RESEARCH

Open Access

Single-cell multi-omics delineates the dynamics of distinct epigenetic codes coordinating mouse gastrulation



Mingzhu Fu^{1,2†}, Long Pang^{2†}, Zhenwei Wu^{2†}, Mei Wang^{2*}, Jin Jin^{4*}, Shanshan Ai^{2,3*} and Xin Li^{1*}

Abstract

Gastrulation represents a crucial stage in embryonic development and is tightly controlled by a complex network involving epigenetic reprogramming. However, the molecular coordination among distinct epigenetic layers entailing the progressive restriction of lineage potency remains unclear. Here, we present a multi-omics map of H3K27ac and H3K4me1 single-cell ChIP-seq profiles of mouse embryos collected at six sequential time points. Significant epigenetic priming, as reflected by H3K27ac signals, is evident, yet asynchronous cell fate commitment of each germ layer at distinct histone modification levels are observed. Integrated scRNA-seq and single-cell ChIP-seq analysis unveil a "time lag" transition pattern between enhancer activation and gene expression during germ-layer specifica-tion. Notably, by utilizing the H3K27ac and H3K4me1 co-marked active enhancers, we construct a gene regulatory network centered on pivotal transcription factors, highlighting the potential critical role of *Cdkn1c* in mesoderm line-age specification. Together, our study broadens the current understanding of intricate epigenetic regulatory networks governing mouse gastrulation and sheds light on their relevance to congenital diseases.

Keywords Gastrulation, Epigenetic, Single-cell, Enhancer, Cell fate

[†]Mingzhu Fu, Long Pang and Zhenwei Wu contributed equally to this work.

*Correspondence: Mei Wang wangmei94@smu.edu.cn Jin Jin jinjin1065@smu.edu.cn Shanshan Ai aishanshan233@smu.edu.cn Xin Li xinli2022@i.smu.edu.cn

¹ Department of Cardiology, Heart Center, Zhujiang Hospital, Southern Medical University, 253 Industrial Avenue, Guangzhou 510280, China

² Department of Physiology, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China

³ Department of Cardiology, Heart Center, First People's Hospital of Shunde, Southern Medical University, ShundeDistrict, Foshan 528300, China

⁴ Zhujiang Hospital, Southern Medical University, 253 Industrial Avenue, Guangzhou 510280, China

Introduction

Gastrulation of the early mammalian embryo is a process characterized by the gradual acquisition of specialized lineage and morphogenetic movement of individual cells, which orchestrates the formation of the basic body plan [1-3]. In mice, gastrulation occurs during approximately E6.0-E7.5, when the pluripotent epiblast undergoes lineage restriction, giving rise to the ectoderm, mesoderm and definitive endoderm [1, 4-6]. Transplantation and clone tracing experiments have unveiled the global pattern of lineage specification during gastrulation [7, 8]. Specifically, the anterior epiblast differentiates into ectodermal cells, including neuroectoderm and surface ectoderm. Meanwhile, the posterior epiblast cells segregate proximally and anteriorly to contribute to the mesoderm and endoderm lineages through the primitive streak (PS) [**9**].



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Numerous prior studies have provided valuable insights into the activity of various molecular mechanisms including epigenetic regulation, transcription factors and signaling pathways, governing the precise developmental process and ensuring its robustness [9-11]. For example, key developmental genes in embryonic cells and extraembryonic ectoderm (ExE) cells are silenced by H3K27me3 and DNA methylation during gastrulation, respectively [12]. Furthermore, the dynamics of both proximal and distal chromatin states during the specification of the three primary germ layers exhibit asynchronous features as revealed by the epigenomic profiling of sub-regions of the mouse gastrula [13]. The function of epigenetic regulators in the mouse gastrula has been thoroughly dissected through perturbation and single-cell RNA-seq analysis [14]. Xiang et al. revealed robust bivalent chromatin states co-marked by H3K4me3 and H3K27me3 at promoters of developmental genes, gradually losing H3K4me3 or H3K27me3 along developmental progression [11]. Altogether, global epigenetic reprogramming is coordinated with the formation of the three germ layers during gastrulation. Nevertheless, the epigenetic basis for cell-fate plasticity and correct lineage allocation remains unresolved.

The advent of single-cell technologies has facilitated high-resolution profiling of multiple molecular layers of mouse gastrula embryos at the single-cell level [6, 8, 15-20]. While single-cell transcriptomic analysis has generated comprehensive atlases and identified diverse developmental trajectories during gastrulation, the role of epigenome in these processes is largely unknown. The application of scNMT-seq in studying mouse gastrulation represents the first multi-omics analysis, encompassing chromatin accessibility, DNA methylation, and gene expression levels. This analysis reveals that the epigenetic states in DNA methylation and chromatin accessibility in cells fated to ectoderm are already established in the early epiblast, whereas cells committed to mesoderm and endoderm undergo extensive epigenetic reprogramming [21]. However, the dynamic patterns of other epigenetic molecular layers, such as histone modifications at the single-cell level during the specification of three germ layers, and whether different primary germ layers utilize distinct histone codes to orchestrate their fate commitment, remain uncharacterized due to technological limitations.

In this study, we utilized our previously developed single-cell ChIP-seq method, CoBATCH [22], to profile the H3K27ac and H3K4me1 states in mouse embryos at six continuous developmental stages, ranging from Pre-Primitive Streak to Early Headfold stages [23]. Given that H3K27ac (marking active enhancers) and H3K4me1 (associated with both poised and active enhancers) enable precise identification of enhancers, characterization of super-enhancers, and improved prediction of promoter-enhancer interactions, reflecting the current and prospective developmental potential of cells [24-28], respectively, the joined analysis of both histone marks enables us to compare the dynamic patterns of two enhancer marks orchestrating the specification of three germ layers during mouse gastrulation. This approach enhances our understanding of the unique epigenetic codes utilized by cells committed to distinct fates. Furthermore, the integration of single-cell ChIP-seq with single-cell RNA data unravels the mechanisms through which enhancer usage coordinates the transition of gene expression along lineage progression, and provides rich resources to explore critical transcription factors actively involved in the fate commitments of individual germ layers.

Results

Single-cell mapping of histone modifications during mouse gastrulation

To elucidate the dynamics of epigenetic programming during gastrulation, we conducted single-cell CoBATCH for H3K27ac and H3K4me1 in the embryonic portion of mouse embryos, with a limited number of extra-embryonic ectoderm cells (Table S1). The samples were collected across six developmental stages spanning from E6.0 to E7.5, including pre-streak (Pre_Ps), Early Streak (ES), Mid-Streak (MS), Late Streak (LS), No Allantoic Bud (OB), and Early Headfold (EHF) stages (Fig. 1A). After quality control, 3,170 (H3K27ac) and 3,225 (H3K4me1) cells were retained, with averages of 8,100/7,888 reads, 95%/94% FRIP, and 94.30%/93.43% mapping rates, respectively (Fig. S1A-S1F; Table S2).

Unbiased iterative clustering of these single cells using the Seurat package [29] identified seven clusters for H3K27ac and five clusters for H3K4me1 single cells (Fig. 1B and C) (Methods). And single cells from H3K27ac and H3K4me1 ChIP-seq signal were positively correlated with developmental progression along the sampled time points (Fig. S1G and S1H). To define the cell identity in each dataset, we examined H3K27ac ChIP signals within ± 100 kb of gene body and H3K4me1 ChIP signals within ± 5 kb of the gene body of the known marker genes. In the H3K27ac dataset, we manually annotated clusters representing Epiblast (Epi), Late nascent mesoderm (LNM), Neural ectoderm 1 and 2 (NE1 and NE2), Exe ectoderm (ExE), Definitive endoderm (DE) and Mesenchyme to Mesoderm (MM) based on the ChIP-seq signals within ± 100 kb of the gene body of cluster-specific marker genes (Table S3), such as Esrrb and Epcam for Epi, Tbx6 and Mesp1 for LNM, Otx2 for NE1, Bdnf for NE2, Bmp8b and Tfap2c for ExE, Hnf1b



Fig. 1 Single-cell epigenomic profiling of mouse gastrula. A: Schematic of the experimental design. Single cells from the embryonic portion of Pre_PS to EHF mouse embryos were collected and subjected to CoBATCH to profile H3K27ac and H3K4me1 states at single-cell levels. The area below the dashed line indicates the sampling location (embryonic portion). Colors represent morphological landmarks during gastrulation. ES to MS: Yellow delineates the extent of the embryonic mesoderm. LS: Light blue marks the node at the distal tip; pink indicates the posterior amniotic fold (not yet fused to the anterior proamniotic fold). OB: Blue denotes the head process extending anterior to the dorsal tip (now visible); green marks the small allantois. EHF: Blue indicates head fold formation; green represents the enlarged allantois. B and C: UMAP showing the clustering of single cells from H3K27ac (B) and H3K4me1 (C) profiles, respectively. D-G: Stacked bar plots showing the percentage of each cell stage (D, E) and cell type (F, G) identified by H3K27ac (D, F) and H3K4me1 (E, G) single-cell ChIP-seq across different developmental stages. Pre_PS, Pre-Primitive Streak; ES, Early Streak; MS, Mid-Streak; LS, Late Streak; OB, No Allantoic Bud; EHF, Early Headfold

and *Sox17* for DE, *Mesp1* and *Hand2* for MM, respectively (Fig. S2A-S2I). Additionally, GREAT [30] analysis confirmed that cluster-specific H3K27ac peaks preferentially reside adjacent genes functioning in corresponding subpopulations (Fig. S2J-K; Table S4).

Nevertheless, we faced challenges in obtaining strong cluster-specific H3K4me1 ChIP-seq signals around the marker genes (Fig. S3; Table S3). This is likely because H3K4me1 primarily marks poised enhancers, which do not necessarily correlate positively with gene transcriptional activities [31]. Consequently, the H3K4me1 single-cell subpopulations were broadly categorized into clusters resembling Epiblast like (Epi_L), Neural ectoderm like (NE_L), Exe ectoderm like (ExE_L), Endoderm like (EN_L) and Mesoderm like (ME_L) cells (Fig. 1C). The percentage of cells at different stages within each cell type provided an overall view of the developmental process from the epiblast to the formation of the three primary germ layers (Fig. 1D-E). Among these, epiblast cells defined by H3K27ac ChIP-seq signals mostly originated from the Pre PS stage (Fig. 1D), while the H3K4me1 dataset showed a more disordered composition of cell stages in epiblast-like cells (Fig. 1E). From another perspective, when assessing the representation of each cluster at different developmental stages, we observed a decline in the percentage of epiblast cells as development progressed (Fig. 1F-G). Notably, the H3K27ac ChIP-seq-defined germ layer-specific subpopulations emerged as early as the Pre_PS stage, with the exception of the extraembryonic lineage (Fig. 1F), Similarly, ectoderm- and mesoderm-like lineages defined by H3K4me1 ChIP-seq were also detectable at Pre_PS stage (Fig. 1G), indicating effective epigenetic priming for lineage specifications in Pre_PS embryos.

Asynchronous germ-layer fate commitment revealed by distinct histone modification dynamics

We proceeded to interrogate the level of epigenetic similarity among clusters during developmental processes across different molecular layers to elucidate cell fate commitment. Principal components analysis (PCA) of the average H3K27ac and H3K4me1 signals in each subpopulation revealed different molecular layers. Across all stages, the ectoderm cluster located closer with mesenchyme to mesoderm, while separating from the endodermal lineages at the H3K27ac level (Fig. 2A). Interestingly, we observed highly similar H3K27ac patterns of NE1 and Epi cells at the Pre_PS stages, suggesting that the fate of NE1 lineage was already primed at the Pre PS stage. Conversely, germ layer-specific differences appeared before H3K4me1 ChIP-seq signal patterns in each subpopulation (Fig. 2B). When characterizing the lineage-specific dynamics of H3K27ac- and H3K4me1-marked enhancers during mouse gastrulation (Fig. 2C-H), we found the majority of lineagespecific H3K27ac peaks in EHF stage were generated at the ES stages and maintained during development (Fig. 2C, E and G). However, most ectodermal-specific H3K4me1 peaks in the EHF stage were already primed at the Pre_PS stage (Fig. 2D), while endodermal- and mesodermal-specific H3K4me1 peaks exhibited higher dynamics at the ES to MS stages. During this period, about half of the peaks in the Pre-PS stage disappeared at the ES stage, and peaks were gradually de novo generated along later development (Fig. 2F, H). The differences in histone modification patterns at distinct developmental stages indicate cells in different subpopulations likely utilized different epigenetic codes to orchestrate developmental functions.

To further characterize how different histone modifications were associated with germ-layer specification during development, we examined the enrichment of germ-layer-specific H3K27ac- and H3K4me1-defined enhancers in epiblast cells. Consistent with the observations that ectodermal cells exhibited the highest epigenetic similarity with epiblast cells at the H3K27ac level (Fig. 2A), the H3K27ac-marked ectoderm enhancers showed higher enrichment in epiblast cells than mesoderm and endoderm enhancers (Fig. 2I). However, all germ-layer-specific H3K4me1-marked enhancers were primed in the epiblast cells, with ectodermal enhancers showing the highest enrichment (Fig. 2J). This

⁽See figure on next page.)

Fig. 2 Dynamics of H3K27ac and H3K4me1 ChIP-seq signals across mouse gastrulation. A and B: PCA analysis of the average H3K27ac (A) and H3K4me1 (B) ChIP-seq signals of each subpopulation at different developmental stages except for the ExE cluster. C and D: Alluvial plots showing the global dynamics of H3K27ac (C) and H3K4me1 (D) ChIP-seq signals during ectoderm lineage specification. The number of total ectoderm-specific H3K27ac (C) and H3K4me1 (D) peaks in all stages was shown. GO enrichment analysis of the genes adjacent to EHF-specific ectoderm peaks was shown on the right panel. The colorful codes in the vertical columns represent specific ChIP-seq peaks at each stage, and the green ribbons indicate the flow of ectoderm-specific ChIP-seg signals from the previous stage along the sampled time points. E and F: Alluvial plots showing the global dynamics of H3K27ac (E) and H3K4me1 (F) ChIP-seq signals during endoderm lineage specification. The number of total endoderm-specific H3K27ac (E) and H3K4me1 (F) peaks in all stages was shown. GO enrichment analysis of the genes adjacent to EHF-specific endoderm peaks was shown on the right panel. The colorful codes in the vertical columns represent specific ChIP-seq peaks at each stage, and the blue ribbons indicate the flow of endoderm-specific ChIP-seq signals from the previous stage along the sampled time points. G and H: Alluvial plots showing the global dynamics of H3K27ac (G) and H3K4me1 (H) ChIP-seg signals during mesoderm lineage specification. The number of total mesoderm-specific H3K27ac (G) and H3K4me1 (H) peaks in all stages was shown. GO enrichment analysis of the genes adjacent to EHF-specific mesoderm peaks was shown on the right panel. The colorful codes in the vertical columns represent specific ChIP-seq peaks at each stage, and the red ribbons indicate the flow of mesoderm-specific ChIP-seq signals from the previous stage along the sampled time points. I and J: Line plot showing cluster-specific H3K27ac (I) and H3K4me1 (J) ChIP-seq signals in Epiblast cells. ChIP-seq signals were calculated within 50 bp window around the ±1 kb of the peak centers. Total 61, 3,174, 1,264, 2,226 H3K27ac peaks in the epiblast (Epi cluster), ectoderm (both NE1 and NE2 clusters), endoderm (DE cluster), and mesoderm (MM cluster), respectively, and 7,137, 176, 548, 21,567 H3K4me1 peaks in the epiblast (Epi_L cluster), ectoderm (NE_L), endoderm (EN_L), and mesoderm (ME_L), respectively, were used for signal calculations



Fig. 2 (See legend on previous page.)

observation aligns with previous reports that H3K4me1 marks inactive enhancers poised for future activation [32].

Altogether, the distinctions in different histone modification patterns during the specification of three germ layers validate the asynchronous fate commitment mechanisms of different lineages.

The regulatory trajectory of germ-layer specification

We then aimed to investigate whether our singlecell H3K27ac and H3K4me1 ChIP-seq data could be employed to reconstruct cellular developmental trajectories in an unbiased manner. We employed ForceAtlas2-based PAGA (partition-based graph abstraction) to construct a force-directed graph integrating six H3K27ac and four H3K4me1 clusters. Pseudotime analysis coupled with lineage-specific marker genes revealed progressive chromatin remodeling patterns that precisely matched known developmental transitions (Fig. 3A-D; Fig. S4; Table S5).To facilitate the following analysis, we combined the NE1 and NE2 subpopulations in the H3K27ac dataset into one ectoderm differentiation trajectory (Fig. 3B). From the above, it can be seen that the active marker can effectively fit the actual developmental trajectory, providing a reliable basis for our subsequent analysis.

Cell fate commitment is a dynamic process generally orchestration by the activity of cell type-specific transcription factors. We thus employed the single-cell regulatory trajectory to identify germ-layer specific activities of transcription factors enriched in distinct cis-elements during lineage specification. The H3K27ac- and H3K4me1-marked cis-elements with dynamic activities across the endoderm, mesoderm and ectoderm trajectories were identified (Table S6), including enhancers near known transcription factors of each lineage. For instance, genes marked by H3K27ac enhancers that were active early in the epiblasts including Foxo1 [33] and Arh, consistent with their roles in maintaining epiblast identity. While H3K27ac-marked enhancers near germ layer-specific transcription factors were selectively activated late in individual trajectories (such as Tbx3 [34], Gatat6 [35] and Hnf4a [36] in endoderm lineage; Meox1 [37], Mesp1 [38] and Hand2 [39] in mesoderm lineage; Pou3f1 [40], Otx2 [41] and Sox1 [42] in ectoderm lineage) (Fig. 3E). Notably, while intergroup differences were subtle, germ layer-specific enhancers maintained largely consistent dynamic patterns. Additionally, H3K4me1marked enhancers exhibited similar lineage-specific activities (Fig. 3F). Taken together, these analyses demonstrated that the regulatory trajectory of scH3K27ac and scH3K4me1 data could be used to identify distinct cis-elements controlling lineage-specific transcription factors essential for germ layer specifications.

Integrating scRNA-seq and scChIP-seq data reveals regulatory bases underly germ-layer specification

To correlate dynamic cis-elements with gene expression profiles, we integrated our single-cell ChIP-seq with previously published scRNA-seq datasets [19]. To better match the sample stage and cell identity, we extracted Epiblast, Mesoderm, Mesenchyme, Ectoderm, Endoderm and ExE cells across E5.5 to E7.5 from the scRNA-seq dataset generated by Pijuan-Sala et al. for integration using canonical correlation analysis (CCA) by Seurat V3 [43] (Fig. S5A; Table S7).

The identity of assigned cell types through integration in the H3K27ac and H3K4me1 datasets was verified by the enhancer activities of selected cluster-specific mark genes in each of the annotated cell types (Fig. S5B and S6A), suggesting that both of the two modalities (i.e., H3K27ac ChIP-seq and RNA-seq; H3K4me1 ChIP-seq and RNA-seq) are generally correlated.

Overall, 32.6% of H3K27ac scChIP-seq cells were assigned to the Ectoderm cluster, 26.2% to Mesoderm, 18.47% to Mesenchyme, 9.37% to Epiblast, 9.85% to Endoderm, and 4.51% to ExE (Fig. 4A; Table S8). Cluster annotations of the cells in the H3K27ac dataset were broadly consistent between ChIP signal defined and scRNA-seq assigned through integration, except for the scRNA-seq annotated Mesoderm cluster, which included cells from Epi, MM, and NE1/2 subpopulation in the H3K27ac ChIP-seq dataset (Fig. 4B). Additionally, we found intermixing of different types of cells defined by scRNA-seq within the same H3K27ac defined cluster (Fig. 4C), indicating extensive chromatin priming occurs among scRNA-seq defined cell types. Similarly, 29.09% of H3K4me1 scChIP-seq cells were assigned to the Mesoderm cluster, 28.34% to Epiblast, 24.19% to Ectoderm, 9.89% to ExE, 7.26% to Endoderm, and 1.24% to Mesenchyme (Fig. S6B, Table S8). However, cluster annotation of the scH3K4me1 and scRNA exhibited weak correlation except for endoderm and ExE (Fig. S6C), coincident with the role of H3K4me1 at enhancers, which function in fine-tuning, rather than tightly control enhancer activity and function [44].

We then utilized the integrated H3K27ac ChIP-seq and RNA-seq datasets to investigate the lineage-specific dynamics of H3K27ac-defined enhancers along germ layer specification, considering its better integration results. Pseudotime analysis of scRNA-seq revealed three differentiation trajectories starting from the epiblast to endoderm, mesoderm or ectoderm lineages (Fig. 4D-K and S7A-S7D; Table S9), in which cell identities were corroborated by the progressive loss of epiblast markers



Fig. 3 Regulatory trajectory analysis of epiblast differentiation according to H3K27ac and H3K4me1 ChIP-seq signals. **A**: Force-directed graph (FDG) showing the differentiation trajectories of H3K27ac single cells from Fig. 1B. **B**: Partition-based approximate graph abstraction (PAGA) showing pseudotime trajectories of cells from (**A**). The size of the dots represents the relative number of cells in each cluster. **C**: FDG showing the differentiation trajectories of H3K4me1 single cells from Fig. 1C. **D**: PAGA showing pseudotime trajectories of cells from (**C**). The size of the dots represents the relative number of cells in each cluster. **E** and **F**: Heatmap (the upper panel) and line plot (the lower panel) showing the H3K27ac (**E**) and H3K4me1 (**F**) signals around lineage-specific genes along three lineage specifications in (**A** and **C**)

and gain of lineage-specific genes (e.g., *Mesp1* for mesoderm, *Sox17* for endoderm) (Fig. 4E, I and S7B). These marker dynamics align with established murine gastrulation studies [6]. The dynamic trends along pseudotime in both gene expression and H3K27ac ChIP signals around ± 100 kb of corresponding variable genes were well correlated, highlighting the critical role of lineagespecific H3K27ac defined enhancers in the orchestration of cell fate commitment during gastrulation (Fig. 4F, G, J, K and S7C-S7D).

Comparing the dynamic patterns of gene expression versus enhancer activity during germ-layer specification, we found that the shutdown of epiblast-specific gene expression slightly precedes the repression of corresponding enhancers at the early stage of endoderm differentiation trajectory (Fig. 4G), and a more obvious pattern was observed during the mesoderm differentiation trajectory (Fig. 4K). On the contrary, the initiation of enhancer activity of germ layer-specific genes occurs before their gene expression at late developmental stages, indicating extensive epigenetic priming before cell fate commitments. However, the enhancer reprogramming for silencing or activation preceded the gene expression at both early and later ectodermal lineage specification (Fig. S7D), suggesting that the dynamics of transcriptional states and epigenetic modifications are far from coherent and synchronous among different layers. As a control, we also analyzed the dynamics of RNA and H3K27ac signals on unchanged genes along three lineage differentiation (Fig. S7E-S7G; Table S9).

Collectively, the above data uncovered distinct regulatory bases underlying ectoderm lineage specification compared to mesoderm and endoderm fate commitment through our integration analysis. Moreover, we revealed a "time lag" transition pattern between H3K27ac-marked enhancer and gene expression during germ-layer specification, which guarantees the correct priming of the epiblast cells towards individual lineages.

Active enhancer-related gene regulatory networks across mesoderm lineage specification

Enhancers can be defined as poised or active states based on the presence or absence of H3K27ac and H3K4me1 modifications [45]. Thus, we aim to further clarify the dynamics of active enhancers co-marked by H3K27ac and H3K4me1 during lineage specification and identify pivotal transcription factors essential for cell fate commitments.

Inspired by our previous integration strategy (Fig. 4 and S5-6), we tended to integrate the H3K27ac and H3K4me1 ChIP-seq datasets through scRNA-seq as a linker. To achieve this goal, we first merged single RNAseq cells with their closest K-nearest neighbors to generate Micro-sc (Micro single cell) utilizing Vision [46], and performed integration of Micro-sc in scRNA-seq dataset with single cells from H3K27ac and H3K4me1 datasets, respectively (Fig. 5A). To avoid potential contamination during integration, we only kept cell pairs with the same identities defined by scRNA-seq and scChIP-seq (Fig. 5B; Table S10). Successful integration of three datasets thus can be achieved through the scRNA-seq dataset as the linker. To further increase the integration efficiency, we performed a second round of Vision analysis to merge Micro-sc into Pseudobulk-sc (Pseudobulk single cell) to get more paired cells with the same identities defined by three modalities (Fig. 5A and B). The identities of correctly linked cells in each dataset were confirmed by the expression and enhancer

Fig. 4 Integrative analysis of single-cell H3K27ac ChIP-seq and single-cell RNA-seq datasets of the mouse gastrula. A: Donut plot showing the percentage of scH3K27ac ChIP-seq cells assigned to different types of scRNA-seq cells. B: Heatmap displaying the fraction of cells in each scRNA-seq cluster linked to corresponding H3K27ac ChIP-seq clusters through integration by Seurat. C: Barplot showing the percentage of scRNA-seq cells within each annotated scH3K27ac ChIP-seq cluster. Cluster colors match the cell type in (A). D: Pseudotime analysis of the scRNA-seq cells of endoderm lineage by Monocle 3. E: Scatter plot showing the cell-type specific gene expression. The representative marker genes were selected for the validation of the epiblast and endoderm cells, respectively. F: Heatmap showing normalized RNA signals of 193 DEGs across 109 scRNA-seq cells of endoderm lineage (left), and normalized H3K27ac ChIP-seq signals at ± 100 kb of gene body across corresponding linked scH3K27ac ChIP-seq cells of endoderm lineage (right). The DEGs were clustered into three modules. The heatmap in the left panel represents scRNA-seq cells ordered by pseudotime, and the heatmap in the right panel represents scH3K27ac ChIP-seq cells linked to the same scRNA-seq cells but annotated with identities defined by the H3K27ac ChIP-seq. G: Aggregate curves showing the scaled RNA and H3K27ac ChIP-seq signals in (\mathbf{F}) of three gene modules, two model signals were max-min normalized to $0 \sim 1$ for comparison. **H**: Pseudotime analysis of the scRNA-seq cells of mesoderm lineage by Monocle 3. I: Scatter plot showing the cell-type specific feature gene expression. The representative marker genes were selected for the validation of epiblast and mesoderm cells, respectively. J: Heatmap showing normalized RNA signals of 159 DEGs across 217 scRNA-seq cells of mesoderm lineage (left), and normalized H3K27ac ChIP-seq signals at ± 100 kb of gene body across corresponding linked scH3K27ac ChIP-seq cells of mesoderm lineage (right). The DEGs were clustered into three modules. The heatmap in the left panel represents scRNA-seq cells ordered by pseudotime, and the heatmap in the right panel represents scH3K27ac ChIP-seq cells linked to the same scRNA-seq cells in the left but annotated with identities defined by H3K27ac ChIP-seq. K: Aggregate curves showing the scaled RNA and H3K27ac ChIP-seq signals in (J) of three gene modules, two model signals were max-min normalized to $0 \sim 1$ for comparison

⁽See figure on next page.)



Fig. 4 (See legend on previous page.)



Fig. 5 Lineage dynamics of chromatin and gene expression identify molecular signatures for mesoderm specification. A: Schematic representation of the triple-omics integration strategy for scRNA-seq, H3K27ac and H3K4me1 scChIP-seq datasets. B: Sankey diagram highlighting cell pairs with the same cell identity defined by three modalities. C: Heatmaps showing the dynamics of gene expression (left) and active enhancer score (right) of 87 variable genes along mesoderm differentiation trajectory. D: GO enrichment analysis of genes represented in three modules of mesoderm differentiation trajectory. E: Enrichment of TF motifs in active enhancers represented in three modules of mesoderm differentiation trajectory. F: Dot plot showing gene discordance scores using TRIAGE analysis, with genes ordered by their expression. TFs with discordance scores above the threshold were displayed. Venn diagram showing the overlap of TFs enriched in TRIAGE and those enriched in active enhancers. G: TF-gene regulatory network showing the regulation of TFs enriched in active enhancers and its targeted genes in mesoderm trajectory. The width of an edge indicates the expression correlation between TFs and targeted genes, the size of a dot indicates the number of nodes in the network.
H: Dynamics of gene expression along mesoderm differentiation trajectory. The color bar represents the z-score standardized value for each row. I: Schematic of stepwise activation of *Cdkn1c* expression in mesoderm specification

activities of selected cluster-specific mark genes in individual subtypes (Fig. S8A).

Using the integrated dataset, we analyzed the dynamics of gene expression and the enhancer score of their corresponding active enhancers defined by both H3K27ac and H3K4me1 signals during mesoderm specification (Table S11) (Methods). Pseudobulk-scs were ordered along pseudotime and 87 variable genes were grouped into three clusters: decreased (cluster A), intermediate (cluster B), and increased (cluster C) (Fig. 5C, S8B and S8C). Cluster A genes exhibited gene ontology (GO) enrichments for embryonic morphogenesis and embryonic organ development, whereas genes represented in cluster C were enriched for mesoderm development, gastrulation, and vasculature development-related terms. Interestingly, cluster B genes particularly participated in the negative regulation of neurogenesis, indicating that the inhibition of neurogenesis is essential for the initiation of mesoderm differentiation during gastrulation (Fig. 5D; Table S12). Notably, the signals of active enhancers defined by the co-occurrence of H3K27ac and H3K4me1 at±5 kb of the gene body of corresponding genes exhibited overall similar dynamic patterns as gene expression (Fig. 5C), highlighting the positive correlation between gene expression and the activity of their corresponding active enhancers.

Consistent with previous reports that enhancers co-marked by both H3K4me1 and H3K27ac exhibit stronger tissue specificity and are prone to be relevant to cell lineage commitments [47], we observed that the gene activity of active enhancers was higher than of the H3K4m1-marked primed enhancers among the pseudobulk mesodermal cells (Fig. S8D). Furthermore, we identified TF motifs enriched in the active enhancers specific to each module (Fig. 5E). The TFs enriched early in the epiblast cells including the Krüppel-like factor and specificity protein (KLF/SP) family transcription factors, have been demonstrated to participate in biological processes including stem cell maintenance, embryonic development, and tissue differentiation [48]. TFs with known function in mesoderm specification, including MEF2C [49], FOXC2 [50], MESP1 [38], HAND2 [51], TWIST2 [52], were specifically enriched in module C (Fig. 5E; Table S13). Additionally, we also identified some unknown TFs with potential roles in mesoderm specification, such as TEAD3 (Fig. 5E; Table S13). Intersecting these TFs, enriched in variable active enhancer regions along mesoderm specification, TFs identified through TRIAGE [53] revealed three key TFs: T, FOXC2, and HAND2. This finding aligns with their well-established critical roles in mesoderm fate commitment [15, 54, 55] (Fig. 5F).

We then explored the TF-driven regulatory networks associated with mesoderm lineage specification by linking relating TFs enriched in active enhancers to their putative target genes (Fig. 5G; Fig. S8E; Table S14). Analysis revealed a stepwise transcriptional cascade involving cell cycle regulator Cdkn1c during mesoderm specification (Fig. 5H and I). Specifically, the expression of Hoxa1 was activated early during lineage trajectory, which then activated Meis1 expression together with TBX1/5 through binding the active enhancers of Meis1. The expressed MEIS1 then activated the expression Cdkn1c together with TBX1/5 (Fig. 5H and I). The above mesoderm-specific transcription factor (HOXA1, TBX1/5, MEIS1) involved regulatory circuit suggests that Cdkn1c may play a coordinated role in mesoderm development, though its functional significance relative to other downstream effectors requires further validation. Thus, the integration of single-cell RNA-seq and single-cell ChIPseq data can be used to construct regulatory networks and explore particular regulatory circuits essential for specific lineage specification.

Discussion

Understanding the epigenetic basis of gastrulation is crucial as it addresses a fundamental question in both stem cell biology and developmental biology. Disruptions in this intricate process can lead to developmental abnormalities and congenital disorders. However, exploring the molecular activities contributing to the intricate choreography of gene expression that underlies successful gastrulation remains technically challenging due to the complexity and plasticity of this process. In this study, we present the first single-cell resolution maps of histone modifications, specifically H3K27ac and H3K4me1, during the development of mouse gastrula across six stages. Although both H3K27ac and H3K4me1 make enhancers, more subclusters were identified by H3K27ac ChIP-seq signals than H3K4me1 signals. One plausible explanation is that H3K27ac ChIP-seq signals demonstrate greater variability than H3K4me1 signals in gastrulating embryos, thereby enabling better discrimination of cell subtypes. Previous studies have established that H3K27ac exhibits superior cell-type specificity compared to H3K4me1, showing dramatic variation across different cellular states [24]. This difference is further amplified by their distinct chromatin signatures: H3K27ac forms narrow, sharp peaks that facilitate clear cluster separation, while H3K4me1 generates broader, more uniform peaks that resist subdivision into discrete clusters. Additionally, H3K27ac-marked enhancers reflect cells' immediate developmental potential through their rapid response to signaling and differentiation cues, whereas H3K4me1 primarily marks poised enhancers that maintain relatively stable chromatin states, making cell identity determination more challenging when relying solely on H3K4me1 signals.

Our findings reveal substantial epigenetic priming for the specification of different germ layers in the Pre_PS embryos, as reflected by H3K27ac signals. Nevertheless, there are asynchronous features in cell fate commitment among each germ layer at distinct histone modification levels. When comparing the dynamic patterns of H3K27ac and H3K4me1 during gastrulation, we observed that cells destined for distinct lineages employ unique epigenetic codes to orchestrate their developmental functions. Notably, cells in early mesoderm and ectoderm lineages exhibit more similar H3K27ac modification patterns compared to the endoderm lineage, which is consistent with the observation that inhibition of neurogenesis restricts the differentiation of mesodermal lineage during gastrulation [56]. However, the three germ layers display more obvious lineage-specific H3K4me1 patterns in each subpopulation.

Integrative scRNA-seq and single cell H3K27ac ChIP-seq analysis revealed a "time lag" transition patterns between H3K27ac-marked enhancer and gene expression during germ-layer specification, showcasing extensive chromatin priming within transcriptionally homogeneous subpopulations. Previous studies have also reported a time lag between epigenetic modifications (e.g., H3K27ac) and transcriptional dynamics. This lag is attributed to enhancer priming for cell fate decisions, the release of paused RNA polymerase II, and the progressive recruitment of co-activator complexes prior to productive transcription initiation [57, 58]. Notably, our results showed that the downregulation of epiblast-specific genes precedes that of enhancer repression in mesoderm and endoderm, but not in ectoderm, suggesting that epiblast cells are epigenetically primed for an ectodermal fate earlier than for mesodermal and endodermal fates, as previously reported [21]. This finding also supports the idea that epiblast cells are more similar to ectoderm at the transcriptional and epigenetic levels [13, 59]. Further trimodality integrative analysis uncovered active enhancerrelated gene regulatory networks defined by H3K27ac and H3K4me1, along with pivotal transcription factors essential for mesoderm lineage specification. While these TFs are not genomically clustered, their co-enrichment within mesoderm-specific enhancers suggests potential functional convergence through shared regulatory hubs or chromatin interactions. Notably, HAND2 and TWIST family, which co-regulate cardiac development via enhancer sharing, exemplify a broader mechanism where lineage-specific TFs may bind spatially proximal enhancers to coordinate gene expression [60]. Exploration of the regulatory network identified a stepwise activation of *Cdkn1c* by a combination of mesoderm-specific transcription factors, emphasizing the potential essential role of Cdkn1c in mesoderm cell differentiation and specification. During mouse gastrulation, Cdkn1c displays dynamic spatiotemporal expression patterns [19], with sustained activity throughout embryonic development that potentially regulates cell proliferation and differentiation [19, 61]. The expression and function of *Cdkn1c* are evolutionarily conserved across species (e.g., zebrafish and Xenopus), underscoring its essential role in developmental processes [61]. Recent studies have demonstrated that cell cycle regulators are involved in cell fate decisions not only through their roles in cell cycle regulation and signaling modulation but also directly as chromatin regulatory factors [62-65]. Moreover, loss-of-function experiments have shown that G1 phase regulators play a critical role in efficient mesoderm formation [66], providing strong evidence that the G1 phase inhibitor *Cdkn1c* is integral to mesoderm specification. Furthermore, the role of *Cdkn1c* in regulating cell division and differentiation in diseases like Beckwith-Wiedemann syndrome (BWS) also suggests its potential influence on lineage specification [67]. While these findings hint at a plausible role for *Cdkn1c* in mesoderm specification, its exact mechanistic contributions remain to be fully resolved. Taken together, our comprehensive single-cell multi-omics profiling of the mouse gastrula provides a new molecular framework for understanding the distinction and cooperation among distinct epigenetic layers in controlling cell fate allocation during this critical developmental stage.

However, our study has certain limitations due to its reliance on widely used computational tools. While these tools are well-established, their inherent may introduce biases. Further studies using multi-omics strategies such as Paired-Tag [27] and CoTECH [68] to directly measure transcriptomic and epigenetic modalities in the same single cell, as well as integrating multi-algorithm consensus, rather than relying on stand-alone bioinformatic analysis-based data integration, would provide more direct and thorough insights into the role and dynamic patterns of distinct epigenetic codes in the lineage specification of the three germ layers.

Materials and methods

Research animals

All mice experiments were performed following the protocol approved by the Institutional Animal Care and Use Committee of Southern Medical University (SMUL2023045). All mice used in this study were are on a C57BL/6 J background and were purchased from the Institutional Animal Care and Use Committee of Southern Medical University. Euthanasia was performed using the CO₂ inhalation method, in accordance with

the American Veterinary Medical Association's Euthanasia Guidelines. Briefly, mice were placed in a euthanasia chamber, and CO_2 was gradually introduced at the recommended flow rate. After no detectable breathing was observed for more than five minutes, pregnant mice were considered deceased and used for dissection. Mouse embryos were dissected at time points of E6.0, E6.5, E7.0, E7.25 and E7.5. The embryos were staged according to their morphology (Downs and Davies staging) [23], and classified as pre-streak (Pre_Ps), Early Streak (ES), Mid-Streak (MS), Late Streak (LS), No Allantoic Bud (OB), and Early Headfold (EHF) stages. No anesthetic agents were used in this study.

Mouse embryo single-cell isolation

E6.0, E6.5, E7.0, E7.25 and E7.5 embryos were dissected from the uterus in 10% FBS-KOSM medium. The individual embryo was micro-dissected and staged as Pre-PS, ES, MS, LS, OB, and EHF respectively according to the morphological criteria of Downs and Davies [23]. Pre-PS, ES, MS and LS embryos were dissected to remove extraembryo tissues. All the embryos were digested into single cells by incubating in Collagenase I dissociation buffer (1% Collagenase I and 10% FBS) at 37 °C for 15-60 min. The dissociated single cells were collected by centrifugation at 600 g for 3 min and the cell pellets were resuspended with 50 µl 0.1 BSA/PBS. The cell suspension was further incubated with $1-5 \mu$ l pre-activated Con-A beads (the volume of Con-A beads depends on the cell number) at RT for 15 min. Finally, the cell-bead mixture was fixed with methyl alcohol and incubated on ice for 5 min before stored at -80 °C for later use.

Purification of PAT

PAT (His-pA-Tn5) was purified as previously described [22]. Briefly, the PAT expression vector was transformed into BL21 (DE3) chemically competent cells following the manufacturer's protocol. One single clone was inoculated with 20 ml LB medium and grew at 37 °C overnight. 5 ml of the culture medium was transferred into 500 ml LB medium and grew at 37 °C, 220 rpm for about 4 h until it reached O.D.~0.8. The culture was chilled on ice for 20 min and fresh 0.2 mM IPTG was added to induce PAT expression. The culture was incubated at 23 °C, 100 rpm for 5 h for induction. Bacteria were collected by centrifugation at 5,000 rpm, 4 °C for 5 min. The pellet can be used directly or stored at -80 °C for later use. To purify PAT, the cell pellet was suspended with 20 ml HGX buffer and sonicated for 5 min at 8 s ON, 16 s OFF, 20% amplitude (Sonics). The sonicated lysate was centrifuged at 10,000 rpm, 4 °C for 30 min. 50 µl 10% PEI (Sigma P3143) was added to precipitate bacterial DNA, and the precipitate DNA was removed by centrifugation at 10,000 rpm, 4 °C for 10 min and the supernatant fraction was further purified by 0.45 μ M filter (Millipore SLHV033RB). The purified lysate was loaded onto the pre-equilibrated Ni–NTA column (QIAGEN) and the Ni–NTA column was further washed with 100 ml HGX- imidazole buffer I (20 mM imidazole). Finally, PAT was eluted with 5 ml HGX- imidazole buffer II (250 mM imidazole) and dialyzed with 1000 ml 2X dialysis buffer (100 HEPES–KOH at pH7.2, 0.2 M NaCl, 0.2 mM EDTA, 2 mM DTT, 0.2% Triton X-100, 20% glycerol). The dialyzed protein was concentrated using a 50 kDa cutoff ultracentrifuge column (Millipore UFC905096) and an equal volume of glycerol was added to the purified PAT.

PAT assembly and activity quantification

The PAT transposase was assembled according to the Tn5 transposase assemble protocol [69–71]. PAT oligos were mixed with an equal volume of MErev oligo and incubated at 95 °C for 5 min and then cooled down at a rate of 0.1 °C/min for annealing [72]. To generate 25 μ M PAT adapter, 25 μ M purified PAT and 25 μ M annealed adaptor were mixed with storage buffer (50 mM HEPES pH7.2, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 0.1% Triton X-100, 60% glycerol) and incubated at 25 °C for 60 min. The activity of the PAT transposase was determined by tagmentation of genomic DNA as previously described [69–71]. The assembled PAT transposase can be stored at -20 °C for about half a year.

CoBATCH experimental procedure

The CoBATCH experiment was performed as previously described with several modifications [22]. Briefly, the methyl alcohol fixed cell was incubated on ice for 5 min and placed onto the magnetic stand to get rid of the methyl alcohol, followed by washing with 0.1 BSA/PBS Thrice. After wash, the cell was resuspended with 100 µl Antibody Buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 µM spermidine, 10 mM sodium butyrate, 0.04 mM EDTA, 0.01% Digitonin, 0.05% TX-100, cocktail) with 0.5 µg antibody after the last wash and incubate at 4 °C for 4 h at 15 rpm. The cells were washed with 180 µl Antibody wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 µM spermidine, 10 mM sodium butyrate, 0.01% Digitonin) twice, followed by incubation with Secondary antibody at 4 °C for 10 min. After antibody binding, the cells were FACS sorted into 96-well plates with 200-2000 cells/well containing 3 µg/ml PAT-T5 and 3 µg/ml PAT-T7. Cells were washed with 180 µl PAT-wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 µM spermidine, 10 mM sodium butyrate, 0.01% Digitonin, 0.05% TX-100, cocktail) twice. Tagmentation was activated by 10 µl reaction buffer (10 mM TAPS-NaOH pH 8.3, 5 mM MgCl2, 10% DMF and supplemented with cocktail,

10 mM sodium butyrate)) and incubated at 25 °C for 1 h. Tagmentation was stopped by adding 8 μ l 50 mM EDTA and incubating at RT for 15 min. Cells were combined after adding 20 μ l 2% BSA/PBS to each well and stained by DAPI. Finally, 20–25 cells were sorted into each well of a new 96-well plate containing 4 μ l nuclear lysis buffer (10 mM Tris–HCl, PH 8.0, 0.05% SDS, 0.1 mg/ml Proteinase K) and the cells were lysed at 55 °C for 3 h, followed by incubation at 85 °C for 15 min to deactivate proteinase K.

Library preparation for CoBATCH

The CoBATCH library was prepared directly in the same tube according to the 2-round Truseq library preparation workflow [72]. Briefly, the first step was performed by the addition of 0.5 µL 50 µM Truseq connector primer mix (Connector primer F:5'-ACACTCTTTCCCTACACG ACGCTCTTCCGATCTTCGTCGGCAGCGTCTCCA CGC-3', Connector primer R:5'-GACTGGAGTTCA GACGTGTGCTCTTCCGATCTGTCTCGTGGGCTC GGCTGTCCCTGT-3'), 10 μ L 5×Q5 reaction buffer, 10 μ L 5×Q5 high GC enhancer, 1 μ L 10 mM dNTP, 23.5 μ L 1 mM MgCl2, 0.3 µL Q5 polymerase (NEB M0491S) to the 3 µL CoBATCH cell lysis in each tube and the reaction was set up by incubation at 72 °C for 5 min, 95 °C for 5 min, 13–16 cycles of amplification (95 °C for 30 s, 63 °C for 30 s, 72 °C for 1 min), and final 72 °C extension for 5 min. Then 0.5 µL 20 U/µl ExoI (NEB M0293S) was added and the reaction was incubated at 37 °C for 30 min, and 72 °C for 20 min., The second PCR enrichment was performed by the addition of 1 μ l 10 mM Truseq index P5, 1 µl Truseq index P7, 2 µl 5×Q5 reaction buffer, 2 μ l 5×Q5 high GC enhancer, 0.5 μ l 10 mM dNTP, 3.5 µl 1 mM MgCl2, 0.1 µl Q5 polymerase (NEB M0491S) to the mixture, and the reaction was set up by incubation at 95 °C for 5 min, 5–7 cycles of amplification (95 °C for 30 s, 63 °C for 30 s, 72 °C for 1 min), and final 72 °C extension for 5 min. Finally, the PCR products were purified and selected by SPRI beads for 200-1000 bp.

ChIP-seq data pre-processing

CoBATCH data were demultiplexed by custom scripts as previously described [22]. Reads were distinguished by different barcodes with a Perl script, and trimmed for constant sequence and barcodes with Cutadapt [73] with default parameters. Reads were aligned to mm10 reference genome through Bowtie2 [74] with the following parameters: –dovetail –very-sensitive-local –no-unal – no-mixed –no-discordant, and then sam files were sorted and converted to bam files with Samtools [75]. Picard (https://broadinstitute.github.io/picard/) with default process was used to remove duplications. The cells with reads > 1500, reads < 50,000 and Frip > 0.1 were retained for downstream analysis.

Dimension reduction and subcluster annotation

Peaks were called by MACS2 for all stages separately [76], additional parameters "–nomodel –nolambda -broad" were used for obtaining broad peaks, and then merged by a 5 kb window, which was used to generate a cell-peak matrix. Doublets were removed with more than 50,000 reads, and then LSD dimensionality reduction and clustering were performed by the Signac package [77]. The single-cell subclusters were manually anno-tated by inspecting of our ChIP-seq datasets and counting the peak signals within \pm 100 kb (H3K27ac) and \pm 5 kb (H3K4me1) of the gene body of cell type-specific mark genes (Table S3).

Trajectory analysis

In order to infer potential differentiation trajectories, the Seurat objects were converted to the AnnData file with the SeuratDisk package (https://mojaveazure.github. io/seurat-disk). The generated h5ad files were applied to PAGA and Force-Directed Graph (FDG) to infer the development trajectories with Scanpy [78]. We recalculated the neighborhood graph (*neighbors* function with n_neighbors equal to 30 for H3K27ac and 10 for H3K4me1) on the latent space to exploit the data, and computed the PAGA graph (*paga* function with a model equal to v1.0). The epiblast cluster was set as the root of the trajectories for diffusion pseudotime based on PAGA maps.

As for the trajectory of integrated data of scRNA-seq and scChIP-seq, we used the pseudotime trajectory of scRNA-seq as a reference inferred through Monocle3 [79]. Cells from different lineages in the scRNAseq dataset were converted into a cds object as inputs for Cicero [80]. The function of detect_genes was used to calculate the number of cells each gene expressed in, and genes that were detected in more than 50 cells were kept for further analysis. Based on the previous UMAP dimensionality reduction, the *learn_graph* function fits a principal graph within each partition. After setting the epiblast cluster as the root point, we used the order_cells function to calculate the pseudotime of each cell along each trajectory. The gene matrix was normalized by the total counts for each cell and multiplied by the scale.factor 10,000 and natural-log transformed using log1p and then z-score scaled. We summed the gene activity score of lineage-specific genes according to pseudotime, and used the loess method to fit the curve. The overlapped signals of H3K27ac and H3K4me1 around ±5 kb of the gene body of lineage-specific genes were used to define the enhancer score of active enhancers. The TF motif enrichment in three modules along the mesoderm lineage was calculated by Fisher's exact test with LOLA package [81] based on the genome-wide motif PWM matrix generated by motifmatchr package (https://doi.org/10. 18129/b9.bioc.motifmatchr). The mouse RTS values were obtained from the human reference RTS table by directly mapping genes between the human and mouse databases using the biomaRt package. The mean expression of mesoderm-specific genes enriched for active enhancers was calculated and used as input for the TRIAGE tool to infer the discordance scores of the genes [53]. Genes with a log10 discordance score above 0.5 were considered specific regulatory factors. Three key transcription factors were identified based on the discordance scores and their overlap with the enriched active enhancers.

Integration of scRNA-seq and scChIP-seq datasets

We extracted scRNA-seq cells for integration from the scRNA-seq dataset generated by Pijuan-Sala et al. [19], which includes mouse embryo samples from the E6.5 to E7.5, and randomly sampled them for each lineage according to the cell proportion of each cluster in the ChIP-seq datasets. Following the same procedures used for integrating our ChIP-seq atlas, we obtained a highquality single-cell reference atlas comprising 3,201 cells for H3K27ac and 3,550 cells for H3K4me1 from 19 samples. The integration was conducted using Seurat [43]. Briefly, the scRNA-seq data were quoted as the reference dataset to assign the cell identity to the ChIP-seq cells (Table S7). We used the *FindVariableFeatures* function to extract top 2000 variable genes in the scRNA-seq dataset, and generated gene activity matrices for our ChIPseq datasets by counting the peak signals around ± 100 kb (H3K27ac) and $\pm 5 kb$ (H3K4me1) of the gene body of the variable genes. FindTransferAnchors (query.assay equal to gene activity, features equal to VariableFeatures (RNA object)) function was used to establish the anchors of the cells in the two datasets by utilizing canonical correlation analysis (CCA) as an dimension reduction method, and the TransferData function was applied to assign the cell identity from scRNA-seq to the ChIP-seq cells. The AnnotateAnchors function was used to extract the anchor pairs of the cells from the integrated dataset, and the anchor pairs with the largest anchor.core for ChIP-seq cells were retained.

For triple-omics integration, the RNA cells were submerged into micro-sc (micro single cell) by VISION package [46], and the identity of the micro-scs was determined by the cell type with the highest proportion in the micro-sc. The micro-sc was used for integration with ChIP-seq datasets as above described. After the first integration, we kept cell pairs with the same cell identity defined by three omics. As for the other cell pairs, we merged two micro-scs with the nearest distance calculated by the *FindNeighbors* function into one pseudobulksc (pseudobulk single cell) and also selected cell pairs with the same cell identity defined by three omics. The selected cell pairs from both micro-scs and pseudobulkscs were used to generate new RNA Seurat object.

GRN construction

To build the mesoderm-specific Gene Regulatory Networks (GRNs), we employed a multi-step approach to identify lineage-critical transcription factors (TFs) with high confidence. Mesoderm-specific active enhancers were defined by co-occurrence of H3K27ac and H3K4me1 signals within ±5 kb of mesoderm marker genes. This ensured enhancer relevance to mesoderm differentiation. Motifmatchr package was used to scan the motifs within these mesoderm-specific active enhancers and the values were converted to a binary matrix. The getBackgroundPeaks function from chromVAR [82] was used to calculate background signals, filtering out ubiquitous motifs to focus on mesoderm-enriched TFs. The motifPeakZest from the FigR package [83] was applied to calculate motif enrichment by z.test of TF motif-topeak match in active enhancers, relative to the expected frequency based on matches to a background peak set. Based on the scRNA-seq dataset, the expression correlation was computed by Hmisc package between TFs and their targeted genes. The GRNs were constructed based on the TF motif enrichment and the correlation. The TFtargeted gene pairs with log10 motif enrichment scores higher than 0.32 and significant correlation absolute values higher than 0.25 were selected to construct the mesoderm core regulatory network,, ensuring functional relevance. Eigenvector centrality was calculated by the evcent function in igraph (https://igraph.org/) to subset network that were displayed in cytoscape [84].

PCA analysis

The single cells from H3K27ac and H3K4me1 scChIPseq datasets were aggregated into 36 and 24 pseudobulk groups according to their stage and cluster information except for the ExE cluster, respectively. The ChIP-seq signals of each group were calculated by summing the value of all single cells in each group according to the singlecell cell-peak matrix. Principal Component Analysis (PCA) was applied to dimensionality reduction on the top 100 variable genes.

Dynamics of H3K27ac andH3K4me1 signals in each germ layer

The single cells from subpopulations of Epiblast, Ectoderm (both NE1 and NE2 for H3K27ac dataset), Endoderm and Mesoderm were aggregated separately to generate four pseudobulk groups. The peaks of the four pseudobulk groups were called by MACS2 [76] and specific peaks for each group were obtained through hierarchical clustering in pheatmap (https://CRAN.R-proje ct.org/package=pheatmap). The four pseudobulk groups were further subdivided into 24 subgroups according to six developmental stages and peaks were called separately for the 24 subgroups. Bedtools [85] were used to calculate the number of overlapped peaks between 24 subgroups and four pseudobulk groups with intersect function. GO term enrichment for specific peaks was performed by GREAT [86] based on UCSC mm10 genome.

The enrichment of germ layer-specific ChIP-seq signals in epiblast cells

We used bedtools [85] intersect with "-v" parameter to obtain germ layer-specific peaks for four pseudobulk groups (Epiblast, Ectoderm, Endoderm and Mesoderm). ComputeMatrix was further applied to calculate the ChIP-seq signals of Epiblast cells at ± 1 kb of germ layerspecific peak centers with the reference-point model and 50 bin sizes for averaging the score.

Statistical information

Parameters and results for Statistical analysis were described in the figure legend for each experiment, and more details were shown in specific method descriptions. For all statistical tests, the 0.05 *P*-value < 0.05 was considered statistically significant.

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s12864-025-11619-5.

Supplementary Material 1.

Declaration on animal experimentation

This study is reported in accordance with ARRIVE guidelines.

Authors' contributions

S.A. and X.L. designed and conceived the study. M.F. and L.P. performed the bioinformatic analyses. S.A., X.L. and Z.W. conducted the experiments. X.L., S.A., M.F., J.J. and M.W. participated in data discussion and interpretation, as well as paper writing.

Funding

This study was supported by grants from the National Key R&D Program of China (2021YFA1102700 and 2022YFA1106200), the National Natural Science Foundation of China (82270307, 32200660 and 3240060016), the Natural Science Foundation of Guangdong Province (2024B1515020058), China Postdoctoral Science Foundation (2024M751309) and Young Talent Support Project of Guangzhou Association for Science and Technology.

Data availability

The single-cell H3K27ac and H3K4me1 CoBATCH datasets generated in this study have been deposited in the GEO database at accession code

GSE249551, also seen at https://www.ncbi.nlm.nih.gov/geo/query/acc. cgi?acc = GSE249551 (secure token: spkzksamtjuxjod). All custom code used in this study is available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was conducted in strict accordance with ethical guidelines and received approval from the Institutional Animal Care and Use Committee of Southern Medical University.

Competing interests

The authors declare no competing interests.

Received: 31 December 2024 Accepted: 21 April 2025 Published online: 08 May 2025

References

- Arnold SJ, Robertson EJ. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat Rev Mol Cell Biol. 2009;10(2):91–103.
- Rivera-Perez JA, Hadjantonakis AK. The Dynamics of Morphogenesis in the Early Mouse Embryo. Cold Spring Harb Perspect Biol. 2014;7(11):a015867.
- Peng G, Suo S, Chen J, Chen W, Liu C, Yu F, Wang R, Chen S, Sun N, Cui G, et al. Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. Dev Cell. 2020;55(6):802–4.
- Takaoka K, Hamada H. Cell fate decisions and axis determination in the early mouse embryo. Development. 2012;139(1):3–14.
- 5. Tam PP, Behringer RR. Mouse gastrulation: the formation of a mammalian body plan. Mech Dev. 1997;68(1–2):3–25.
- Peng G, Suo S, Cui G, Yu F, Wang R, Chen J, Chen S, Liu Z, Chen G, Qian Y, et al. Molecular architecture of lineage allocation and tissue organization in early mouse embryo. Nature. 2019;572(7770):528–32.
- Lawson KA, Meneses JJ, Pedersen RA. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. Development. 1991;113(3):891–911.
- Chan MM, Smith ZD, Grosswendt S, Kretzmer H, Norman TM, Adamson B, Jost M, Quinn JJ, Yang D, Jones MG, et al. Molecular recording of mammalian embryogenesis. Nature. 2019;570(7759):77–82.
- Bardot ES, Hadjantonakis AK. Mouse gastrulation: Coordination of tissue patterning, specification and diversification of cell fate. Mech Dev. 2020;163:103617.
- Atlasi Y, Stunnenberg HG. The interplay of epigenetic marks during stem cell differentiation and development. Nat Rev Genet. 2017;18(11):643–58.
- Xiang Y, Zhang Y, Xu Q, Zhou C, Liu B, Du Z, Zhang K, Zhang B, Wang X, Gayen S, et al. Epigenomic analysis of gastrulation identifies a unique chromatin state for primed pluripotency. Nat Genet. 2020;52(1):95–105.
- Yang X, Hu B, Hou Y, Qiao Y, Wang R, Chen Y, Qian Y, Feng S, Chen J, Liu C, et al. Silencing of developmental genes by H3K27me3 and DNA methylation reflects the discrepant plasticity of embryonic and extraembryonic lineages. Cell Res. 2018;28(5):593–6.
- Yang X, Hu B, Liao J, Qiao Y, Chen Y, Qian Y, Feng S, Yu F, Dong J, Hou Y, et al. Distinct enhancer signatures in the mouse gastrula delineate progressive cell fate continuum during embryo development. Cell Res. 2019;29(11):911–26.
- Grosswendt S, Kretzmer H, Smith ZD, Kumar AS, Hetzel S, Wittler L, Klages S, Timmermann B, Mukherji S, Meissner A. Epigenetic regulator function through mouse gastrulation. Nature. 2020;584(7819):102–8.
- Mittnenzweig M, Mayshar Y, Cheng S, Ben-Yair R, Hadas R, Rais Y, Chomsky E, Reines N, Uzonyi A, Lumerman L, et al. A single-embryo, single-cell time-resolved model for mouse gastrulation. Cell. 2021;184(11):2825-2842 e2822.
- Cheng S, Pei Y, He L, Peng G, Reinius B, Tam PPL, Jing N, Deng Q. Single-Cell RNA-Seq Reveals Cellular Heterogeneity of Pluripotency Transition

and X Chromosome Dynamics during Early Mouse Development. Cell Rep. 2019;26(10):2593-2607 e2593.

- Ibarra-Soria X, Jawaid W, Pijuan-Sala B, Ladopoulos V, Scialdone A, Jorg DJ, Tyser RCV, Calero-Nieto FJ, Mulas C, Nichols J, et al. Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. Nat Cell Biol. 2018;20(2):127–34.
- Mohammed H, Hernando-Herraez I, Savino A, Scialdone A, Macaulay I, Mulas C, Chandra T, Voet T, Dean W, Nichols J, et al. Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. 2017;20(5):1215–28.
- Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, Mulas C, Ibarra-Soria X, Tyser RCV, Ho DLL, et al. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature. 2019;566(7745):490–5.
- Scialdone A, Tanaka Y, Jawaid W, Moignard V, Wilson NK, Macaulay IC, Marioni JC, Gottgens B. Resolving early mesoderm diversification through single-cell expression profiling. Nature. 2016;535(7611):289–93.
- Argelaguet R, Clark SJ, Mohammed H, Stapel LC, Krueger C, Kapourani CA, Imaz-Rosshandler I, Lohoff T, Xiang Y, Hanna CW, et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. Nature. 2019;576(7787):487–91.
- 22. Wang Q, Xiong H, Ai S, Yu X, Liu Y, Zhang J, He A. CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. Mol Cell. 2019;76(1):206-216 e207.
- Downs KM, Davies T. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. Development. 1993;118(4):1255–66.
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010;107(50):21931–6.
- Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307–19.
- Liu M, Yue Y, Chen X, Xian K, Dong C, Shi M, Xiong H, Tian K, Li Y, Zhang QC, et al. Genome-coverage single-cell histone modifications for embryo lineage tracing. Nature. 2025;640(8059):828–39.
- Zhu C, Zhang Y, Li YE, Lucero J, Behrens MM, Ren B. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nat Methods. 2021;18(3):283–92.
- Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012;9(3):215–6.
- Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-3587 e3529.
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28(5):495–501.
- Hong CP, Choe MK, Roh TY. Characterization of Chromatin Structure-associated Histone Modifications in Breast Cancer Cells. Genomics Inform. 2012;10(3):145–52.
- Koenecke N, Johnston J, He Q, Meier S, Zeitlinger J. Drosophila poised enhancers are generated during tissue patterning with the help of repression. Genome Res. 2017;27(1):64–74.
- Zhang X, Yalcin S, Lee DF, Yeh TY, Lee SM, Su J, Mungamuri SK, Rimmele P, Kennedy M, Sellers R, et al. FOXO1 is an essential regulator of pluripotency in human embryonic stem cells. Nat Cell Biol. 2011;13(9):1092–9.
- Mukherjee S, French DL, Gadue P. Loss of TBX3 enhances pancreatic progenitor generation from human pluripotent stem cells. Stem Cell Reports. 2021;16(11):2617–27.
- Heslop JA, Pournasr B, Liu JT, Duncan SA. GATA6 defines endoderm fate by controlling chromatin accessibility during differentiation of humaninduced pluripotent stem cells. Cell Rep. 2021;35(7):109145.
- DeLaForest A, Nagaoka M, Si-Tayeb K, Noto FK, Konopka G, Battle MA, Duncan SA. HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. Development. 2011;138(19):4143–53.
- Mankoo BS, Skuntz S, Harrigan I, Grigorieva E, Candia A, Wright CV, Arnheiter H, Pachnis V. The concerted action of Meox homeobox genes is required upstream of genetic pathways essential for the

- Chan SS, Shi X, Toyama A, Arpke RW, Dandapat A, Iacovino M, Kang J, Le G, Hagen HR, Garry DJ, et al. Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. Cell Stem Cell. 2013;12(5):587–601.
- Tsuchihashi T, Maeda J, Shin CH, Ivey KN, Black BL, Olson EN, Yamagishi H, Srivastava D. Hand2 function in second heart field progenitors is essential for cardiogenesis. Dev Biol. 2011;351(1):62–9.
- Zhu Q, Song L, Peng G, Sun N, Chen J, Zhang T, Sheng N, Tang W, Qian C, Qiao Y, et al. The transcription factor Pou3f1 promotes neural fate commitment via activation of neural lineage genes and inhibition of external signaling pathways. Elife. 2014;3:e02224.
- Vernay B, Koch M, Vaccarino F, Briscoe J, Simeone A, Kageyama R, Ang SL. Otx2 regulates subtype specification and neurogenesis in the midbrain. J Neurosci. 2005;25(19):4856–67.
- 42. Pevny LH, Sockanathan S, Placzek M, Lovell-Badge R. A role for SOX1 in neural determination. Development. 1998;125(10):1967–78.
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888-1902 e1821.
- 44. Rada-Iglesias A. Is H3K4me1 at enhancers correlative or causative? Nat Genet. 2018;50(1):4–5.
- Spicuglia S, Vanhille L. Chromatin signatures of active enhancers. Nucleus. 2012;3(2):126–31.
- DeTomaso D, Jones MG, Subramaniam M, Ashuach T, Ye CJ, Yosef N. Functional interpretation of single cell similarity maps. Nat Commun. 2019;10(1):4376.
- 47. Zwaka TP. Unraveling the score of the enhancer symphony. Proc Natl Acad Sci U S A. 2010;107(50):21240–1.
- Presnell JS, Schnitzler CE, Browne WE. KLF/SP Transcription Factor Family Evolution: Expansion, Diversification, and Innovation in Eukaryotes. Genome Biol Evol. 2015;7(8):2289–309.
- Lin Q, Schwarz J, Bucana C, Olson EN. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. Science. 1997;276(5317):1404–7.
- Wilm B, James RG, Schultheiss TM, Hogan BL. The forkhead genes, Foxc1 and Foxc2, regulate paraxial versus intermediate mesoderm cell fate. Dev Biol. 2004;271(1):176–89.
- George RM, Firulli AB. Hand Factors in Cardiac Development. Anat Rec (Hoboken). 2019;302(1):101–7.
- 52. Chen ZF, Behringer RR. twist is required in head mesenchyme for cranial neural tube morphogenesis. Genes Dev. 1995;9(6):686–99.
- Shim WJ, Sinniah E, Xu J, Vitrinel B, Alexanian M, Andreoletti G, Shen S, Sun Y, Balderson B, Boix C, et al. Conserved Epigenetic Regulatory Logic Infers Genes Governing Cell Identity. Cell Syst. 2020;11(6):625-639 e613.
- Aramaki S, Hayashi K, Kurimoto K, Ohta H, Yabuta Y, Iwanari H, Mochizuki Y, Hamakubo T, Kato Y, Shirahige K, et al. A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. Dev Cell. 2013;27(5):516–29.
- Schule KM, Weckerle J, Probst S, Wehmeyer AE, Zissel L, Schroder CM, Tekman M, Kim GJ, Schlagl IM, Arnold SSJ, et al. Eomes restricts Brachyury functions at the onset of mouse gastrulation. Dev Cell. 2023;58(18):1627-1642 e1627.
- Sasai N, Yakura R, Kamiya D, Nakazawa Y, Sasai Y. Ectodermal factor restricts mesoderm differentiation by inhibiting p53. Cell. 2008;133(5):878–90.
- Wang A, Yue F, Li Y, Xie R, Harper T, Patel NA, Muth K, Palmer J, Qiu Y, Wang J, et al. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. Cell Stem Cell. 2015;16(4):386–99.
- Martin EW, Rodriguez YBA, Reggiardo RE, Worthington AK, Mattingly CS, Poscablo DM, Krietsch J, McManus MT, Carpenter S, Kim DH, et al. Dynamics of Chromatin Accessibility During Hematopoietic Stem Cell Differentiation Into Progressively Lineage-Committed Progeny. Stem Cells. 2023;41(5):520–39.
- Kojima Y, Kaufman-Francis K, Studdert JB, Steiner KA, Power MD, Loebel DA, Jones V, Hor A, de Alencastro G, Logan GJ, et al. The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. Cell Stem Cell. 2014;14(1):107–20.

- Osterwalder M, Speziale D, Shoukry M, Mohan R, Ivanek R, Kohler M, Beisel C, Wen X, Scales SJ, Christoffels VM, et al. HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. Dev Cell. 2014;31(3):345–57.
- Alaiz Noya M, Berti F, Dietrich S. Comprehensive expression analysis for the core cell cycle regulators in the chicken embryo reveals novel tissue-specific synexpression groups and similarities and differences with expression in mouse, frog and zebrafish. J Anat. 2022;241(1):42–66.
- Pauklin S, Vallier L. The cell-cycle state of stem cells determines cell fate propensity. Cell. 2013;155(1):135–47.
- Pauklin S, Madrigal P, Bertero A, Vallier L. Initiation of stem cell differentiation involves cell cycle-dependent regulation of developmental genes by Cyclin D. Genes Dev. 2016;30(4):421–33.
- Singh AM, Sun Y, Li L, Zhang W, Wu T, Zhao S, Qin Z, Dalton S. Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. Stem Cell Reports. 2015;5(3):323–36.
- Singh AM, Chappell J, Trost R, Lin L, Wang T, Tang J, Matlock BK, Weller KP, Wu H, Zhao S, et al. Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. Stem Cell Reports. 2013;1(6):532–44.
- Yiangou L, Grandy RA, Osnato A, Ortmann D, Sinha S, Vallier L. Cell cycle regulators control mesoderm specification in human pluripotent stem cells. J Biol Chem. 2019;294(47):17903–14.
- Matsuoka S, Thompson JS, Edwards MC, Bartletta JM, Grundy P, Kalikin LM, Harper JW, Elledge SJ, Feinberg AP. Imprinting of the gene encoding a human cyclin-dependent kinase inhibitor, p57KIP2, on chromosome 11p15. Proc Natl Acad Sci U S A. 1996;93(7):3026–30.
- Xiong H, Luo Y, Wang Q, Yu X, He A. Single-cell joint detection of chromatin occupancy and transcriptome enables higher-dimensional epigenomic reconstructions. Nat Methods. 2021;18(6):652–60.
- 69. Chen X, Shen Y, Draper W, Buenrostro JD, Litzenburger U, Cho SW, Satpathy AT, Carter AC, Ghosh RP, East-Seletsky A, et al. ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. Nat Methods. 2016;13(12):1013–20.
- Picelli S, Bjorklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 2014;24(12):2033–40.
- Amini S, Pushkarev D, Christiansen L, Kostem E, Royce T, Turk C, Pignatelli N, Adey A, Kitzman JO, Vijayan K, et al. Haplotype-resolved wholegenome sequencing by contiguity-preserving transposition and combinatorial indexing. Nat Genet. 2014;46(12):1343–9.
- Ai S, Xiong H, Li CC, Luo Y, Shi Q, Liu Y, Yu X, Li C, He A. Profiling chromatin states using single-cell itChIP-seq. Nat Cell Biol. 2019;21(9):1164–72.
- Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. J Comput Biol. 2017;24(11):1138–43.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
- Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. Nat Methods. 2021;18(11):1333–41.
- Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15.
- Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature. 2019;566(7745):496–502.
- Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, et al. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. Mol Cell. 2018;71(5):858-871 e858.
- Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics. 2016;32(4):587–9.
- Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods. 2017;14(10):975–8.

- Kartha VK, Duarte FM, Hu Y, Ma S, Chew JG, Lareau CA, Earl A, Burkett ZD, Kohlway AS, Lebofsky R, et al. Functional inference of gene regulation using single-cell multi-omics. Cell Genom. 2022;2(9):100166.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
- 85. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.
- Tanigawa Y, Dyer ES, Bejerano G. WhichTF is functionally important in your open chromatin data? PLoS Comput Biol. 2022;18(8):e1010378.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.