was limited to hematopoiesis, leaving the dynamics of whole-organism GRNs unexplored. To address these gaps, we introduce Zebrahub-Multiome, a comprehensive, time-resolved single-cell multi-omic atlas of zebrafish development (10 to 24 hpf), which resolves distinct cellular states and trajectories across early developmental stages.

To expand the accessibility of this resource and facilitate the generation of hypotheses, we developed an interactive web portal that allows the exploration of dynamic cis-regulatory elements and gene regulatory networks for genes and cell types of interest. Most analyses and examples presented in this work are generalized to all genes and cell types in our online *interactive figures* accessible at zebrahub.org/epigenomics.

Results

A time-resolved single-cell multi-omic atlas of zebrafish development reveals temporal dynamics of chromatin accessibility and gene expression at the whole-embryo level.

To investigate the epigenomic and transcriptomic profiles during cell fate specification in zebrafish embryogenesis, we conducted a single-cell Multiome assay (see Methods), which simultaneously captures gene expression (RNA) and chromatin accessibility (ATAC) from the same cells. We sampled six time points between 10 and 24 hours post-fertilization (hpf), a critical developmental window spanning the transition from the end of gastrulation to the onset of organogenesis (Fig.1a; see Methods and Fig.S1 for the experimental protocol and data processing workflow). After quality control, the dataset comprised 94,562 cells (post quality control), with a median of 1,400 detected genes per cell (median 3,700 UMI counts per cell) and a median of 15,000 peaks per cell (Fig. S2).

To visualize the epigenomic and transcriptomic profiles of these cells, we computed uniform manifold approximation and projections (UMAPs) integrating all developmental stages (Fig. 1a; see Methods for details). Cell type annotation was per-

* Correspondence: yang-joon.kim, merlin.lange, and loic.royer @czbiohub.org

formed in two steps: first, the RNA modality was used for label transfer from a reference atlas[10], followed by expert curation using RNA marker genes and zebrafish ontology terms. We generated three types of UMAPs for visualization: (i) RNA, (ii) ATAC, and (iii) Joint (RNA+ATAC), as shown in Fig. 1b, c, and d, and available for exploration online. As shown in Fig. 1b, c, and d, the joint UMAP showed a clearer separation of major cell types, such as mesoderm, endoderm, neuro-ectoderm, and hematopoietic vasculature, as well as more continuous developmental trajectories compared to either RNA or ATAC modalities alone (see Fig. 1e and Methods).

To investigate the dynamics of chromatin accessibility and gene expression during development, we quantified changes in average chromatin accessibility and gene expression at the whole-embryo level for each time point. As a proxy for chromatin accessibility at each gene locus, we used the gene activity score, which is computed by summing ATAC-seq reads within the gene body and its proximal regulatory regions[22] (see Methods).

As an example, we focused on two key transcription factors involved in myogenesis: *myf5* and *myl1*. *myf5* promotes early myogenesis, acting before our initial time point of 10 hpf, while *myl1* initiates muscle differentiation at later stages[23–26]. We observed that *myf5* exhibits a decreasing trend in both chromatin accessibility and gene expression (Fig. 1f and g), whereas *myl1* shows an increasing trend, with changes in gene activity preceding changes in gene expression (Fig. 1g). This suggests a temporal delay between epigenetic and transcriptional regulation at the whole-embryo level[27]. Similar analyses for all zebrafish genes are available online as an interactive resource.

To systematically assess how chromatin accessibility and gene expression change together at the whole embryo scale, we computed the Pearson correlation between chromatin accessibility and gene expression across developmental stages. Correlation coefficients over 2000 highly variable genes showed that there is a positive correlation between chromatin accessibility and gene expression dynamics overall (Fig. S3). Highly correlated genes, such as *myf5* and *tbx16*, exhibit a linear relationship between chromatin accessibility and gene expression over time. These genes are crucial early in development (e.g., during myogenesis), where chromatin becomes accessible, and gene expression is activated before the first measured time point at 10 hpf. After this, both chromatin accessibility and gene expression decrease in parallel over time (Fig. 1h, top). In contrast, genes with lower correlation display more complex temporal dynamics, often with chromatin accessibility changes preceding gene expression. This pattern is evident in *myl1* and *hbbe3*, which are important for late muscle differentiation and hematopoiesis, respectively[28]. These genes are likely required during the later stages of organogenesis (19 to 24 hpf), where gene expression is regulated by more intricate dynamics (Fig. 1h, bottom). This analysis is available online as an interactive resource for all genes.

Temporal dynamics of correlation between chromatin accessibility and gene expression is associated with the emergence of cell types.

Our whole embryo analysis revealed temporal changes in the coupling between chromatin accessibility and gene expression. However, analyzing all cells together can obscure important signals and mask differences between individual cell types. We hypothesized that this temporal coupling may vary across specific cell types, particularly in those where the corresponding genes are highly expressed.

To investigate the relationship between chromatin accessibility and gene expression at the cell-type level, we employed a metacell approach using SEACells[29] (see Methods). Metacells are clusters of epigenomically or transcriptionally similar cells that act as aggregated representatives of distinct cellular states. This method addresses the sparsity and noise inherent in single-cell data while preserving the heterogeneity of cell states[29]. For each metacell, we aggregated gene expression counts (RNA) and gene activity scores (ATAC), visualized through UMAP (Fig. 2a;Fig. S4a; see Methods). Using this metacell approach, we uncovered cell-type-specific patterns that were previously obscured in single-cell analyses. By calculating the correlation between gene expression and chromatin accessibility for each metacell at multiple time points, we were able to track how these relationships evolved temporally across different cell types.

Our analysis of over 3,000 highly variable genes revealed distinct temporal patterns in the correlation between chromatin accessibility and gene expression (Fig. 2b). Gene set enrichment analysis (Enrichr)[30,31] showed that genes with increasing correlation are enriched in pathways related to "Heme Biosynthesis" and "Calcium Regulation in Cardiac Cells", both critical for hematopoiesis and cardiac development (see Methods). This aligns with the timing of these lineages' differentiation during mid-gastrulation[32]. Conversely, genes with decreasing correlation are enriched in pathways such as "Nodal/Wnt signaling", "Cytoplasmic ribosomal proteins", and "Neural crest development", which are typically activated earlier to support rapid translation during early developmental fate determination (Fig. 2b). Fig. 2c illustrates the averaged trends of two distinct clusters of genes based on their temporal dynamics: those with increasing correlation and those with decreasing correlation over time.

We observed that the two correlation trends were specific to certain genes and cell types in relation to developmental timing. For example, *slc4a1a*, a gene involved in hematopoiesis and erythrocyte differentiation, showed a clear increase in correlation between chromatin accessibility and gene expression over time (Fig. 2d, top)[33]. The chromatin region became accessible early (10 hpf) and remained open, with transcriptional activation peaking at 24 hpf in the hemangioblast population. In contrast, *cdx1a*, a homeobox gene involved in anterior-posterior axis formation, exhibited a high correlation at 10 hpf, which decreased over time (Fig. 2d, bottom).

We identified specific cell types driving strong chromatin accessibility and gene expression for each gene, and re-clustered the heatmap in Fig. 2b based on cell types and dynamic correlation patterns (Fig. 2e). The re-clustered heatmap reveals how correlation dynamics change over time for each cell type. Notably, cell types prevalent during earlier development (e.g.,

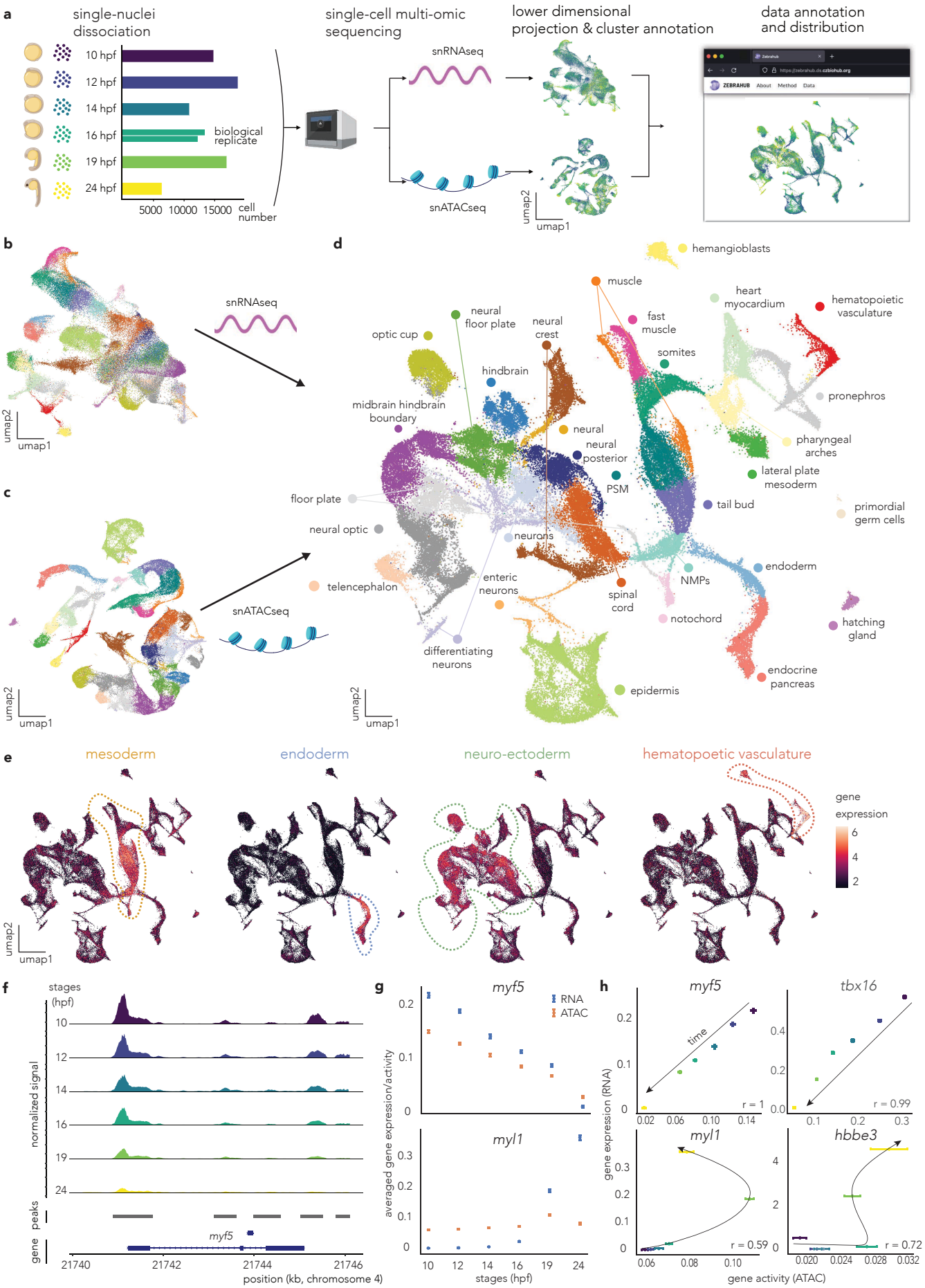
a
single-nuclei dissociation
10 hpf
12 hpf
14 hpf
16 hpf
19 hpf
24 hpf
biological replicate
5000 10000 15000 cell number
single-cell multi-omic sequencing
snRNAseq
snATACseq
lower dimensional projection & cluster annotation
umap2
umap1
data annotation and distribution
ZEBRAHUB
About
Method
Data
b
snRNAseq
umap2
umap1
c
snATACseq
umap2
umap1
d
hemangioblasts
muscle
heart myocardium
hematopoietic vasculature
neural floor plate
neural crest
fast muscle
optic cup
somites
hindbrain
pronephros
midbrain hindbrain boundary
neural
neural posterior
pharyngeal arches
lateral plate mesoderm
PSM
floor plate
tail bud
primordial germ cells
neural optic
neurons
endoderm
telencephalon
enteric neurons
NMPs
spinal cord
notochord
hatching gland
differentiating neurons
endocrine pancreas
epidermis
umap2
umap1
e
mesoderm
endoderm
neuro-ectoderm
hematopoetic vasculature
gene expression
6
4
2
umap2
umap1
f
stages (hpf)
10
12
14
16
19
24
normalized signal
peaks
gene
myf5
21740
21742
21744
21746
position (kb, chromosome 4)
g
myf5
RNA
ATAC
0.2
0.1
0
myl1
0.3
0.2
0.1
0
averaged gene expression/activity
10 12 14 16 19 24
stages (hpf)
h
myf5
time
r = 1
0.02 0.06 0.10 0.14
tbx16
0.4
0.2
0.0
r = 0.99
0.1 0.2 0.3
myl1
0.3
0.2
0.1
0
r = 0.59
0.06 0.08 0.10
hbbe3
4
2
0
r = 0.72
0.020 0.024 0.028 0.032
gene expression (RNA)
gene activity (ATAC)

Figure 1: **A time-resolved single-cell multi-omic atlas of zebrafish embryonic development. a,** A schematic of the experimental and computational analysis workflow. From left to right, we show: First, the number of cells per time point and replicates. Second, single-nuclei dissociation from pooled zebrafish embryos at 10, 12, 14, 16, 19, and 24 hours post fertilization (hpf), followed by single-cell Multiome assay for both snRNA-seq and snATAC-seq. Third, computational analysis of single-cell multiome datasets, including dimensionality reduction, clustering, and cell type annotation. Lastly, a web portal allows the exploration of single-cell Multiome datasets online. **b-d,** UMAP representation of **b,** RNA modality, **c,** ATAC modality, and **d,** joint (RNA+ATAC) colored by annotated cell types. **e,** A joint UMAP representation of gene expression patterns for major developmental lineages: mesoderm, endoderm, neuro-ectoderm, and hematopoietic vasculature. (See Methods for the list of marker genes for each tissue or lineage.) **f,** Genome browser view of the *myf5* locus, showing chromatin accessibility profiles across developmental stages. **g,** Pseudo-bulk analysis of *myf5* and *myl1* gene expression and chromatin accessibility over developmental stages. **h,** Correlation between gene expression and chromatin accessibility at the pseudo-bulk level for *myf5*, *tbx16*, *myl1*, and *hbbe3* genes across developmental stages. (Viridis color palette was used for the developmental stages as shown in **a**.

NMPs, tail bud, hatching gland) showed decreasing correlation, where chromatin remained accessible but gene expression levels declined over time, as seen with *cdx1a* (Fig. 2f). In contrast, cell types emerging later (e.g., hematopoietic vasculature, hemangioblasts, muscle, fast muscle) exhibited increasing correlation, exemplified by *slc4a1a*, where chromatin accessibility was followed by transcriptional activation (Fig. 2f).

Altogether, these findings suggest that the dynamic correlation between chromatin accessibility and gene expression is closely linked to the emergence of specific cell types during development. This analysis is available online as an interactive resource for all genes.

Cell-type- and stage-specific gene regulatory networks reveal conserved and dynamic gene regulatory modules over developmental lineages.

Our previous findings highlight the importance of characterizing gene regulatory programs in both a time- and cell-type-specific manner to fully understand the complex regulatory mechanisms driving zebrafish embryonic development. To achieve this, we mapped the relationships between transcription factors (TFs), the *cis*-regulatory elements they access, and the target genes they regulate, forming what is known as a gene regulatory network (GRN). We then examined how these GRNs are rewired across different developmental stages and lineages (Fig. 3a).

To compute time- and cell-type-specific GRNs, we used the CellOracle framework[17], which integrates scATAC-seq and scRNA-seq data (see Methods). CellOracle predicts interactions between transcription factors (TFs), *cis*-regulatory elements, and their target genes, constructing GRNs that account for dynamic changes in regulatory elements and gene expression across developmental stages (Fig. S6). This approach allowed us to model gene expression as a function of TF activity, capturing both developmental stage- and cell-type-specific regulatory programs (Fig. 3a).

To assess the temporal dynamics of GRNs, we computed whole-embryo GRNs for each developmental stage. Notably, scATAC-seq datasets from each time point were used to compute the *cis*-regulatory elements required for the GRN inference for each time point, reflecting the temporal dynamics of *cis*-regulatory elements (Fig. S5). These GRNs were visualized as heatmaps, with transcription factors (TFs) as rows, target genes as columns, and interaction strengths as elements. This allowed us to identify key regulatory modules across time points (Fig. 3b).

To identify the most important transcription factors (TFs) at each time point, we used *degree centrality*, a graph theory metric that quantifies how central a node is within a network. The top five central TFs in each GRN are shown in Fig. 3b, with the list of central TFs evolving gradually over the developmental stages. *hmga1a*, a chromatin modifier[34], remains the most central TF at early time points (10, 12, and 14 hpf). Several TFs from the Sox family, which are key developmental regulators, also appear prominently[35]. For instance, *sox5*, present at 14 and 16 hpf, plays an important role in spinal cord and neural crest development[36, 37].

Notably, the interaction strengths (GRN edges) appear to weaken significantly by 24 hpf compared to earlier stages. This may result from the increasing heterogeneity at the whole-embryo level, where the GRNs of diverse cell types are averaged out at later stages.

To investigate cell-type specificity in GRNs, we focused on one cell type, the spinal cord, a critical structure in vertebrates responsible for coordinating movement and sensory integration[38]. At 24 hpf, more detailed GRN patterns were observed, supporting the hypothesis that cell-type-specific GRNs were obscured when computed at the whole-embryo level (Fig. 3c).

Our analysis uncovered dynamic regulatory modules that change over time, reflecting the shifting priorities of the developing embryo (Fig. 3d-left). For example, a module active in earlier stages (10 hpf), driven by the chromatin modifier *hmga1a*, was responsible for activating genes encoding ribosomal subunits. This module was prominent in whole-embryo GRNs at early time points (Fig. 3b). In contrast, later-stage modules were enriched with transcription factors specific to neuronal development, as identified through gene ontology curation from ZFIN[38] (Fig. 3d-right).

The early activation of modules related to ribosomal biogenesis and translation suggests that the embryo prioritizes building cellular machinery to support cell cycle progression and protein production, which are crucial for cell fate specification. The involvement of *hmga1a* in regulating these early processes highlights the role of chromatin modifiers in orchestrating early developmental events[34] (Fig. 3b and d).

To further explore cell-type-specific gene regulatory modules, we examined GRN dynamics across the differentiation of

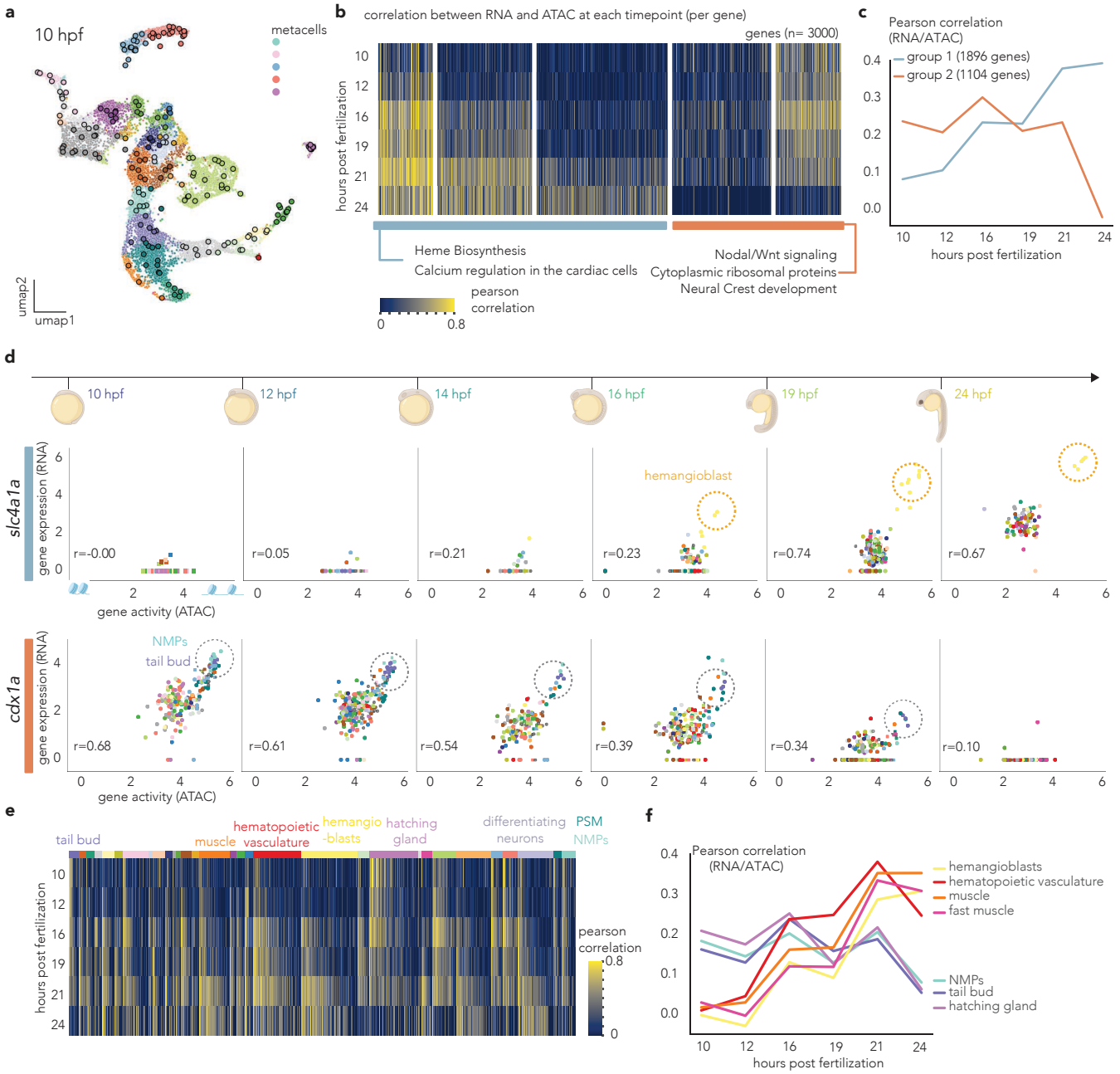

Figure 2: **Dynamics of correlation between chromatin accessibility and gene expression resolved by cell types during zebrafish embryogenesis. a,** A UMAP representation of joint modalities (RNA+ATAC) at 10 hpf showing single-cells with metacells colored by the cell type annotation in Fig. 1d. **b,** A heatmap where the rows are time points (10 hpf to 24 hpf), the columns are 3000 highly variable genes (over time), and the elements are Pearson correlation coefficients between gene activity (ATAC) and gene expression (RNA) across metacells for each time point. Two distinct clusters were identified based on their temporal dynamics. Cluster 1 (blue) represents the genes whose correlation coefficients increase over time, and cluster 2 (orange) represents the genes whose correlation coefficients decrease over time. **c,** A line plot showing the averaged correlation coefficients per each cluster for each time point. **d,** Scatter plots of gene activity (ATAC) versus gene expression (RNA) for *slc4a1a* and *cdx1a* genes across developmental timepoints (10 to 24 hpf). Each dot represents a metacell, colored by the most prevalent cell type. Pearson correlation coefficients (r) are shown for each timepoint. **e,** Heatmap showing the correlation between gene activity (ATAC) and gene expression (RNA) for all genes across developmental timepoints. Clustering was done based on the cell-types where the gene accessibility and gene expression are the highest. **f,** A line plot showing the temporal dynamics of correlation for the cell types from the left heatmap, highlighting different patterns of chromatin accessibility and gene expression coordination for each cell-type.

two major lineages: neuro-ectoderm and mesoderm (Fig. 3e and f). We focused on Neuro-Mesodermal Progenitors (NMPs), a bipotent group of progenitors conserved in vertebrates, which are essential for axial elongation[39]. These progenitors can differentiate into either neuro-ectoderm (spinal cord and neural posterior) or mesoderm (presomitic mesoderm, somites, and fast muscle)[10]. At the early stage (10 hpf), we found that the GRN architecture was highly similar across these developmental lineages, suggesting that the regulatory network structure is largely shared during the initial phase of development (Fig. 3e).

By 16 hpf, we observed distinct GRN modules activated during the differentiation of each lineage, with neural-specific modules highlighted in red and mesodermal-specific modules in green as shown in Fig. 3f. Further analysis of the genes within these modules (Fig. 3g) revealed that they are crucial for neural and mesodermal development, respectively[39].

Overall, our results show that in early development, GRNs are largely conserved across cell types (Fig. 3e). However, as development progresses, these shared regulatory modules give way to lineage-specific gene regulatory programs (Fig. 3f), illustrating the shift from a common regulatory architecture to specialized modules over time.

In silico genetic perturbation reveals dynamic roles of transcription factors across developmental stages.

Our analysis of the temporal dynamics of GRNs indicates that the role of transcription factors in cell differentiation evolves significantly across developmental stages. For instance, the regulatory module governing neuronal differentiation (Fig. 3g-left) remains inactive during early development (10 hpf) but becomes active at around 14 hpf (see Interactive Web Portal). This temporal shift suggests that transcription factors such as *pax6a* and *foxp4* play increasingly pivotal roles in neuronal lineage specification as development advances. To further elucidate the dynamic functions of these transcription factors in cell fate determination, we performed *in silico* genetic perturbations, systematically assessing their impact on both neuro-ectodermal and mesodermal lineages at different developmental stages.

We employed the CellOracle library to perform *in silico* perturbation simulations based on our time- and cell-type-resolved GRNs computed from single-cell multi-omic data[17]. In these simulations, transcription factors were "knocked out" *in silico* by setting their expression to zero, and the downstream effects on gene expression were estimated. These simulated transcriptomic changes were then used to infer cellular state transitions and calculate cell-cell transition probabilities.

To interpret the resulting transition probabilities, we employed two complementary methods. First, we projected the transition probabilities onto 2D UMAP embeddings, enabling a visual representation of how cellular states shift in response to transcription factor knockouts (Fig. 4a). Second, we quantitatively compared wild-type and knockout conditions using cosine similarity (Fig. 4b), facilitating both qualitative and quantitative evaluations of perturbation effects. By leveraging the temporal resolution of our data, we conducted systematic *in silico* knockouts across multiple developmental stages and quantified perturbation scores for each gene (Fig. 4b). This analysis uncovered a key trend: during early development (10 hpf), many transcription factors exerted significant influence over both mesodermal and neuro-ectodermal lineages, reflecting a high degree of regulatory plasticity (Fig. 4c).

As development progressed, transcription factors became more committed to specific lineages. By 24 hpf, most transcription factors affected either the mesodermal or neuro-ectodermal lineage, but rarely both (Fig. 4c). For example, *pax6a*, which is known to be important for neuronal progenitors during eye development[40, 41], had strong effects on both lineages at 10 hpf but primarily influenced the neuro-ectodermal lineage by 24 hpf.

To illustrate this shift, we visualized the cell-cell transition probabilities for *pax6a* knockouts compared to wild-type conditions at both 10 and 19 hpf (Fig. 4d and e). At 10 hpf, the 2D UMAP projections showed that perturbed transitions affected both mesodermal and neuro-ectodermal trajectories. However, by 19 hpf, the effects were largely confined to the neuro-ectodermal lineage, with minimal impact on mesodermal trajectories (Fig. 4d and e).

These time-resolved *in silico* perturbations provide strong evidence that transcription factors become increasingly lineage-specific as development progresses.

A web resource for data exploration and timecourse _in silico_ genetic perturbation.

To facilitate access to our time-resolved single-cell Multiome datasets and computational analyses, we developed an interactive web portal, Zebrahub-Multiome (Fig. 5), which has been integrated into the existing Zebrahub resource[10] (Fig. 5a). Users can explore processed data via the CZ CELLxGENE browser[42], allowing them to visualize gene activity (ATAC) or gene expression (RNA) across various tissues and time points (Fig. 1a). In addition, gene-level analysis of *cis*-regulatory elements, such as distal and proximal enhancers, is available at each time point across the entire transcriptome (Fig. 5b). Moreover, the portal also includes gene-level correlation analysis between gene activity and expression, as shown in Fig. 2, allowing users to explore their genes of interest across the full transcriptome (Fig. 5c). Furthermore, the gene regulatory networks (GRNs) for the axial cluster, as presented in Fig. 3, can be explored across developmental lineages or time points (Fig. 5d). Zoom-in features enable detailed exploration of specific GRN modules. Finally, the portal also includes *in silico* genetic perturbation tools for all transcription factors captured within the GRNs at each time point (Fig. 4). Users can systematically perform *in silico* genetic perturbations and monitor their effects on different cell types or tissues, comparing them to wild-type conditions for any transcription factor found in the GRNs (Fig. 5e).

This user-friendly resource provides developmental biologists with convenient access to comprehensive datasets, facilitating deeper biological discovery.

Discussion

We introduced Zebrahub-Multiome, a high-quality comprehensive time-resolved single-cell multi-omic atlas of zebrafish embryonic development. We simultaneously profiled chro-

a
GRN reconfiguration over time
10 hpf
12 hpf
14 hpf
16 hpf
19 hpf
24 hpf
time
...
GRN reconfiguration over cell types
celltype 1
celltype 2
celltype 3
b
GRNs at different timepoints for whole embryo
genes
TFs
hmga1a
sox19a
vox
irx7
dlx3b
0.1
0.2
degree centrality
meox1
sox3
sox11a
sox5
sox13
rxraa
meis1b
sox6
fli1b
fli1a
hoxb3a
klf1
gata1a
sox11b
c
GRNs at different timepoints for spinal cord
decreasing over time
increasing over time
d
0.1
0
−0.1
TFs
ved
vox
meox1
sox19a
hmga1a
pax6b
pax6a
sox3
foxp4
sox11a
sox13
rfx4
meis1b
e
GRNs at different cell types for one timepoint (10 hpf & 16 hpf)
neural posterior
spinal cord
NMP
tail bud
PSM
somites
10 hpf
f
16 hpf
neuronal
mesodermal
g
TFs
nr2f5
thrab
foxp1b
pax6b
pax6a
foxp4
meis2a
meis3
sox13
rfx4
sub GRN (neuronal, 16 hpf)
Gene Ontology (pathways)
ribosomal biogenesis/translation
homeobox genes
RA signaling pathway
mesoderm/somitogenesis
nervous system development
sub GRN (mesodermal, 16 hpf)

Figure 3: **Dynamic Gene Regulatory Networks (GRNs) across developmental stages and cell types in zebrafish embryogenesis. a,** A schematic representation of GRN reconfiguration (1) over developmental time and (2) across different cell types. **b,** GRNs for whole embryo across developmental stages (10 to 24 hpf). Heatmaps show edge strengths between transcription factors (TFs) and target genes. Top five transcription factors with the highest degree centrality at each time point are shown with their degree centrality score. **c,** GRNs for spinal cord development across different developmental stages (10 to 24 hpf). Heatmaps show edge strengths between transcription factors (TFs) and target genes, with decreasing and increasing patterns highlighted. **d,** Detailed view of the GRN heatmaps for spinal cord, showing specific TF-gene interactions. Edge strength is represented by color intensity, with red indicating positive and blue indicating negative regulation. **e, f,** GRNs for different cell types at the (e) 10 hpf, and (f) 24 hpf stages, including Neural Posterior, Spinal Cord, Neuromesodermal Progenitors (NMP), Tail Bud, Presomitic Mesoderm (PSM), and Somites. **g,** Sub-GRNs of the 15-somite stage GRN for neuronal (left) and mesodermal (right) lineages. Gene Ontology pathways associated with specific modules are annotated. Notable TFs and their target genes are highlighted, demonstrating lineage-specific regulatory patterns.

matin accessibility and gene expression across key developmental stages. Combined with our previously published datasets[10], our Atlas of zebrafish development,Zebrahub, offers a rich and unique resource integrating transcriptomics, epigenomics, and imaging. Moreover, the reconstructed GRNs, together with the *in silico* framework, enable time- and cell-type-specific understanding of differentiation.

First, we uncovered a complex and often discordant relationship between chromatin accessibility and gene expression. Our analysis revealed diverse temporal patterns in this relationship, including cases where poised chromatin precedes gene activation and where chromatin accessibility persists even after gene expression decreases. These findings not only highlight the complexity of the regulatory patterns but also emphasize the need for integrated multi-omic approaches and further experimental work to fully understand the intricate dynamics between chromatin accessibility and gene expression.

Secondly, the analysis of time-resolved and cell-type-specific gene regulatory networks (GRNs) revealed both conserved and variable regulatory modules across developmental stages and cell lineages. We identified that the early regulatory modules, such as those involving *hmga1a*, that prioritize translational upregulation, are shared across different cell types at early stages. As development advances, we observed a shift towards more lineage-specific regulatory modules, reflecting the increasing specialization of cell types. The interactive visualization of our data confirmed that the cell-type specific GRNs were activated gradually, not abruptly, along both temporal and developmental lineage axes (Fig. 5).

Finally, our *in silico* genetic perturbation analysis provided insights into the evolving roles of transcription factors during development. We observed a general trend where transcription factors transition from having broad influences on multiple lineages in early stages to more specialized, lineage-specific roles at later stages. This computational approach allows for the generation of fine-grained hypotheses about transcription factor function across developmental time, complementing traditional experimental methods.

However, it is important to acknowledge the limitations of our study. The *in silico* genetic perturbation approach, while powerful for generating hypotheses, relies on computational models and assumptions that may not fully capture the complexity of biological systems. The experimental validation of key predictions will be crucial to confirming the biological relevance of our findings. Additionally, while our time-resolved approach provides valuable insights, it still represents discrete time points in a continuous developmental process. Future studies with even higher temporal resolution could further refine our understanding of regulatory dynamics.

The Zebrahub-Multiome web portal (Fig. 5), which was built as an extension of Zebrahub[10], serves as a valuable resource for the scientific community. By making our data and analysis tools readily accessible, we aim to facilitate new discoveries and drive further research in developmental biology and related fields[10,43–45]. Our study complements existing atlases, particularly the recent work by Liu et al.[44], which characterized a single-cell multi-omic atlas of early zebrafish embryos from 3.3 to 12 hpf. Our data spans from 10 to 24 hpf, extending the temporal coverage of zebrafish embryonic development. We envision that integrating our atlas with other existing and expanding multi-modal datasets will contribute to a more comprehensive understanding of zebrafish embryonic development, benefiting the broader zebrafish research community. In the broader context of developmental biology, our study contributes to the growing body of work emphasizing the dynamic and context-dependent nature of gene regulation. By providing a comprehensive resource for zebrafish development, we enable comparative studies with other model organisms, potentially revealing conserved and divergent regulatory mechanisms across vertebrates.

In conclusion, our study not only advances our understanding of TF regulation and GRNs during embryogenesis but also establishes a framework for investigating the dynamics of gene regulation during development in other systems. By demonstrating the power of integrating multi-omic data with advanced computational analyses, we have paved the way for future studies that will further unravel the complexities of vertebrate development. As these approaches are refined and applied to diverse biological systems, we move closer to a comprehensive understanding of the regulatory principles governing embryonic development and cellular differentiation.

Looking ahead, we envision that the framework we've demonstrated, combining time-resolved GRNs with computational models, will enable more comprehensive *in silico* genetic perturbation studies at the whole-embryo scale. This approach builds upon and extends previous genetic perturbation studies in zebrafish embryos[9], offering the potential to generate quantitative, testable predictions of phenotypic outcomes resulting

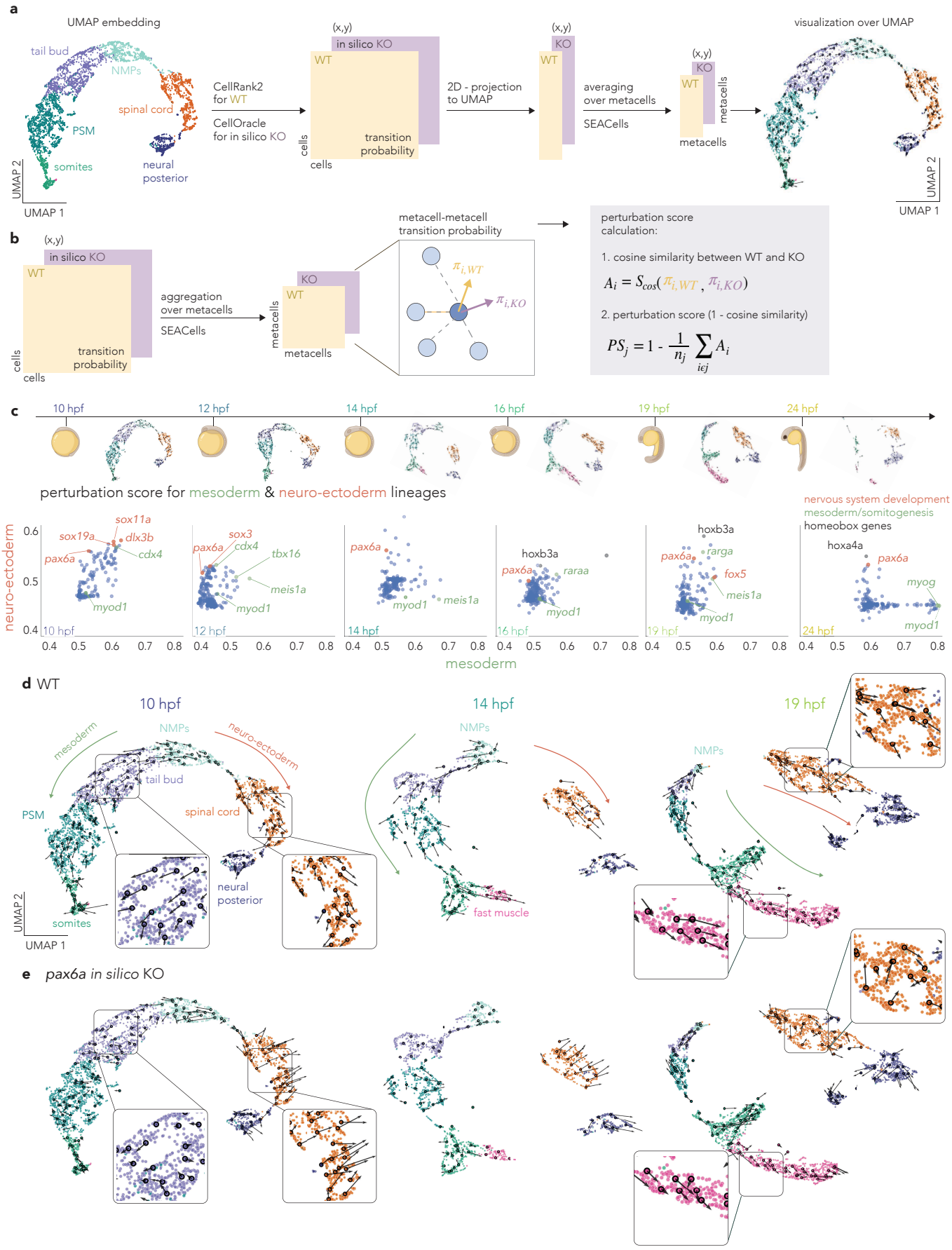

Figure 4: **Cell-cell transition probability analysis and *in silico* perturbation using CellOracle**. See the caption on the next page.

Figure 4: **Cell-cell transition probability analysis and *in silico* perturbation using CellOracle a,** Workflow for computing and visualizing cell-cell transition probabilities. Starting from 2D UMAP embedding, transition probabilities are calculated for wild-type (WT) and knock-out (KO) conditions, projected onto UMAP, and averaged over metacells for visualization. **b,** Schematic of metacell-metacell transition probability calculation and perturbation score computation using cosine similarity between WT and KO transition probabilities. **c,** Perturbation scores for mesoderm/neuro-ectoderm lineages across developmental time points (10 to 24 hpf). Each dot represents a transcription factor, with higher scores indicating stronger perturbation effects. **d,** UMAP visualizations of WT zebrafish embryo development from 10 to 24 hpf, with major cell types and lineages annotated. Arrows indicate predicted cell state transitions. **e,** UMAP visualizations of *in silico pax6a* knockout at 10, 14, and 19 hpf, showing altered cell state transitions compared to WT.

from genetic perturbations.

Acknowledgements

We thank the entire Computational Biolog & Genomics Platforms, and the Royer group of the Chan Zuckerberg Biohub-SF for helpful discussion and support. We thank the Scientific Computing team for their support for streamlining the data processing workflows. We appreciate the brainstorming and feedback from Luca Pinello, Max Lombardo, Sara Simmonds, Alec Tarashansky, Pablo Garcia-Nieto, Guillaume Le Treut, and Andy Zhou. We thank Sandra Schmid for the critical reading and feedback on the manuscript. Funding for this work was provided by Chan Zuckerberg Biohub San Francisco (CZB SF). We thank the CZB SF donors, Priscilla Chan and Mark Zuckerberg, for their generous support.

Competing interests

The authors declare no competing interests.

Data availability

Raw sequencing data are available at NCBI's SRA BioProject (PRJNA1164307). The processed h5ad objects as well as Oracle objects (for GRNs) are available in Google Drive. The dataset can be interactively explored at zebrahub.org.

Code availability

The code/notebooks used for data processing and analysis are in github repository: `https://github.com/czbiohub-sf/zebrahub-multiome-analysis`

Methods

Zebrafish husbandry. The zebrafish care and experimental protocols adhered to guidelines approved by the institutional animal care and use committee (IACUC) at the University of California San Francisco (UCSF). Wild-type EKW embryos were utilized for the multiome experiments. Adults zebrafish were bred and maintained at a temperature of 28.5°C, receiving food three times daily through an automatic feeder. Embryos were raised and staged according to the somite count along with the hours post fertilization (hpf) serving as temporal phenotypic markers during somitogenesis.

Single-cell multiome sample preparation. For multiome experiments, 60 embryos were pooled for each developmental stage. We have one biological replicate for 15 somites (16 hpf) as an internal control. First, the embryos were dechorionated using pronase solution (1mg/ml). The embryos were incubated in pronase solution for a maximum of 2 minutes. This is followed by a series of washes using egg water till the chorions are removed and any remanent pronase solution is washed off. Next, 60 embryos of the required stage were dissociated and nuclei were isolated using the 10x nuclei isolation kit (PN-1000494). The nuclei were resuspended in 25 ul of resuspension buffer. The nuclei count was determined using a dilution (1:4) of the cell suspension and DAPI solution (1:1000 dilution). The cells are incubated in DAPI for 15 minutes at room temperature in the dark. This is followed by counting nuclei manually using a hemocytometer.

Single-cell multiome library preparation and sequencing. The single-nuclei suspensions were transposased such that the transposase enters the nuclei and preferentially fragments the DNA in open regions of the chromatin. This is then processed into droplet emulsions using the Chromium Single Cell Controller (10x Genomics), and both scRNA-seq libraries as well as ATAC libraries are generated for each cell using the Chromium Next GEM Single Cell Multiome ATAC + Gene expression reagents. The single-nuclei suspensions were assessed using the hemocytometer to determine the nuclei count, we aimed for a maximum recovery of 10,000 nuclei wherever possible. Fragmentation, ligation, and PCR amplification were performed using the BioRad C1000 Touch Thermal Cycler with thermal profiles specified by the 10X Genomics User Guide CG000338. Amplified cDNA and final ATAC, gene expression libraries were assessed using the Agilent 4200 TapeStation/High Sensitivity D5000 Screentape and qPCR with the Kapa Library Quantification kit for Illumina. Each library was normalized to 4 nM and pooled for each NovaSeq sequencing run. The pools were sequenced with Novaseq run kits, using either single or dual index plates (Illumina 20028313), with a 1% PhiX control library spiked in. Sequencing was performed on the NovaSeq 6000 Sequencing System (Illumina) aiming for 1 billion reads per library.

Single-cell Multiome data processing. Sequencing reads were aligned to the Ensembl reference genome (GRCz11) using Cellranger-arc (version 2.0.2, 10x Genomics), generating count matrices for both RNA (Gene Expression) and ATAC modalities. Subsequent processing of these raw count matrices was performed using the Seurat (version 4.4.0) and Signac (version 1.11.0) packages[22,46]. For each developmental stage, we created Seurat objects incorporating both "RNA" and "ATAC" assays. Low-quality nuclei were filtered out based on the following quality control (QC) metrics:

- RNA modality: Cells were retained if they contained be-

Figure 5: **An overview of interactive web portal for single-cell multi-omics atlas. a,** Zebrahub-Multiome main tabs: Visualizing the distal *cis*-regulatory elements over time per gene, Correlation between chromatin accessibility and gene expression, Visualization of GRNs, and *in silico* genetic perturbation. **b,** Example visualization of the distal *cis*-regulatory elements over time for *myf5* gene. Features were highlighted with arrows and annotated texts. **c,** Example scatter plots for chromatin accessibility and gene expression at metacell level resolved by developmental stages. **d,** GRN visualization with highlighted modules of activation and repression. **e,** Left: UMAP visualizations of 2D-projected metacell transition probabilities for WT (black arrows) and *in silico* (red arrows) *meox1* knockout at 16 hpf, Right: Perturbation scores for mesoderm/neuro-ectoderm lineages across developmental timepoints (10 to 24 hpf). Each dot represents a transcription factor, with higher scores indicating stronger perturbation effects.

tween 1,000 and 25,000 UMIs.

- ATAC modality: Cells were retained if they contained between 1,000 and 100,000 fragments, exhibited a nucleosome signal below 2, and demonstrated a TSS enrichment signal above 1.

The RNA modality was processed first by normalizing the raw transcript counts using counts per million (CPM), followed by scaling and principal component analysis (PCA). This process resulted in the "SCT" assay, generated using the *scTransform* function from the Seurat package.

Cell type annotation. Cell type annotations were predicted using Zebrahub[10] as a reference dataset. The *FindAnchors* function in Seurat was employed to transfer labels from the reference to our dataset, resulting in predicted cell type annotations at the tissue level, as illustrated in Figure 1 of Lange et al.[10].

Peak calling in scATAC-seq datasets. In scATAC-seq analysis, the features of interest are "peaks," which represent discrete regions of open chromatin in the genome, in contrast to the "genes" analyzed in scRNA-seq. Peak calling in scATAC-seq data presents unique challenges, particularly in determining how to group cells with similar chromatin accessibility profiles to achieve robust peak identification[22]. We employed MACS2, as implemented in Signac, to call peaks using two different grouping strategies: (1) "cell-type" grouping, based on annotations transferred from Zebrahub, and (2) "bulk" grouping, considering all cells collectively. This approach, combined with the peaks called by Cellranger-arc, resulted in three distinct peak profiles (Figure S8a). To integrate these profiles, we adopted the "iterative peak calling" protocol from ArchR, which prioritizes peaks with higher confidence. To create a comprehensive atlas of cell-type-specific regulatory elements, we implemented the following strategy: We first established cell type-specific peaks as the baseline. We then incorporated non-overlapping peaks from the bulk analysis. Finally, we iteratively added Cellranger-arc peaks that were absent from the previous two profiles (Fig. S8b).

Iterative peak calling strategy. Our analysis revealed that MACS2-called peaks were generally shorter than those identified by Cellranger-arc, suggesting superior detection sensitivity for MACS2 (Fig. S8c)[22]. The number of peaks added at each iteration (Fig. S8d) indicated substantial overlap between cell type-specific and bulk peaks called by MACS2 with those detected by Cellranger-arc. This iterative peak-calling approach offers two key advantages: it captures peaks specific to rare cell types that might be missed in bulk-level analyses, and it identifies specific regulatory regions that enhance transcription factor binding motif identification in subsequent analyses.

Gene activity score. Gene activity scores were computed using Signac, which sums up the scATAC-seq reads within the gene body and 2kb upstream of the transcription start site for each gene[22]. The resulting gene activity scores were log-normalized to account for the sequencing depth for downstream analyses.

Tissue-specific marker gene visualization. To visualize and validate the cell type annotations in our single-cell atlas, we selected sets of well-established marker genes for four major tissue types: endoderm, mesoderm, neuroectoderm, and hematopoietic lineages. The expression patterns of these markers were overlaid on the UMAP embeddings to confirm the identity of annotated cell clusters and to illustrate the spatial organization of different cell types in the low-dimensional space. For endodermal tissues, we used the following marker genes: *fgfrl1b*, *col2a1a*, *ptprfa*, *emid1*, *nr5a2*, *ism2a*, *pawr*, *mmp15b*, *foxa3*, and *onecut1*. Mesodermal markers included *msgn1*, *meox1*, *tbx6*, *tbxta*, *fgf8a*, and *her1*. To identify neuroectodermal tissues, we employed *pax7a*, *pax6a*, *pax6b*, *col18a1a*, *en2b*, *znf536*, *gpm6aa*, *gli3*, and *chl1a* as marker genes. Hematopoietic lineages were identified using *lmo2*, *etv2*, *tal1*, and *sox17*. For each tissue type, we calculated the average log-normalized expression of its respective marker genes for each cell. These average expression values were then projected onto the UMAP embedding, with color intensity representing the level of marker gene expression. This approach allowed us to visualize the spatial distribution of different tissue types within the UMAP space and validate our cell type annotations.

Metacell inference. To account for the noise and sparsity of single-cell Multiome datasets for the correlative analysis of chromatin accessibility and gene expression, we used SEACells[29] to compute the metacells. Specifically, for each dataset, we used ATAC modality to compute the latent semantic indexing (LSI) as the dimensionality reduction. We used the 2nd to 40th LSI components to compute the metacells, and we excluded the first LSI component as the it is usually correlated with the sequencing depth. After the metacell computation, we aggregated the gene expression (RNA) and gene activity (ATAC) values at the metacell level, creating a set of aggregated count matrices used for Fig. 2. The number of cells per metacell was set to 75 following the SEACells tutorial[29]. Note that we used 30 cells/metacell for the Fig. 4 analysis for a fine-grained visualization of the cell-cell transition probability in 2D embeddings.

Heatmap clustering. To visualize the heatmap of correlation coefficients between gene activity (ATAC) and gene expression (RNA) per gene per time point, we used seaborn's clustermap function with "ward" algorithm and "euclidian distance" metric[47]. From visual inspection, we identified two large clusters as identified in Figure. 2 b as increasing or decreasing over time.

Gene set enrichment analysis. To identify the biological pathways enriched in a group of genes identified from clustering, we used FishEnrichr, a web-based gene set enrichment analysis tool, specialized for zebrafish[30,31]. We used the list of genes identified from the clustering and referred to the gene ontology terms of enriched pathways from WikiPathways database[48].

Subsetting for the Neuro-Mesodermal Progenitor population. To investigate key transcription factors in Neuro-Mesodermal Progenitors (NMPs), we created a subset of our data focusing on NMPs and their subsequent developmental trajectories. This subset included cells from mesodermal (tail bud, presomitic mesoderm (PSM), somites, and fast muscle) and neuroectodermal (spinal cord, neural posterior) lineages.

Gene Regulatory Networks inference. We used CellOracle to infer cell-type specific Gene Regulatory Networks (GRNs) for the NMP-subsetted objects at each time point[17]. The process followed the CellOracle protocol as illustrated in Fig. S6. Briefly, CellOracle computes the distal and proximal *cis*-regulatory regions for each gene using cicero[49]. Then, transcription factor binding motifs were scanned within these *cis*-regulatory regions, constructing a base GRN, where the rows are all genes, columns are all transcription factors, and binarized elements for the presence or absence of TF binding motif within the *cis*-regulatory element for that gene. Importantly, these base GRNs were computed for each time point, noting that the *cis*-regulatory elements change dynamically over developmental stages (Fig. S5). Using the base GRN, CellOracle constructed linear equations modeling each gene's expression level as a combination of regulatory transcription factors from the base GRN. Coefficients for these models were determined through regularized linear regression, using gene expression values to approximate transcription factor activity. The resulting cell-type specific GRNs were further refined by filtering out weak edges based on edge strength and/or regression p-values (Fig. S6).

Chromatin co-accessibility computation. We used Cicero[49] to calculate chromatin co-accessibilities for all peak-peak pairs within each chromosome. The resulting dataframe contained peak pairs (*i* and *j*) and their co-accessibility scores. We applied CellOracle's filtering criteria:

a) Removed peak-peak pairs with co-accessibility scores below 0.8.
b) Retained pairs where at least one peak corresponded to a transcription start site (TSS, as annotated by Ensembl).

Aligned UMAPs of individual time points. To visualize the temporal progression of cellular trajectories in NMP trajectories across time points, we used AlignUMAP to align the UMAPs from individual time points using pairwise anchors between time points. Briefly, the pairwise anchors were computed by nearest neighbors using the dimensionality reduction from the ATAC modality (Latent Semantic Indexing, LSI, from 2nd to 40th components). To avoid over-alignment, we relaxed the strength of the alignment regularization parameter to 0.001. The output is a series of UMAPs for each time point, which shows conserved orientation and configurations for easier visual inspection of the WT and *in silico* KO transitions.

***in silico* genetic perturbation.** We performed *in silico* genetic perturbation following CellOracle's documentation. Briefly, CellOracle uses cell-type-specific gene regulatory networks and developmental trajectories to predict the shift in cellular differentiation vectors. First, it simulates the transcriptomic change upon knockout of the gene of interest by using signal propagation. Then, the changed transcriptome is compared with the transcriptome of cells in the local neighborhood to estimate the cell-cell transition probabilities.

cell-cell transition probability inference for wild-type embryos. Notably, we used pySlingshot, a python implementation of Slingshot[50]) to compute the pseudotime with "NMPs" defined as the root cells[10]. Next, we used CellRank2[51]'s "PseudotimeKernel" to compute the cell-cell transition probability matrix and 2D projections onto the embeddings (UMAP) for the wild-type (WT) trajectory inference.

Bibliography

1. Levine, M. & Davidson, E. H. Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4936–4942 (2005).
2. Davidson, E. H. & Levine, M. S. Properties of developmental gene regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20063–20066 (2008).
3. Griffiths, J. A., Scialdone, A. & Marioni, J. C. Using single-cell genomics to understand developmental processes and cell fate decisions. *Mol. Syst. Biol.* **14**, e8046 (2018).
4. Farrell, J. A. *et al.* Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).
5. Cao, J., Zhou, W., Steemers, F., Trapnell, C. & Shendure, J. Characterizing the temporal dynamics of gene expression in single cells with sci-fate (2019).
6. Packer, J. S. *et al.* A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution. *Science* **365**, eaax1971 (2019).
7. Li, H. *et al.* Fly cell atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022).
8. Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
9. Saunders, L. M. *et al.* Deep molecular, cellular and temporal phenotyping of developmental perturbations at whole organism scale. *bioRxiv* 2022.08.04.502764 (2022).
10. Lange, M. *et al.* Zebrahub – multimodal zebrafish developmental atlas reveals the state transition dynamics of late vertebrate pluripotent axial progenitors. *bioRxiv* 2023.03.06.531398 (2023).
11. Tabula Muris Consortium *et al.* Single-cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature* **562**, 367–372 (2018).
12. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).

13. Argelaguet, R. *et al*. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019).

14. Argelaguet, R. *et al*. Decoding gene regulation in the mouse embryo using single-cell multi-omics (2022).

15. Jimenez, E. *et al*. A regulatory network of sox and six transcription factors initiate a cell fate transformation during hearing regeneration in adult zebrafish. *Cell Genom*. **2**, 100170 (2022).

16. Badia-i Mompel, P. *et al*. Gene regulatory network inference in the era of single-cell multi-omics. *Nat. Rev. Genet*. **24**, 739–754 (2023).

17. Kamimoto, K. *et al*. Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751 (2023).

18. Bravo González-Blas, C. *et al*. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).

19. Wang, L. *et al*. Dictys: dynamic gene regulatory network dissects developmental continuum with single-cell multiomics. *Nat. Methods* **20**, 1368–1378 (2023).

20. Kartha, V. K. *et al*. Functional inference of gene regulation using single-cell multi-omics. *Cell Genom*. **2**, 100166 (2022).

21. Ramirez, R. N. *et al*. Dynamic gene regulatory networks of human myeloid differentiation. *Cell Syst*. **4**, 416–429.e3 (2017).

22. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with signac. *Nat. Methods* **18**, 1333–1341 (2021).

23. Relaix, F., Rocancourt, D., Mansouri, A. & Buckingham, M. A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature* **435**, 948–953 (2005).

24. Kassar-Duchossoy, L. *et al*. Pax3/Pax7 mark a novel population of primitive myogenic cells during development. *Genes Dev*. **19**, 1426–1431 (2005).

25. Berkes, C. A. & Tapscott, S. J. MyoD and the transcriptional control of myogenesis. *Semin. Cell Dev. Biol*. **16**, 585–595 (2005).

26. Lin, C.-Y., Lee, H.-C., Chen, H.-C., Hsieh, C.-C. & Tsai, H.-J. Normal function of Myf5 during gastrulation is required for pharyngeal arch cartilage development in zebrafish embryos. *Zebrafish* **10**, 486–499 (2013).

27. Ma, S. *et al*. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116.e20 (2020).

28. Sprague, J., Doerry, E., Douglas, S. & Westerfield, M. The zebrafish information network (ZFIN): a resource for genetic, genomic and developmental research. *Nucleic Acids Res*. **29**, 87–90 (2001).

29. Persad, S. *et al*. SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data. *Nat. Biotechnol*. **41**, 1746–1757 (2023).

30. Chen, E. Y. *et al*. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

31. Kuleshov, M. V. *et al*. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. **44**, W90–7 (2016).

32. Bradford, Y. M. *et al*. Zebrafish information network, the knowledgebase for danio rerio research. *Genetics* **220** (2022).

33. Davidson, A. J. & Zon, L. I. The 'definitive' (and 'primitive') guide to zebrafish hematopoiesis. *Oncogene* **23**, 7233–7246 (2004).

34. Olan, I. *et al*. HMGA1 orchestrates chromatin compartmentalization and sequesters genes into 3D networks coordinating senescence heterogeneity. *Nat. Commun*. **15**, 6891 (2024).

35. Hu, Y., Wang, B. & Du, H. A review on *sox* genes in fish. *Rev. Aquac*. **13**, 1986–2003 (2021).

36. Baroti, T. *et al*. Transcription factors Sox5 and Sox6 exert direct and indirect influences on oligodendroglial migration in spinal cord and forebrain. *Glia* **64**, 122–138 (2016).

37. Scott, K., O'Rourke, R., Winkler, C. C., Kearns, C. A. & Appel, B. Temporal single-cell transcriptomes of zebrafish spinal cord pMN progenitors reveal distinct neuronal and glial progenitor populations. *Dev. Biol*. **479**, 37–50 (2021).

38. England, S. J. *et al*. Transcriptional regulators with broad expression in the zebrafish spinal cord. *Dev. Dyn*. (2024).

39. Wymeersch, F. J., Wilson, V. & Tsakiridis, A. Understanding axial progenitor biology in vivo and in vitro. *Development* **148**, dev180612 (2021).

40. Thummel, R. *et al*. Pax6a and Pax6b are required at different points in neuronal progenitor cell proliferation during zebrafish photoreceptor regeneration. *Exp. Eye Res*. **90**, 572–582 (2010).

41. Takamiya, M. *et al*. Pax6 organizes the anterior eye segment by guiding two distinct neural crest waves. *PLoS Genet*. **16**, e1008774 (2020).

42. Megill, C. *et al*. cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices. *bioRxiv* 2021.04.05.438318 (2021).

43. Sur, A., Wang, Y., Capar, P., Margolin, G. & Farrell, J. A. Single-cell analysis of shared signatures and transcriptional diversity during zebrafish development. *bioRxiv* (2023).

44. Liu, J. *et al*. Dissecting the regulatory logic of specification and differentiation during vertebrate embryogenesis. *bioRxivorg* (2024).

45. Wan, Y. *et al*. Whole-embryo spatial transcriptomics at subcellular resolution from gastrulation to organogenesis. *bioRxiv* (2024).

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

47. Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).

48. Slenter, D. N. *et al*. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* **46**, D661–D667 (2018).

49. Pliner, H. A. *et al*. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871.e8 (2018).

50. Street, K. *et al*. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19** (2018).

51. Weiler, P., Lange, M., Klein, M., Pe'er, D. & Theis, F. CellRank 2: unified fate mapping in multiview single-cell data. *Nat. Methods* **21**, 1196–1205 (2024).

Supplementary Information

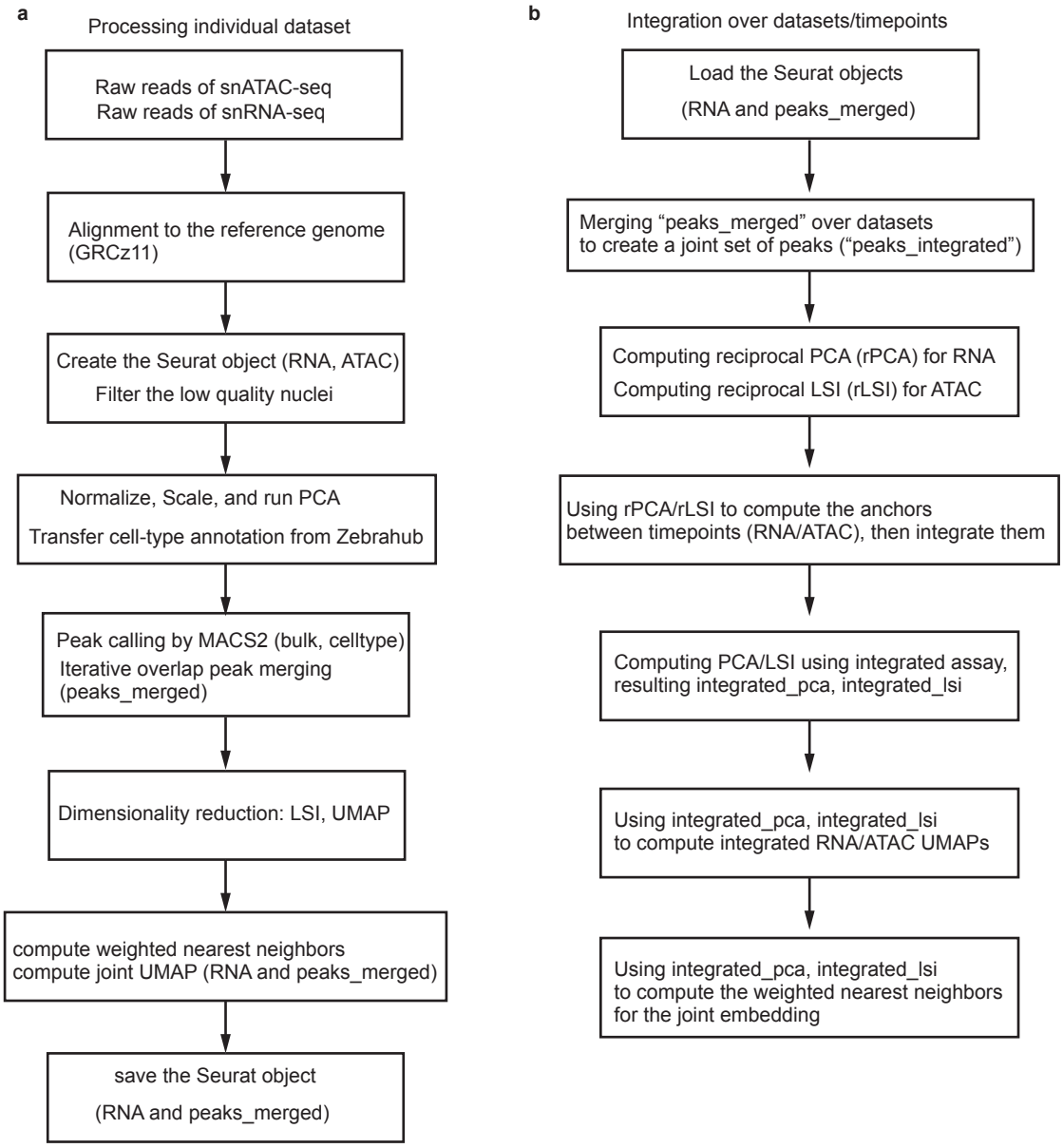

Figure S1: **Workflow for processing and integrating snATAC-seq and snRNA-seq data Workflow for processing and integrating snATAC-seq and snRNA-seq data. a,** Processing individual datasets: The workflow begins with raw reads from snATAC-seq and snRNA-seq, which are aligned to the zebrafish genome (DanRer11). Seurat objects are then created for both RNA and ATAC data, with low-quality nuclei filtered out. The data undergoes normalization, scaling, and PCA, followed by cell-type annotation transfer from Zebrahub. Peak calling is performed using MACS2, with subsequent iterative overlap peak merging. Dimensionality reduction techniques, including LSI and UMAP, are applied. The process concludes with the computation of weighted nearest neighbors and a joint UMAP for RNA and merged peaks, culminating in a final Seurat object containing both RNA and merged peaks data. **b,** Integration over datasets/timepoints: This phase begins by loading the previously saved Seurat objects. A joint set of peaks ("peaks_integrated") is created by merging across datasets. The workflow then employs reciprocal PCA (rPCA) for RNA and reciprocal LSI (rLSI) for ATAC data. These are used to compute anchors between timepoints and integrate the data. Following integration, PCA/LSI is computed using the integrated assay. The process culminates in generating integrated RNA/ATAC UMAPs using the integrated PCA and LSI, and the computation of weighted nearest neighbors for the joint embedding.

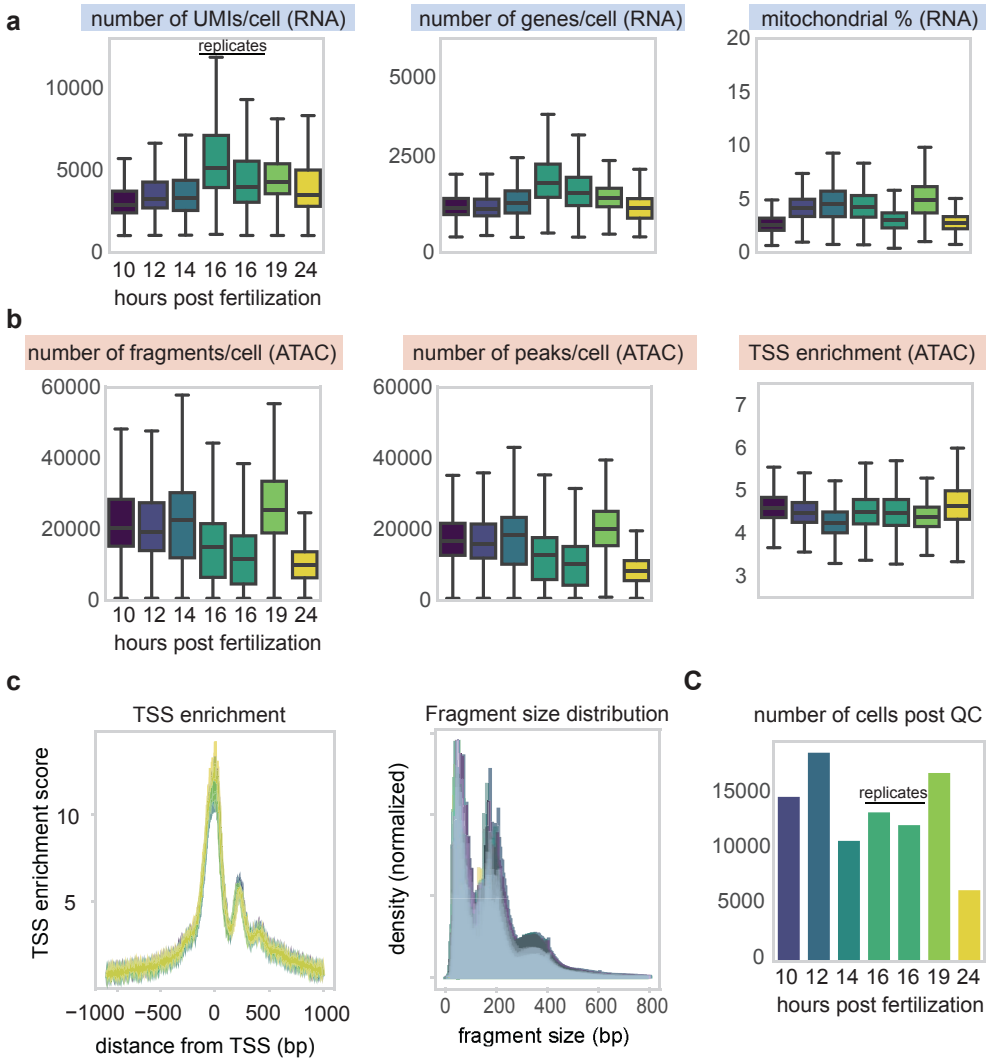

Figure S2: **Quality control metrics for scRNA-seq and scATAC-seq datasets across zebrafish developmental timepoints. a,** scRNA-seq metrics: Box plots show the distribution of (left) number of UMIs per cell, (middle) number of genes detected per cell, and (right) percentage of mitochondrial reads across developmental timepoints from 10 to 24 hours post fertilization (hpf). These metrics indicate the quality and depth of RNA sequencing data. **b,** scATAC-seq metrics: Box plots display (left) the number of fragments per cell, (middle) the number of peaks detected per cell, and (right) the transcription start site (TSS) enrichment score across the same developmental timepoints. These metrics reflect the quality and information content of ATAC sequencing data. **c,** Additional ATAC-seq quality metrics: (Left) TSS enrichment profile showing the aggregated ATAC-seq signal around transcription start sites. (Middle) Fragment size distribution of ATAC-seq reads, typically showing nucleosome-free regions and mono-nucleosome peaks. (Right) Bar plot of the number of cells passing quality control for each developmental timepoint.

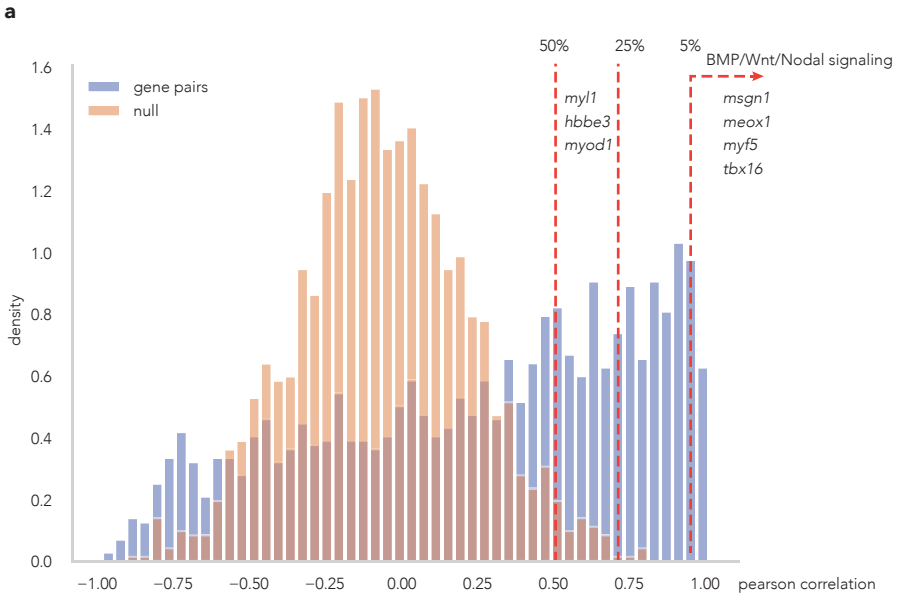

Figure S3: Distribution of correlation between chromatin accessibility and gene expression at the whole-embryo level across developmental stages. **a,** A histogram showing the distribution of Pearson correlation coefficients between chromatin accessibility and gene expression at the whole-embryo level over developmental stages. Two thousand highly variable genes were selected using RNA expression level. The top 5, 25, and 50 percent cutoffs were shown with dotted red vertical lines. The enriched pathways for the top 5% genes are shown above the red curves along with some example genes. Some genes between top 25-50% are shown as an example.

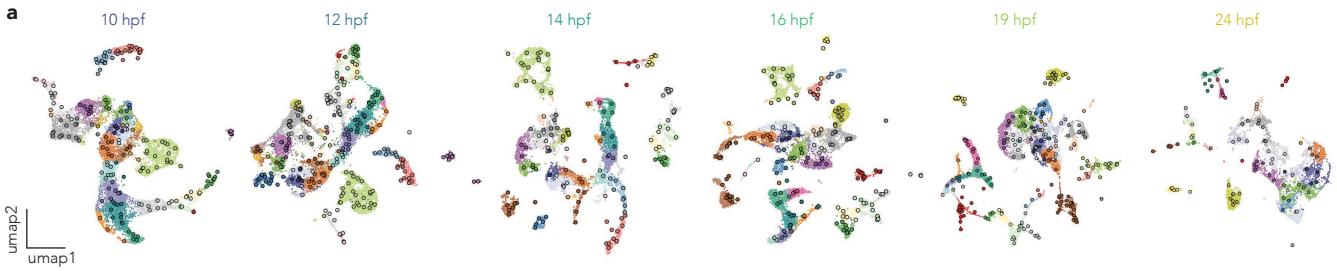

Figure S4: Metacells computed from each time point. **a,** A series of UMAPs overlaid with metacells for each time point over six developmental stages.

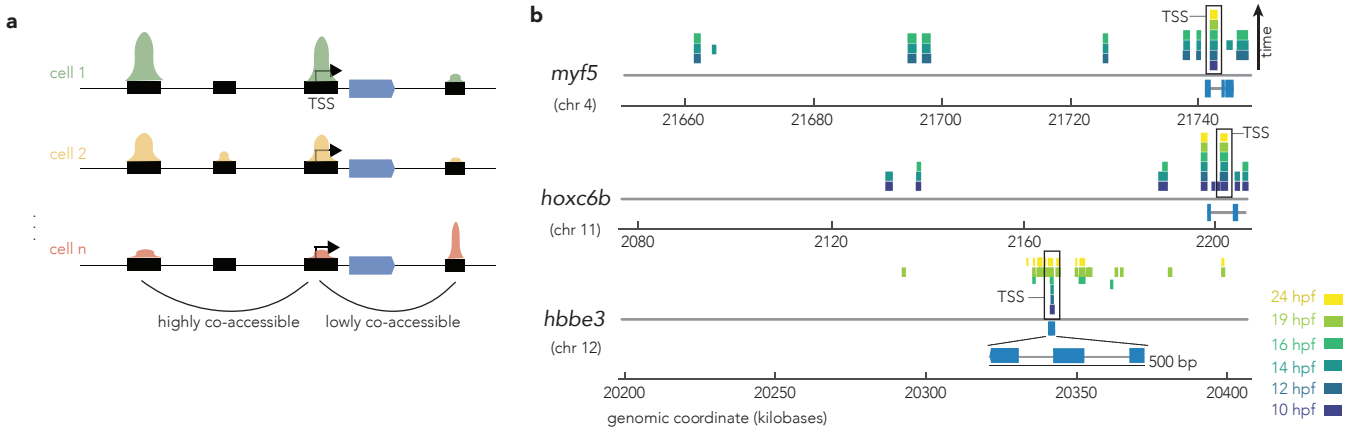

Figure S5: Temporal dynamics of chromatin co-accessibility. **a,** A schematic illustrating chromatin co-accessibility. **b,** Highly co-accessible peaks with transcription start site for *myf5*, *hoxc6b*, and *hbbe3* for each developmental stage along the genomic tracks.

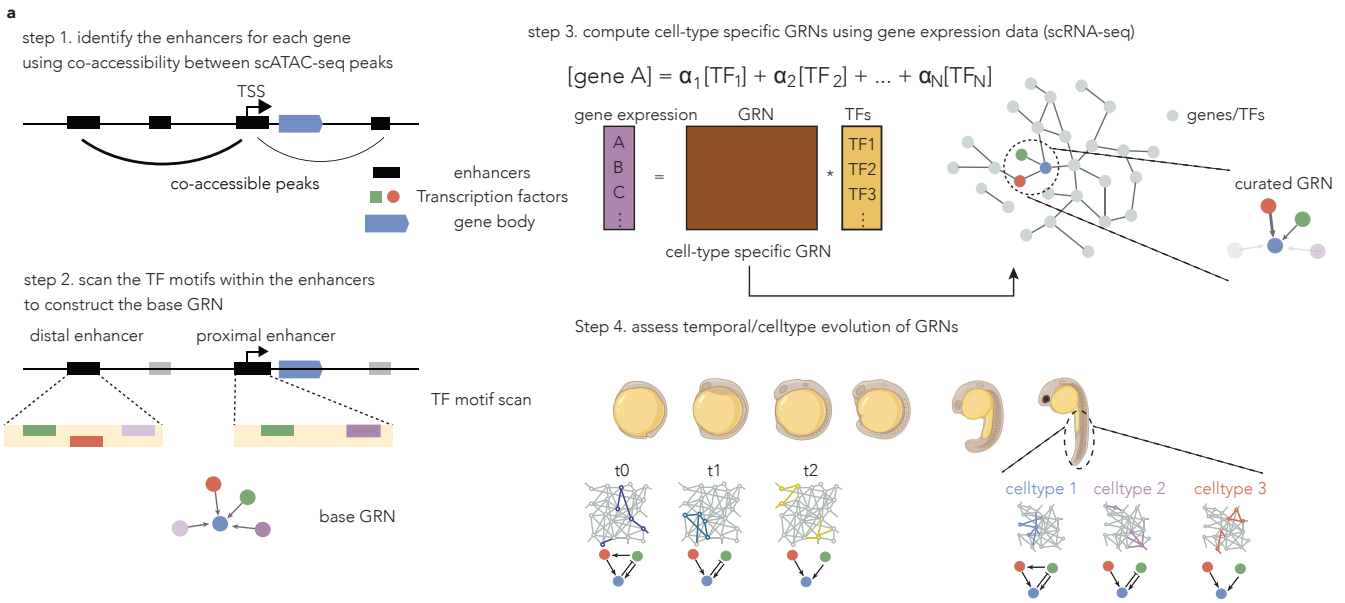

Figure S6: CellOracle workflow schematics. **a,** A schematic of CellOracle workflow to compute the gene regulatory network.

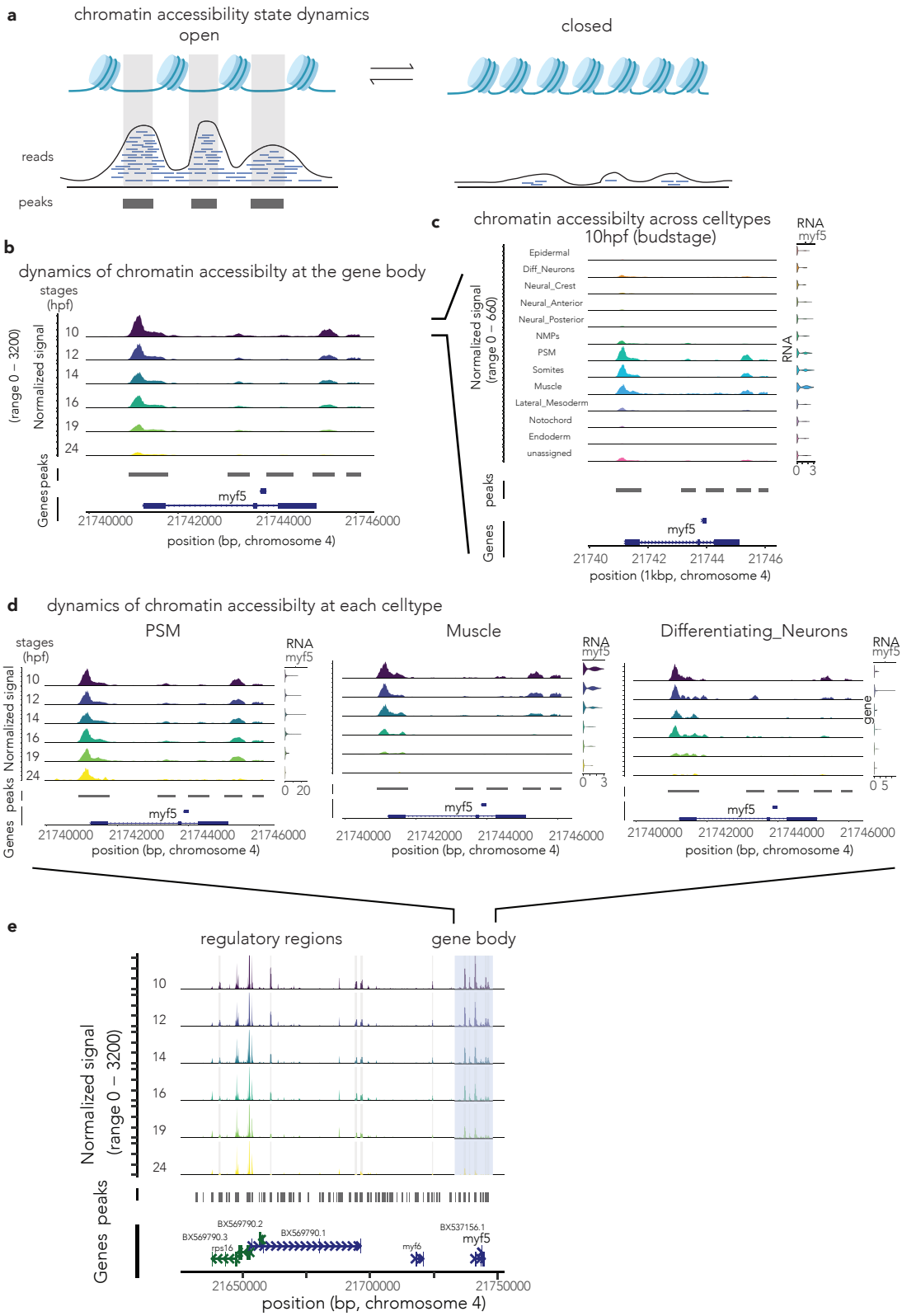

Figure S7: **Dynamics of chromatin accessibility during zebrafish development a** Schematic representation of chromatin accessibility states and their corresponding ATAC-seq signal profiles. **b** Chromatin accessibility dynamics at the *myf5* gene locus across developmental stages (10 to 24 hpf). **c** Cell-type-specific chromatin accessibility patterns at the *myf5* locus at 10 hpf (budstage). **d** Temporal dynamics of chromatin accessibility at the *myf5* locus in specific cell types: presomitic mesoderm (PSM), muscle, and differentiating neurons. **e** Detailed view of chromatin accessibility changes at the *myf5* locus and surrounding regulatory regions across developmental stages (10 to 24 hpf).

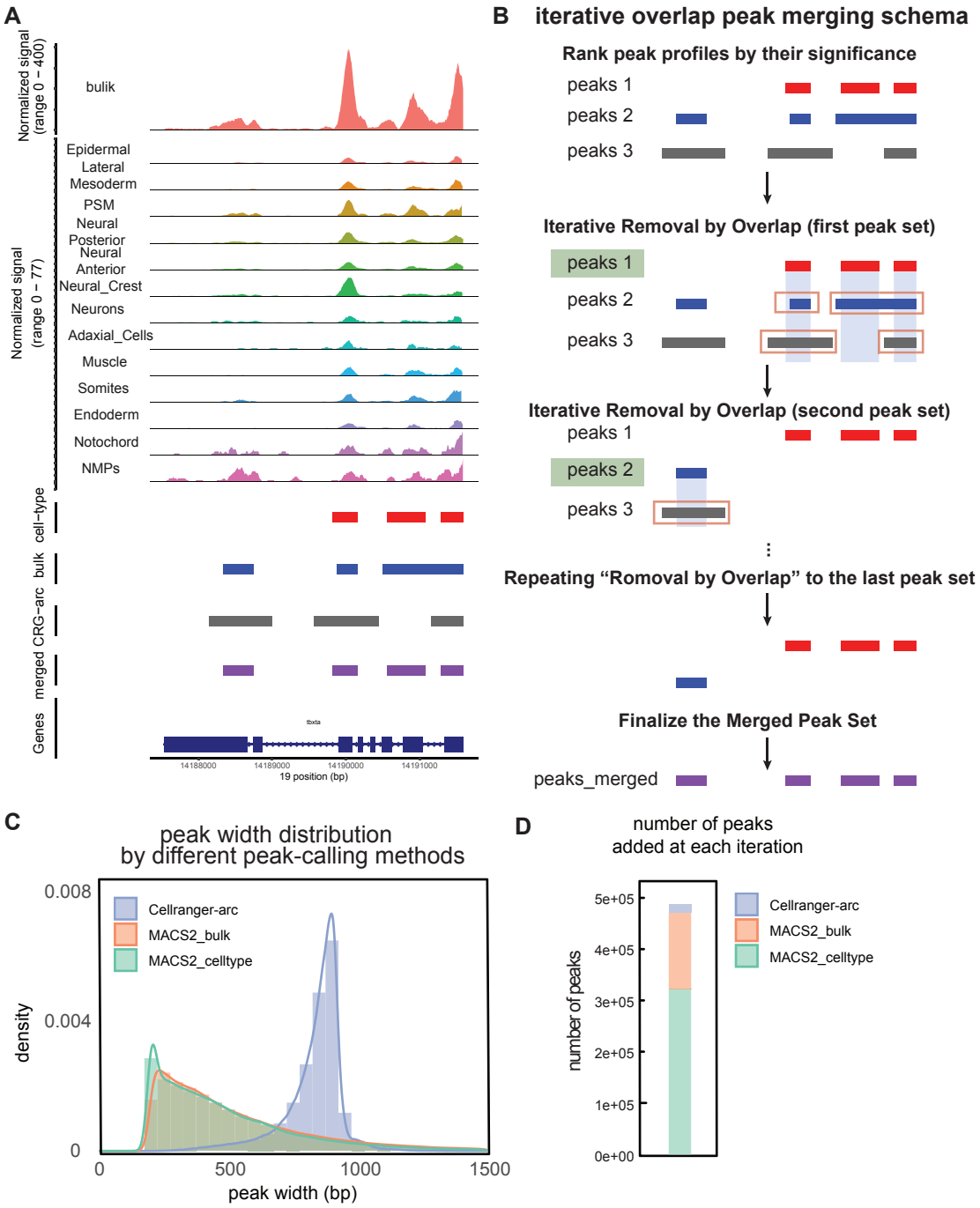

Figure S8: Peak calling strategy for scATAC-seq datasets. **a,** An example genomic coverage plot for *tbxta* gene. The normalized signal on the y-axis represents the pseudo-bulk counts for *tbxta* gene, normalized by the total counts from the specific group of cells. Here, the pseudobulk groups are shown either as bulk (all cells) or cell types, respectively. Four peak profiles are shown below the coverage plot - (1) by MACS2 for each cell type (red), (2) by MACS2 for bulk (blue), (3) by Cellranger-arc for bulk (grey), and (4) by iteratively merging the overlaps for three aforementioned peak profiles. **b,** Iterative overlap peak merging schematics. Peak profiles from different peak-calling metrics were first ranked by their significance. Then, less-significant peaks that overlap with more significant peaks are removed. Lastly, all the remaining peaks are finalized as the unique peak set. **c,** Histogram for the peak width from different peak-calling methods. **d,** Number of peaks added at each iteration in **b**.