ARTICLE

Genome-wide profiling of histone modifications and transcription factor binding at single-cell resolution by DeChIC-seq

Zhifei Shi[1,2,3,6], Xiyang Chen[1,6], Yijia Yang[1], Ang Wu[1], Heng Wang[1], Kai Chen[1], Chong Li[1], Lina Zou[4], Zhipeng Qu[4], Yuyan Zhao[1], Wenjing Gan[1], Shuhui Li[1], Jiayu Chen[1], Wenqiang Liu[1], Jiejun Shi[2], Hong Wang[1,3], Jia-min Zhang[2,3]✉, Chenfei Wang[2,3,5]✉, Shaorong Gao[1,3]✉ and Xiaoyu Liu[1,3]✉

Mapping of protein-DNA interactions at single-cell resolution remains a central challenge in epigenomics, particularly for transcription factors (TFs), whose sparse binding limits reliable detection. Here, we establish DeChIC-seq (DNA Deaminase-based Chromatin Immuno-Conversion sequencing), a conversion-based strategy that uses a protein A-DddA$_{tox}$ fusion to directly record protein-DNA interactions by inducing localized C-to-U conversions near antibody-bound chromatin. Retaining genome-wide background sequence information without immunoprecipitation, DeChIC-seq enables profiling of histone modifications and sensitive detection of TF binding. Integration with single-cell whole-genome amplification extends DeChIC-seq to single-cell applications (scDeChIC-seq), enabling chromatin profiling of individual cells. Applied to mouse embryogenesis, scDeChIC-seq resolves lineage-specific chromatin states through profiling of H3K4me3, CTCF, and RAD21 and sensitively detects TF binding, including that of NR5A2, TFAP2C, and KLF5, from extremely limited blastomere inputs. This underscores its strong potential for detecting TF-binding sites in scarce biological samples. DeChIC-seq establishes a conversion-based framework for chromatin profiling that enables mechanistic dissection of TF-driven gene regulation across rare cells, developmental systems, and disease contexts.

Cell Research (2026) 0:1–18; https://doi.org/10.1038/s41422-026-01275-z

INTRODUCTION

Precise gene regulation is orchestrated through the coordinated interplay of chromatin modifications and transcription factor (TF) binding at specific genomic loci.[1–3] Genome-wide, single-cell mapping of histone modifications and TF binding is crucial for understanding transcriptional regulation during development and disease progression.[4] This underscores the need for technologies that can sensitively profile chromatin features at single-cell resolution, particularly in early embryonic development, when cell numbers are extremely limited.

A variety of next-generation sequencing technologies have been developed and widely adopted, with chromatin immunoprecipitation sequencing (ChIP-seq), CUT&RUN, and CUT&Tag being particularly prominent.[5–7] These methods have been adapted for single-cell applications, especially histone modification profiling, including scChIP-seq, scDrop-ChIP, scCUT&RUN, and transposase-based techniques like scCUT&Tag, CoBATCH, nano-CUT&Tag, and TACIT.[8–15]

However, these approaches typically require pooling large numbers of cells and rely on immunoprecipitation-based enrichment or chromatin fragmentation.[4,16] This reliance on fragment enrichment poses particular challenges for TF profiling, especially for low-abundance factors, as efficiency and sensitivity are inherently limited at the single-cell level. In addition, the use of bioinformatics methods to assign single-cell data can obscure the original fragment abundance, thereby reducing data resolution and limiting accuracy.[17]

In parallel, enzyme-mediated DNA-labeling approaches such as DamID[18,19] have emerged and been adapted for single-cell applications, including sc-DamID and EpiDamID.[20,21] These methods depend on the methylation of adenine at GATC sequences, which are sparsely distributed across the genome, ultimately limiting resolution. Strategies like SpDamID and 3D-seq,[22,23] which involve the fusion of TFs with DNA-modifying enzymes, present challenges in controlling enzymatic activity and require the construction of specific cell lines, both of which restrict their application. Single-molecule long-read sequencing methods

[1]Shanghai Key Laboratory of Maternal Fetal Medicine, Shanghai Institute of Maternal-Fetal Medicine and Gynecologic Oncology, Clinical and Translation Research Center, Shanghai First Maternity and Infant Hospital, School of Life Science and Technology, Tongji University, Shanghai, China. [2]Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration of the Ministry of Education, Department of Orthopedics, Tongji Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China. [3]Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai, China. [4]Vazyme Biotech Co., Ltd., Nanjing, Jiangsu, China. [5]Sycamore Research Institute of Life Sciences, Shanghai, China. [6]These authors contributed equally: Zhifei Shi, Xiyang Chen. ✉email: zhangjiamin@tongji.edu.cn; 08chenfeiwang@tongji.edu.cn; gaoshaorong@tongji.edu.cn; liuxiaoyu@tongji.edu.cn

Received: 10 May 2025 Accepted: 4 June 2026
Published online: 13 July 2026

such as DiMeLo-seq,[24] nanoHiMe-seq,[25] and BIND & MODIFY[26] have recently been reported. However, these methods require substantial amounts of input DNA for sequencing, as methylation signals must be detected directly without amplification, limiting their applicability in low-input or single-cell contexts.

Significant progress has been made in optimizing single-cell methods for histone modification profiling, but improvements in resolution and accuracy are still needed.[16] More critically, robust and direct detection of TF binding events starting from a single cell remains a technical challenge.[13] These limitations hinder our ability to fully understand transcriptional regulation in contexts such as early embryonic development, where cell numbers are limited, and chromatin dynamics are highly heterogeneous.

To address these limitations, we developed a single-cell DNA-deaminase-based chromatin immunoconversion sequencing (scDeChIC-seq) method, which provides an efficient solution, particularly for TF binding, at single-cell resolution. Using scDeChIC-seq, we performed genome-wide profiling of H3K4me3, CTCF, and RAD21 in mouse pre-implantation embryos during the first lineage differentiation. This approach enabled us to obtain detailed insight into the chromatin landscape and transcriptional regulation in more research areas without the limitations of cell number.

RESULTS

Rationale and optimization of DeChIC-seq for profiling histone modifications and protein-DNA interactions

By leveraging a fusion of protein A with the double-stranded DNA (dsDNA) deaminase $DddA_{tox}$ (hereafter referred to as pA-DddA),[22,27] we developed DeChIC-seq for the profiling of histone modifications and chromatin-associated proteins (CAPs), including TFs, across the genome. In DeChIC-seq, the fusion protein pA-DddA is anchored to specific chromatin regions by an antibody that targets the histone modification or CAP of interest. DddA catalyzes the deamination of cytosine (C) to uracil (U) in local DNA, creating a sequence-specific conversion event that marks the protein-bound regions. During PCR amplification, U is converted to T, enabling the detection of these sites through sequencing-based analysis of DNA conversion patterns (Fig. 1a; Supplementary information, Fig. S1a). Furthermore, by integrating DeChIC-seq with single-cell whole-genome amplification (WGA) technologies, scDeChIC-seq enables the mapping of histone modifications and protein-DNA interactions at the single-cell level, facilitating accurate analysis of epigenetic dynamics and cellular heterogeneity during development (Supplementary information, Fig. S1a).

A critical aspect of DeChIC-seq development was achieving reversible control over the enzymatic activity of pA-DddA. We first confirmed that the pA-DddA fusion retains its deaminase activity (Supplementary information, Fig. S1b).[27] We then verified that pA-DddA preserves the intrinsic TC sequence preference of DddA, achieving conversion rates approaching 100% at TC sites (Supplementary information, Fig. S1c). Further optimization revealed that the deaminase activity of pA-DddA could be effectively repressed in a pH 8.0 Tris-HCl buffer and reactivated in a pH 6.4 MES buffer, enabling precise control of deaminase activity (Supplementary information, Fig. S1d). This prevents nonspecific activity during chromatin binding while ensuring efficient conversion during the deamination process. By contrast, other dsDNA deaminases tested under identical conditions, including Ddd_Ss, Ddd_Fa,[28] and $SsdA_{tox}$,[29] did not show fully repressible activity, underscoring the distinctive controllability of pA-DddA (Supplementary information, Fig. S1e).

To confirm that our method could be used to detect histone modifications and CAP binding sites, we profiled H3K4me3 and CTCF using DeChIC-seq in mouse embryonic fibroblast (MEF) cells, achieving high deamination efficiency at ChIP-seq-defined peak centers (Fig. 1b), with a consistent TC sequence preference (Supplementary information, Fig. S1f). We further optimized key reaction conditions for pA-DddA binding and deamination, including pH, ion concentration, surfactant level, and enzyme concentration, to enhance reaction efficiency and improve the signal-to-noise ratio (Supplementary information, Fig. S1g, h). In addition, we found that a rigid linker between protein A and $DddA_{tox}$ improved performance compared with a flexible linker (Supplementary information, Fig. S1i). We confirmed that DeChIC-seq signals showed low correlation with ATAC-seq peaks and that DeChIC-seq peaks exhibited limited overlap with open chromatin regions, indicating that DeChIC-seq does not exhibit a pronounced bias toward such regions (Supplementary information, Fig. S1j, k). To further exclude potential background interference, we performed DeChIC-seq experiments using IgG and no-antibody controls to assess the potential for nonspecific antibody binding and the occurrence of stochastic deamination events during the incubation steps. At transcription start sites (TSS) regions, the average conversion rate in IgG DeChIC-seq was below 0.005, markedly lower than that observed with H3K4me3 and CTCF antibodies. Notably, the conversion rate in the no-antibody control was even lower, less than 0.002. These findings provide independent evidence for the specificity and overall reliability of DeChIC-seq signals (Supplementary information, Fig. S1l).

Mapping the distribution of histone modifications by DeChIC-seq

To further assess the ability of DeChIC-seq to reliably identify the locations of H3K4me3, we performed peak calling using H3K4me3 DeChIC-seq data from MEF cells, with rabbit IgG antibody serving as a negative control (Supplementary information, Table S1). We found that DeChIC-seq and ChIP-seq shared 26,092 overlapping peaks, accounting for 88.02% of all peaks identified by DeChIC-seq and 79.2% of those identified by ChIP-seq (Fig. 1c, d; Supplementary information, Fig. S2a). Accordingly, a heatmap also showed high similarity in the distribution and intensity of H3K4me3 between the DeChIC-seq and ChIP-seq data (Fig. 1e). The conversion signals of individual peaks showed strong correlations between replicates, and the conversion rates of DeChIC-seq peaks were well correlated with the ChIP-seq signals (Fig. 1d; Supplementary information, Fig. S2b). The distribution of peak features indicated that most DeChIC-seq H3K4me3 peaks were located near promoter regions (Supplementary information, Fig. S2c).[30,31] Notably, track views of H3K4me3 target regions exhibited patterns consistent with the ChIP-seq data and showed a favorable signal-to-noise ratio, highlighting the sensitivity and robustness of DeChIC-seq (Fig. 1f). These analyses suggest that DeChIC-seq signals can be effectively enriched near regions of H3K4me3 modification. Moreover, the strong correlation with transcriptomic data further supports the reliability of this method (Supplementary information, Fig. S2d).[32]

To further assess the versatility of DeChIC-seq, we extended its application to the mapping of additional histone modifications, including H3K4me1, H3K9ac, and H3K27ac, in MEF cells. As with H3K4me3, we observed that DeChIC-seq showed high deamination efficiency at the corresponding ChIP-seq peaks, with strong correlation and overlap with ChIP-seq data, particularly ULI-NChIP-seq data (Fig. 1f; Supplementary information, Fig. S2e–g). These results indicate that DeChIC-seq can be used to profile a range of histone modifications across the genome.

We next asked whether DeChIC-seq could be applied to repressive histone modifications such as H3K27me3 and H3K9me3. Elevated conversion signals were detectable at ChIP-defined peak regions, but with reduced peak-centered contrast relative to active marks, likely reflecting the compact chromatin architecture of heterochromatin (Supplementary information, Fig. S2h). Notably, depletion of the H3K9me3 writers Suv39h1 and Suv39h2[33] led to a global reduction in conversion levels at

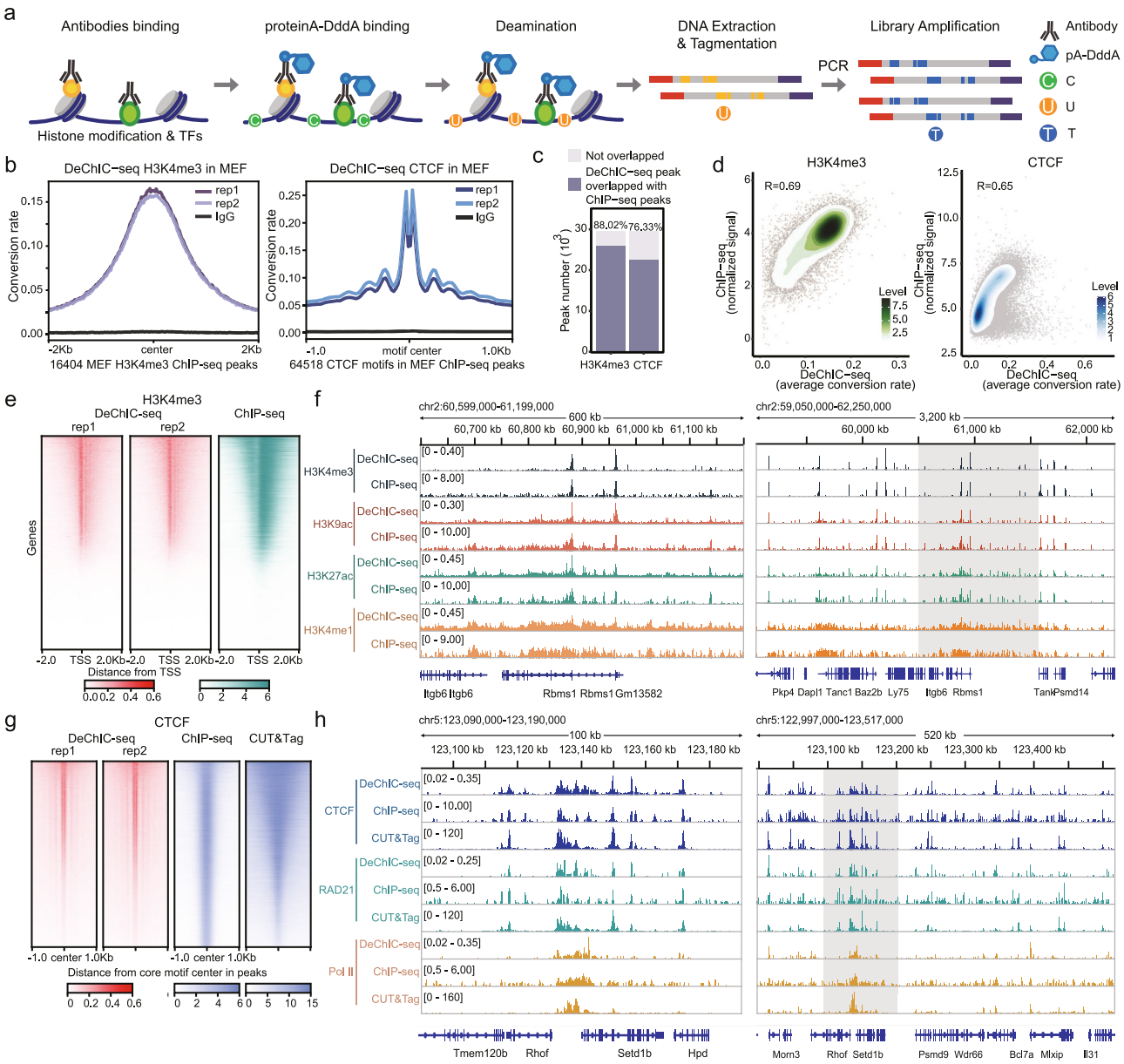


Fig. 1 DeChIC-seq enables high-resolution, genome-wide profiling of histone modifications and TFs. a Schematic representation of the DeChIC-seq workflow. pA-DddA is anchored in situ to the target histone modification or TF via an antibody, catalyzing the deamination of C to U. Genomic DNA is then extracted and fragmented by Tn5 transposase. During library amplification, U is converted to T. **b** Average conversion rates of H3K4me3 DeChIC-seq at ChIP-seq-defined H3K4me3 peaks (left) and CTCF DeChIC-seq at ChIP-seq-defined CTCF peaks (aligned to the CTCF core motif) (right), with IgG as a negative control. Different curves represent biological replicates. **c** Bar plots showing the numbers of H3K4me3 and CTCF peaks identified from DeChIC-seq data, with the percentages of peaks that overlapped with ChIP-seq-defined peaks. **d** Correlation of H3K4me3 (left) and CTCF (right) signals between DeChIC-seq and ChIP-seq. Each dot represents a single ChIP-seq peak. **e** Heatmap of H3K4me3 signals from DeChIC-seq and ChIP-seq at transcription start sites (TSSs) of individual genes (rows). Profiles are centered on the TSS. **f** Representative genome browser tracks of H3K4me3, H3K9ac, H3K27ac, and H3K4me1 from DeChIC-seq and ChIP-seq (left), with corresponding zoomed-in regions (right; shaded boxes indicate the track view shown in the left panel, and only a subset of representative genes is shown). DeChIC-seq signals are presented as average conversion rates, and ChIP-seq signals are normalized as $\log_2$fold change (negative values are hidden). **g** Heatmaps of CTCF signals from DeChIC-seq, ChIP-seq, and CUT&Tag at the CTCF core motif within ChIP-seq-defined peaks. **h** Representative genome browser tracks of CTCF, RAD21, and Pol II from DeChIC-seq, ChIP-seq, and CUT&Tag (left), with corresponding zoomed-in regions (right, shaded boxes indicate the track view shown in the left panel). DeChIC-seq signals are presented as average conversion rates, ChIP-seq signals are normalized as $\log_2$(RPKM of TF/input) (negative values are hidden), and CUT&Tag signals are normalized as RPKM.

H3K9me3-enriched regions, indicating that conversion signals remain mark dependent (Supplementary information, Fig. S2i). Together, these results show that DeChIC-seq robustly profiles active histone modifications and retains mark-dependent signal behavior in repressive chromatin, despite reduced peak-centered contrast.

Mapping protein-DNA interactions by DeChIC-seq

Accurate mapping of protein-DNA interactions is essential for understanding gene regulation and chromatin organization.[34,35] The favorable performance of DeChIC-seq in profiling active histone modification marks provided a basis for its subsequent use in profiling TF binding. To assess the ability of DeChIC-seq to profile protein binding sites, we performed peak calling on CTCF DeChIC-seq data from 5×10^3 MEF cells, identifying a total of 29,638 peaks. The DeChIC-seq results exhibited substantial overlap with those of ChIP-seq and CUT&Tag: 76.33% of the DeChIC-seq peaks coincided with CTCF ChIP-seq peaks (Fig. 1c) and 74.46% with CUT&Tag peaks (Supplementary information, Fig. S3a). Moreover, the peaks detected by both DeChIC-seq and ChIP-seq tended to have stronger signal intensities than the non-overlapping peaks, suggesting that DeChIC-seq effectively captures high-confidence binding sites (Supplementary information, Fig. S3b). DeChIC-seq signals were highly consistent between replicates, as reflected by the strong correlation between peak intensities. The intensities of DeChIC-seq peaks also correlated well with ChIP-seq signals (Fig. 1d; Supplementary information, Fig. S3c). A heatmap centered around the CTCF motifs within ChIP-seq peaks revealed a similar pattern of signal distribution and intensity (Fig. 1g). Notably, DeChIC-seq conversion rates around the CTCF motif demonstrated a phased nucleosome pattern within ±1 kb, reflecting the chromatin organization around CTCF binding sites[36] and indirectly highlighting the high resolution of DeChIC-seq (Fig. 1g). Genome browser track views of specific regions also showed that DeChIC-seq detects protein-DNA interactions with a signal pattern comparable to those of ChIP-seq and CUT&Tag (Fig. 1h).

To explore the broader applicability of DeChIC-seq, we extended the method to other TFs and performed DeChIC-seq on RAD21 and Pol II in MEF cells. Both exhibited high conversion rates around their respective ChIP-seq peak centers, suggesting that DeChIC-seq can capture their genomic localization with enriched signals (Supplementary information, Fig. S3d). Furthermore, track views and peak analysis revealed a high degree of similarity between the DeChIC-seq data and the ChIP-seq and CUT&Tag data (Fig. 1h; Supplementary information, Fig. S3e, f). These results indicate that DeChIC-seq can be used for genome-wide mapping of various protein-DNA interactions with good consistency and reliability.

DeChIC-seq enables proportional and semi-quantitative detection of chromatin-associated binding signals

Current methods such as ChIP-seq enrich chromatin fragments bound by the target protein, often losing valuable non-targeted genomic background data.[37,38] Such enrichment-based strategies can alter the effective dynamic range of the signal, and normalization procedures may further obscure differences in signal intensity between samples, thereby limiting the quantitative potential.[39] By recording localized C-to-U conversions at antibody-bound chromatin while preserving genome-wide background information, DeChIC-seq is expected to generate signals that scale proportionally with local protein-associated chromatin features, thereby enabling a faithful readout with semi-quantitative potential.

As a technical validation, we first tested whether the proportion of U could be stably detected during PCR amplification using a synthesized *Klf2* sequence enriched with H3K4me3 from MEF cells[40] in which different numbers of cytosines were replaced with uracils. We found that the proportion of U in the DNA template remained stable before and after PCR, indicating that conversion rates were maintained during library construction (Supplementary information, Fig. S4a).

To evaluate whether DeChIC-seq signals scale proportionally with chromatin-feature abundance, we performed H3K4me3 profiling of mixed populations of R1 and MEF cells at defined ratios, then quantified the relationship between the measured signal intensity and the expected cell proportion. The same experimental design was also used with ChIP-seq and CUT&Tag for parallel benchmarking. H3K4me3 peaks identified in R1 cells by DeChIC-seq showed strong overlap with ChIP-seq-defined peaks (Supplementary information, Fig. S4b–d). Using integrated peak sets from the three methods, we defined MEF- and R1-specific H3K4me3 regions for downstream quantitative comparisons. DeChIC-seq signals at integrated cell type-specific peaks changed proportionally with the relative proportions of each cell type. As the proportion of R1 cells decreased and that of MEF cells increased, the conversion signal at R1-specific peaks gradually declined ($R^2 = 0.89$), and the signal at MEF-specific peaks increased accordingly ($R^2 = 0.94$) (Fig. 2a–c). This trend was consistently observed across biological replicates and remained robust when peaks were defined either by integrated consensus peak sets or by each platform independently (Supplementary information, Fig. S4e–g), indicating that DeChIC-seq can reliably capture relative signal changes across mixed populations.

By contrast, ChIP-seq signals at the same peaks showed limited variation across mixed samples, particularly at intermediate conditions (3:1 to 1:3), indicating reduced sensitivity in resolving differences in signal intensity between cell types (R1-specific peaks $R^2 = 0.70$; MEF-specific peaks $R^2 = 0.84$) (Fig. 2a–c; Supplementary information, Fig. S4e–g). CUT&Tag showed intermediate performance for R1-specific peaks ($R^2 = 0.80$) and performance comparable to that of DeChIC-seq for MEF-specific peaks ($R^2 = 0.96$) (Fig. 2a–c; Supplementary information, Fig. S4e–g). Heatmaps further illustrated the graded signal intensities of both DeChIC-seq and CUT&Tag across the mixture series, whereas ChIP-seq signals showed limited variation at intermediate ratios (Fig. 2c). Genome browser views of representative loci likewise showed progressive shifts in DeChIC-seq signal intensity that corresponded to the known cell-type composition (Fig. 2d). Collectively, these findings indicate that DeChIC-seq enables more proportional and reproducible semi-quantitative comparison of histone-modification signal intensities between distinct cell types or samples than conventional enrichment-based methods.

scDeChIC-seq enables single-cell detection of histone modifications and protein-DNA interactions

The ability to detect histone modifications and protein-DNA interactions at the single-cell level is critically important for unraveling cellular heterogeneity and facilitating the study of rare samples.[37,41] Unlike conventional methods that require fragment enrichment, which often results in significant signal loss, DeChIC-seq preserves genome-wide contextual information by directly marking protein-bound DNA through C-deamination. Building on this feature, we developed scDeChIC-seq, which can be performed with as few as one cell by integrating DeChIC-seq with single-cell whole-genome sequencing in an all-in-one-tube workflow.[42,43] This design enables high-resolution, genome-wide detection of protein-DNA interactions at the single-cell level (Fig. 3a).

To assess the applicability of scDeChIC-seq for profiling histone modifications and CAPs, we used it to map H3K4me3 modifications and CTCF-binding sites in MEF and R1 cells. For H3K4me3, conversion rates around peak centers and TSSs in individual cells showed consistent patterns and high correlations with bulk DeChIC-seq data, demonstrating the robustness and consistency of scDeChIC-seq (Fig. 3b; Supplementary information, Fig. S5a–d

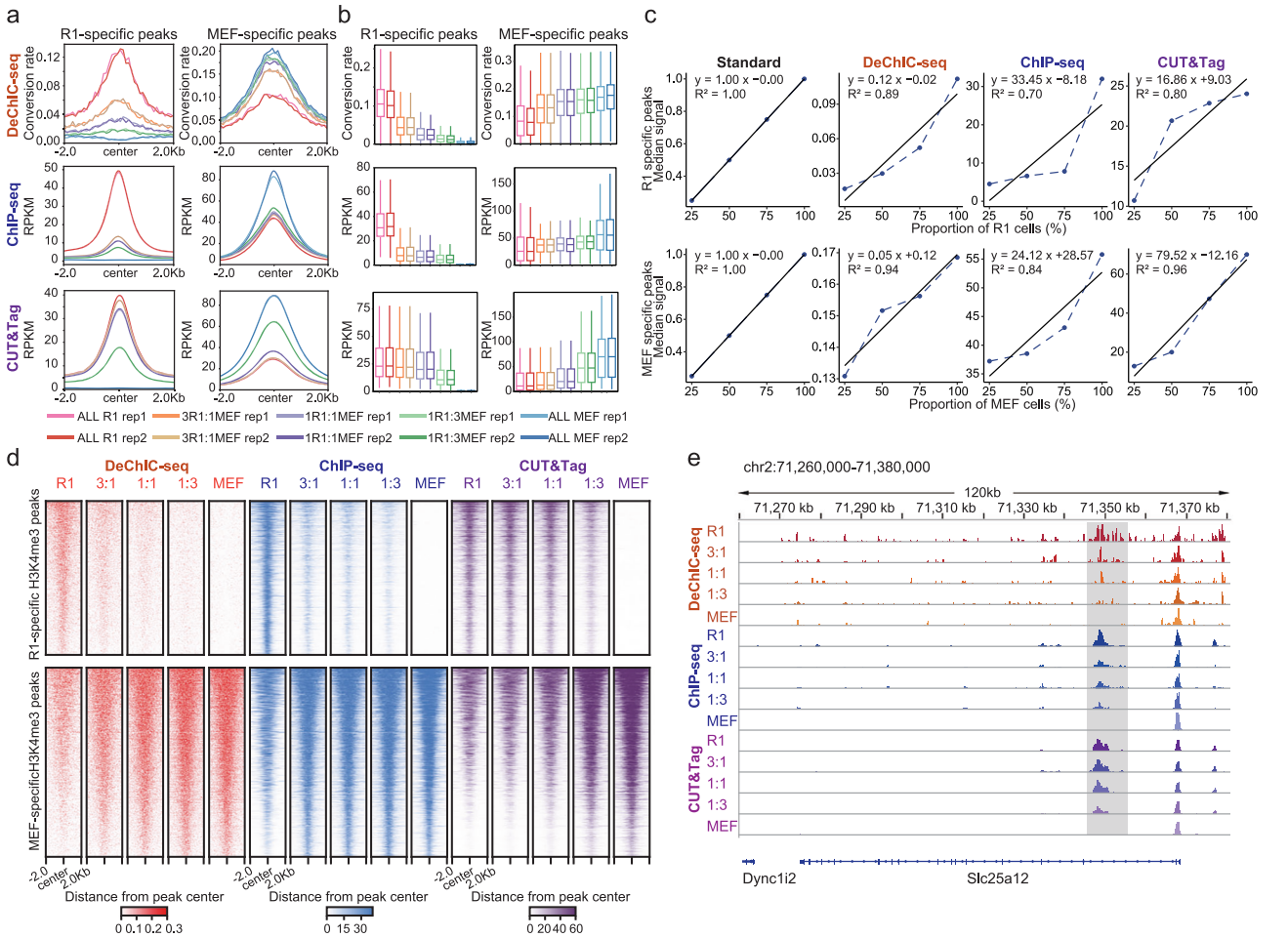
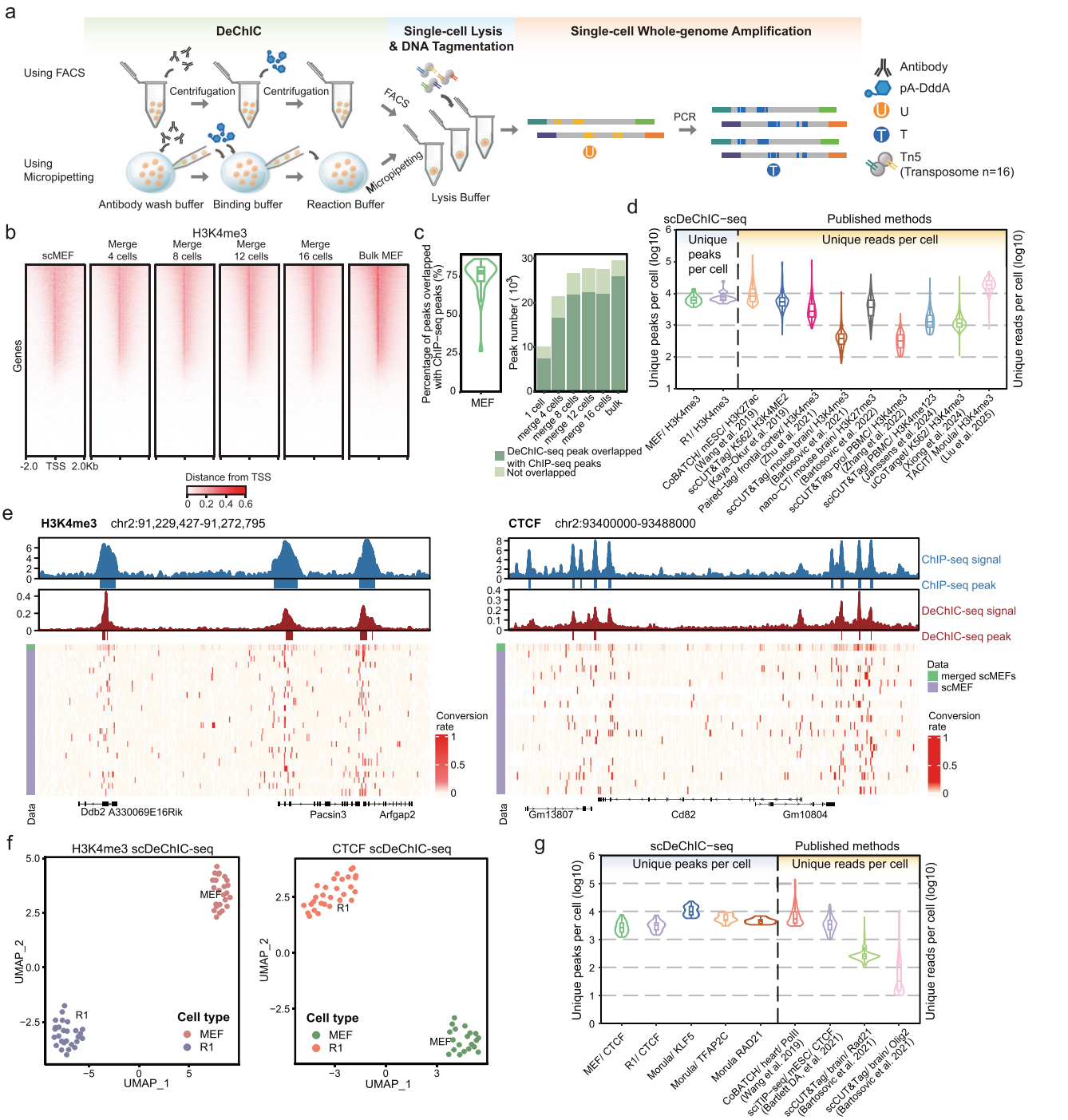


Fig. 2 DeChIC-seq demonstrates the potential for highly accurate profiling. a, **b** H3K4me3 DeChIC-seq (top), ChIP-seq (middle), and CUT&Tag (bottom) were performed on mixtures of R1 and MEF cells at various ratios. Line plots show the average conversion rates (for DeChIC-seq) and RPKM values (for ChIP-seq and CUT&Tag) at R1- and MEF-specific peaks across different groups (**a**). Boxplots show the signals within R1- and MEF-specific peaks across different groups, with colors representing different mixing ratios (**b**). **c** Linear relationship between cell-mixing ratio and assay signals at R1-specific (top) and MEF-specific (bottom) peaks. Median signals are plotted against the proportion of R1 cells (top) or MEF cells (bottom) for DeChIC-seq (average conversion rate), ChIP-seq (RPKM), and CUT&Tag (RPKM). Dashed blue lines indicate the observed medians, and solid black lines show linear regression fits, with corresponding equations and R^2 values. The standard shows the expected input relationship and serves as the reference line. **d** Heatmaps showing H3K4me3 signals from DeChIC-seq, ChIP-seq, and CUT&Tag of mixed cells at top R1-specific (top) and MEF-specific (bottom) peaks. The proportion indicates the R1-to-MEF ratio in the mixed cell sample. Profiles are centered on the center of the R1- or MEF-specific H3K4me3 peak. **e** Representative genome browser track view of H3K4me3 data from DeChIC-seq (top), ChIP-seq (middle), and CUT&Tag (bottom) performed with mixtures of R1 and MEF cells at various ratios.

and Table S1). Independent peak calling for individual cells led to the identification of approximately 10,000 unique H3K4me3 peaks per cell, a high proportion of which overlapped with ChIP-seq peaks, yielding higher per-cell peak numbers among the analyzed single-cell protein-DNA interaction datasets (Fig. 3c, d; Supplementary information, Fig. S5e, f).[7,11,13–15,44–47] By merging data from different numbers of cells, we found that 8–12 cells were sufficient to generate profiles comparable to ChIP-seq and bulk DeChIC-seq datasets (Fig. 3c; Supplementary information, Fig. S5g). Notably, scDeChIC-seq could detect H3K4me3 peaks at single loci within individual cells (Fig. 3e), supporting its applicability for assessment of histone modification patterns in heterogeneous samples. We subsequently generated a peak-by-cell matrix for the scDeChIC-seq datasets based on conversion rates. This matrix was used for latent semantic indexing (LSI)-based dimensionality reduction and uniform manifold approximation and projection (UMAP)-based visualization using Signac.[48] Dimensionality reduction and clustering of the H3K4me3 data revealed two main cell clusters, with all cells matching the identities manually annotated based on sample information (Fig. 3f).

For CTCF, scDeChIC-seq also produced distinct single-cell conversion signals around the CTCF motif within ChIP-seq peaks in MEF and R1 cells, showing comparable signal patterns and consistent correspondence with bulk DeChIC-seq data (Fig. 3e; Supplementary information, Fig. S5f, h, i). We used scDeChIC-seq to detect additional TF-binding sites and obtained more unique peaks per cell than were obtained with most previously reported single-cell methods (Fig. 3g).[11,15,49] Merging data from only 8–12 cells was sufficient to generate profiles comparable to bulk datasets (Supplementary information, Fig. S5i). scDeChIC-seq of CTCF also enabled clear separation of R1 and MEF into distinct clusters (Fig. 3f). Together, these results indicate that scDeChIC-seq is a feasible and robust approach for profiling histone modifications and protein-DNA interactions at the single-cell level and can be applied to rare or limited biological samples such as pre-implantation embryos.



scDeChIC-seq reveals single-cell H3K4me3 heterogeneity and remodeling during early embryonic lineage specification

Previous studies of chromatin modifications in pre-implantation embryos have provided important insights. However, systematic single-cell characterization of histone modification dynamics during the first lineage segregation remains limited.[50] scDeChIC-seq overcomes these limitations by enabling detection of H3K4me3 peaks in individual cells, making it ideal for studying H3K4me3 dynamics during early lineage differentiation. We applied scDeChIC-seq to mouse Morula, E4.0 inner cell mass (ICM), and trophectoderm (TE) cells and obtained single-cell H3K4me3 profiles across these developmental stages. The conversion rates at TSSs were highly consistent across individual and merged cell samples (Fig. 4a; Supplementary information, Fig. S6a). Approximately 10^4 unique peaks were independently captured for each single cell, and merging of single-cell data for each lineage led to the identification of ~40,000–50,000 peaks, a large fraction of which overlapped with ChIP-seq peaks (Fig. 4b; Supplementary information, Fig. S6b).

To assess whether scDeChIC-seq-derived H3K4me3 profiles provide sufficient resolution to distinguish different cell types, we performed dimensionality reduction and clustering analysis. This approach effectively separated the Morula, ICM, and TE populations, consistent with their predefined identities (Fig. 4c, d; Supplementary information, Fig. S6c). Further validation using the average conversion rate from H3K4me3 data in the promoter

Fig. 3 scDeChIC-seq enables high-quality profiling of genome-wide protein-DNA interactions using a single-cell input. a Schematic of the scDeChIC-seq workflow, which consists of three main modules: DeChIC, single-cell lysis & DNA tagmentation, and single-cell WGA. During the DeChIC step, different strategies were used during sample collection: for scDeChIC-seq of R1 and MEF cells, centrifugation was used for buffer exchange and FACS or mouth pipetting for single cell isolation; for pre-implantation embryo samples, both steps were performed by mouth pipetting. For single-cell genomic DNA tagmentation, Tn5 transposomes ($n = 16$) were used to minimize DNA loss. Finally, single-cell samples underwent WGA followed by sequencing. **b** Heatmaps showing H3K4me3 signals around the TSS from individual and merged MEF scDeChIC-seq data, with bulk H3K4me3 DeChIC-seq data included as a reference. **c** Violin plot showing the percentage of peaks in each MEF scDeChIC-seq dataset that overlapped with ChIP-seq-defined peaks (left). Bar plot showing the number of H3K4me3 peaks identified from individual and merged MEF scDeChIC-seq data, as well as their overlap with ChIP-seq peaks (right). **d** Comparison of different methods for single-cell protein-DNA interaction profiling, showing unique peaks per cell for scDeChIC-seq and unique reads per cell for other published methods. The cell types and histone modifications profiled differ across the datasets generated by each method. To distinguish these conditions, for published methods, the *x*-axis is labeled with the method/cell or tissue type/histone modification. Note that even for the same histone modification, different antibodies may have been used. **e** Heatmaps showing tracks for H3K4me3 (left) and CTCF (right) from scDeChIC-seq, bulk DeChIC-seq, and ChIP-seq data. Each row represents a data type or an individual scDeChIC-seq single cell. **f** UMAP embedding of H3K4me3 (left, MEF $n = 24$, R1 $n = 27$) and CTCF (right, MEF $n = 20$, R1 $n = 29$) scDeChIC-seq data. **g** Comparison of different methods for single-cell profiling of TF-binding sites, showing unique peaks per cell for scDeChIC-seq and unique reads per cell for other published methods. The cell types and CAPs or TFs profiled differ across the datasets generated by each method. To distinguish these conditions, for published methods, the *x*-axis is labeled with the method/cell or tissue type/CAP or TF. Datasets for the same cell line and target protein are indicated using the same color.

regions of lineage-specific markers (*Pou5f1* for ICM; *Gata2* for TE)[51–53] on the UMAP embedding confirmed correct assignment to their respective cell types (Supplementary information, Fig. S6d). We next integrated scDeChIC-seq data with mouse embryo scRNA-seq data and found that scDeChIC-seq data from the three cell types co-clustered with their corresponding scRNA-seq data, demonstrating high concordance in cell-identity annotations (Supplementary information, Fig. S6e, f).[54] To characterize H3K4me3 heterogeneity at single-cell resolution, we defined the Peak Presence Percentage (PPP) as the proportion of cells in which a given locus exhibits a detectable peak, thereby explicitly preserving both presence and absence information. This enabled a more intuitive and faithful measurement of histone modifications and protein-DNA interactions across individual cells, providing a novel framework for assessing cell-to-cell heterogeneity.

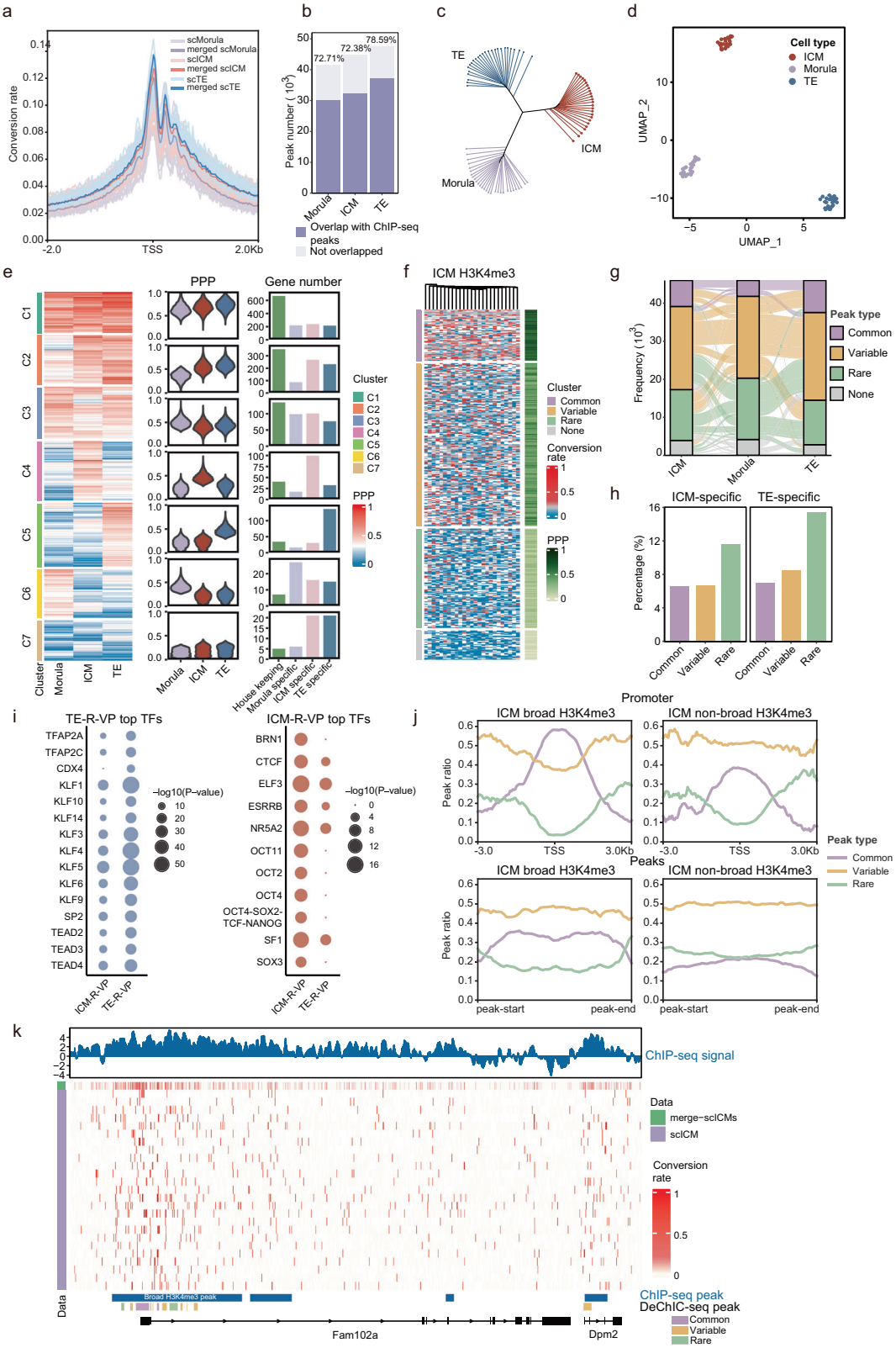
We first used the PPP to cluster H3K4me3 peaks across Morula, ICM, and TE cells, identifying seven distinct groups based on the PPP score of individual peaks in the three cell types (Fig. 4e). Clusters 1 and 2 contained peaks with a high PPP in all three cell types, whereas cluster 7 exhibited the lowest PPP. Clusters 3–6 displayed various PPPs across different cell types, reflecting dynamic chromatin features during early embryonic lineage specification. Functional analysis of associated genes revealed that clusters 4, 5, and 6 were enriched for lineage-specific genes of the ICM, TE, and Morula, respectively, whereas clusters 1 and 2 were highly enriched for housekeeping genes,[55–59] consistent with their strong H3K4me3 signal and promoter-proximal enrichment (Fig. 4e; Supplementary information, Fig. S7a, b).

To further explore H3K4me3 heterogeneity during lineage differentiation, we classified peaks into three categories: Common (PPP ≥ 60%), Variable (PPP 30%–60%), and Rare (PPP < 30%, excluding peaks under 10%) (Fig. 4f). Common peaks were predominantly located at promoters, strongly enriched for housekeeping genes, and associated with higher gene expression levels, indicating their role in the maintenance of core cellular functions (Supplementary information, Fig. S7c–e). Gene Ontology (GO) analysis confirmed that Common peaks of the three stages were linked to essential biological processes such as mRNA processing (e.g., *Hnrnpa2b1*, *Sf3a1*, *Xrn2*), histone modification (e.g., *Rbbp4*, *Sin3a*), and cell cycle regulation (e.g., *Vcp*, *Ube2n*, *Psmd4*, *Psmd7*) (Supplementary information, Fig. S7f). When UMAP clustering was performed separately using each of the three peak sets, Common peaks also yielded a more compact clustering of single cells across the three developmental stages (Supplementary information, Fig. S7g). By contrast, Variable peaks exhibited a clearer separation between ICM and TE than the other peak categories (Supplementary information, Fig. S7g), consistent with their frequent association with lineage-specific regulatory programs and the enrichment of genes characteristic of distinct cell fates. In the ICM, this gene set included canonical pluripotency regulators such as *Pou5f1* (*Oct4*), *Sox2*, and *Nanog*, as well as ICM and naïve pluripotency-associated factors, such as *Utf1*, *Dppa5a*, *Zfp281*, *Zfp57*, *Zscan10*, *Otx2*, and *Tdgf1*.[60] Conversely, Variable peaks in the TE were enriched for established TE markers, including the Hippo pathway effector *Tead2*, the epithelial markers *Krt8*, *Krt18*, and *Tacstd2*, and key signaling and lineage regulators such as *Fgfr2*, *Stat3*, *Ovol1*, and *Tinagl1*.[61] In addition, Variable and Rare peaks were more frequently located in intergenic and distal regions and showed a higher tendency to co-localize with distal H3K27ac-marked enhancer regions,[62,63] suggesting a potential role in cell type-specific regulatory activity (Supplementary information, Fig. S7c, h).

Tracking H3K4me3 peak transitions across the Morula, ICM, and TE revealed extensive chromatin remodeling during lineage commitment (Fig. 4g; Supplementary information, Fig. S7i). We found substantial overlap between Morula-stage rare peaks and TE- or ICM-specific peaks, suggesting that these rare peaks may represent pre-established, lineage-specific peaks that are primed for subsequent developmental stages (Fig. 4h). We then focused on peaks that transitioned from Rare in Morula to Variable in ICM or TE. To identify potential regulators of these transitions, we performed TF motif enrichment analysis on ICM- or TE-specific Rare-to-Variable peaks (ICM-R-VPs and TE-R-VPs). Notably, ICM-R-VPs were enriched for pluripotency factors such as OCT4, SOX2,[51,64] and NANOG,[65] whereas TE-R-VPs contained motifs for the TE lineage-specifying factors TEAD,[66] KLF,[67] and CDX2 (Fig. 4i).[68] These findings suggest that rare peaks may serve as chromatin footholds for lineage-specific TFs to initiate differentiation programs.

Previous studies identified broad H3K4me3 domains (> 5 kb) as crucial for lineage maintenance in pre-implantation embryos.[69] However, it remained unclear whether these domains were universally present in all cells or arose from overlapping signals in different cells. Leveraging scDeChIC-seq PPP data, we found that variable peaks accounted for the largest proportion of broad domains, whereas common peaks were enriched at peak centers or TSS regions (Fig. 4j, k). This suggests that broad H3K4me3 domains form through a combination of universally present core signals and variable flanking regions contributed by different cells.

Collectively, these findings indicate that scDeChIC-seq provides a powerful framework for resolving chromatin state heterogeneity at single-cell resolution, offering new insight into its coordinated dynamics during early embryonic development.



scDeChIC-seq resolves the single-cell architecture of non-canonical H3K4me3 in early embryos

Leveraging the unique capabilities of scDeChIC-seq, we were able to further interrogate the authentic chromatin landscape of the non-canonical form of H3K4me3 (ncH3K4me3) observed in early embryos, particularly within cleavage-stage blastomeres prior to the late 2-cell (L2C) stage.[62,69,70] These ncH3K4me3 regions are typically represented in ChIP-seq as broad, low-amplitude peaks, a feature that may obscure their true signal intensity when analyzed using bulk-cell data. We therefore performed H3K4me3 scDeChIC-

Fig. 4 scDeChIC-seq provides new insight into dynamic changes in H3K4me3 during the first lineage differentiation. a H3K4me3 scDeChIC-seq at the Morula and blastocyst stages. Line plots show the average conversion rates at the TSS. Different curves represent individual single cells or merged cells. Colors represent different cell types. **b** Bar plots showing the overlap between peaks obtained from merged H3K4me3 scDeChIC-seq data from single Morula, TE, or ICM cells and ChIP-seq-defined peaks. **c**, **d** Dimensionality reduction and clustering analysis of scDeChIC-seq data from Morula ($n = 29$), ICM ($n = 25$), and TE ($n = 26$) cells using hierarchical clustering (**c**) and UMAP (**d**). **e** Heatmap showing differential clustering of H3K4me3 PPP among Morula, ICM, and TE cells (left). All H3K4me3 peak regions (rows, $n = 46{,}175$) defined by these cell types are grouped into seven main clusters. Violin plots (middle) depict the PPP distribution within each cluster, and bar plots (right) show the number of corresponding genes. **f** Heatmap showing the average conversion rates of ICM H3K4me3 peaks categorized by PPP into Common peaks (PPP ≥ 60%), Variable peaks (PPP 30%–60%), and Rare peaks (PPP < 30%, with under 10% excluded) across the ICM scDeChIC-seq data (columns, $n = 25$). **g** Alluvial plot showing dynamic changes in the three peak types (based on PPP) during the first lineage differentiation in H3K4me3 scDeChIC-seq data from Morula, ICM, and TE cells. **h** Bar plot showing the percentage of Common, Variable, and Rare H3K4me3 scDeChIC-seq peaks in Morula cells — specifically those detected in at least three cells — that overlap with ICM- or TE-specific H3K4me3 scDeChIC-seq peaks. **i** Analysis of enriched TF-binding motifs in ICM-specific Rare-to-Variable peaks (ICM-RVP) and TE-specific Rare-to-Variable peaks (TE-RVP), co-localized with putative enhancer regions. **j** Line plots showing the distribution of three peak types defined by ICM scDeChIC-seq across promoter regions (top) and ChIP-seq-defined H3K4me3 broad peaks (bottom). A set of randomly selected H3K4me3 non-broad peaks served as the control. **k** Comparison of scDeChIC-seq and ChIP-seq tracks at ChIP-seq-defined broad H3K4me3 peaks. Each row represents a data type or an individual scDeChIC-seq single cell; each column represents a TC site. Color intensity indicates the conversion rate; scDeChIC-seq-defined peak types are shown below.

seq profiling at the PN5 late 1-cell (L1C), L2C, and 4C stages. Notably, genomic regions defined as ncH3K4me3 peaks at the L1C stage and that subsequently transitioned into canonical, sharp H3K4me3 peaks at the L2C stage exhibited significantly greater peak lengths and higher average conversion rates at L1C compared with L2C and 4C in our DeChIC-seq dataset (Supplementary information, Fig. S7j, k). These results indicate that the broad and low-amplitude appearance of ncH3K4me3 in bulk ChIP-seq does not simply reflect reduced H3K4me3 abundance at the early stage. Instead, scDeChIC-seq reveals extended domains with greater modification density that subsequently reorganize into canonical peaks during development. Accordingly, representative track views of these regions reveal continuous stretches greater than 5 kb with elevated conversion rates, providing strong support for this interpretation (Supplementary information, Fig. S7l).

Together, these findings highlight the ability of scDeChIC-seq to resolve the single-cell architecture of non-canonical H3K4me3 in early embryos, clarifying the basis of non-canonical H3K4me3 domains observed by bulk approaches.

scDeChIC-seq resolves single-cell CTCF binding dynamics during early embryonic lineage differentiation

Detection of protein-DNA interactions at the single-cell level remains a major challenge in mammalian early embryos, where sample availability is inherently limited. Most previous studies have relied on bulk-cell analyses requiring hundreds to thousands of cells, which pose substantial constraints on sample collection in embryos and limit the scope of investigation.[50] However, by incorporating single-cell WGA technology, scDeChIC-seq enables library construction from a single cell, overcoming limitations posed by low TF abundance and sample-collection constraints. This enables high-resolution, single-cell profiling of protein-DNA interactions, facilitating the characterization of CAP binding dynamics during early embryo development.

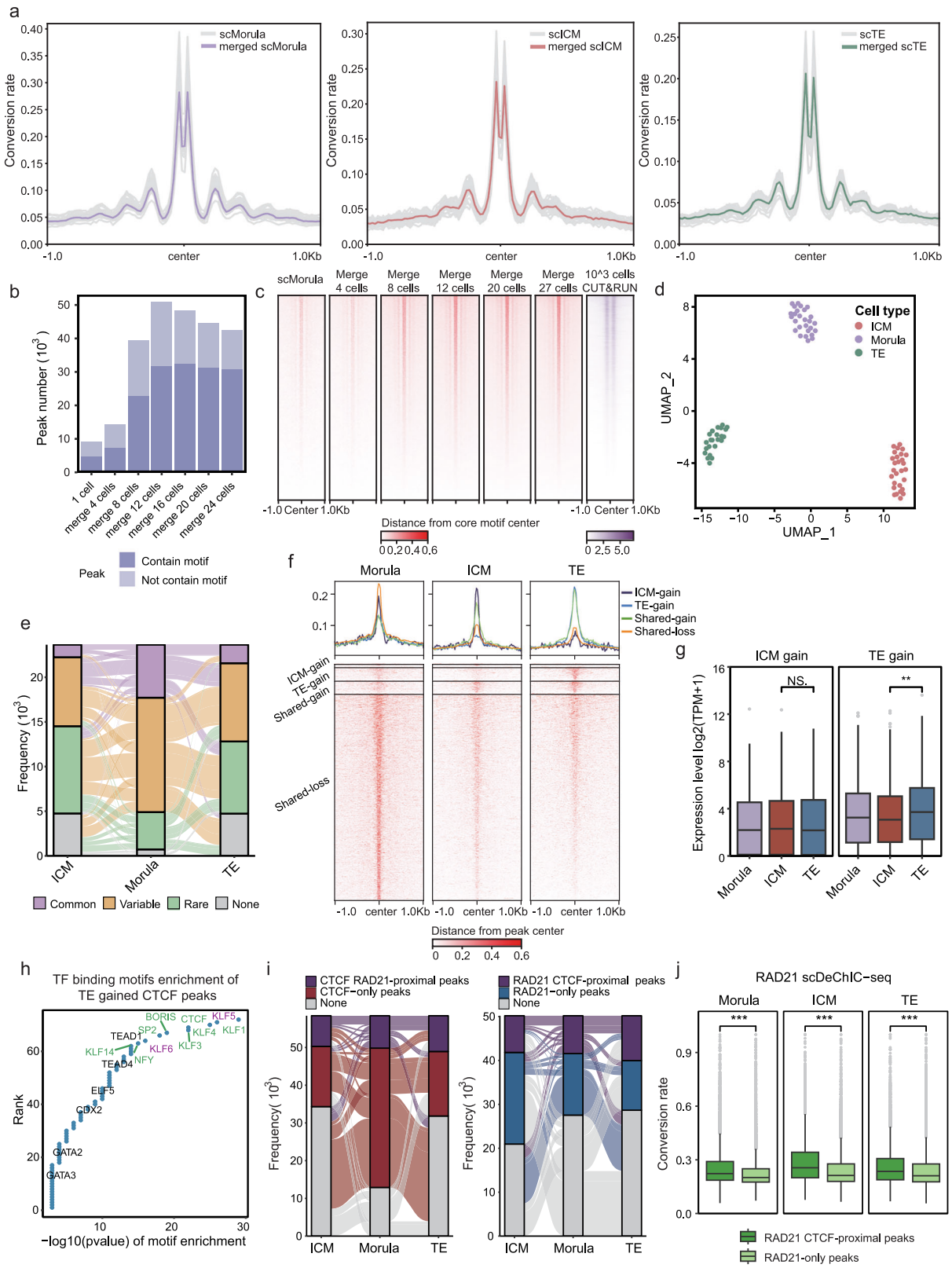
We used scDeChIC-seq to profile CTCF binding in Morula, TE, and ICM cells at the single-cell level. The conversion rate exhibited a distinct pattern around the CTCF core motif within CUT&RUN-defined peaks,[71] with a sharp decrease at the motif center, indicative of CTCF binding and protection from deamination (Fig. 5a). On average, 11,408 CTCF peaks per cell were independently identified using scDeChIC-seq, whereas merging of single-cell datasets yielded 41,644 peaks, a high proportion of which contained the CTCF motif (Supplementary information, Fig. S8a). Notably, merging as few as 12 Morula cells yielded a peak count and motif content comparable to those of the merged dataset (Fig. 5b, c), indicating that scDeChIC-seq captures robust CTCF binding information from a limited number of cells. The heatmap of CTCF core-motif signals in Morula cells further reinforced the robust performance of scDeChIC-seq (Fig. 5c). To determine whether scDeChIC-seq preserves single-cell identity, we performed dimensionality reduction and clustering analysis, successfully distinguishing Morula, ICM, and TE cells as three distinct populations, without a significant batch effect (Fig. 5d; Supplementary information, Fig. S8b). This highlights the effectiveness of scDeChIC-seq in capturing the dynamic landscape of CTCF binding during early embryo development.

We compared conversion rates and peak distributions across Morula, TE, and ICM cells. A global reduction in CTCF binding was observed during the Morula-to-blastocyst transition (Fig. 5a; Supplementary information, Fig. S8a). To further characterize CTCF binding dynamics, we performed PPP-based peak classification, grouping scDeChIC-seq-defined CTCF peaks into Common, Variable, and Rare categories. PPP-based dynamics provided further support for a global reduction in CTCF binding during the Morula-to-blastocyst transition (Fig. 5e).

Previous studies identified cleavage-specific CTCF-binding sites (cs-CBSs) that disappear in both ICM and TE, whereas reserved CTCF binding sites (r-CBSs) persist from the 8-cell stage to the blastocyst.[71,72] To track these dynamic changes, we examined scDeChIC-seq signals at r-CBSs and cs-CBSs during the Morula-to-blastocyst transition. We found that r-CBSs were composed mainly of Common peaks and remained largely stable across developmental stages, whereas cs-CBSs exhibited greater variability and a progressive decline, consistent with prior findings (Supplementary information, Fig. S8c–f). To provide independent validation, we incorporated scDeChIC-seq profiling of RAD21, a core subunit of the cohesin complex,[73] at the corresponding developmental stages (Supplementary information, Fig. S8g, h). Consistent with expectations, RAD21 exhibited largely stable signal levels at r-CBSs, whereas its binding was markedly reduced at cs-CBSs (Supplementary information, Fig. S8i).

Building upon these findings, we performed a comparative analysis of changes in CTCF-binding sites during lineage differentiation between TE and ICM (Fig. 5f). Interestingly, the decay of cs-CBSs differed between the ICM and TE lineages, with TE exhibiting a more rapid loss of cs-CBSs (Supplementary information, Fig. S8j). We also identified a subset of TE-gained CBSs, which either showed an increased PPP relative to ICM or were newly gained during TE differentiation (Fig. 5f). Genes associated with TE-gained CBSs showed higher expression levels and were enriched at TE-specific TF motifs, including CDX2 and GATA3 (Fig. 5g, h).[68,74–76] These patterns suggest a lineage-asymmetric remodeling of CTCF binding during the first lineage specification.

Finally, to investigate the relationship between CTCF and cohesin dynamics, we stratified the CTCF and RAD21 peaks



according to whether they were proximal to one another (within 500 bp), then visualized their transitions using alluvial plots (Fig. 5i). RAD21 peaks proximal to CTCF were preferentially retained, whereas RAD21-only peaks were frequently lost during differentiation, indicating that CTCF binding dynamics were largely independent of RAD21 proximity but that RAD21 retention may partly depend on CTCF association (Fig. 5i). The average conversion rates of the corresponding peak categories across the three embryonic stages further supported this observation (Fig. 5j; Supplementary information, Fig. S8k).

Fig. 5 scDeChIC-seq detects changes in CTCF binding during morula differentiation at single-cell resolution. a CTCF scDeChIC-seq for Morula (left), ICM (middle), and TE (right) cells. Line plots show the average conversion rates at the CTCF core motif in CUT&RUN-defined CTCF peaks for each cell type. Different curves represent individual single cells or merged cells. **b** Bar plot showing the number of CTCF peaks identified in individual and merged Morula scDeChIC-seq data, categorized by the presence or absence of the CTCF motif. **c** Heatmaps showing CTCF signals derived from scDeChIC-seq data of individual and merged Morula cells, together with published Morula CTCF CUT&RUN data. Signals are centered on the top 50,000 CTCF motifs with the highest core motif score. **d** Dimensionality reduction and clustering analysis of CTCF scDeChIC-seq data for Morula ($n = 27$), ICM ($n = 28$), and TE ($n = 23$) cells using UMAP. **e** Alluvial plot showing dynamic changes in the three types of CTCF peak (based on PPP) during the first lineage differentiation in Morula, ICM, and TE cells. Only peaks covered by more than 50% of the cells (23,645 in total) were used. **f** Heatmap showing the average conversion rates at CBSs in Morula, ICM, and TE cells, categorized as ICM-gain, TE-gain, shared-gain, or shared-loss based on their dynamic changes during differentiation. Corresponding line plots depict the average conversion rates across the three cell types at CTCF peak centers within these CBS categories. **g** Boxplots showing the expression levels of genes associated with ICM-gained (left) and TE-gained (right) peaks across Morula, ICM, and TE cells. *P*-values were calculated by unpaired two-sided Student's *t*-test. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$; NS, not significant. **h** Analysis of enriched TF-binding motifs in the TE-gained CTCF peaks. Each point represents a TF. The top 10 enriched TFs (excluding those without detectable expression) are labeled in green; TFs significantly associated with TE lineage differentiation are labeled in black; and TFs that met both criteria are labeled in purple. **i** Alluvial plot showing dynamic changes in CTCF (left) and RAD21 (right) peaks during the first lineage differentiation in Morula, ICM, and TE cells. Peaks are stratified by whether or not they are proximal to the other TF (within 500 bp). Only peaks covered by more than 50% of the cells were used. **j** Box plots showing the average conversion rates of RAD21 scDeChIC-seq peaks at the Morula, ICM, and TE stages, stratified by their proximity to CTCF (within 500 bp). *P* values were calculated by a two-tailed Wilcoxon rank-sum test. *** $P < 0.001$.

Collectively, these results demonstrate that scDeChIC-seq enables single cell-level profiling of CTCF binding dynamics during early embryonic differentiation, suggesting coordinated yet asymmetric regulation of CTCF and cohesin during early embryonic development.

The broad potential of scDeChIC-seq for single-cell detection of TF binding sites

To evaluate the broader applicability of scDeChIC-seq for profiling TF-DNA interactions, we tested its ability to detect chromatin binding of NR5A2, TFAP2C, and KLF5 during early embryogenesis.[77–79] scDeChIC-seq revealed clear enrichment of deamination signals flanking TF motifs within aggregated peaks, with a pronounced reduction at the motif center, indicating direct TF binding and local protection from deamination (Supplementary information, Fig. S9a–c). Merging scDeChIC-seq data from approximately 20 cells was sufficient to identify tens of thousands of peaks, the majority of which contained the corresponding motifs (Fig. 6a). Heatmaps further demonstrated that scDeChIC-seq captured localized deamination enrichment around TF motifs even at the single-cell level. Notably, aggregating as few as 12 cells yielded clear and high-quality signals (Fig. 6b). These patterns were comparable to CUT&RUN profiles generated from about 1000 cells but exhibited sharper motif-centered signals, highlighting the improved resolution of scDeChIC-seq for detection of TF binding events (Fig. 6b). Notably, individual cells could yield 2714–11,158 unique peaks while retaining a substantial fraction of motif-containing peaks (Fig. 6c, d; Supplementary information, Fig. S9d, e and Table S1).

Building on these observations, we next assessed the performance of scDeChIC-seq in a biologically and technically challenging context by interrogating NR5A2 binding in *Obox* knockout (*Obox*-KO) L2C embryos. OBOX is a key regulator of zygotic genome activation (ZGA) in mouse embryos, and its loss leads to reduced expression of major ZGA genes, including *Nr5a2*, ultimately resulting in developmental arrest. Notably, *Nr5a2* knockdown does not significantly affect *Obox* expression, nor does it prevent OBOX from activating ZGA genes, suggesting that NR5A2 may function downstream of OBOX.[80] Using scDeChIC-seq, we observed a substantially reduced NR5A2 occupancy in *Obox*-KO L2Cs, independent of proximity to OBOX binding sites (Fig. 6e, f). These results indicate that NR5A2 occupancy is broadly influenced by *Obox*-KO, although this analysis did not enable us to determine whether the observed reduction was driven primarily by reduced NR5A2 protein abundance or by changes in binding capacity.[80]

Collectively, these findings demonstrate that scDeChIC-seq enables sensitive mapping of TF-DNA interactions in extremely low-input samples and offers mechanistic insight into transcriptional regulation during early embryonic development.

DISCUSSION

We developed scDeChIC-seq, a deaminase-based method for detection of histone modifications and TF-binding sites, starting with as few as one cell. This approach addresses several critical limitations of existing chromatin-profiling technologies and opens new avenues for studying epigenetic heterogeneity at the single-cell level.

By eliminating the immunoprecipitation step required in traditional ChIP-based approaches, DeChIC-seq enables genome-wide detection of chromatin modifications without enrichment. Conventional methods suffer from antibody-dependent enrichment biases, which can lead to incomplete coverage of binding sites and limit confidence in assessing the presence of binding signals.[37] Such biases can also obscure relative differences in signal intensity across genomic regions.[17] By dispensing with enrichment, DeChIC-seq preserves genome-wide background information and enables more faithful comparison of chromatin modification signals across loci. This feature is particularly advantageous for studying dynamic processes characterized by substantial changes in epigenetic modifications.

The ability to detect TF-binding sites at single-cell resolution represents a significant advancement in the field. Current single-cell ChIP-seq-based methods struggle with TF profiling owing to the sparse nature of TF binding events and the signal dilution caused by limited immunoprecipitation efficiency. DeChIC-seq circumvents these issues through its whole-genome sequencing approach, whereby detection sensitivity becomes independent of target abundance. In principle, TF binding events may be interrogated from individual cells when antibodies are available to direct deaminase recruitment. This capability is helpful for studying the dynamics of transcriptional regulation, particularly in processes like cell-fate determination or cancer evolution, in which stochastic TF activation drives phenotypic diversification.

A key innovation of DeChIC-seq lies in its ability to profile chromatin features from individual cells without requiring cell pooling. Traditional single-cell sequencing methods, such as scCUT&tag and uCoTarget, rely on combinatorial barcoding strategies that aggregate data from hundreds or thousands of cells, limiting their application to rare samples.[11,47] By contrast,

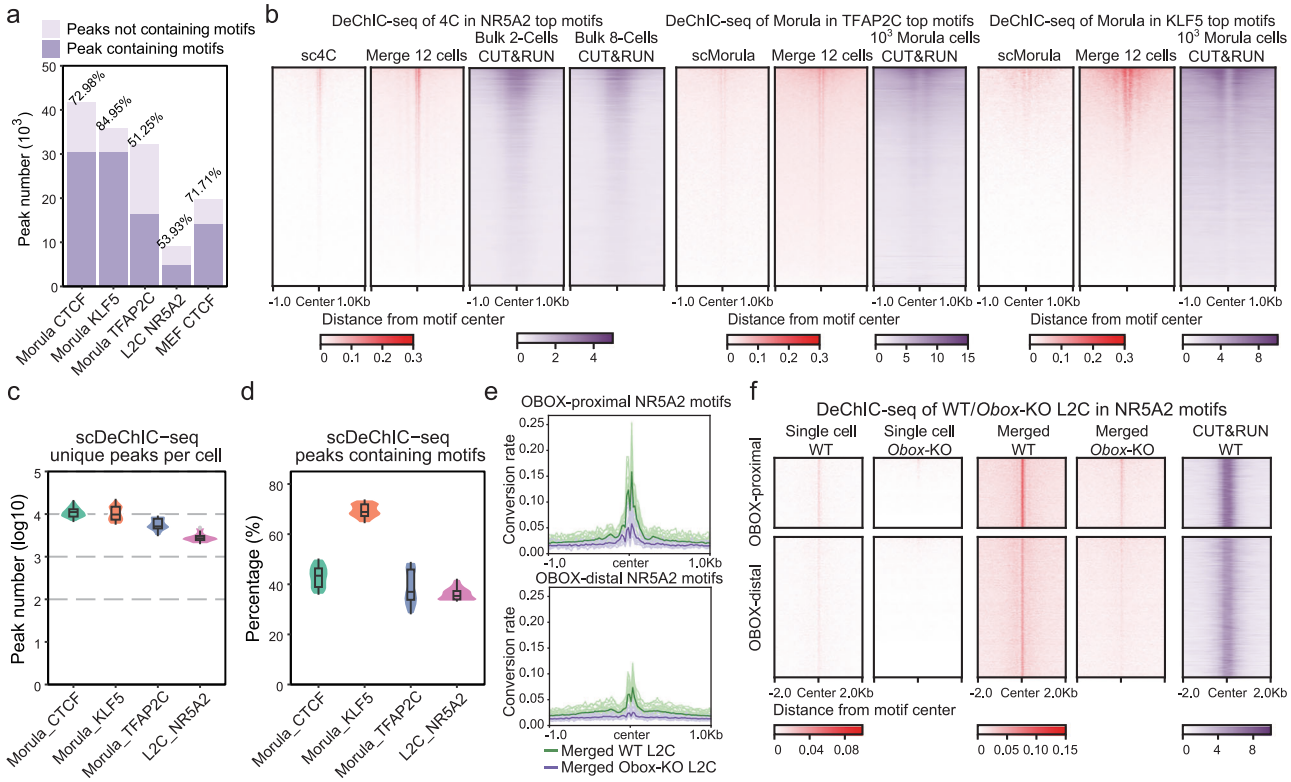


Fig. 6 scDeChIC-seq enables robust detection of TF binding in early embryos, underscoring its broad application potential. a Bar plots showing the number of peaks and the proportion of peaks that contain or are proximal to the corresponding motifs (within 400 bp) in merged TF scDeChIC-seq data, using Morula data as an example, with merged MEF CTCF scDeChIC-seq data shown for comparison. **b** Heatmaps showing NR5A2, TFAP2C, and KLF5 signals around their respective top motifs from individual and merged scDeChIC-seq data in single-cell samples from pre-implantation embryos, shown alongside published CUT&RUN data. **c**, **d** Violin plots showing the number of unique peaks per cell (**c**) and the proportion of peaks that contain or are proximal to the corresponding motif (within 400 bp) (**d**) in TF scDeChIC-seq data from early embryonic samples. **e** Line plots showing the average conversion rates around NR5A2 motifs within L2C NR5A2 CUT&RUN peaks of merged WT (green) and *Obox*-KO (purple) L2C NR5A2 scDeChIC-seq datasets. NR5A2 motifs in the peaks are stratified into proximal (top; within 100 bp) and distal (bottom) groups based on their distance to OBOX motifs within OBOX CUT&Tag peaks. **f** Heatmaps showing NR5A2 scDeChIC-seq signals around NR5A2 motifs in NR5A2 CUT&RUN peaks from individual and merged single-cell datasets for WT and *Obox*-KO L2C embryos, shown alongside published CUT&RUN data. NR5A2 motifs are stratified into proximal (top) and distal (bottom) groups based on their distance to OBOX motifs within OBOX CUT&Tag peaks.

DeChIC-seq enables direct library preparation from individual nuclei, avoiding the need for combinatorial indexing or large cell pools. This feature is particularly valuable for the analysis of rare cell populations, such as primordial germ cells or circulating tumor cells, as well as dynamic processes, such as embryonic development or immune responses, in which cellular heterogeneity plays a crucial functional role.

Although DeChIC-seq demonstrates strong potential, further optimization of antibody specificity and sequencing depth may enhance its robustness across a broader range of targets. Throughout the DeChIC-seq workflow, cellular architecture and chromatin integrity are preserved, enabling the method to be readily adapted for barcoding or other multiplexing strategies to increase throughput. The modular design of DeChIC-seq allows potential integration with spatial transcriptomics or multi-omics profiling approaches, enabling the correlation of chromatin states with gene expression or spatial context. As single-cell epigenomics moves toward clinical applications, DeChIC-seq may provide valuable insights into epigenetic heterogeneity in diseases such as cancer and neurodegeneration, where pathogenic cell subpopulations are often obscured in bulk analyses.

Despite its advantages, several technical considerations may influence the performance of DeChIC-seq. First, because DeChIC-seq relies on C-to-U conversion at TC dinucleotides, extremely low local TC density could reduce detection sensitivity at a small subset of genomic loci. Second, because deaminase recruitment depends on antibody binding, signal quality is influenced by antibody specificity, affinity, and epitope accessibility within chromatin. Third, although DeChIC-seq preserves genome-wide background information, adequate sequencing depth is still required to reliably capture low-frequency or weak binding events, particularly in single-cell applications. At the same sequencing depth, its detection performance for TF binding may be weaker than that for histone modifications. In addition, owing to the localized and dynamic nature of TF binding and the distinct, conversion-based signal representation of DeChIC-seq, which produces narrower and less peak-centered profiles, its agreement with enrichment-based reference datasets is expected to be reduced, reflecting differences in measurement rather than reduced biological relevance. Also, for features of local chromatin architecture, nucleosome positioning may influence the spatial spread of deamination signals and should be considered when interpreting signal patterns. Finally, in its current version, DeChIC-seq may be less suited for high-resolution profiling of repressive histone marks such as H3K27me3 and H3K9me3. This limitation likely reflects the compact chromatin architecture associated with these marks, which results in reduced peak-centered contrast and more diffuse conversion signals compared with those of active histone modifications, although further methodological optimization may extend its applicability in heterochromatin landscapes.

MATERIALS AND METHODS

Cell culture

MEFs were isolated from E13.5 embryos and cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS, 2 mM L-glutamine, 1 mM non-essential amino acids, 0.1 mM β-mercaptoethanol, and 1% penicillin/streptomycin. Before being used as feeder cells, MEFs were inactivated by mitomycin C treatment for embryonic stem cell (ESC) culture.

R1 ES cells were cultured on mitomycin C-treated MEF feeders in ES cell medium composed of DMEM, 15% FBS, 2 mM L-glutamine, 1 mM non-essential amino acids, 0.1 mM β-mercaptoethanol, 1% penicillin/streptomycin, and LIF. The medium was changed every 24 h.

Collection of mouse pre-implantation embryos

Specific-pathogen-free mice were housed in the animal facility at Tongji University, Shanghai, China. All animal maintenance and experimental procedures were performed according to the Tongji University Guide for the use of laboratory animals. Mice were housed under a 12-h light/dark cycle and pathogen-free conditions at 22 ± 2 °C and given free access to standard mouse chow and tap water. The female mice used in this study were 6–8-week-old B6D2F1 (C57BL/6n × DBA2, BDF1) mice, and the male mice were 12-week-old BDF1 mice. To obtain fertilized mouse embryos, female mice were super-ovulated by injection with 7 IU of pregnant mare serum gonadotropin (PMSG), followed by 5 IU of human chorionic gonadotropin (hCG) (San-Sheng Pharmaceutical) 48 h later. Immediately after hCG injection, the super-ovulated female mice were mated with male mice. After vaginal plugs appeared, the zygotes were collected from the oviducts of the female mice at 20 h post hCG (h.p.h.) injection and were cultured in G-1 PLUS medium until the blastocyst stage. Samples were collected at the L1C (28 h.p.h.), L2C (46 h.p.h.), 4C (54 h.p.h.), Morula (80 h.p.h.), and ICM/TE (104–108 h.p.h.) stages for DeChIC-seq.

Antibodies

Antibodies used in this study were anti-H3K4me3 (1:100, Abcam, ab213224; 1:100, Cell Signaling Technology, 9751S), anti-H3K4me1 (1:100, Cell Signaling Technology, 5326S), anti-H3K9ac (1:100, Abcam, ab4441), anti-H3K27ac (1:100, Active Motif, 39133; 1:100, Diagenode, C15410196), anti-CTCF (1:150, ABclonal, A1133), anti-PolII (1:100, Active Motif, 91151), anti-RAD21 (1:100, GeneTex, GTX106012; 1:100, ABclonal, 19749), anti-NR5A2 (1:100, R&D Systems, PPH232500), anti-KLF5 (1:100, Proteintech, 21017-1-AP), anti-TFAP2C (1:100, Santa Cruz Biotechnology, sc-12762), normal rabbit IgG (1:100, Cell Signaling Technology, 2729S), and goat anti-rabbit secondary antibody (1:100, Vazyme, AB207).

X-ChIP-seq and ULI-N-ChIP-seq

For X-ChIP-seq, 3×10^4 cells were used per reaction for H3K4me3 and 10^5 cells per reaction for CTCF; three replicates were performed for each sample. All isolated cells were washed three times with 0.5% BSA-PBS, then fixed with 1% formaldehyde for 10 min at room temperature to cross-link proteins to DNA. Chromatin was then extracted and sheared using a sonicator to generate DNA fragments (~200–500 bp). For each immunoprecipitation reaction, 1 μg of histone H3K4me3 antibody (Abcam, ab213224) or CTCF antibody (ABclonal, A1133) was used. Antibody-bound protein-DNA complexes were pulled down by protein A/G beads, and the DNA was eluted and purified.

For ULI-N-ChIP-seq, 500 cells were used per reaction for H3K9ac, H3K27ac, and H3K4me1, and two or three replicates were performed for each sample. All isolated cells were washed three times with 0.5% BSA-PBS, and the ULI-NChIP procedure was performed as described previously.[81]

Sequencing libraries were generated using the KAPA Hyper Prep Kit (KAPA, KK8504, Switzerland) for the Illumina platform, following the manufacturer's instructions. Paired-end sequencing with a 150-bp read length was performed on the Illumina NovaSeq platform.

CUT&Tag

All CUT&Tag data were generated using the Hyperactive Universal CUT&Tag Assay Kit for Illumina Pro (Vazyme, TD904) following the manufacturer's protocol. For CTCF, Pol II, and RAD21, each replicate was performed with 50,000 MEF cells. For the MEF and R1 mixed-cell H3K4me3 CUT&Tag, the two cell types were combined at the specified ratios, totaling 40,000 cells. Primary antibodies were used at a final concentration of 0.01 μg/μL, and the secondary antibody was goat anti-rabbit (Vazyme, AB207) at a 1:100 dilution. Paired-end sequencing with a 150-bp read length was performed on the Illumina NovaSeq platform.

Purification of pA-DddA

Purification of 6× His-DddA$_{tox}$-A(EAAAK)3A-protein A and 6× His-DddA$_{tox}$-DDDEFG(G4S)4-protein A (both hereafter referred to as pA-DddA) was performed as outlined in Mok et al.[27] with the modifications described below. DNA manipulation and transformation in bacterial studies were performed with standard methods, and all amino acid sequences of pA-DddA in this study are provided in Supplementary information, Table S2.

Rosetta 2(DE3) cells were obtained from EMD Millipore and transformed with pETDuet-1::DddA$_{tox}$-proteinA_DddIA. Single colonies of transformed Rosetta 2(DE3) were used to inoculate 2 L of LB in a 1:50 dilution and cultured overnight. When cultures had grown to about OD600 of 0.6, 0.5 mM IPTG was added to induce plasmid expression. The cultures were then grown for 16 h at 18 °C in a shaking incubator. The bacteria were collected by centrifugation at 4000× *g* for 20 min at 4 °C and washed twice with 50 mL of DPBS, followed by lysing in 50 mL of lysis buffer (50 mM Tris-HCl, pH 7.4, 500 mM NaCl, 30 mM imidazole, 1 mM DTT, 1 mg/mL lysozyme, 1% Triton X-100, and 2 mM PMSF). Pellets were lysed sufficiently with a high-pressure cell crusher, and the supernatant was separated by centrifugation at 25,000× *g* for 30 min at 4 °C.

The His-tagged DddA$_{tox}$-proteinA_DddIA complex was purified from the cell-lysate supernatant using 2 mL of Ni-NTA agarose beads. The supernatant was loaded into a gravity-flow column, and Ni-NTA agarose beads were washed with 50 mL of wash buffer (50 mM Tris-HCl, pH 7.4, 500 mM NaCl, 30 mM imidazole, 1 mM DTT, 1% Triton X-100, and 2 mM PMSF).

Next, 50 mL of 8 M urea denaturing buffer (50 mM Tris-HCl, pH 7.4, 30 mM imidazole, 500 mM NaCl, 1 mM DTT, and 1% Triton X-100) was added to the beads in the column and incubated for 2 h at 4 °C to remove any remaining DddIA. The 8 M urea denaturing buffer was discarded from the column. The pA-DddA was sequentially washed with 50 mL denaturing buffer with decreasing concentrations of urea (6 M, 4 M, 2 M, 1 M, 0 M, 0 M) to remove the remaining urea, then eluted with 4 mL of elution buffer (50 mM Tris-HCl, pH 7.5, 300 mM imidazole, 500 mM NaCl, 1 mM DTT, 1% Triton X-100, and 2 mM PMSF).

The purified samples were dialyzed twice with 10 K MWCO Slide-A-Lyzer Dialysis Devices (Thermo Fisher Scientific) in dialysis buffer (20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 1 mM DTT, 5% (w/v) glycerol, and 2 mM PMSF) for 6 h at 4 °C, and then concentrated to 200 μL. For long-term storage, 200 μL of glycerol was added, and the samples were stored at −80 °C.

Enzyme activity of pA-DddA

All DNA substrates, with a 6-FAM fluorophore for visualization, were obtained from Tsingke Biotech. Reactions were performed in a total volume of 10 μL in Reaction buffer (20 mM MES, pH 6.4, 50 mM NaCl, 1 mM DTT, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 0.02% digitonin, 1× PIC), Dig-HEPES wash buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA, 0.02% digitonin, 1× PIC), or Dig-Tris8.0 buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA, 0.02% digitonin, 1× PIC) with 1 μM substrate and pA-DddA at the indicated concentrations. Samples were incubated for 1 h at the indicated temperature, followed by the addition of 0.5 μL of UDG (NEB, 5000 units/mL) and 1 μL of 10× UDG Reaction Buffer, then incubated for 30 min at 37 °C. Denaturation of substrates was induced by the addition of 0.5 μL of 2 mM NaOH and incubation at 95 °C for 10 min, followed by the addition of 10 μL of formamide and 2 μL of Gel Loading Dye (NEB) and incubation at 95 °C for 10 min. The samples were then placed immediately on ice and analyzed by 15% denaturing acrylamide gel electrophoresis; fluorescent DNA fragments were detected by fluorescence imaging with Biorad ChemiDoc Imaging systems.

U PCR bias detection

All ssDNA substrates (with 0 Us, 2 Us, 4 Us, 6 Us, or 8 Us) and primers for PCR-bias detection were obtained from Sangon Biotech. The sequences are provided in Supplementary information, Table S3.

The ssDNA substrates were mixed in different proportions and transformed into dsDNA by PCR. Detailed information on the mixed ssDNA substrates is provided in Supplementary information, Table S3. Preamplification mix was prepared as follows: 0.25 μL of 0.5 μM ssDNA substrates, 5 μL of 5× Phusion HF Buffer, 0.5 μL of 10 mM dNTP, 0.25 μL each of 50 μM U-PCR-bias-detection primer-F and 50 μM primer-R, 0.25 μL of Phusion U Hot Start DNA Polymerase (Thermo Fisher Scientific), and

18.5 μL of ddH_2O. Preamplification was performed as follows: 98 °C for 30 s; 10 cycles of 98 °C for 10 s, 53 °C for 30 s, and 72 °C for 30 s; and 72 °C for 5 min. Primers were removed by the addition of 1 μL of exonuclease I (NEB; M0293), followed by incubation at 37 °C for 30 min and 80 °C for 20 min. The preamplification product was added to 2.0× AMPure XP beads, and the final library was eluted in 25 μL of EB.

Sequencing libraries were generated using the KAPA Hyper Prep Kit for the Illumina platform, following the manufacturer's instructions. Paired-end sequencing with a 150-bp read length was performed on the Illumina NovaSeq platform.

Bulk DeChIC-seq

MEF and R1 cells were isolated by trypsin at 37 °C for 5 min. The cells were washed once with DPBS and twice with 1 mL of room-temperature Wash buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA, and 1× PIC), then resuspended in 600 μL of room-temperature Wash buffer.

Next, 20 μL of CoA beads were transferred into 60 μL of pre-cooled CoA binding buffer (20 mM HEPES-KOH, pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$), washed twice with CoA binding buffer, and resuspended in 300 μL of CoA binding buffer. Beads were incubated with cells for 10 min at room temperature with rotation, blocked with 1 mL of Blocking buffer (obtained by addition of 2 mM EDTA and 0.02% digitonin to Wash buffer) for 5 min, washed with 1 mL of Antibody wash buffer (obtained by addition of 0.02% digitonin to Wash buffer), and resuspended in 300 μL of Antibody wash buffer. The antibody was added, and the mixture was incubated at 4 °C for 2−6 h with rotation. The beads were then washed twice with Antibody wash buffer to remove residual antibody and resuspended in 300 μL of Binding buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA, 0.02% digitonin, and 1× PIC); 0.08 μM pA-DddA was added, and the mixture was incubated at 4 °C for 1 h. After two washes with Binding buffer, the beads were resuspended in 300 μL of Reaction buffer (20 mM MES, pH 6.4, 50 mM NaCl, 1 mM DTT, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 0.02% digitonin, and 1× PIC) at 37 °C for 30 min.

The DNA was then extracted and fragmented according to the protocol of the TruePrep DNA library prep kit V2 for Illumina (Vazyme, TD501) and eluted in 10 μL of ddH_2O. Preamplification mix was prepared as follows: 10 μL of fragmented DNA, 10 μL of 5× Phusion HF Buffer, 1 μL of 10 mM dNTP, 5 μL each of TruePrep index kit V2 for Illumina adaptor (Vazyme, TD202) N5XX and N7XX primers, 0.5 μL of Phusion U Hot Start DNA Polymerase, and 18.5 μL of ddH_2O. Preamplification was performed as follows: 72 °C for 3 min; 98 °C for 45 s; 10−12 cycles of 98 °C for 15 s, 62 °C for 30 s, and 72 °C for 1 min 30 s; and 72 °C for 5 min. The preamplification product was added to 0.6× + 0.2× AMPure XP beads, and the final library was eluted in 20 μL of ddH_2O. The library was sent to the Nanjing Jiangbei New Area Biopharmaceutical Public Service Platform for paired-end sequencing with a 150-bp read length on the Illumina NovaSeq platform.

scDeChIC-Tn5 loading

scDeChIC-Tn5 loading was performed according to the protocol of TruePrep Tagment Enzyme (Vazyme, S601). Primers for preparation of the adapter mix were fully referenced to META-CS and obtained from Sangon Biotech.

META-CS transposons were dissolved in Annealing Buffer and annealed with primer A (5′-phos-CTGTCTCTTATACACATCT-NH2-3′), and then incubated at 75 °C for 15 min, 60 °C for 10 min, 50 °C for 10 min, 40 °C for 10 min, and 25 °C for 30 min. All adapters were mixed together.

Coupling of scDeChIC-Tn5 was performed immediately. Coupling mix was prepared as follows: 4 μL of TruePrep Tagment Enzyme (500 ng/μL), 7 μL of Adapter Mix, and 39 μL of Coupling Buffer. The coupling mix was incubated at 30 °C for 1 h. The reaction product was named TTE mix and stored at −30 °C to −15 °C.

scDeChIC-seq

scDeChIC was performed as described for bulk DeChIC with the modifications described below. Instead of CoA beads, centrifugation or mouth-pipetting was used to wash or transfer cells; no CoA bead binding or blocking operations were performed.

For MEFs and mouse ESCs, buffer replacement was achieved by centrifugation (4 °C, 600× *g* for 5 min). Single cells (10^3) were washed once with DPBS and twice with 1 mL of room temperature Wash buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA, and 1× PIC), and then resuspended in 300 μL of cold Antibody wash buffer (obtained by addition of 0.02% digitonin to Wash buffer). The antibody was added, and the mixture was incubated at 4 °C for 2−6 h with rotation. The cells were then washed twice with 0.5 mL cold Antibody wash buffer and resuspended in 300 μL of cold Binding buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA, 0.02% digitonin, and 1× PIC); 0.08 μM pA-DddA was added, and the mixture was incubated at 4 °C for 1 h. After two washes with 0.5 mL cold Binding buffer, cells were resuspended with 300 μL in Reaction buffer (20 mM MES, pH 6.4, 50 mM NaCl, 1 mM DTT, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 0.02% digitonin, and 1× PIC) at 37 °C for 30 min. Cells were centrifuged and resuspended in DPBS. Finally, single cells were individually distributed into PCR tubes with 2 μL of META-CS lysis buffer (20 mM Tris-HCl, pH 8.0, 20 mM NaCl, 0.15% Triton X-100, 25 mM DTT, 1 mM EDTA, 1.5 mg/mL Qiagen protease, 500 nM carrier ssDNA 5′-TCAGGTTTTCCTGAA-3′) by FACS or mouth-pipetting. Twenty to thirty single cells were randomly selected for the construction of sequencing libraries.

For mouse pre-implantation embryos, all incubation and enzymatic reactions were performed within droplets, and cell transfers were performed manually using mouth pipettes. Single cells (one cell is sufficient) were washed 3 times with 50 μL of room temperature Wash buffer droplets, transferred to 50 μL of cold Antibody wash buffer (obtained by addition of 0.02% digitonin and 0.01% Triton X-100 to Wash buffer) droplets, and incubated for 2 min. The cells were transferred to 30 μL of Antibody wash buffer droplets with antibody. The droplets were overlaid with mineral oil and incubated undisturbed at 4 °C for 6 h. The cells were then washed 3 times with 50 μL of Antibody wash buffer and transferred to 30 μL of cold Binding buffer (with 0.01% Triton X-100 and 0.08 μM pA-DddA) droplets. The droplets were overlaid with mineral oil and incubated undisturbed at 4 °C for 1 h. The cells were then washed 3 times with 50 μL of Binding buffer and transferred to 200 μL of Reaction buffer at 37 °C for 30 min. The single cells were spaced out into PCR tubes with 2 μL of META-CS lysis buffer by mouth-pipetting.

Single-cell DNA extraction and library preparation were performed as outlined in Xing et al.[42] with the modifications described below. After cells were incubated in Reaction buffer for 30 min, single cells were transferred to 0.2-mL tubes with 2 μL of META-CS lysis buffer (20 mM Tris-HCl, pH 8.0, 20 mM NaCl, 0.15% Triton X-100, 25 mM DTT, 1 mM EDTA, 1.5 mg/mL Qiagen protease, 500 nM carrier ssDNA 5′-TCAGGTTTTCCTGAA-3′) and lysed at 50 °C for 1 h and 70 °C for 15 min.

The single-cell lysate was transposed using TruePrep Tagment Enzyme (Vazyme, S601). Transposition mix was prepared as follows: 2 μL of single-cell lysate, 2 μL of 5× Tagment Buffer L, 0.76 nM scDeChIC-Tn5, and ddH_2O to a final volume of 10 μL. The mixture was incubated at 55 °C for 10 min. scDeChIC-Tn5 was removed by the addition of 2 μL 5× Stop buffer (Vazyme, TD501) and 0.5 μL of 12.5 mg/mL Qiagen protease, followed by incubation at 50 °C for 30 min and 70 °C for 15 min.

First-strand tagging was performed by adding 12.5 μL of Strand Tagging Mix 1 (5 μL of 5× Phusion HF Buffer, 0.5 μL of 10 mM dNTP, 0.8 μL of 50 μM (total) Adp1 primer mix, 1 μL of 100 mM $MgCl_2$, 4.95 μL of water, and 0.25 μL of Phusion U Hot Start DNA Polymerase) and incubating at 72 °C for 3 min, 98 °C for 30 s, 62 °C for 5 min, 60 °C for 5 min, and 72 °C for 5 min. Adp1 primers were removed by the addition of 1 μL of exonuclease I, followed by incubation at 37 °C for 30 min and 80 °C for 20 min.

Second-strand tagging was performed by adding 4 μL Strand Tagging Mix 2 (1 μL of 5× Phusion HF Buffer, 0.1 μL of 10 mM dNTP, 1 μL of 50 μM (total) Adp2 primer mix, 1.85 μL of water, and 0.05 μL of Phusion U Hot Start DNA Polymerase) and incubating at 72 °C for 3 min, 98 °C for 30 s, 62 °C for 5 min, 60 °C for 5 min, and 72 °C for 5 min. Adp2 primers were removed by the addition of 1 μL of exonuclease I, followed by incubation at 37 °C for 30 min and 80 °C for 20 min.

Strand tagging products were then amplified by adding 19 μL of PCR mix (5 μL of NEBNext Multiplex Oligos Universal Primer, 5 μL of Index Primers (NEB; E7335L, E7500L), 4 μL of 5× Phusion HF Buffer, 0.4 μL of 10 mM dNTP mix, 4.4 μL of water, and 0.2 μL of Phusion U Hot Start DNA Polymerase) and incubating at 98 °C for 30 s; 12–14 cycles of 98 °C for 10 s, 61.5 °C for 30 s, and 72 °C for 1 min 30 s; and 72 °C for 5 min.

The preamplification product was added to 0.6× + 0.5× AMPure XP beads, and the final library was eluted in 20 μL of ddH_2O. The library was sent to the Nanjing Jiangbei New Area Biopharmaceutical Public Service Platform for paired-end sequencing with a 150-bp read length on the Illumina NovaSeq platform.

Bulk DeChIC-seq data processing

The raw reads were filtered using Trim Galore (v0.6.7, https://github.com/FelixKrueger/TrimGalore) with Cutadapt (v4.1)[82] and the parameters "--trim-n --clip_R1 10 --clip_R2 10 --three_prime_clip_R1 10 --three_prime_clip_R2 10 --length 50 -q 20 --paired". The trimmed reads were aligned to the mm10 reference genome using Basal (v1.2)[83] in paired-end mode with the parameters "-r 1 -v 0.06 -s 16 -S 1 -n 1 -g 1 -M C:T". The avgmod function of Basal was used to calculate the C count and T count for each covered cytosine site with the parameters "-r -m 1 -i correct". Slight modifications were made to the original code (https://github.com/XiyangChen12/DeChIC/tree/main/SCENE/scripts/basalkit_filterTC.py) to convert the output of sequence context from CG/CHG/CHH to TC/AC/GC/CC. To exclude interference from SNPs, those from C to T and from G to A were excluded from the analysis. The SNP information for C57BL/6n, BDF1 (offspring of C57BL × DBA), and 129 mice was downloaded from https://www.mousegenomes.org/. For the analysis of bulk DeChIC-seq data, TC sites with a depth of at least 2× were retained for downstream analysis.

All the profile line plots and heatmaps around the TSS, peak center, or motif core center were generated using deepTools (v 3.5.1).[84] In the Integrative Genomics Viewer[85] (IGV)-generated track view, DeChIC-seq signal visualization incorporated sliding window-based smoothing. Specifically, a 200-bp sliding window with a step size of 5 bp was used to calculate the average conversion rate for each window. The smoothed conversion rate for each TC site represents the mean value derived from all windows spanning the site.

scDeChIC-seq data processing

The first 45 bases of raw reads were trimmed to remove the META adapter, and the last 3 bases were also removed owing to their instability; trimming was performed using TrimGalore (v0.6.7, https://github.com/FelixKrueger/TrimGalore) with Cutadapt[82] in paired-end mode. The mapping workflow and calculation of conversion rate were similar to those described for the bulk data analysis. For the analysis of scDeChIC-seq data, TC sites with at least 1× depth were retained for downstream analysis. Cells were discarded if they met any of the following criteria: (1) number of covered C sites $\leq 1 \times 10$,[8] (2) duplicate rate > 30%, and (3) signal-to-noise ratio (average conversion rate in peak region/average conversion rate of all covered TC sites) < 8 (except for L1C H3K4me3 scDeChIC-seq, owing to the widespread presence of ncH3K4me3 domains).

Quantification of H3K4me3 modification and CTCF binding

The conversion rate for each covered TC site was calculated as the number of converted reads (T count) divided by the sum of converted and unconverted reads (C count + T count). The conversion rate of each peak was determined by averaging the conversion rates of all covered TC sites within the region.

Identification of blacklist regions

To address the TC bias of DddA, which can cause sites with high TC density to exhibit abnormally high conversion rates and lead to false positives, we defined a custom blacklist region. First, TC density was calculated in 20-bp bins, and 4,558,691 regions with high TC density (TC count ≥ 6 per 20-bp bin) were identified. Next, because of the presence of TC-rich and low-complexity sequences in simple repeats and low-complexity repeats, these two types of repeat were also added to the blacklist. Finally, the blacklist regions of mm10 downloaded from the ENCODE database were incorporated.[86] These three components formed the blacklist region used in our study.

Peak calling for bulk DeChIC-seq data

Peak regions in the bulk DeChIC-seq data were defined by comparing the histone modification or TF DeChIC-seq dataset to the IgG DeChIC-seq dataset. First, TC sites located in blacklist regions were removed. Differentially converted regions between the histone modification or TF and the IgG datasets were identified using Metilene (v0.2-8)[87] with the parameters "-M 50 -m 6". Three cutoff values — *P*-value, difference in conversion rate, and conversion rate of IgG — were used to define potential peak regions. Peaks with $P < 0.05$, a difference in conversion rate > 0.15, and a conversion rate of IgG < 0.04 were retained as putative histone-modification peaks. Peaks within 200 bp were then merged using the merge function of BEDTools (v2.30.0).[88] Finally, peaks that overlapped with blacklist regions by more than 50% were excluded from downstream analyses. Given that the signal-to-noise ratio is typically higher around TF binding regions, peaks with $P < 0.05$, a difference in conversion rate > 0.25, and a conversion rate of IgG < 0.04 were retained as TF peaks.

Peak calling for scDeChIC-seq data

For peak calling in the single-cell data, all cells were first merged within each stage (pseudo-bulk), followed by removal of TC sites located in blacklist regions. A pseudo-IgG dataset for each merged single-cell group was simulated by averaging the conversion rates across the whole genome. Metilene[87] (v0.2-8) was then used to call peaks by comparing the pseudo-bulk data with the pseudo-IgG data. The cutoff for histone modification datasets was the same as for the bulk level, and peaks with $P < 0.05$ and a difference in conversion rate > 0.2 were retained as merged single-cell TF peaks.

A higher cutoff for the difference in conversion rate is recommended for peak calling in single-cell data. We used 0.2 or 0.25 as the cutoff value. For RAD21, given its relatively low signal, we applied a lower cutoff of 0.15.

Cell clustering and annotation of DeChIC-seq data

Guided by previously published analytical methods for single-cell conversion-based sequencing data,[83,89] we adopted a matrix-based strategy in which a conversion-rate matrix and the Signac workflow were used to cluster cells. First, a peak-by-cell matrix was generated by calculating the average conversion rate for each peak in each single cell. The matrix was then converted into a binary matrix, with cutoffs of 0.3 and 0.35 used for histone modification and TFs, respectively. Single-cell clustering was performed using Signac[48] (v1.10.0). Specifically, the peak-by-cell matrix was converted into a Seurat object[90] using CreateSeuratObject and analyzed with Signac.[48] Peaks were normalized using TF-IDF, and top features were selected with FindTopFeatures. LSI was used to reduce dimensionality, and UMAP was used for visualization (dims = 2:7). Cell clustering was performed using the FindNeighbors and FindClusters functions. Cell type annotation was based on conversion rate in the promoter region (TSS ± 1 kb).

Integration of H3K4me3 scDeChIC-seq data and published scRNA-seq data

scRNA-seq data for mouse Morula, ICM, and TE cells were obtained from a previous publication (GSE136718).[54] The gene expression count matrix was obtained using the procedure described in the paper. Seurat (v5.1.0)[90] was used to perform scRNA-seq analysis based on the matrix. The gene activity score of H3K4me3 scDeChIC-seq for each gene was generated by quantifying the average conversion rate within 1 kb upstream and downstream of the TSS. The FindTransferAnchors function from the Seurat package was used to identify anchors between the H3K4me3 scDeChIC-seq and scRNA-seq datasets based on gene activity scores and the gene expression count matrix using the parameter reduction = "cca" (cca, canonical correlation analysis). Annotations from the scRNA-seq dataset were transferred to the scDeChIC-seq data using the TransferData function with the parameter "dims = 2:5", and the combined object was visualized using UMAP.

ChIP-seq data processing

The raw paired-end reads were trimmed to remove adapter sequences and low-quality bases using Trim Galore (v0.6.7, https://github.com/FelixKrueger/TrimGalore) with Cutadapt[82] (v4.8). The filtered reads were then aligned to mm10 using bwa[91] (v0.7.17) with default parameters, and only high-quality mapped reads (MAPQ > 30) were retained. PCR duplicates were removed using Picard (v2.27.4, https://broadinstitute.github.io/picard/). Peak calling for each replicate was performed using MACS2[92] (v2.2.7.1), followed by peak merging using BEDTools[88] (v2.30.0).

RNA-seq data processing

The raw reads were filtered using TrimGalore (v0.6.7, https://github.com/FelixKrueger/TrimGalore) with Cutadapt[82] (v4.8). The filtered reads were then mapped to mm10 using HISAT2[93] (v2.2.1) with the parameters "--no-discordant --no-mixed --dta --no-unal". The TPM values and gene read count matrix were calculated using StringTie[94] (v2.2.1).

Identification of Common, Variable, and Rare peaks

To ensure the accuracy of peak classification, only peaks covered by > 50% of the cells were retained for downstream analysis. The proportion of H3K4me3 occupancy in the cell population for each peak was calculated using a cutoff of 0.15. Peaks with a proportion ≥ 60% were identified as "Common peaks"; peaks with a proportion ≤ 60% and > 30% were classified as "Variable peaks"; peaks with a proportion ≤ 30% and > 10% were identified as "Rare peaks"; and peaks with a proportion ≤ 10% were classified as "None peaks". The identification of CTCF peaks followed the same procedure, but with a cutoff of 0.2 for the proportion of CTCF binding in the cell population.

Identification of housekeeping genes and stage-specific genes in embryos

We generated a gene expression matrix using RNA-seq data, including published RNA-seq data for 17 mouse tissues from 6-week-old mice,[55] as well as data for the ICM and TE stages (GSE66582 and GSE195762).[56,57] We generated our own Morula RNA-seq data. To identify putative housekeeping genes in the mouse tissues and embryos, we applied the following criteria: (1) genes expressed robustly in all samples (TPM > 3); (2) genes with low variance across all samples (standard deviation of $\log_2$TPM < 1); (3) genes not specifically expressed in certain tissues (0.5 < fold change ($\log_2$TPM in one sample/average $\log_2$TPM) < 2); (4) genes that overlap with three published housekeeping gene lists.[55,58,59] As a result, 1487 putative housekeeping genes were identified.

To identify stage-specific genes for the Morula, ICM, and TE, differential expression analysis was performed using DESeq2[95] (v1.38.3), comparing one tissue with the other two. Genes with a $\log_2$fold change > 1 and $P_{adj} < 0.05$ were selected as stage-specific genes.

Identification of ICM- and TE-specific H3K4me3 scDeChIC-seq peaks

H3K4me3-specific peaks for the ICM and TE were identified by comparative analysis between the two lineages. Specifically, ICM-specific H3K4me3 peaks were identified by excluding any regions that overlapped with TE H3K4me3 peaks using BEDTools[88] (v2.30.0), followed by further filtering to retain only those peaks located at least 1 kb away from any TE H3K4me3 peaks. A similar approach was used to identify TE-specific H3K4me3 peaks.

Identification of ncH3K4me3 domains in L1C embryos using ChIP-seq data

Using data from our previously published study,[69] we identified ncH3K4me3 domains with ChromHMM,[96] guided by the analytical methods described in that work. In brief, alignment files were binarized into 200-bp bins using the BinarizeBam command, and ChromHMM was first run with default parameters to generate genome-wide state annotations. We then used ChromHMM to identify high-enrichment H3K4me3 bins that showed fold enrichment > 4 (using the -f 4 option) and excluded the corresponding high-enrichment H3K4me3 domains from the initial segmentation. Finally, the remaining bins were merged within 1 kb to define the final ncH3K4me3 domains.

Peak genomic feature annotation and identification of CBS-associated genes

Peaks were mapped to genomic features using ChIPseeker[97] (v1.34.1). The nearest genes to CBS peaks were identified as CBS-associated genes using ChIPseeker.

GO enrichment analysis

Gene set enrichment analysis for GO terms was performed using clusterProfiler[98] (v4.6.2).

TF motif enrichment analysis

The TF motif enrichment analysis was performed using the findMotifsGenome.pl script in Homer[99] (v4.11.1).

TF motif site detection

Position weight matrices of TF motifs, retrieved from the JASPAR[100] and Cis-BP databases,[101] were used to scan the mouse mm10 genome sequences using FIMO[102] (v5.5.7).

Identification of R1- and MEF-specific H3K4me3 peaks

To identify H3K4me3 peaks specific to MEF or R1 cells based on DeChIC-seq, ChIP-seq, and CUT&Tag data, we first downsampled each ChIP-seq and CUT&Tag replicate to 20 million read pairs. We defined the consensus peak set as those peaks that overlapped between the CUT&Tag and ChIP-seq data. For each assay (DeChIC-seq, ChIP-seq, and CUT&Tag), we quantified H3K4me3 signals over the consensus peaks in R1 and MEF cells and calculated the fold change (R1/MEF). Peaks were classified as R1- or MEF-specific if their fold change ranked within the top 15% across all three assays.

In parallel, we derived assay-specific cell type-specific peaks independently for each platform. For each assay, we computed the fold change between R1 and MEF cells using peaks called from that assay and selected the top 1000 peaks as the R1- or MEF-specific peaks for that assay.

Linear regression analysis of assay signals and cell mixing proportions

To assess the linear relationships of DeChIC-seq, ChIP-seq, and CUT&Tag signals with cell input proportions, we calculated the signal at each R1-specific peak for each assay, and then used the median value for each group for linear regression against the corresponding R1 cell proportion (25%, 50%, 75%, and 100%). Linear regression was performed using the lm() function in R (v4.2.1), and R^2 values were obtained from the model summary. The same regression analysis was used for MEF-specific peaks and corresponding MEF cell proportions (25%, 50%, 75%, and 100%).

DATA AVAILABILITY

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive[103] at the National Genomics Data Center,[104] China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA021319) and are publicly accessible at https://ngdc.cncb.ac.cn/gsa. The H3K4me3 ChIP-seq and RNA-seq datasets for MEFs were obtained from a previous publication (GSE117909).[32] The H3K4me3 ChIP-seq data for pre-implantation mouse embryos were downloaded from our previous work (GSE73952 and GSE97778).[69,105] The CTCF CUT&RUN datasets for Morula, ICM, and TE were downloaded from CRA011730.[71] Published CTCF peaks of MEFs were downloaded from the Cistrome database[106] (CistromeDB ID 84049). The ATAC-seq data for MEF cells were obtained from previous publications (GSE109395).[107] The RNA-seq data for the ICM and TE stages were obtained from previous publications (GSE66582 and GSE195762).[56,57] The scRNA-seq data for pre-implantation mouse embryos were obtained from a previous publication (GSE136718).[54] The NR5A2 CUT&RUN datasets for 2C and 8C were downloaded from GSE229740.[77] TFAP2C and KLF5 CUT&RUN datasets for the Morula were obtained from previous publications (GSE212568).[79] H3K27ac ChIP-seq peaks for the Morula, ICM, and TE were obtained from previous publications (CRA018821).[108]

CODE AVAILABILITY

We developed the SCENE package (a scDeChIC-seq data analysis pipeline) for data analysis. The SCENE package and processing code used in this study have been deposited in GitHub (https://github.com/XiyangChen12/DeChIC).

REFERENCES

1. Iwafuchi, M. et al. Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nat. Genet.* **52**, 418–427 (2020).
2. Xia, W. et al. Resetting histone modifications during human parental-to-zygotic transition. *Science* **365**, 353–360 (2019).
3. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
4. Kelsey, G., Stegle, O. & Reik, W. Single-cell epigenomics: recording the past and predicting the future. *Science* **358**, 69–75 (2017).
5. Mundade, R., Ozer, H. G., Wei, H., Prabhu, L. & Lu, T. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle* **13**, 2847–2852 (2014).
6. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856 (2017).
7. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
8. Rotem, A. et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).

9. Grosselin, K. et al. High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* **51**, 1060–1066 (2019).
10. Patty, B. J. & Hainer, S. J. Transcription factor chromatin profiling genome-wide using uliCUT&RUN in single cells and individual blastocysts. *Nat. Protoc.* **16**, 2633–2666 (2021).
11. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat. Biotechnol.* **39**, 825–835 (2021).
12. 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* **18**, 652–660 (2021).
13. Bartosovic, M. & Castelo-Branco, G. Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. *Nat. Biotechnol.* **41**, 794–805 (2023).
14. Liu, M. et al. Genome-coverage single-cell histone modifications for embryo lineage tracing. *Nature* **640**, 828–839 (2025).
15. Wang, Q. et al. CoBATCH for high-throughput single-cell epigenomic profiling. *Mol. Cell* **76**, 206–216.e7 (2019).
16. Moreno-Gonzalez, M., Sierra, I. & Kind, J. A hitchhiker's guide to single-cell epigenomics: methods and applications for cancer research. *Int. J. Cancer* **158**, 291–304 (2026).
17. Amezquita, R. A. et al. Orchestrating single-cell analysis with bioconductor. *Nat. Methods* **17**, 137–145 (2019).
18. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424–428 (2000).
19. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat. Protoc.* **2**, 1467–1478 (2007).
20. de Luca, K. L. & Kind, J. Single-cell DamID to capture contacts between DNA and the nuclear lamina in individual mammalian cells. *Methods Mol. Biol.* **2157**, 159–172 (2021).
21. Rang, F. J. et al. Single-cell profiling of transcriptome and histone modifications with EpiDamID. *Mol. Cell* **82**, 1956–1970.e14 (2022).
22. Gallagher, L. A. et al. Genome-wide protein-DNA interaction site mapping in bacteria using a double-stranded DNA-specific cytosine deaminase. *Nat. Microbiol.* **7**, 844–855 (2022).
23. Hass, M. R. et al. SpDamID: Marking DNA bound by protein complexes identifies notch-dimer responsive enhancers. *Mol. Cell* **59**, 685–697 (2015).
24. Altemose, N. et al. DiMeLo-seq: a long-read, single-molecule method for mapping protein-DNA interactions genome wide. *Nat. Methods* **19**, 711–723 (2022).
25. Yue, X. et al. Simultaneous profiling of histone modifications and DNA methylation via nanopore sequencing. *Nat. Commun.* **13**, 7939 (2022).
26. Weng, Z. et al. BIND&MODIFY: a long-range method for single-molecule mapping of chromatin modifications in eukaryotes. *Genome Biol.* **24**, 61 (2023).
27. Mok, B. Y. et al. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583**, 631–637 (2020).
28. Mi, L. et al. DddA homolog search and engineering expand sequence compatibility of mitochondrial base editing. *Nat. Commun.* **14**, 874 (2023).
29. Yin, L. L. et al. Structural basis for sequence context-independent single-stranded DNA cytosine deamination by the bacterial toxin SsdA. *Nat. Commun.* **16**, 8841 (2025).
30. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10**, 161–172 (2009).
31. Chen, X. et al. Structures of +1 nucleosome-bound PIC-mediator complex. *Science* **378**, 62–68 (2022).
32. Zhang, Y. et al. Targets and genomic constraints of ectopic Dnmt3b expression. *Elife* **7**, e40757 (2018).
33. Peters, A. H. F. M. et al. Loss of the histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
34. Weidemuller, P., Kholmatov, M., Petsalaki, E. & Zaugg, J. B. Transcription factors: bridge between cell signaling and gene regulation. *Proteomics* **21**, e2000034 (2021).
35. Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes Dev.* **30**, 881–891 (2016).
36. Sun, X., Zhang, J. & Cao, C. CTCF and its partners: shaper of 3D genome during development. *Genes* **13**, 1383 (2022).
37. Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**, 669–680 (2009).
38. Sheaffer, K. L. & Schug, J. ChIP-Seq: library preparation and sequencing. *Methods Mol. Biol.* **1402**, 101–117 (2016).
39. Liang, K. & Keles, S. Normalization of ChIP-seq data with control. *BMC Bioinform.* **13**, 199 (2012).
40. Polo, J. M. et al. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632 (2012).
41. Xiong, H., Wang, R. & He, A. Single-cell protein-DNA interactomics and multiomics tools for deciphering genome regulation. *Natl. Sci. Open* **2**, 20220057 (2023).
42. Xing, D., Tan, L., Chang, C. H., Li, H. & Xie, X. S. Accurate SNV detection in single cells by transposon-based whole-genome amplification of complementary strands. *Proc. Natl. Acad. Sci. USA* **118**, e2013106118 (2021).
43. Hoang, M. L. et al. Genome-wide quantification of rare somatic mutations in normal human tissues using massively parallel sequencing. *Proc. Natl. Acad. Sci. USA* **113**, 9846–9851 (2016).
44. Zhu, C. et al. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat. Methods* **18**, 283–292 (2021).
45. Zhang, B. et al. Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro. *Nat. Biotechnol.* **40**, 1220–1230 (2022).
46. Janssens, D. H. et al. Scalable single-cell profiling of chromatin modifications with sciCUT&Tag. *Nat. Protoc.* **19**, 83–112 (2024).
47. Xiong, H., Wang, Q., Li, C. C. & He, A. Single-cell joint profiling of multiple epigenetic proteins and gene transcription. *Sci. Adv.* **10**, eadi3664 (2024).
48. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with signac. *Nat. Methods* **18**, 1333–1341 (2021).
49. Bartlett, D. A. et al. High-throughput single-cell epigenomic profiling by targeted insertion of promoters (TIP-seq). *J. Cell Biol.* **220**, e202103078 (2021).
50. Hainer, S. J., Boskovic, A., McCannell, K. N., Rando, O. J. & Fazzio, T. G. Profiling of pluripotency factors in single cells and early embryos. *Cell* **177**, 1319–1329.e11 (2019).
51. Rizzino, A. & Wuebben, E. L. Sox2/Oct4: a delicately balanced partnership in pluripotent stem cells and embryogenesis. *Biochim. Biophys. Acta* **1859**, 780–791 (2016).
52. Home, P. et al. Genetic redundancy of GATA factors in the extraembryonic trophoblast lineage ensures the progression of preimplantation and post-implantation mammalian development. *Development* **144**, 876–888 (2017).
53. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
54. Wang, Y. et al. Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nat. Commun.* **12**, 1247 (2021).
55. Li, B. et al. A comprehensive mouse transcriptomic bodyMap across 17 tissues by RNA-seq. *Sci. Rep.* **7**, 4200 (2017).
56. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
57. Xu, R. et al. Unreprogrammed H3K9me3 prevents minor zygotic genome activation and lineage commitment in SCNT embryos. *Nat. Commun.* **14**, 4807 (2023).
58. Hounkpe, B. W., Chenou, F., de Lima, F. & De Paula, E. V. HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids Res.* **49**, D947–D955 (2021).
59. Zeng, J. et al. Identification and analysis of house-keeping and tissue-specific genes based on RNA-seq data sets across 15 mouse tissues. *Gene* **576**, 560–570 (2016).
60. Bergsmedh, A., Donohoe, M. E., Hughes, R. A. & Hadjantonakis, A. K. Understanding the molecular circuitry of cell lineage specification in the early mouse embryo. *Genes* **2**, 420–448 (2011).
61. Wu, G. et al. Initiation of trophectoderm lineage specification in mouse embryos is independent of Cdx2. *Development* **137**, 4159–4169 (2010).
62. Dahl, J. A. et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).
63. Liu, B. et al. Mapping putative enhancers in mouse oocytes and early embryos reveals TCF3/12 as key folliculogenesis regulators. *Nat. Cell Biol.* **26**, 962–974 (2024).
64. Li, L. et al. Multifaceted SOX2-chromatin interaction underpins pluripotency progression in early embryos. *Science* **382**, eadi5516 (2023).
65. Rodda, D. J. et al. Transcriptional regulation of nanog by OCT4 and SOX2. *J. Biol. Chem.* **280**, 24731–24737 (2005).
66. Yagi, R. et al. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* **134**, 3827–3836 (2007).
67. Lin, S. C., Wani, M. A., Whitsett, J. A. & Wells, J. M. Klf5 regulates lineage formation in the pre-implantation mouse embryo. *Development* **137**, 3953–3963 (2010).
68. Strumpf, D. et al. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093–2102 (2005).
69. Liu, X. et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558–562 (2016).
70. Zhang, B. et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).

71. Wang, W. et al. ADNP modulates SINE B2-derived CTCF-binding sites during blastocyst formation in mice. *Genes Dev.* **38**, 168–188 (2024).
72. Du, Z. et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
73. Li, Y. et al. The structural basis for cohesin-CTCF-anchored loops. *Nature* **578**, 472–476 (2020).
74. Huang, D. et al. The role of Cdx2 as a lineage specific transcriptional repressor for pluripotent network during the first developmental cell lineage segregation. *Sci. Rep.* **7**, 17156 (2017).
75. Home, P. et al. GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. *J. Biol. Chem.* **284**, 28729–28737 (2009).
76. Ralston, A. et al. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* **137**, 395–403 (2010).
77. Lai, F. et al. NR5A2 connects zygotic genome activation to the first lineage segregation in totipotent embryos. *Cell Res.* **33**, 952–966 (2023).
78. Gassler, J. et al. Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378**, 1305–1315 (2022).
79. Gao, R. et al. Defining a TFAP2C-centered transcription factor network during murine peri-implantation. *Dev. Cell* **59**, 1146–1158.e6 (2024).
80. Ji, S. et al. OBOX regulates mouse zygotic genome activation and early development. *Nature* **620**, 1047–1053 (2023).
81. Brind'Amour, J. et al. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.* **6**, 6033 (2015).
82. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
83. Xu, M. et al. BASAL: a universal mapping algorithm for nucleotide base-conversion sequencing. *Nucleic Acids Res.* **53**, gkae1201 (2025).
84. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
85. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
86. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
87. Juhling, F. et al. metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res.* **26**, 256–262 (2016).
88. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
89. Tegowski, M., Flamand, M. N. & Meyer, K. D. scDART-seq reveals distinct m(6)A signatures and mRNA methylation heterogeneity in single cells. *Mol. Cell* **82**, 868–878.e10 (2022).
90. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
91. Li, H. & Durbin, R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
92. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
93. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, stringTie and ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
94. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
95. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
96. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
97. Wang, Q. et al. Exploring epigenomic datasets by ChIPseeker. *Curr. Protoc.* **2**, e585 (2022).
98. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
99. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
100. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).
101. Weirauch, M. T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
102. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
103. Zhang, S. et al. The GSA family in 2025: a broadened wharing platform for multi-omics and multimodal data. *Genomics Proteomics Bioinformatics* **23**, qzaf072 (2025).
104. CNCB-NGDC Members and Partners. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. *Nucleic Acids Res.* **53**, D30–D44 (2025).
105. Wang, C. et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.* **20**, 620–631 (2018).
106. Zheng, R. et al. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* **47**, D729–D735 (2019).
107. Jegu, T. et al. Xist RNA antagonizes the SWI/SNF chromatin remodeler BRG1 on the inactive X chromosome. *Nat. Struct. Mol. Biol.* **26**, 96–109 (2019).
108. Zhu, Q. S. et al. Distal enhancers regulate mammalian early embryonic lineage differentiation through long-range interactions. *Nucleic Acids Res.* **54**, gkag026 (2026).

ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (2022YFA1106000), the National Natural Science Foundation of China (32488101, 32270775, 81902583, 32330030, 32370868, 32222026), the National Key R&D Program of China (2021YFA1102000, 2025YFA1805500, 2022YFC2702200, 2021YFA1102900), the Natural Science Foundation of Shanghai (24ZR1492800), the Science and Technology Commission of Shanghai Municipality (23JC1403700) and the Fundamental Research Funds for the Central Universities (22120240023), the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM508). The *Obox*-KO mice were generously provided by Wei Xie (Tsinghua University). We also thank members of the Zhang, Gao, Wang and Liu laboratories for their helpful suggestions and discussions on the subject during the course of the study.

AUTHOR CONTRIBUTIONS

X.L., S.G., C.W., and J.-M.Z. designed the study and guided the project. Z.S. performed the majority of the experiments. X.C. analyzed the majority of the data. Heng W., K.C., Y.Y., C.L., Y.Z., W.G., Hong W., L.Z., S.L. and Z.Q. performed part of the experiments. A.W. performed part of the data analysis. J.S. guided the development of the data analysis method. W.L. and J.C. provided critical comments and suggestions on the revised manuscript. X.L., S.G., C.W., J.-M.Z., Z.S., and X.C. wrote the manuscript.

COMPETING INTERESTS

The authors have filed an international patent application (PCT_CN2024_110080) and a national patent application in China (202410834221.2) related to the DeChIC-seq method described in this manuscript.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41422-026-01275-z.

Correspondence and requests for materials should be addressed to Jia-min Zhang, Chenfei Wang, Shaorong Gao or Xiaoyu Liu.