

Advances in scCUT&Tag and computational analysis for single-cell gene regulatory element mapping

Jun Wu^{1,‡}, Md Wahiduzzaman^{1,‡}, Pengfei Yin^{1,‡}, Puxuan Sun¹, Haoping Chen¹, Yongwen Ding¹, Jiankang Wang^{1,2,*}

¹School of Biomedical Sciences, Hunan University, Fubehe Road, Yuelu District, Changsha 410082, Hunan, China

²Shenzhen Research Institute, Hunan University, Yuehai Road, Nanshan District, Shenzhen 518000, Guangdong, China

*Corresponding author. School of Biomedical Sciences, Hunan University, Fubehe Road, Yuelu District, Changsha 410082, Hunan, China. E-mail: wangjk321@hnu.edu.cn

[‡]Jun Wu, Md Wahiduzzaman, and Pengfei Yin contributed equally.

Abstract

Histone modifications (HMs) and transcription factors (TFs) are central to chromatin dynamics and transcriptional regulation. Conventional bulk approaches like ChIP-seq require large cell populations, limiting applicability to heterogeneous studies and tissue samples. In contrast, single-cell cleavage under targets and tagmentation (scCUT&Tag) and its variants have enabled high-resolution profiling of HMs and TFs for investigating gene regulatory mechanisms in individual cells, transformatively broadening single-cell epigenomics beyond chromatin accessibility measured by scATAC-seq. Despite rapid advances in scCUT&Tag-related methods and the accumulation of ~21 public datasets, a systematic overview of the current research status, especially the forefront of computational analysis and ensuing challenges, remains lacking. Here, we comprehensively overview current scCUT&Tag studies from a bioinformatics perspective. We catalog representative applications spanning diverse chromatin features, experimental designs, and data characteristics. We delineate a typical computational workflow from matrix generation to downstream functional annotations, emphasizing distinctions from scATAC-seq analysis, and highlighting critical analytical considerations. We extensively survey commonly used computational tools and key algorithms, compare analytical features between scCUT&Tag and scATAC-seq, and discuss major challenges in integrative analysis. This work provides a structured reference for understanding the current research landscape of scCUT&Tag and offers computational perspectives for researchers aiming to explore gene regulatory machinery at single-cell resolution.

Keywords scCUT&Tag, bioinformatics, single-cell epigenomics, computational analysis, transcriptional regulation

Introduction

From bulk profiling to single-cell profiling of TFs, HMs, and beyond

Gene expression is regulated by epigenetic modifications and chromatin remodeling [1, 2]. Histone modifications (HMs) and transcription factors (TFs) coordinate chromatin organization and orchestrate gene transcription [3, 4]. Modifications on histone tails represent distinct regulatory programs [5], as histone (de-)acetyltransferases and (de-)methyltransferases deposit post-translational modifications at various regulatory elements, such as promoters (e.g. H3K4me3) and enhancers (e.g. H3K4me1) [5, 6]. Typically, active HMs such as H3K27ac are associated with chromatin relaxation and promote gene expression [6, 7], whereas repressive marks like H3K27me3 and H3K9me3 correspond to distinct forms of heterochromatin and silence transcription [7, 8]. On the other hand, TFs regulate chromatin accessibility and transcription by recognizing specific DNA sequences [9]. Pioneer TFs, in particular, bind to their target sites within closed

or inactive chromatin, initiating chromatin remodeling and thereby establishing an open chromatin landscape that becomes accessible to other TFs [10]. TF-mediated transcriptional regulation is usually dynamic and sequence specific [9]. Moreover, the same TFs can regulate distinct target genes across different cell types, and some TFs can function as both activators and repressors, depending on their specific chromatin context [11, 12]. The interplay between HMs and TFs is a central focus of tissue-specific epigenomics, as their coordinated regulation of gene transcription underlies various biological processes including disease progression, development, and cancers [13–16].

Previous studies have utilized bulk methods such as chromatin immunoprecipitation with sequencing (ChIP-seq) [17–19], cleavage under targets and release using nuclease (CUT&RUN) [20, 21], and cleavage under targets and tagmentation (CUT&Tag) [13] to capture the chromatin landscape of TFs and HMs in various cell types. ChIP-seq has significantly contributed to biological and epigenomic research over the past decades, providing extensive insights into

Received: August 13, 2025. **Revised:** December 29, 2025. **Accepted:** January 6, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

epigenetic modifications and transcriptional regulation [19, 22, 23]. However, ChIP-seq and its variants suffer from high background noise, epitope masking from cross-linking, low throughput, and the necessity for large cell numbers [24, 25]. CUT&RUN operates within the cell nucleus by using antibodies targeting specific TFs or HMs. It employs micrococcal nuclease (MNase) to selectively cleave DNA at target sites, requiring a moderate cell number while maintaining a low background noise level due to the specificity of MNase digestion [21]. However, the release of MNase-cleaved fragments into the supernatant is not well suited for applications involving few cells, such as single-cell sequencing [26, 27]. CUT&Tag follows a comparable principle but utilizes protein A/G-fused (pA/pG) transposase (Tn5) to cut DNA at target sites [13]. Thus, CUT&Tag requires an even smaller cell number and does not involve fixation and crosslinking, thereby minimizing nonspecific binding and further reducing background noise [13].

Similar to mRNA profiling, bulk sequencing of HMs and TFs generates only population-averaged signals, hindering the resolution of cellular heterogeneity of epigenetic modifications in complex tissues [28]. The advancements in single-cell technologies designed for other modalities has paved the way for single-cell epigenomics, making it feasible to profile HMs and TFs at single-cell resolutions and facilitating a deeper understanding of cell type-specific regulatory landscapes [29–31]. A decade ago, scChIP-seq attempted to overcome the limitations of bulk sequencing by leveraging droplet-based microfluidic technology to analyze chromatin at single-cell resolution [28]. Despite its high specificity, scChIP-seq's low genome coverage underscores inherent limitations related to input sample constraints [28, 32]. Meanwhile, CUT&Tag with pA-Tn5 transposase addresses these challenges by delivering greater efficiency and sensitivity for single-cell applications, especially when compared to the low throughput and high background noise of ChIP-seq, as well as the fragmentation limitations of CUT&RUN [13, 33]. Consequently, the widely adopted single-cell CUT&Tag (scCUT&Tag) has emerged as a powerful approach for profiling HMs and TFs at single-cell resolution, enabling high-throughput epigenomic analysis with minimal input cell requirements [5, 13, 14]. scCUT&Tag has fundamentally broadened the paradigm of single-cell epigenomics beyond chromatin accessibility as detected by scATAC-seq.

Over the past few years, scCUT&Tag has undergone rapid advancements, leading to the development of various variants for profiling TFs, HMs, and more, such as scCUT&Tag [5], scCUT&Tag-pro [14], and scNanoSeq-CUT&Tag (nano-CT) [16, 34]. Other notable innovations include nanobody-tethered transposition followed by sequencing (NTT-seq) [35], and single-nucleus (sn)-m6A-CT, specifically designed for m⁶A methylation profiling [36]. Furthermore, inspired by single-cell joint profiling of ATAC and RNA [37], methods such as ultrahigh-throughput CoTarget (uCoTarget) method enables the concurrent detection of tagmentation events and transcriptomic profiles within the same cells [6]. Collectively, these advances facilitate the integrated analysis of multiple HMs, TFs, accessibilities, and transcription within individual cells, fundamentally transforming the study of gene regulatory mechanisms at single-cell resolution for diverse biological contexts (Table 1). However, despite the rapid innovations of scCUT&Tag-related methods (hereafter we refer to all related methods as scCUT&Tag for brevity) and the accumulation of ~21 publicly available datasets, there remains no systematic review summarizing the current research status of scCUT&Tag for profiling HMs and TFs, particularly concerning the state of the art in

computational analysis and ensuing challenges. Here, we present a comprehensive overview of scCUT&Tag and related technologies from a bioinformatics perspective. We summarized representative studies and data characteristics, described experimental designs and typical computational workflows, discussed critical analytical considerations, surveyed commonly used computational tools and key algorithms, highlighted distinctions from scATAC-seq analysis, and discussed major integrative analysis challenges. This work aims to offer valuable insights into the evolving landscape of single-cell chromatin profiling.

Main context

Representative scCUT&Tag studies and available datasets

scCUT&Tag and related technologies have rapidly evolved to enable high-resolution profiling of HMs, TFs, and other regulators at single-cell levels [32], operating within intact cells to allow low-input applications such as rare or clinical samples. Leveraging this advantage, to date, 21 representative datasets have been published, covering a range of epigenomic marks and serving diverse biological research purposes (Table 1). The literature was retrieved in accordance with the PRISMA 2020 guidelines (Fig. S1), covering the period from January 2019 to December 2025. In 2019, Kaya-Okur *et al.* first introduced scCUT&Tag [13] by employing a pA-Tn5 fusion protein in human K562 and H1 cell lines using the Takara ICELL8 nano-well platform. Subsequent sequencing on the Illumina platform generated high-resolution single-cell epigenomic mapping for H3K27me3 and H3K4me2 [13, 38]. So far, the 21 representative studies have explored a diverse array of chromatin features, including active marks (H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K36me3) [14, 39–41], repressive marks (H3K9me3 and H3K27me3) [14, 16, 39, 42, 43], as well as a histone variant (H2A.Z) [39], chromatin architectural proteins (CTCF, RAD21) [16], transcriptional components such as RNA Polymerase II (Pol II) [35, 44, 45], TFs (e.g. NR2F1 and OLIG2) [5, 46], RNA methylation marks (m⁶A) [36], and nucleic acid secondary structure G-quadruplexes (G4) [47, 48].

A variety of tissues and cell lines have been utilized in those scCUT&Tag studies, including human-derived K562, HEK293T, GM12878, HG002, H9, HFF1, H1, MCF7, U2OS, and cancer cells, as well as mouse-derived samples including peripheral blood mononuclear cells (PBMCs), Kasumi-1 cells, mouse embryonic fibroblasts (MEFs), NIH-3T3, mouse embryonic stem cells (mESCs), mouse trophoblast stem cells (mTSCs), and neuronal cells (Table 1). The number of captured single cells ranges from ~2000 to >100 000 across studies, starting with the initial profiling of H3K4me2 and H3K27me3 in K562 and H1 cells [13], and extending to recent applications such as sn-m6A-CT in mESCs and MEFs [36]. The sequencing depth, represented by fragments per cell, varies widely from 500 to 20 000 fragments [7, 42]. The fraction of reads in peaks (FRiP), a quality metric for epigenomic data, ranges from 0.1 to 0.81 across those datasets [35, 42]. Additionally, genomic bin sizes used for analysis vary from 200 bp to 50 kb [7, 14]. These variations in sample types, cell numbers, sequencing depth, FRiP, and binning resolutions reflect both the diverse objectives of individual studies and the evolving technical capabilities of scCUT&Tag platforms. Currently, scCUT&Tag has achieved a comparable number of cells and moderate FRiP scores; however, sequencing coverage remains a significant challenge. It

Table 1 Currently available resources of published scCUT&Tag-related datasets.

Accession	Target	Tissue/cell type	Cell number	Frag-ments per cell	Method	FRiP	Bin size	Tag antibody	Platform	Ref
Split-pool based										
<i>GSE299567</i>	H3K4me3, H3K36me3, H3K27me3, CTCF, Pol II	Human K562 cells, H1 cells, HEK293 T cells, H1-derived MSC; mouse mESCs	~10 000	~10 000	IT-scC&T-seq	~0.3–0.8	2 kb, 10 kb, 50 kb	pA-Tn5	ANOROAD GENOME	[44]
<i>GSE249799</i>	H3K4me3, H3K27ac, H3K36me3, H3K27me3, H3K9me3, CTCF, RAD21	Human K562 cells, 293 T cells, H9 cells, GM12878 cells, HG002 cells, HFF1 cells; mouse PBMCs	~17 000	~10 000	scNanoSeq-CUT&Tag	0.25–0.8	5 kb	pG-Tn5	Oxford Nanopore	[16]
<i>GSE220193</i>	H3K27ac, H3K4me3, H3K4me1, H3K36me3, H3K27me3	Human K562 cells, Kasumi-1 cells	~20 000	~1100–3700	Nano-CUT&Tag	0.58–0.77	5 kb	pA-Tn5	Illumina NovaSeq	[6]
Droplet based										
<i>GSE291468</i>	G4 structure	Murine embryonic fibroblast (MEF, NIH-3 T3), mESCs, brain	~9000	~1500–3500	scG4	0.52	10 kb	pA-Tn5	10x Genomics, Illumina NextSeq 2000	[47]
<i>GSE254088</i>	H3K27ac	Mouse cranial motor neuron (cMN)	NA	NA	scCUT&Tag	NA	NA	pAG-Tn5	10x Genomics, Illumina NextSeq 500	[54]
<i>CNP0003538 (CSA)</i>	H3K27me3	Human K562 cells, hepatocellular carcinoma (HCC)	~10 000	NA	scCPA-Tag	~0.5	5 kbp	pG-Tn5	10x Genomics, BGISEQ-500	[43]
<i>GSE223271</i>	NR2F1	Mouse embryonic neurons	~9000	NA	scCUT&Tag	NA	NA	pAG-Tn5	10x Genomics	[46]
<i>GSE217255</i>	m6A	HEK293 T cells, mESCs, MEF cells	~100 000	~500	sn-m6A-CT	0.5–0.75	200 bp	pAG-Tn5	10x Genomics, Illumina NovaSeq 6000	[36]
<i>GSE212588</i>	H3K27me3, H3K27ac, Pol II	Human K562 cells, HEK293 cells, PBMCs, BMMCs	~14 000	~500–700, 3000	NTT-seq	0.4–0.6	10 kb	Nanobody-Tn5	10x Genomics, Illumina NextSeq 550	[35]
<i>GSE198467</i>	H3K27ac, H3K27me3	Mouse brain	~20 000	~10 000	Nano-CUT&Tag	0.36–0.52	1 kb	pA-Tn5	10x Genomics, Illumina NovaSeq v1.5	[34]
<i>GSE195725</i>	H3K4me1, H3K4me2, H3K4me3, H3K9me3, H3K27ac, H3K27me3	Human PBMCs	~65 000	~500	scCUT&Tag	0.81	200 bp	pAG-Tn5	10x Genomics, Illumina NextSeq 550	[14]
<i>GSE185271</i>	H3K27me3	mESCs	~11 000	~300–20 000	scCUT&Tag	>0.1	NA	pAG-Tn5	10x Genomics, Illumina NovaSeq 6000	[42]
<i>GSE181373</i>	G4 structure	Human MCF7 cells, U2OS cells	~3000	~600–1200	snG4-CUT&Tag	NA	NA	pA-Tn5	10x Genomics, Illumina NextSeq 500 sequencer	[48]
<i>GSE171554</i>	H3K27ac, H3K27me3	mESCs, mouse trophoblast stem cells (mTSCs)	~2000	NA	scMulti-CUT&Tag	0.4–0.53	3 kbp	pA-Tn5	10x Genomics	[41]
<i>GSE157637</i>	H3K4me3, H3K27ac, H3K36me3, H3K27me3, OLIG2, RAD21	Mouse brain	~50 000	~600	scCUT&Tag	0.39–0.86	5 kbp	pA-Tn5	10x Genomics	[5]

(continued)

Table 1 Continued.

Accession	Target	Tissue/cell type	Cell number	Frag-ments per cell	Method	FRiP	Bin size	Tag antibody	Platform	Ref
Nano-well based										
GSE235109	H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K36me3, H3K9me3, H2A.Z	Mouse embryos, mESC cells	~4400	NA	TACIT, CoTACIT	0.46–0.69	5 kb, 15 kb	MEA/B-Tn5	Illumina NovaSeq 6000, Nextera index (96-well)	[39]
GSE224579	H3K4me1, H3K4me2, H3K4me3, H3K27me3	Human K562 cells, H1 cells, PBMCs, mouse 3T3 cells	~40 000	~500	sciCUT&Tag	0.5–0.8	500 bp, 5 kb, 50 kb	pA-Tn5	ICELL8, Illumina NextSeq 2000	[7]
GSE183032	H3K27me3, Pol II	Human K562 cells, H1 cells	~1000	NA	scCUT&Tag2for1	<0.5	NA	pAG-Tn5	ICELL8, Illumina HiSeq 2500	[45]
GSE159608	H3K4me3, H3K27me3, H3K36me3	Homo K562 cells, leukemia	~6000	~4000–13 000	scCUT&Tag	0.4–0.6	5 kb	pA-Tn5	ICELL8, Illumina HiSeq 2500	[40]
GSE157910	H3K27me3	Human endoderm, glioblastoma, PBMCs	~5000	~300–6000	scCUT&Tag	0.57	5 kb	pA-Tn5	10x Genomics; ICELL8, Illumina NovaSeq 6000	[55]
GSE124557	H3K4me2, H3K27me3	Human K562 cells, H1 cells	~2000	NA	scCUT&Tag	NA	500 bp	pA-Tn5	ICELL8, Illumina sequencing	[13]

NA, not disclosed in the original study.

typically yields $<10^3$ fragments per cell [5], which is still lower than the $>10^4$ fragments per cell commonly detected in scATAC-seq [49–51].

Current scCUT&Tag studies employ a variety of Tn5-based transposase complexes, including pA-Tn5, pAG-Tn5, pG-Tn5, nano-Tn5, and nanobody-Tn5, depending on the target protein or epigenetic mark (Fig. 1 and Table 1). For instance, pAG-Tn5 provides broader compatibility with IgG antibodies. Zhang *et al.* developed the scCUT&Tag-pro platform, which integrates pAG-Tn5 with the 10x Genomics system to simultaneously profile protein–DNA interactions and cell surface protein abundance, enabling chromatin state inference through combinatorial HM patterns [14].

Three major strategies for single-cell partitioning and barcoding in scCUT&Tag are nano-well-based (e.g. ICELL8), droplet-based (e.g. 10x Genomics), and split-pool-based methods (Fig. 1). Among them, droplet-based 10x Genomics is now the most widely used, offering compatibility with pA/G-Tn5 and nanobody-Tn5 [5, 14, 35]. Especially, a very recent study applied droplet-based scCUT&Tag to profile G4 structures, non-canonical DNA secondary structures increasingly recognized for their crucial roles in transcription [52], at single-cell resolution in mESCs, MEFs, and neural cells [47]. Integration of single-cell G4 with gene expression profiles revealed a strong correlation between G4 formation and transcriptional activity, highlighting the potential for dissecting G4 heterogeneity in complex tissues and illuminating the epigenomic landscape of the developing brain. The nano-well strategy, typically implemented via the ICELL8 platform, allows physical isolation of single cells and imaging-based confirmation prior to library preparation [26]. Kaya-Okur *et al.* used this approach in the original scCUT&Tag study [13], while Janssens

et al. later introduced single-cell combinatorial indexing CUT&Tag (sciCUT&Tag), which streamlined the ICELL8 workflow to enhance throughput, resolution, and cost-efficiency [7]. Split-pool barcoding, originally developed for scRNA-seq, involves successive rounds of distributing cells into 96-well plates with sequential barcode incorporation [53]. Reads sharing identical barcode combinations are computationally assembled to reconstruct single-cell transcriptomes. More recently, this strategy has been extended to scCUT&Tag applications, e.g. the uCoTarget employs split-pool barcoding to achieve ultrahigh-throughput, multimodal profiling of HMs (e.g. H3K27ac, H3K4me3, H3K4me1, H3K36me3, H3K27me3) in nearly 20 000 single cells [6]. This versatile approach enables comprehensive reconstruction of the single-cell epigenomic and transcriptomic landscape.

Typical experimental workflow for scCUT&Tag

The general experimental workflow used across 21 reported scCUT&Tag studies is summarized in Fig. 1. The workflow begins with the isolation of cells from cultured human or mouse cell lines or tissue samples. Cells are incubated with antibodies specific to chromatin features of interest, including HMs, TFs, Pol II, or DNA structural elements such as G4s. Tagmentation is subsequently performed using Tn5 transposase-based complexes (e.g. pA-Tn5, pAG-Tn5, or nanobody-Tn5) to insert sequencing adapters at target sites. Single-cell barcoding and partitioning are achieved through one of three main strategies: nano-well platforms (e.g. ICELL8), droplet-based systems (e.g. 10x Genomics), or split-pool combinatorial indexing. Libraries are PCR amplified, sequenced, and mapped to reference genomes

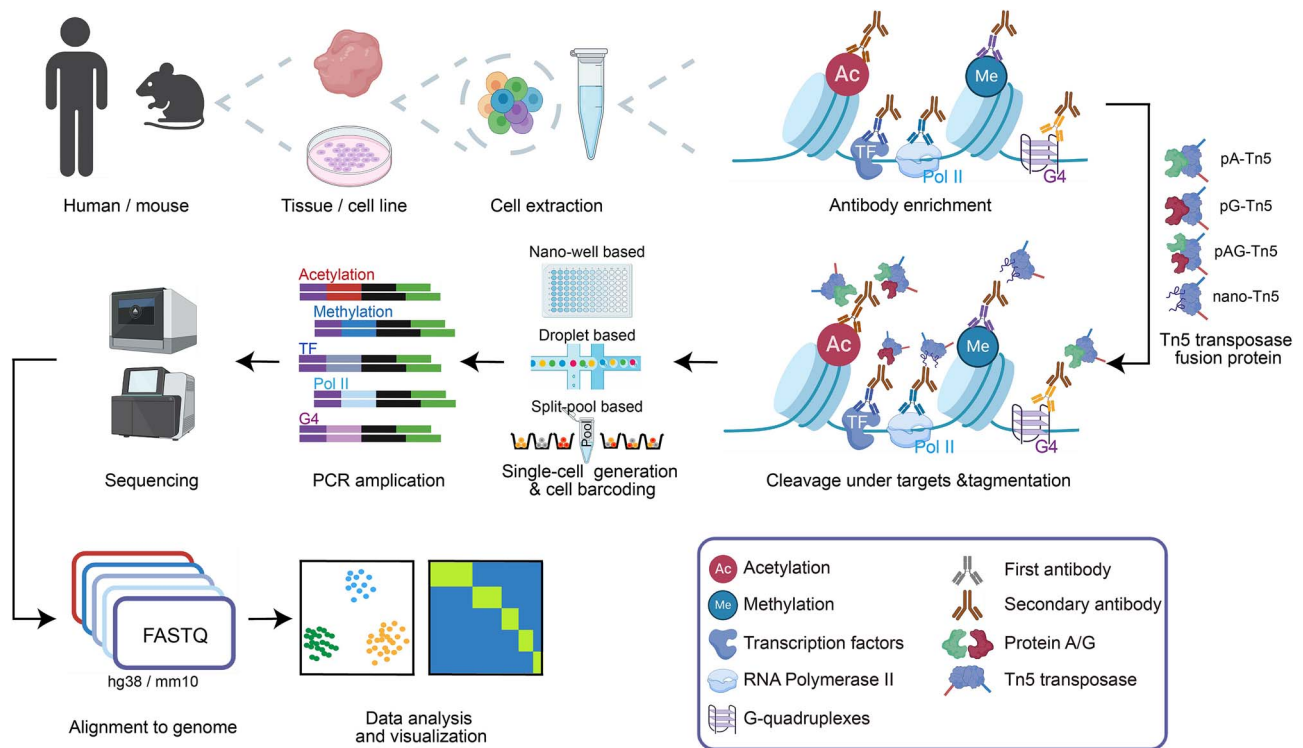


Figure 1 Overview of the general experimental workflow of scCUT&Tag. The process typically begins with the extraction of cells from human or mouse tissues or from established cell lines. Cells are then incubated with antibodies targeting specific chromatin features such as HMs, TFs, pol II, or G4 structures. Tagmentation is carried out using Tn5-based transposase complexes (e.g. pA/G-Tn5 or nanobody-Tn5). Single-cell generation and barcoding are achieved using nano-well, droplet-based, or split-pool platforms. The libraries are then amplified via PCR, sequenced, and aligned to reference genomes (e.g. hg38 or mm10), yielding data for downstream analysis and visualization.

(e.g. hg38 or mm10), providing high-resolution single-cell chromatin profiles for downstream analysis.

Typical computational analysis for scCUT&Tag

Despite substantial differences in their biological objectives, current scCUT&Tag data analyses still predominantly rely on computational strategies originally developed for scATAC-seq [14, 34, 36]. Therefore, a typical scCUT&Tag analysis pipeline usually includes matrix generation, quality control and filtering, peak calling, global correlation analysis, dimensionality reduction and cell clustering, multiomics integration, and additional downstream analyses (Fig. 2).

Data matrix generation

The 10x Genomics platform is currently one of the most widely used systems for scCUT&Tag library construction [14, 54]. In most studies, the chromium single-cell ATAC sequencing pipeline is adapted for scCUT&Tag, with paired-end sequencing performed on Illumina platforms such as the NextSeq 550 or 6000 system [14, 42]. After demultiplexing, cell barcodes are extracted from index reads to assign sequencing reads to individual cells [16, 34, 55]. Adapter sequences are then removed using trimming tools such as Cutadapt [16], followed by read alignment to the reference genome (e.g. hg38 or mm10) using aligners such as minimap2 or BWA [16, 55]. Reads of low quality (e.g. $Q < 7$ or $MAPQ < 30$) and short fragments (e.g. < 100 bp) are filtered out using tools like NanoFilt or SAMtools [16]. After alignment, PCR duplicates are removed, and the remaining reads are converted into

fragment files that record the genomic coordinates of DNA insertion sites tagged by each cell barcode [14, 16]. These fragment files are then used for downstream matrix construction, peak calling, and chromatin accessibility quantification [5, 34, 55]. Alternatively, by using a streamlined pipeline such as cellranger-atac, the raw sequencing data (FASTQ files) can be automatically processed to generate the data matrices required for downstream analyses [56].

Peak calling

In epigenomic analysis, peaks denote genomic regions with statistically significant enrichment of aligned fragments, marking potential TFs/HMs loci, and serving as focal points for downstream analyses (Fig. 2). In scCUT&Tag, peak calling internally relies on MACS2 [57], often executed via Signac or ArchR [58, 59], e.g. CallPeaks function in Signac with either default or customized parameters. Due to the sparsity of scCUT&Tag data, peak calling is generally performed in a pseudobulk manner by aggregating fragments across cells, rather than at the single-cell level. Peaks can be called globally across all cells, e.g. the nano-CT study identified H3K27ac peaks using merged pseudobulk BAM files by MACS2 with custom parameters ($-\text{keep-dup} = 1$ $-\text{llocal} 100000$ $-\text{min-length} 1000$ $-\text{max-gap} 1000$ $-\text{broad-cutoff} = 0.1$) [34]. Other broad HMs such as H3K27me3 were also defined using MACS2's $-\text{broad}$ mode and merged within a 1-kb window [6]. Alternatively, cluster-level peak calling enables the identification of cell type-specific regulatory regions. For example, the scG4 study applied MACS2 separately to each cluster with default settings [47], while Zhang *et al.* analyzed peaks for six distinct HMs across multiple cell

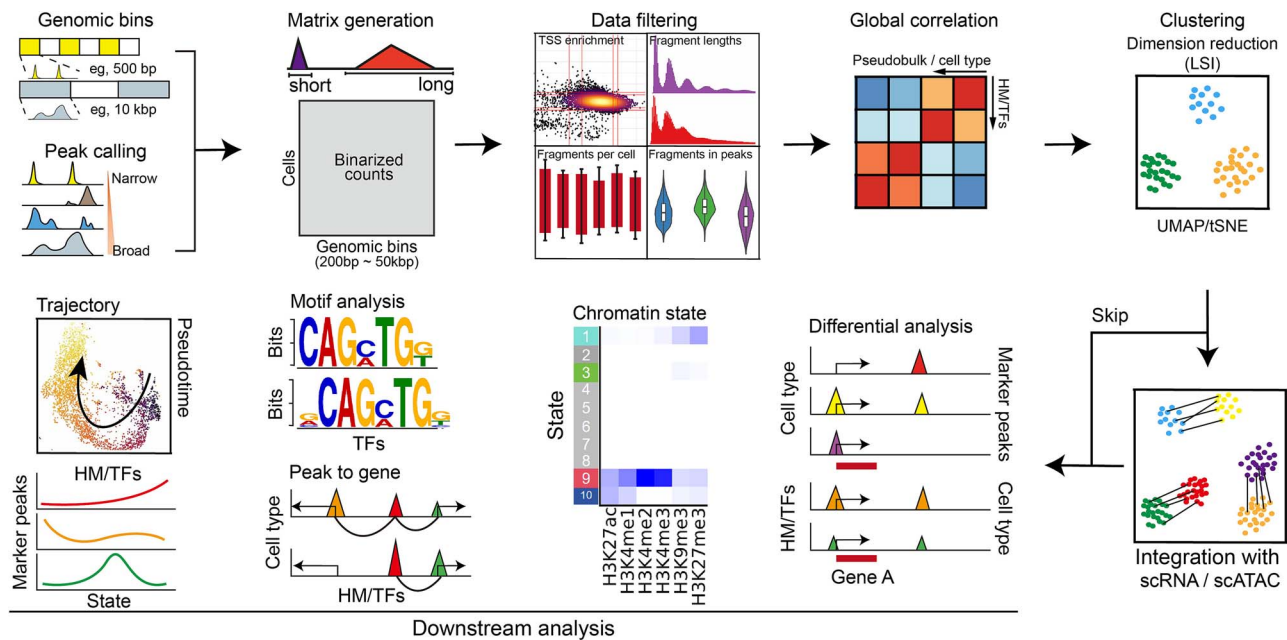


Figure 2 Overview of the computational workflow for scCUT&Tag data analysis. The analysis pipeline comprises seven major steps. Step 1: Feature definition. Genomic regions are defined either by fixed-size bins or through peak calling (e.g. MACS2) to construct candidate features for quantification. Step 2: Data matrix generation. Fragment files (e.g. from cell ranger ATAC or custom workflows) are used to count fragments per cell per feature (peak or bin), resulting in a sparse cell-by-feature matrix. Step 3: Data filtering. Quality control is performed to remove low-quality cells or outliers based on metrics such as nucleosomal signal, TSS enrichment score, and fragment counts. Step 4: Global correlation. Pseudobulk-level Pearson correlation analyses are often used to evaluate consistency between histone marks, cell types, or between scCUT&Tag and bulk datasets. Step 5: Dimensionality reduction and clustering. Normalization (e.g. TF-IDF) and reduction methods such as LSI or PCA are followed by clustering (e.g. UMAP or Louvain algorithm). Step 6: Integration with scRNA-seq or scATAC-seq (optional). Multiomic alignment can facilitate accurate cell-type annotation and regulatory interpretation. Step 7: Downstream analyses. This includes five major modules: (i) differential peak analysis; (ii) chromatin state inference; (iii) motif enrichment and TF activity inference; (iv) peak-to-gene linkage analysis; (v) pseudotime trajectory analysis.

types [14]. However, peak calling is not strictly required and can be substituted by a genome-wide binning approach using fixed-size windows (e.g. 5 kb), similar to strategies used in Hi-C analysis [60]. When peak calling is used to generate the feature matrix, it must be performed beforehand. Otherwise, with a bin-based approach, peak calling can still support downstream analyses.

Quality control and filtering

Quality control (QC) assessment is essential prior to downstream analysis. Analysis tools originally developed for scATAC-seq, such as the R-based Signac/Seurat and the Python-based SnapATAC, are commonly used to evaluate scCUT&Tag data [58, 61] (Table S1 and Fig. S2). Key QC metrics include insert size distribution, transcription start site (TSS) enrichment scores, fragment counts per cell, and the fraction of fragments in peaks (FRiP). Cells with outlier values in these QC metrics are filtered out to ensure data reliability [58]. For instance, FRiP reflects the proportion of mapped reads located within identified peak regions. A higher FRiP score indicates a better signal-to-noise ratio and overall data quality. QC thresholds for scCUT&Tag are often empirically determined and may differ from those used for standard scATAC-seq. Notably, advanced QC metrics developed for bulk ChIP-seq, such as the cross-correlation-based normalized strand coefficient and relative strand correlation [62], have not been adopted for scCUT&Tag, primarily due to the limited sequencing depth. Here, we performed a demo analysis using the widely adopted Signac package on GSE195725 (Fig. 3A), which employs scCUT&Tag to profile six different HMs in human PBMC [14]. Different HMs exhibited substantial variation in

total fragment counts (left), non-zero fragments per cell (middle), and unique non-zero fragments per cell (right), indicating the need for mark-specific QC thresholds. Six HMs all showed median FRiP scores >0.7, indicating high data quality (Fig. 3B) [14].

Global correlation analysis

Biological evaluation of scCUT&Tag features often involves global correlation analysis to assess relationships among different HMs/TFs [5, 6, 14, 16, 35, 55], examine the concordance of the same HMs/TFs across different cell types [5, 35], or evaluate the consistency between scCUT&Tag pseudobulk profiles and corresponding bulk datasets (e.g. ChIP-seq from ENCODE) in the same cell types [5, 6, 14, 16, 35, 55]. A common approach involves calculating Pearson [35, 39] or Spearman [35, 39] correlation coefficients on pseudobulk Bigwig files [16, 34, 39, 55] to validate scCUT&Tag data quality, find associations between HMs/HFs, and characterize cell type-specific regulatory patterns.

In the CoTACIT study [39], global Pearson correlations were used to compare pseudobulk signals of H3K27ac, H3K4me3, H3K36me3, and H3K27me3 with the corresponding bulk ENCODE data. Moreover, correlations for H3K4me3, H3K9me3, and H3K27ac across embryonic stages (2-cell, 4-cell, 8-cell, morula, and blastocyst stages) revealed dynamic chromatin remodeling during development [39]. Xiong *et al.* demonstrated that CoTarget-generated H3K27ac and H3K27me3 profiles exhibited high concordance with their respective bulk ChIP-seq data, whereas H3K27ac and H3K27me3 themselves were negatively correlated [6]. The scCUT&Tag-pro study [14] found high concordance between pseudobulk profiles of six HMs and corresponding



Figure 3 Representative analysis for scCUT&Tag data. (A) Data quality metrics for 6 HMs. (B) Fraction of reads in peaks (FRiP) values across 6 HMs. (C) Global correlations among the 6 HMs. (D) UMAP plots of the 6 HMs using latent semantic indexing (LSI, left), weighted nearest neighbors (WNN) integration with scRNA-seq (middle), and predicted cell types (right). (E) Visualization of scCUT&Tag signal intensity together with mRNA and protein levels, at the CD8A locus across B cell, CD4⁺ T cell, CD8⁺ T cell, dendritic cell (DC), and monocyte groups. (F) Fragment heatmaps show H3K27ac occupancy of five cell populations (x-axis) at the cell type-specific H3K27ac peaks (y-axis). (G) Chromatin states inferred by ChromHMM, which was applied to pseudobulk tracks for six histone marks across all cell types, including CD8⁺ T cells, CD4⁺ T cells, and monocytes. The states are broadly categorized into five functional groups. (H) Motif enrichment analysis for T/NK cells (top) and monocytes (bottom), highlighting distinct transcription factor binding signatures. (I) Peak-to-gene linkage analysis of the CD5 gene, showing associated H3K27ac signals across cell types.

bulk CUT&Tag data through global correlations, and assessed cell-type specificity by comparing H3K4me1 pseudobulk profiles across cell types [14]. Similarly, the scG4 study [47] applied Pearson correlation to compare G4 pseudobulk signals between mESC and MEFs cell types, and with corresponding bulk G4 datasets, demonstrating cell-type specificity and data robustness. An example correlation analysis of six HMs (Fig. 3C) showed that the active mark H3K27ac displayed strong positive correlations with H3K4 methylation (H3K4me1/2/3) but negatively correlated with the repressive mark H3K27me3 [5, 14].

Normalization, dimensionality reduction, and clustering

Prior to dimensionality reduction, the peak-by-cell or bin-by-cell count matrix should be normalized. Given the similar data structure to scATAC-seq, normalization strategies developed for scATAC-seq are generally applicable to scCUT&Tag analysis. One of the most commonly used normalization approaches is “term frequency-inverse document frequency” (TF-IDF), originally adopted from information retrieval, which adjusts for both sequencing depth and chromatin signals (Fig. 4 and Table S2) [63]. Following TF-IDF normalization, singular value decomposition (SVD) is performed for dimensionality reduction [63]. This two-step process, TF-IDF followed by SVD, is commonly referred to as latent semantic indexing (LSI) (Table S3) [58]. Within the Signac framework, LSI is widely utilized to reduce the dimensionality of scATAC-seq peak matrices and is equally applicable to scCUT&Tag data [58]. This approach can also be implemented in ArchR, another popular scATAC-seq analysis toolkit (Table S1) [59]. Once the reduced dimensions are obtained, clustering is typically performed using algorithms such as Louvain or Leiden (Table S4), which identify groups of cells with similar chromatin profiles. Subsequently, UMAP (Uniform Manifold Approximation and Projection) is commonly used to project the data into a low-dimensional space for visualization. UMAP is widely applied in both scRNA-seq and scATAC-seq analyses, and is similarly effective in visualizing scCUT&Tag-derived clustering structures (Table S3) [58, 61]. Most analysis frameworks integrate dimensionality reduction and clustering into a streamlined workflow. For instance, ArchR provides functions such as `addIterativeLSI`, `addClusters`, and `addUMAP` to perform these steps sequentially and efficiently [59].

Integration with scRNA-seq or scATAC-seq

To improve the accuracy of cell-type annotation, scCUT&Tag data can be optionally integrated with other modalities, particularly when both datasets originate from the same tissue or cell population [5, 16]. This multimodal integration of transcriptional level or chromatin accessibility aids in refining cell identity and validating regulatory elements inferred from scCUT&Tag data [36]. For example, in the scNanoSeq-CUT&Tag study by Li *et al.* [16], scRNA-seq data from mouse PBMCs were used to assist with cell-type annotation for scCUT&Tag profiling of five HMs, CTCF, and RAD21. Similarly, integration with scATAC-seq data can provide complementary insights into scCUT&Tag chromatin features [34]. In a recent study, distinct epigenomic states at the *Foxg1* locus were identified between astrocyte subtypes, with H3K27ac enrichment and chromatin openness observed in telencephalon astrocytes, but H3K27me3 deposition in non-telencephalon astrocytes. Integration with scATAC-seq confirmed differential accessibility at the same locus, supporting the scCUT&Tag-derived inference

and illustrating how multimodal data can enhance the interpretation of epigenetic regulation across cell types [34]. Although integration with scRNA-seq or scATAC-seq is not required for every study, incorporating such multimodal data when compatible datasets or matched samples are available may enhance biological interpretation and improve annotation accuracy. Figure 3D visualizes the UMAP-based clustering derived from chromatin profiles of six HMs (left). By integrating scCUT&Tag with scRNA-seq data, we can project single-cell HM signals onto the weighted nearest neighbor (WNN) embedding (middle). The right panel showed the corresponding cell-type annotations inferred from scRNA-seq, aligned within the same WNN space.

Downstream analyses

scCUT&Tag downstream analysis varies depending on the biological objectives but typically encompasses the following five major modules (Fig. 2). To clarify, the term “peak” in downstream analysis can refer either to MACS2-called peaks or to fixed-size genomic bins [14].

Differential peak analysis is the common application in both scCUT&Tag and scATAC-seq, involving the comparison of signal intensities across different cell types or conditions, i.e. identifying cell type-specific HMs/TFs peaks. An example region of differential peaks is shown in Fig. 3E, where multimodal features were plotted across five major cell types to illustrate the epigenomic (six HMs), transcriptomics, and proteomic profiles around CD8A gene. CD8A exhibited enriched peaks for active HMs, high RNA expression, and elevated protein abundance in CD8⁺ T cells, supporting cell type-specific regulation at this locus. Differential scCUT&Tag peaks can be visualized in a particularly informative way [5, 14], by plotting HMs/TFs signals from N cell clusters against the differential peaks identified in each cell type, generating an $N \times N$ aggregate signal profile [5]. Our demo analysis showed H3K27ac scCUT&Tag intensity across five cell populations at the loci corresponding to the top 50 differential peaks for each cell cluster, revealing distinct enrichment (Fig. 3F). For example, CD4⁺ and CD8⁺ T cells shared a subset of marker peaks, while CD8⁺ T cells and NK cells also exhibited partial overlap. In contrast, monocyte-specific marker peaks displayed weak or absent signals in T cells or NK cells, and low enrichment in dendritic cells, supporting their classification into distinct differentiation lineages.

Chromatin state inference relies on the simultaneous profiling of multiple HMs and structural proteins such as CTCF, thus providing capabilities unique to scCUT&Tag but not attainable with scATAC-seq. Applying the chromatin state discovery tool ChromHMM to scCUT&Tag facilitates the classification of combinatorial chromatin states (e.g. active enhancers or repressive domains), thereby enabling a refined interpretation of epigenetic landscapes (Table S1 and Fig. S2) [14]. Zhang *et al.* demonstrated that pseudobulk scCUT&Tag-pro data of six HMs support robust ChromHMM training to define consistent, cell type-specific chromatin states [14]. Similarly, CoTACIT applied ChromHMM on synthetic cells constructed along developmental pseudotime to reveal embryonic state transitions of promoters, enhancers, heterochromatin, and multivalent domains [39]. These studies highlight either static or dynamic regulatory features revealed by scCUT&Tag chromatin state annotations. To illustrate single-cell regulatory element resolution from scCUT&Tag data, we applied ChromHMM to human PBMC scCUT&Tag data. (Fig. 3G). Four major chromatin states were robustly identified: promoters, enhancers, repressive domains, and heterochromatin, consistent with those reported in the original study [14]. Furthermore, chromatin state

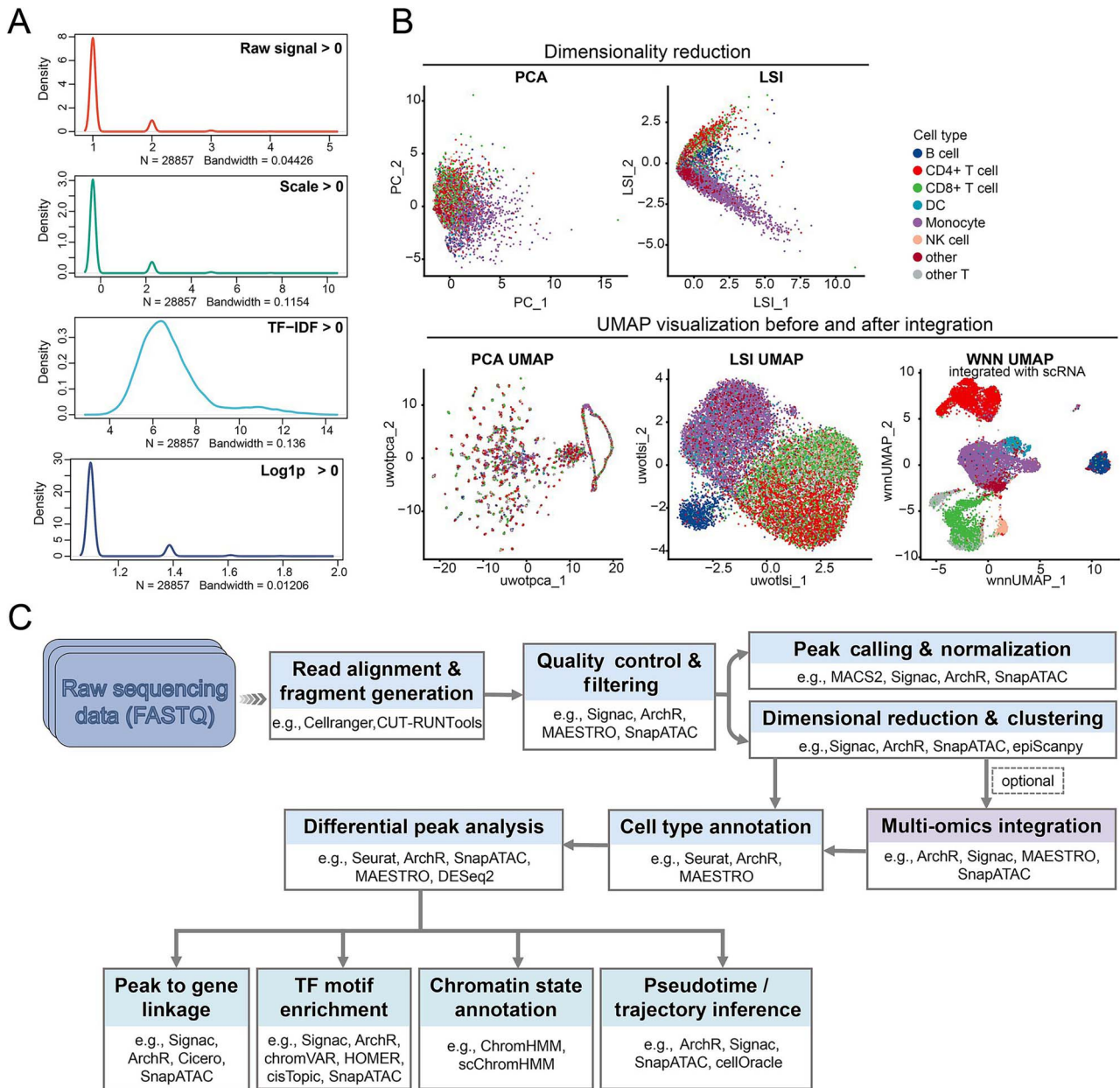


Figure 4 Illustration of normalization, dimensionality reduction for scCUT&Tag data. (A) Density plots showing the distribution of non-zero peak counts in the original raw matrix (top left) and after applying three different normalization strategies: term frequency-inverse document frequency (TF-IDF, top right), scale normalization (bottom left). Each feature (peak or bin) is centered and scaled by subtracting the mean and dividing by the SD across cells (i.e. scale (x, center = TRUE, scale = TRUE) in R), and log-transformation [log1p(x + 1), bottom right]. The analysis was performed on non-zero elements of the peak-by-cell matrix. Each method reveals distinct transformation effects on signal distribution. (B) Dimensionality reduction of scCUT&Tag data with principal component analysis (PCA, top left) and latent semantic indexing (LSI, top middle), and UMAP visualization before (bottom left and middle) and after weighted nearest neighbor (WNN) integration with scRNA-seq (bottom right). Cell-type annotations are indicated as labeled in the figure. (C) Overview of analytical modules in scCUT&Tag and their corresponding commonly used tools. This schematic summarizes key analysis steps from preprocessing to downstream analyses in scCUT&Tag, listing representative tools for each stage.

profiles varied across specific cell types. CD8⁺ T cells, CD4⁺ T cells, and monocytes exhibited distinct enrichment patterns in promoter and enhancer regions, whereas monocytes showed broader coverage of repressive domains, suggesting cell type-specific epigenomic regulation (Fig. 3G).

Motif analysis enables the identification of potential TF binding sites associated with cell type-specific chromatin accessibility. scCUT&Tag-pro modified the chromVAR algorithm to correlate TF motifs with inferred chromatin states [14]. In CoTACIT, ChromHMM-defined

regions were intersected with ATAC-seq peaks, followed by HOMER motif enrichment to refine TF predictions [39]. Wu *et al.* applied ArchR's addDeviationsMatrix function to calculate motif deviation scores [55], while Bartosovic *et al.* used the MEME suite and matched results to JASPAR or published ChIP-seq datasets [5]. These approaches exemplify how motif detection in scCUT&Tag can effectively link HMs to TF-driven regulatory programs. In Fig. 3H, differential H3K27ac peak analysis between T/NK cells and monocytes revealed common enrichment of E2F1/3/4, as well as distinct enrichment of HOXC8

in T/NK cells and IRF6 in monocytes, suggesting cell type-specific regulatory programs.

Peak-to-gene link analysis infers potential regulatory relationships between chromatin features (typically peaks from HM data) and nearby genes, typically by correlating signal intensity at regulatory regions with gene expression or activity. This approach is particularly powerful in scCUT&Tag, enabling identification of putative enhancer-promoter interactions and cell type-specific regulatory elements. Bartosovic *et al.* employed Cicero [64] to predict distal enhancers of the Sox10 gene in oligodendrocyte lineage cells using H3K27ac scCUT&Tag data, linking distal H3K27ac-enriched peaks to the Sox10 promoter and illustrating enhancer activity mapping at single-cell resolution [5]. In the uCoTarget study, Cicero was used to predicted enhancer co-binding patterns using H3K27ac and H3K4me3 scCUT&Tag data [6], revealing mesoderm-specific SMAD3 enhancers with strong co-accessibility signals that were diminished in HSPCs. The study also showed that H3K27ac outperformed H3K4me3 in predicting enhancer-gene regulatory links, highlighting mark-specific sensitivity [6]. We illustrated the peak-to-gene linkage approach using CD5, an immune-promoting gene, showing H3K27ac peaks at its regulatory regions and their signal strength across five cell types (Fig. 3I) [65–67].

Pseudotime trajectory analysis uncovers dynamic HMs/TF binding changes during lineage development and differentiation (Fig. 2) [5, 6, 14, 34, 35, 55]. Wu *et al.* investigated H3K27me3 scCUT&Tag in glioblastoma using pseudotime trajectories to uncover progressive silencing of TF motifs. Distinct silencing patterns along two differentiation paths were revealed, including shared (e.g. DNMT1) and trajectory-specific TF motifs (e.g. HES5 versus GATA6), indicating epigenetic reprogramming during tumor evolution [55]. Bartosovic *et al.* applied pseudotime analysis to H3K4me3 scCUT&Tag in oligodendrocyte precursor cells and mature oligodendrocytes (MOLs). They observed a gradual broadening of H3K4me3 signals at MOL promoters along the trajectory, reflecting chromatin maturation during lineage progression [5]. The uCoTarget study used pseudotime analysis of H3K27ac scCUT&Tag during mesoderm-to-HSPC differentiation to identify dynamic enhancers linked to lineage-specific TFs (e.g. RUNX1), showing that enhancer activation following TF binding reflects the temporal sequence of epigenomic changes in hematopoiesis [6]. Thus, scCUT&Tag trajectories reveal chromatin dynamics underlying lineage specification, enhancer activation, and disease progression.

scChIP-seq analyses

scChIP-seq follows an analytical workflow that partially overlaps with that of scCUT&Tag [68, 69]. Briefly, reads are demultiplexed, aligned to a reference genome, and deduplicated to generate a cell-by-bin count matrix [68]. Low-quality cells are filtered and normalized for sequencing depth, followed by peak calling to identify enriched chromatin regions [28]. Dimensionality reduction and clustering methods such as PCA or t-SNE are then applied to identify subpopulations [68, 69], and differential analysis is performed to detect regions with variable HM levels across clusters or conditions [68].

Critical analytical considerations for scCUT&Tag

Cell-by-peak or cell-by-bin matrix?

In scCUT&Tag analysis, sequencing reads are first converted into fragment coverage profiles across either fixed-size genomic bins or called peaks, to construct a cell-by-feature matrix used for all

downstream analyses. Among published scCUT&Tag studies, the scCUT&Tag-pro [14], CoTACIT [39], NTT-seq [35], sciCUT&Tag [7], and scNanoSeq-CUT&Tag [16] studies have utilized cell-by-bin matrices, typically based on 5-kb bins. In contrast, studies such as nano-CT [34], scG4 [47], uCoTarget [6], and sn-m6A-CT [36] constructed cell-by-peak matrices, where features are defined by peak calling. The primary consideration in choosing between these approaches is data sparsity. For datasets with low read coverage, a cell-by-bin matrix, particularly with larger bin sizes (e.g. 50 kb), can help capture more fragments and improve data density [7, 32]. However, this comes at the expense of reduced single-cell resolution and potentially introduces background noise. Conversely, cell-by-peak matrices are more sensitive to sequencing depth but tend to enhance analysis performance when sufficient coverage is available. This aligns with standard practices in scATAC-seq analysis, where cell-by-peak matrices are generally preferred [49].

Is TSS enrichment still good for scCUT&Tag?

TSS enrichment quantifies the accumulation of single-cell chromatin accessibility or HM/TF signals around transcription start sites [47, 58, 59, 70]. For scATAC-seq, TSS enrichment often positively correlate with transcription, as highly expressed genes have greater accessibility near their TSS [70]. For scCUT&Tag targeting promoter-affiliated markers such as H3K4me3 and G4, TSS enrichment can be interpreted similarly to scATAC-seq. For instance, the scG4 study reported average TSS enrichment scores of G4 ranging from 4 to 5, indicating moderate signal specificity around TSS [47]. However, in repressive-modification-based scCUT&Tag assays, such as H3K27me3, this relationship is reversed. Wu *et al.* have introduced a chromatin silencing score (CSS) based on H3K27me3 levels at TSSs, where active genes are expected to show low CSS values, reflecting an absence of silencing marks at their promoters [55]. In uCoTarget study, a comparative analysis of TSS signals for SOX2 revealed that highly expressed genes were associated with high H3K27ac but low H3K27me3 signals, demonstrating the regulatory balance between active and repressive HMs at gene loci [6]. The TSS enrichment patterns of TFs and structural proteins are more complex, e.g. ESR1, CTCF, and cohesin can function both at promoters and at distal regulatory elements [71, 72]. Therefore, TSS enrichment analysis should be tailored to the specific targets profiled by scCUT&Tag.

Does scCUT&Tag need PCR duplicate removal?

The initial implementation of scCUT&Tag did not explicitly report the removal of PCR duplicates during data preprocessing [13]. However, it has been noted that scCUT&Tag libraries typically contain 30%–50% duplicated fragments, as observed in studies such as nano-CT [34]. Subsequent protocols have addressed this limitation more explicitly. For example, TACIT, CoTACIT [39], and scG4 [47] adopted Picard tools for PCR duplicate removal. sciCUT&Tag [7] and scNano-CUT&Tag [16] employed SAMtools to filter duplicates from aligned BAM files. Additionally, for protocols leveraging the 10x Genomics Single Cell ATAC-seq kit, Cell Ranger ATAC can be used to perform duplicate removal as part of its alignment and filtering pipeline [55]. The following peak calling process also uses “-keep-dup = 1” to remove duplicates. While duplicates in bulk CUT&Tag data may reflect true biological relevance arising from repeated tagmentation at the same sites by chance [73], current scCUT&Tag analyses usually remove PCR duplicates.

How many dimensions for downstream analysis?

LSI [74] is widely employed in single-cell epigenomic analysis [63] for normalization and dimensionality reduction, capturing major variation while mitigating technical noise [58, 63]. LSI serves as a core step in popular tools like Signac [58] and ArchR [59]. In most scCUT&Tag studies, dimensions 2 to 30 are typically retained for cell clustering, visualization, and other downstream analysis, while the first component is usually excluded due to its strong correlation with sequencing depth [5, 55, 59]. This strategy has been adopted in multiple studies, including scCUT&Tag-pro [14], CoTAICIT [39], scG4 [47], nano-CT [34], and uCoTarget [6], where LSI enabled effective separation of cell types and identification of regulatory patterns. This approach is similarly applicable to scCUT&Tag data [58, 59].

Cell-type annotation in scCUT&Tag

Compared to gene expression or chromatin accessibility, which are directly linked to cell identity, HMs and TFs profiled by scCUT&Tag pose greater challenges for cell-type annotation. Currently, two major strategies are employed: (1) peak-based annotation using known functional markers [5, 55] and (2) integration with scRNA-seq data to enhance interpretability [6, 14, 34–36]. Wu *et al.* annotated clusters based on cell type-specific genes characterized by low chromatin silencing score of H3K27me3 at their TSS, including SOX2 for tumor/neural lineages, PTPRC for microglia, and MOBP for oligodendrocytes [55]. This peak-based approach enables direct mapping of known regulatory features to corresponding cell types. Alternatively, multimodal integration has been widely adopted to improve annotation accuracy. Studies including sn-m6A-CT [36], NTT-seq [35], nano-CT [34], and scCUT&Tag-pro [14] applied the WNN algorithm to jointly analyze scCUT&Tag with matched scRNA-seq data. This approach refines cell-type annotation by leveraging both transcriptional and chromatin features of HMs/TFs, particularly useful for distinguishing cell types at varying resolution levels.

Multimodal profiling by scCUT&Tag

Inspired by multimodal single-cell technologies that simultaneously profile chromatin accessibility and gene expression [75–77], recent methods have extended scCUT&Tag to enable multiomics measurements within the same cell. For example, scCUT&Tag-pro [14] and NTT-seq [35] combine CUT&Tag with surface protein profiling, capturing both protein–DNA interactions and cell surface markers at single-cell resolution. Moreover, functional chromatin state annotation relies on integrating multiple HMs. Yet, conventional scCUT&Tag typically measure only one HM per cell and cannot robustly capture combinatorial patterns at the single-cell level [14]. Very recently, technologies like nano-CT have enabled simultaneous profiling of chromatin accessibility and multiple HMs (e.g. H3K27ac and H3K27me3) within the same cell, as demonstrated in the juvenile mouse brain [34]. Furthermore, uCoTarget allows joint measurement of the transcriptome alongside multiple HM targets, providing a comprehensive view of gene regulatory landscapes at single-cell resolution [6].

Optimal parameters for scCUT&Tag analysis

A recent study benchmarked computational pipelines for single-cell HM data, offering the first comprehensive evaluation tailored to scCUT&Tag analysis [32]. Using simulated and public datasets across

multiple HMs (e.g. H3K27me3, H3K4me3, H3K27ac), it compared preprocessing, dimensionality reduction, and clustering strategies. The authors found that cell-by-bin matrices (e.g. 5 kb) are more robust under sparse signals, whereas cell-by-peak matrices perform better at high sequencing depth. TF-IDF normalization with LSI consistently yielded effective dimensionality reduction [58, 59, 78]. They emphasized that scCUT&Tag differs from scATAC-seq in signal distribution and sparsity, requiring HM-specific parameter tuning. The study also benchmarked Signac [58], ArchR [59], cisTopic [79], and SnapATAC [78], outlining their respective strengths and limitations for single-cell HM analysis. Based on these evaluations, the authors recommend starting with fixed-size bin matrices (200 kb for broad marks, 100 kb for narrow marks), avoiding traditional feature selection, limiting barcode filtering to removal of non-cell reads, applying TF-IDF normalization, and then using SVD or NMF for dimensionality reduction. It provides a standardized analytical foundation and highlights the urgent need for dedicated pipelines optimized for scCUT&Tag data.

Computational tools for scCUT&Tag analysis

Interpreting scCUT&Tag data relies on diverse computational tools, mostly adapted from scATAC-seq, covering data preprocessing to regulatory element inference. Seurat and Signac remain the most widely used analytical frameworks for scCUT&Tag analysis, often used together to enable integrated analysis of chromatin features. They have been applied in diverse biological contexts, including mendelian regulatory disorders [54], PBMC cellular heterogeneity [14], pre-implantation development in mouse embryos [39], G4 structure heterogeneity [47], and epigenetic regulation in the central nervous system [5]. Signac has supported the development of novel scCUT&Tag technologies including nano-CT [34], NTT-seq [35], uCoTarget [6], and sn-m6A-CT [36]. ArchR has been employed in studies focused on differentiation and tumor progression [55], providing effective analytical pipeline for advanced variants such as scNanoSeq-CUT&Tag [16] and sciCUT&Tag [7]. Based on primary objective, we categorized current scCUT&Tag tools into four types: data process pipelines and integration, functional element investigation and epigenomic annotations, gene regulatory networks and TF analysis, and extended application. Detailed descriptions and the summary table are provided in Table S1.

Normalization, dimensionality reduction, and cell clustering algorithms

Normalization, dimensionality reduction, and clustering in scCUT&Tag are tightly interconnected processes that work together to partition cells into biologically meaningful groups, addressing data sparsity and complex manifold geometry [58, 59, 78]. Normalization first strips away technical biases such as sequencing depth, library size, and batch effects to reveal true biological variation [58, 59, 78]. Dimensionality reduction then projects this normalized matrix into a low-dimensional space that preserves the major axes of biological variation while minimizing noise [32, 80]. Next, clustering groups cells based on distances or graph topology within the compressed space [32, 80]. Suboptimal performance in any step, e.g. incomplete normalization or noisy dimensionality reduction, can cause clustering to reflect technical artifacts instead of true cell states [49, 81–83]. Detailed descriptions and summary tables for

normalization, dimensionality reduction, and clustering are provided in [Tables S2, S3, and S4](#), respectively.

Comparison of scCUT&Tag and scATAC-seq analyses

Although scCUT&Tag utilizes different library construction protocols compared to scATAC-seq, both share a similar sequencing workflow to generate single-cell chromatin fragments [49, 50, 63, 84]. In terms of data formats, both techniques generate comparable sparse matrices of chromatin features across thousands of single cells, using common preprocessing algorithms for demultiplexing, adapter trimming, read alignment, read filtering, and fragment generation [49, 80, 84, 85]. For upstream analysis, scCUT&Tag and scATAC-seq rely on shared computational tools such as Signac [58], ArchR [59], and SnapATAC [78] ([Table S1](#)). TF-IDF normalization, LSI dimensional reduction, and UMAP visualization are standard components for both data types [58, 80, 86] ([Fig. 4](#)). Their shared analytical logic facilitates cross-technology applications and allows researchers to transfer expertise from scATAC-seq to scCUT&Tag with minimal adjustments [87].

However, a key difference arises because scATAC-seq employs “blind” Tn5 transposition, whereby the enzyme inserts adapters wherever chromatin is accessible. By contrast, scCUT&Tag uses an antibody-tethered pA-Tn5 fusion to target specific HMs/TFs, enriching exclusively for fragments adjacent to those marks [13]. Although CUT&Tag theoretically provides higher signal-to-noise ratios for profiling HMs/TFs compared to ChIP-seq, scCUT&Tag remains limited by relatively low per-cell coverage (typically $10^2\sim 10^4$ fragments) [7, 14, 34, 35], whereas scATAC-seq generally delivers genome-wide accessibility profiles at high resolutions, often yielding $>10^4$ fragments per cells [49–51]. Therefore, scCUT&Tag and scATAC-seq require distinct analytical considerations due to data sparsity and, primarily, the underlying biological differences between HMs/TFs and open chromatin [7].

Current scCUT&Tag analyses are usually based on a cell-by-bin matrix due to insufficient coverage [7, 14, 16, 35, 39], which brings a fundamental difference compared to scATAC-seq, whose cell-by-peak matrix is based on called peaks [80]. This draws attention to the fact that “peaks” mentioned in scCUT&Tag studies may refer to fixed-size rather than true peaks [14]. Mitochondrial genomes (mtDNA) are tiny and highly accessible, resulting in scATAC-seq libraries typically containing a higher fraction of mtDNA-mapped reads than scCUT&Tag libraries [88–90]. These mitochondrial fragments do not inform on nuclear chromatin state and are typically excluded in most studies [13, 91]. scATAC-seq data relies heavily on FRiP and TSS enrichment scores to identify bona fide nuclei, whereas cells with low FRiP (below 0.2) or TSS enrichment (below 6–8×) are considered almost entirely background [58, 91]. By contrast, appropriate thresholds for FRiP and TSS enrichment in scCUT&Tag vary substantially depending on the target factor. In other words, these thresholds are often context specific and must be determined on a case-by-case basis. Certain TFs such as RAD21 may yield highly concentrated signals resulting in elevated FRiP, whereas others may not [5]. The promoter mark H3K4me3 shows high TSS enrichment, while heterochromatin marks like H3K9me3 show minimal enrichment near TSS regions [14, 16]. In addition, footprints in scATAC-seq typically manifest as relatively broad, nucleosome-phased dips in Tn5 insertion frequency centered on bound TF motifs and flanked by peaks over linker DNA [92, 93]. These shallow signals result from widespread Tn5 insertions in

accessible chromatin, with steric hindrance at TF-bound sites. Consequently, robust footprint detection requires deep coverage and rigorous correction for sequence bias (e.g. HINT-ATAC, TOBIAS) [92, 94]. By contrast, TF events in scCUT&Tag appear as narrow, high-amplitude peaks exactly at the binding site. This is because the pA-Tn5 complex is directly tethered to the TFs/HMs, concentrating cleavage events at the edges of the epitope. This targeted cleavage requires only local background normalization and avoids the elaborate bias correction needed in scATAC-seq [13].

Another key analytical distinction between scATAC-seq and scCUT&Tag lies in their functional interpretability. Tn5 in scATAC-seq captures fragments from all open regions, providing a landscape of chromatin accessibility without regard to functional specificity, while scCUT&Tag selectively yields function-focused profiles including active marks (e.g. H3K27ac) [54], repressive marks (e.g. H3K27me3) [55], TFs (e.g. OLIG2, NR2F1) [5, 46], and structural proteins (e.g. RAD21) [5]. Active marks in scCUT&Tag can often be interpreted similarly to scATAC-seq, whereas signals from other regulators, particularly repressive marks, differ substantially and pose additional analytical challenges. Nonetheless, these differences may provide complementary information, enabling multilayered annotation of cell states [95]. One prominent application is chromatin state inference using multiple HMs (typically H3K4me1, H3K4me2, H3K4me3, H3K27ac, and H3K27me3), enabling annotation of active enhancers, repressed regions, or bivalent domains [14]. Tools like ChromHMM and scChromHMM utilize hidden Markov models to perform combinatorial segmentation, providing more nuanced insights into epigenetic regulation than accessibility-based clustering alone at the single-cell level [14]. HM-centric trajectory inference in scCUT&Tag tracks dynamic chromatin landscape changes during development or differentiation [34, 35]. For instance, temporal shifts in H3K27ac or H3K27me3 occupancy along pseudotime trajectories can reveal the emergence of oligodendrocyte lineage-specific regulatory programs [34]. Furthermore, the latest scCUT&Tag variants can detect multiple targets or other modalities within the same cell simultaneously, e.g. mRNA in uCoTarget [6], protein in scCUT&Tag-pro [14], or multiple HMs in nano-CT [34] and CoTACIT [39]. These variants have been applied to cell type-specific analysis of cell fate during pre-implantation development [39] and hematopoietic lineages [58]. These advances eliminate the need for complex cross-modality cell-matching analyses but increase analytical complexity. Recent work also extends scCUT&Tag to spatial epigenomics, profiling HMs *in situ* to reveal the spatial architecture of gene regulation in complex tissues such as the developing brain [90]. These applications showcase scCUT&Tag’s unique potential to dissect chromatin regulation beyond accessibility and advance understanding of transcriptional control in disease [96].

Integrative analysis of scCUT&Tag

Apart from challenges inherent to the scCUT&Tag method itself, computational difficulties also arise in integrative analysis, including conjoining multiple analytical steps, combining data across studies, harmonizing multiomics data, and jointly analyzing multiple markers within scCUT&Tag.

When combining many scCUT&Tag datasets across diverse experimental platforms, variations in output formats, sequencing depth, and the number of unique fragments per cell across different methods can introduce substantial batch effects [16, 34]. Moreover, differences in the FRiP across HMs/TFs, along with inconsistent genomic bin sizes

(ranging from 200 bp to 50 kb), further complicate the harmonization of data [7, 14, 16]. These inconsistencies hinder the development of unified analysis pipelines and limit cross-sample or cross-study comparisons.

Integrating scCUT&Tag with other single-cell omics relies heavily on expert-driven interpretation, which limits reproducibility and broader applications [5, 16]. Methods like WNN aid integration with scRNA-seq for cell-type annotation but may partly obscure native chromatin variation from HMs/TFs alone. As shown in Fig. 3D, the active HMs H3K27ac and H3K4me1/2/3 exhibit more clustered and coherent distribution patterns. These genes are not informative when determining cellular state from scRNA-seq data, yet their repressive chromatin landscape exhibits clear patterns of cell-type specificity [14]. The integration of scCUT&Tag with scRNA-seq, scATAC-seq, CITE-seq, and other single-cell omics is still in its early stages and requires substantial further development [6, 14, 34, 39]. Importantly, single-cell Hi-C is gradually becoming a feasible technology for uncovering 3D chromosomal organizations [97, 98], offering a powerful opportunity to deeply investigate and interpret scCUT&Tag data.

A unique challenge in scCUT&Tag compared to scATAC-seq or scRNA-seq is that, even within a single dataset, there is a need to integrate multiple chromatin features to fully illustrate the complexity of transcriptional machinery [34, 39]. Currently, ChromHMM and scChromHMM appear to be the only promising approaches, indicating substantial room for developing new methods to integrate multiple features within the same datasets [14, 99].

Another key challenge in integrating scCUT&Tag with other omics or multiple HMs/TFs is accurately matching different data types to the same cells. From the experimental development perspective, it can be addressed by simultaneously detecting multiple markers within the same cells, as demonstrated by uCotarget [6] and CoTACIT [39]. From the computational side, the prevailing strategy could involve constructing pseudobulk profiles or more refined “synthetic/digital” metacells to unify multisource signals [39]; however, a method specifically designed for scCUT&Tag is still lacking.

Discussions

Here, we provided a comprehensive overview of current scCUT&Tag studies from a computational perspective. We summarized representative datasets spanning diverse chromatin features, surveyed studies involving various transposases and platforms, and outlined a typical computational workflow from raw data processing to functional downstream analyses. We also discussed critical analytical considerations, described commonly used tools and related algorithms, highlighted analytical features in comparison to scATAC-seq, and discussed challenges in integrative analysis. These foundations provide a framework for understanding current capabilities and inform the development of future analytical strategies. To facilitate practical implementation, we provide a schematic overview of common analytical steps for scCUT&Tag analysis, along with the typical tools used at each stage (Fig. 4C).

Despite rapid advances in scCUT&Tag in detecting HMs, TFs, and other epigenomic features at the single-cell level, the relatively high cost of current protocols and the technical complexity of sample preparation still hinder widespread adoption [7, 16, 35]. Processing small-volume samples, fixed tissues, and precious clinical specimens remain challenging and unreliable [90]. Although scCUT&Tag holds substantial promise for translational research, its clinical applications remain relatively limited. One study reported AutoCUT&Tag, in which

the integration of automated and single-cell CUT&Tag enables the identification of epigenomic heterogeneity within leukemia patient samples and predicts their sensitivity to therapeutic agents [40]. More recently, Wang *et al.* introduced scCPA-Tag, enabling the characterization of heterogeneous cell populations in hepatocellular carcinoma (HCC) [43]. Their epigenetic profiling revealed distinct HCC subtypes associated with patient prognosis, thereby deepening our understanding of chromatin regulatory mechanisms in HCC [43]. With the growing availability of clinical datasets and ongoing improvements in sensitivity and throughput, scCUT&Tag is expected to play an increasingly important role in clinical and translational research. Furthermore, Deng *et al.* reported spatial CUT&Tag, which integrates *in situ* CUT&Tag chemistry, microfluidic deterministic barcoding, and next-generation sequencing to achieve genome-wide spatial mapping of HMs [90]. This approach enabled the dissection of epigenetic regulation during mouse cortical development and revealed spatially organized cell populations defined by HM patterns. More recently, the same group developed spatial CUT&Tag-RNA-protein-seq (spatial CTRP-seq), which enables the simultaneous spatial profiling of HMs (e.g. H3K27me3), the transcriptome, and proteins at spatial single-cell resolution, providing new insights into brain development [100]. These advances highlight the potential of spatial integration to further extend CUT&Tag-based technologies for studying development and disease. However, current spatial CUT&Tag methods require sophisticated instrumentation and remain technically demanding for routine application [96, 100, 101]. Looking ahead, the continued development of scCUT&Tag technologies will benefit from parallel progress in both experimental and computational fronts. Experimentally, innovations aimed at reducing input requirements, increasing throughput, and enhancing compatibility with challenging clinical samples will broaden the accessibility of scCUT&Tag in translational research [46, 54, 55].

Computationally, there is an urgent need for scalable, automated, and user-friendly pipelines capable of accommodating data heterogeneity, supporting integrative multiomics analyses, and facilitating chromatin state annotation. Several computational frameworks have been developed, including Signac [58], ArchR [59], and SnapATAC2 [102], but they largely operate as independent end-to-end ecosystems. These tools differ substantially in computational efficiency, memory usage, and scalability. SnapATAC2, implemented in Python, demonstrates improved computational efficiency and scalability compared with its predecessor and other R-based frameworks. On the other hand, Signac and ArchR provide comprehensive and user-friendly workflows within the R environment but typically require greater computational resources [102]. These differences reflect distinct design philosophies rather than simple performance trade-offs, and tool selection therefore depends on dataset size, computational environment, and analytical objectives. Cross-tool integration remains technically challenging, particularly for new users, highlighting the need for interoperable and modular pipelines that leverage the strengths of existing tools to enhance reproducibility and standardization in scCUT&Tag data analysis. Especially, effective scCUT&Tag tools should be highly customizable to address diverse targets, including active, repressive, and indirect epigenomic regulators with context-specific functions.

Recent advances in foundation models and multimodal learning are reshaping single-cell epigenomic analysis and are likely to influence future scCUT&Tag workflows. Foundation models such as Epi-Agent [103], EpiFoundation [104], and Atacformer [105], pretrained on large-scale single-cell chromatin datasets (primarily scATAC-seq), learn transferable representations of cells and regulatory elements

that enable robust feature extraction, cell-type annotation, denoising or imputation, and cross-dataset mapping. In addition, multimodal transformers can integrate multiple regulatory modalities to improve generalization across cellular contexts and support principled regulatory prediction; e.g. EpiBERT jointly models DNA sequence and cell type-specific chromatin accessibility [106]. Combining large-scale pre-training with multimodal information may improve signal recovery and regulatory interpretation in scCUT&Tag. However, its practical adoption will require larger, standardized scCUT&Tag datasets and careful handling of biases introduced by different marks, antibodies, and experimental platforms [44].

As community standards mature and multimodal single-cell datasets continue to grow, scCUT&Tag is expected to play a central role in the field of single-cell epigenomics, offering a powerful way for investigating gene regulatory mechanisms, cell identity, and disease-associated epigenetic variation.

Key Points

- This work presents the first timely and essential synthesis that provides a comprehensive overview of scCUT&Tag-related studies from a bioinformatics perspective.
- We survey representative studies and their data characteristics of currently available scCUT&Tag research, outlining experimental designs and typical computational workflows.
- Focusing on scCUT&Tag computational analysis, we discuss critical analytical considerations, summarize commonly used computational tools and key algorithms, highlight distinctions from scATAC-seq analysis, and address major challenges in integrative analysis.

Author contributions

Jun Wu led the visualization work, performed demo analysis, participated in data collection and tabulation, contributed to data verification, and was a major contributor to manuscript writing. Md Wahiduzzaman was primarily responsible for data collection and tabulation, contributed to visualization, and co-wrote the manuscript. Pengfei Yin provided suggestions and co-wrote the manuscript. Puxuan Sun supported demo analysis and contributed to data verification. Haoping Chen participated in the visualization. Yongwen Ding assisted with data verification. Jiankang Wang conceived and supervised the project, organized the whole manuscript, wrote the core perspectives, and revised the final manuscript. All authors read and approved the final manuscript.

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Conflicts of interest

None declared.

Funding

This work was supported by the National Natural Science Foundation of China (32400526), Hunan Provincial Natural Science Foundation

of China (2024JJ6131), the Changsha Municipal Natural Science Foundation (kq2402059), the Guangdong Basic and Applied Basic Research Foundation (2023A1515110873, 2025A1515012822), and the Fundamental Research Funds for the Central Universities (541109030087).

Data availability

No original data were generated in this study.

References

1. Kouzarides T. Chromatin modifications and their function. *Cell* 2007;**128**:693–705. <https://doi.org/10.1016/j.cell.2007.02.005>.
2. Wu YL, Lin ZJ, Li CC *et al*. Epigenetic regulation in metabolic diseases: mechanisms and advances in clinical study. *Signal Transduct Target Ther* 2023;**8**:98. <https://doi.org/10.1038/s41392-023-01333-7>
3. Minarovits J, Banati F, Szenthe K *et al*. Epigenetic regulation. *Adv Exp Med Biol* 2016;**879**:1–25. https://doi.org/10.1007/978-3-319-24738-0_1
4. Eustermann S, Patel AB, Hopfner KP *et al*. Energy-driven genome regulation by ATP-dependent chromatin remodellers. *Nat Rev Mol Cell Biol* 2024;**25**:309–32. <https://doi.org/10.1038/s41580-023-00683-y>
5. Bartosovic M, Kabbe M, Castelo-Branco G. Single-cell CUT&tag profiles histone modifications and transcription factors in complex tissues. *Nat Biotechnol* 2021;**39**:825–35. <https://doi.org/10.1038/s41587-021-00869-9>
6. Xiong H, Wang Q, Li CC *et al*. Single-cell joint profiling of multiple epigenetic proteins and gene transcription. *Sci Adv* 2024;**10**:eadi3664. <https://doi.org/10.1126/sciadv.adi3664>
7. Janssens DH, Greene JE, Wu SJ *et al*. Scalable single-cell profiling of chromatin modifications with sciCUT&tag. *Nat Protoc* 2024;**19**:83–112. <https://doi.org/10.1038/s41596-023-00905-9>
8. Haws SA, Simandi Z, Barnett RJ *et al*. 3D genome, on repeat: higher-order folding principles of the heterochromatinized repetitive genome. *Cell* 2022;**185**:2690–707. <https://doi.org/10.1016/j.cell.2022.06.052>
9. Lambert SA, Jolma A, Campitelli LF *et al*. The human transcription factors. *Cell* 2018;**172**:650–65. <https://doi.org/10.1016/j.cell.2018.01.029>
10. Mayran A, Drouin J. Pioneer transcription factors shape the epigenetic landscape. *J Biol Chem* 2018;**293**:13795–804. <https://doi.org/10.1074/jbc.R117.001232>
11. He J, Huo X, Pei G *et al*. Dual-role transcription factors stabilize intermediate expression levels. *Cell* 2024;**187**:2746–2766.e25. <https://doi.org/10.1016/j.cell.2024.03.023>
12. Ameen M, Sundaram L, Shen M *et al*. Integrative single-cell analysis of cardiogenesis identifies developmental trajectories and non-coding mutations in congenital heart disease. *Cell* 2022;**185**:4937–4953.e23. <https://doi.org/10.1016/j.cell.2022.11.028>
13. Kaya-Okur HS, Wu SJ, Codomo CA *et al*. CUT&tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 2019;**10**:1930. <https://doi.org/10.1038/s41467-019-09982-5>
14. Zhang B, Srivastava A, Mimitou E *et al*. Characterizing cellular heterogeneity in chromatin state with scCUT&tag-pro. *Nat Biotechnol* 2022;**40**:1220–30. <https://doi.org/10.1038/s41587-022-01250-0>
15. Rubin AJ, Parker KR, Satpathy AT *et al*. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* 2019;**176**:361–376.e17. <https://doi.org/10.1016/j.cell.2018.11.022>

16. Li Q, Guo Y, Wu Z *et al*. scNanoSeq-CUT&tag: a single-cell long-read CUT&tag sequencing method for efficient chromatin modification profiling within individual cells. *Nat Methods* 2024;**21**:2044–57. <https://doi.org/10.1038/s41592-024-02453-w>
17. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009;**10**:669–80. <https://doi.org/10.1038/nrg2641>
18. Visel A, Blow MJ, Li Z *et al*. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 2009;**457**:854–8. <https://doi.org/10.1038/nature07730>
19. Johnson DS, Mortazavi A, Myers RM *et al*. Genome-wide mapping of in vivo protein-DNA interactions. *Science* 2007;**316**:1497–502. <https://doi.org/10.1126/science.1141319>
20. Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* 2017;**6**:e21856. <https://doi.org/10.7554/eLife.21856>
21. Skene PJ, Henikoff JG, Henikoff S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* 2018;**13**:1006–19. <https://doi.org/10.1038/nprot.2018.015>
22. Dahl JA, Jung I, Aanes H *et al*. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* 2016;**537**:548–52. <https://doi.org/10.1038/nature19360>
23. Yoshida H, Lareau CA, Ramirez RN *et al*. The cis-regulatory atlas of the mouse immune system. *Cell* 2019;**176**:897–912.e20. <https://doi.org/10.1016/j.cell.2018.12.036>
24. Teytelman L, Thurtle DM, Rine J *et al*. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci USA* 2013;**110**:18602–7. <https://doi.org/10.1073/pnas.1316064110>
25. Kasinathan S, Orsi GA, Zentner GE *et al*. High-resolution mapping of transcription factor binding sites on native chromatin. *Nat Methods* 2014;**11**:203–9. <https://doi.org/10.1038/nmeth.2766>
26. Mezger A, Klemm S, Mann I *et al*. High-throughput chromatin accessibility profiling at single-cell resolution. *Nat Commun* 2018;**9**:3647. <https://doi.org/10.1038/s41467-018-05887-x>
27. Zheng GX, Terry JM, Belgrader P *et al*. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017;**8**:14049.
28. Rotem A, Ram O, Shores N *et al*. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* 2015;**33**:1165–72. <https://doi.org/10.1038/nbt.3383>
29. Kalisky T, Quake SR. Single-cell genomics. *Nat Methods* 2011;**8**:311–4. <https://doi.org/10.1038/nmeth0411-311>
30. Nagano T, Lubling Y, Stevens TJ *et al*. Single-cell hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 2013;**502**:59–64. <https://doi.org/10.1038/nature12593>
31. Shalek AK, Satija R, Adiconis X *et al*. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 2013;**498**:236–40. <https://doi.org/10.1038/nature12172>
32. Raimundo F, Prompsy P, Vert JP *et al*. A benchmark of computational pipelines for single-cell histone modification data. *Genome Biol* 2023;**24**:143. <https://doi.org/10.1186/s13059-023-02981-2>
33. Picelli S, Björklund AK, Reinius B *et al*. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* 2014;**24**:2033–40. <https://doi.org/10.1101/gr.177881.114>
34. Bartosovic M, Castelo-Branco G. Multimodal chromatin profiling using nanobody-based single-cell CUT&tag. *Nat Biotechnol* 2023;**41**:794–805. <https://doi.org/10.1038/s41587-022-01535-4>
35. Stuart T, Hao S, Zhang B *et al*. Nanobody-tethered transposition enables multifactorial chromatin profiling at single-cell resolution. *Nat Biotechnol* 2023;**41**:806–12. <https://doi.org/10.1038/s41587-022-01588-5>
36. Hamashima K, Wong KW, Sam TW *et al*. Single-nucleus multiomic mapping of m(6)a methylomes and transcriptomes in native populations of cells with sn-m6A-CT. *Mol Cell* 2023;**83**:3205–3216.e5. <https://doi.org/10.1016/j.molcel.2023.08.010>
37. Chen X, Litzzenburger UM, Wei Y *et al*. Joint single-cell DNA accessibility and protein epitope profiling reveals environmental regulation of epigenomic heterogeneity. *Nat Commun* 2018;**9**:4590. <https://doi.org/10.1038/s41467-018-07115-y>
38. Goldstein LD, Chen YJ, Dunne J *et al*. Massively parallel nanowell-based single-cell gene expression profiling. *BMC Genomics* 2017;**18**:519. <https://doi.org/10.1186/s12864-017-3893-1>
39. Liu M, Yue Y, Chen X *et al*. Genome-coverage single-cell histone modifications for embryo lineage tracing. *Nature* 2025;**640**:828–39. <https://doi.org/10.1038/s41586-025-08656-1>
40. Janssens DH, Meers MP, Wu SJ *et al*. Automated CUT&tag profiling of chromatin heterogeneity in mixed-lineage leukemia. *Nat Genet* 2021;**53**:1586–96. <https://doi.org/10.1038/s41588-021-00941-9>
41. Gopalan S, Wang Y, Harper NW *et al*. Simultaneous profiling of multiple chromatin proteins in the same cells. *Mol Cell* 2021;**81**:4736–4746.e5. <https://doi.org/10.1016/j.molcel.2021.09.019>
42. Wen Q, Zhou J, Tian C *et al*. Symmetric inheritance of parental histones contributes to safeguarding the fate of mouse embryonic stem cells during differentiation. *Nat Genet* 2023;**55**:1555–66. <https://doi.org/10.1038/s41588-023-01477-w>
43. Wang C, Huang W, Zhong Y *et al*. Single-cell multi-modal chromatin profiles revealing epigenetic regulations of cells in hepatocellular carcinoma. *Clin Transl Med* 2024;**14**:e70000. <https://doi.org/10.1002/ctm2.70000>
44. Ma J, Jin W, Rong L *et al*. IT-scC&T-seq streamlines scalable, parallel profiling of protein-DNA interactions in single cells. *Genome Biol* 2025;**26**:196. <https://doi.org/10.1186/s13059-025-03661-z>
45. Janssens DH, Otto DJ, Meers MP *et al*. CUT&Tag2for1: a modified method for simultaneous profiling of the accessible and silenced regulome in single cells. *Genome Biol* 2022;**23**:81. <https://doi.org/10.1186/s13059-022-02642-w>
46. Tenney AP, Di Gioia SA, Webb BD *et al*. Noncoding variants alter GATA2 expression in rhombomere 4 motor neurons and cause dominant hereditary congenital facial paresis. *Nat Genet* 2023;**55**:1149–63. <https://doi.org/10.1038/s41588-023-01424-9>
47. Lyu J, Hetey S, Castelo-Branco G *et al*. G-quadruplex profiling in complex tissues using single-cell CUT&tag. *bioRxiv* 2025; 2025.04.03.647101. <https://doi.org/10.1101/2025.04.03.647101>
48. Hui WWI, Simeone A, Zyner KG *et al*. Single-cell mapping of DNA G-quadruplex structures in human cancer cells. *Sci Rep* 2021;**11**:23641. <https://doi.org/10.1038/s41598-021-02943-3>
49. De Rop FV, Hulselmans G, Flerin C *et al*. Systematic benchmarking of single-cell ATAC-sequencing protocols. *Nat Biotechnol* 2024;**42**:916–26. <https://doi.org/10.1038/s41587-023-01881-x>
50. Buenrostro JD, Wu B, Litzzenburger UM *et al*. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015;**523**:486–90. <https://doi.org/10.1038/nature14590>
51. Bielecki P, Riesenfeld SJ, Hütter JC *et al*. Skin-resident innate lymphoid cells converge on a pathogenic effector state. *Nature* 2021;**592**:128–32. <https://doi.org/10.1038/s41586-021-03188-w>
52. Batra S, Allwein B, Kumar C *et al*. G-quadruplex-stalled eukaryotic replisome structure reveals helical inchworm DNA translocation. *Science* 2025;**387**:eadt1978. <https://doi.org/10.1126/science.adt1978>

53. Rosenberg AB, Roco CM, Muscat RA *et al*. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 2018;**360**:176–82. <https://doi.org/10.1126/science.aam8999>
54. Lee AS, Ayers LJ, Kosicki M *et al*. A cell type-aware framework for nominating non-coding variants in Mendelian regulatory disorders. *Nat Commun* 2024;**15**:8268. <https://doi.org/10.1038/s41467-024-52463-7>
55. Wu SJ, Furlan SN, Mihalas AB *et al*. Single-cell CUT&tag analysis of chromatin modifications in differentiation and tumor progression. *Nat Biotechnol* 2021;**39**:819–24. <https://doi.org/10.1038/s41587-021-00865-z>
56. Zhang H, Song L, Wang X *et al*. Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat Commun* 2021;**12**:6566. <https://doi.org/10.1038/s41467-021-26865-w>
57. Zhang Y, Liu T, Meyer CA *et al*. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;**9**:R137. <https://doi.org/10.1186/gb-2008-9-9-r137>
58. Stuart T, Srivastava A, Madad S *et al*. Single-cell chromatin state analysis with Signac. *Nat Methods* 2021;**18**:1333–41. <https://doi.org/10.1038/s41592-021-01282-5>
59. Granja JM, Corces MR, Pierce SE *et al*. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* 2021;**53**:403–11. <https://doi.org/10.1038/s41588-021-00790-6>
60. Wang J, Nakato R. HiC1Dmetrics: framework to extract various one-dimensional features from chromosome structure data. *Brief Bioinform* 2022;**23**:bbab509. <https://doi.org/10.1093/bib/bbab509>
61. Butler A, Hoffman P, Smibert P *et al*. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;**36**:411–20. <https://doi.org/10.1038/nbt.4096>
62. Wang J, Nakato R. Churros: a Docker-based pipeline for large-scale epigenomic analysis. *DNA Res* 2024;**31**:dsad026. <https://doi.org/10.1093/dnares/dsad026>
63. Cusanovich DA, Daza R, Adey A *et al*. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 2015;**348**:910–4. <https://doi.org/10.1126/science.aab1601>
64. Pliner HA, Packer JS, McFaline-Figueroa JL *et al*. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol Cell* 2018;**71**:858–871.e8. <https://doi.org/10.1016/j.molcel.2018.06.044>
65. Patel RP, Ghilardi G, Zhang Y *et al*. CD5 deletion enhances the antitumor activity of adoptive T cell therapies. *Sci Immunol* 2024;**9**:eadn6509. <https://doi.org/10.1126/sciimmunol.adn6509>
66. Zhu H, Xu J, Wang W *et al*. Intratumoral CD38(+)CD19(+)B cells associate with poor clinical outcomes and immunosuppression in patients with pancreatic ductal adenocarcinoma. *EBioMedicine* 2024;**103**:105098. <https://doi.org/10.1016/j.ebiom.2024.105098>
67. He M, Roussak K, Ma F *et al*. CD5 expression by dendritic cells directs T cell immunity and sustains immunotherapy responses. *Science* 2023;**379**:eabg2752. <https://doi.org/10.1126/science.abg2752>
68. Grosselin K, Durand A, Marsolier J *et al*. High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat Genet* 2019;**51**:1060–6. <https://doi.org/10.1038/s41588-019-0424-9>
69. Marsolier J, Prompsy P, Durand A *et al*. H3K27me3 conditions chemotolerance in triple-negative breast cancer. *Nat Genet* 2022;**54**:459–68. <https://doi.org/10.1038/s41588-022-01047-6>
70. Yu W, Uzun Y, Zhu Q *et al*. scATAC-pro: a comprehensive workbench for single-cell chromatin accessibility sequencing data. *Genome Biol* 2020;**21**:94. <https://doi.org/10.1186/s13059-020-02008-0>
71. Wang J, Bando M, Shirahige K *et al*. Large-scale multi-omics analysis suggests specific roles for intragenic cohesin in transcriptional regulation. *Nat Commun* 2022;**13**:3218. <https://doi.org/10.1038/s41467-022-30792-9>
72. Rodriguez J, Ren G, Day CR *et al*. Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. *Cell* 2019;**176**:213–226.e18. <https://doi.org/10.1016/j.cell.2018.11.026>
73. Abbasova L, Urbanaviciute P, Hu D *et al*. CUT&tag recovers up to half of ENCODE ChIP-seq histone acetylation peaks. *Nat Commun* 2025;**16**:2993. <https://doi.org/10.1038/s41467-025-58137-2>
74. Deerwester S, Dumais ST, Furnas GW *et al*. Indexing by latent semantic analysis. *J Am Soc Inf Sci* 1990;**41**:391–407. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<#x003C;391::AID-ASI1>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<#x003C;391::AID-ASI1>3.0.CO;2-9)
75. Duren Z, Chang F, Naqing F *et al*. Regulatory analysis of single cell multiome gene expression and chromatin accessibility data with scREG. *Genome Biol* 2022;**23**:114. <https://doi.org/10.1186/s13059-022-02682-2>
76. Ma S, Zhang B, LaFave LM *et al*. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* 2020;**183**:1103–1116.e20. <https://doi.org/10.1016/j.cell.2020.09.056>
77. Plongthongkum N, Diep D, Chen S *et al*. Scalable dual-omics profiling with single-nucleus chromatin accessibility and mRNA expression sequencing 2 (SNARE-seq2). *Nat Protoc* 2021;**16**:4992–5029. <https://doi.org/10.1038/s41596-021-00507-3>
78. Fang R, Preissl S, Li Y *et al*. Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat Commun* 2021;**12**:1337. <https://doi.org/10.1038/s41467-021-21583-9>
79. Bravo González-Blas C, Minnoye L, Papasokrati D *et al*. cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat Methods* 2019;**16**:397–400. <https://doi.org/10.1038/s41592-019-0367-1>
80. Chen H, Lareau C, Andreani T *et al*. Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biol* 2019;**20**:241. <https://doi.org/10.1186/s13059-019-1854-5>
81. Tayyebi Z, Pine AR, Leslie CS. Scalable and unbiased sequence-informed embedding of single-cell ATAC-seq data with CellSpace. *Nat Methods* 2024;**21**:1014–22. <https://doi.org/10.1038/s41592-024-02274-x>
82. Martens LD, Fischer DS, Yépez VA *et al*. Modeling fragment counts improves single-cell ATAC-seq analysis. *Nat Methods* 2024;**21**:28–31. <https://doi.org/10.1038/s41592-023-02112-6>
83. Luo S, Germain PL, Robinson MD *et al*. Benchmarking computational methods for single-cell chromatin data analysis. *Genome Biol* 2024;**25**:225. <https://doi.org/10.1186/s13059-024-03356-x>
84. Sinha S, Satpathy AT, Zhou W *et al*. Profiling chromatin accessibility at single-cell resolution. *Genom Proteom Bioinform* 2021;**19**:172–90. <https://doi.org/10.1016/j.gpb.2020.06.010>
85. Baek S, Lee I. Single-cell ATAC sequencing analysis: from data preprocessing to hypothesis generation. *Comput Struct Biotechnol J* 2020;**18**:1429–39. <https://doi.org/10.1016/j.csbj.2020.06.012>
86. Stuart T, Butler A, Hoffman P *et al*. Comprehensive integration of single-cell data. *Cell* 2019;**177**:1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>
87. Cheng S, Miao B, Li T *et al*. Review and evaluate the bioinformatics analysis strategies of ATAC-seq and CUT&tag data. *Genom Proteom Bioinform* 2024;**22**:qzae054. <https://doi.org/10.1093/gpbjnl/qzae054>

88. Montefiori L, Hernandez L, Zhang Z *et al*. Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. *Sci Rep* 2017;**7**:2451. <https://doi.org/10.1038/s41598-017-02547-w>
89. Blumberg A, Danko CG, Kundaje A *et al*. A common pattern of DNase I footprinting throughout the human mtDNA unveils clues for a chromatin-like organization. *Genome Res* 2018;**28**:1158–68. <https://doi.org/10.1101/gr.230409.117>
90. Deng Y, Bartosovic M, Kukanja P *et al*. Spatial-CUT&tag: spatially resolved chromatin modification profiling at the cellular level. *Science* 2022;**375**:681–6. <https://doi.org/10.1126/science.abg7216>
91. Lareau CA, Duarte FM, Chew JG *et al*. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat Biotechnol* 2019;**37**:916–24. <https://doi.org/10.1038/s41587-019-0147-6>
92. Bentsen M, Goymann P, Schultheis H *et al*. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* 2020;**11**:4267. <https://doi.org/10.1038/s41467-020-18035-1>
93. Hu Y, Horlbeck MA, Zhang R *et al*. Multiscale footprints reveal the organization of cis-regulatory elements. *Nature* 2025;**638**:779–86. <https://doi.org/10.1038/s41586-024-08443-4>
94. Li Z, Schulz MH, Look T *et al*. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol* 2019;**20**:45. <https://doi.org/10.1186/s13059-019-1642-2>
95. Zhu C, Zhang Y, Li YE *et al*. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat Methods* 2021;**18**:283–92. <https://doi.org/10.1038/s41592-021-01060-3>
96. Farzad N, Enninfu A, Bao S *et al*. Spatially resolved epigenome sequencing via Tn5 transposition and deterministic DNA barcoding in tissue. *Nat Protoc* 2024;**19**:3389–425. <https://doi.org/10.1038/s41596-024-01013-y>
97. Liu Z, Chen Y, Xia Q *et al*. Linking genome structures to functions by simultaneous single-cell hi-C and RNA-seq. *Science* 2023;**380**:1070–6. <https://doi.org/10.1126/science.adg3797>
98. Li W, Lu J, Lu P *et al*. scNanoHi-C: a single-cell long-read concatenation sequencing method to reveal high-order chromatin structures within individual cells. *Nat Methods* 2023;**20**:1493–505. <https://doi.org/10.1038/s41592-023-01978-w>
99. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 2017;**12**:2478–92. <https://doi.org/10.1038/nprot.2017.124>
100. Zhang D, Rubio Rodríguez-Kirby LA, Lin Y *et al*. Spatial dynamics of brain development and neuroinflammation. *Nature* 2025;**647**:213–27. <https://doi.org/10.1038/s41586-025-09663-y>
101. Li H, Bao S, Farzad N *et al*. Spatially resolved genome-wide joint profiling of epigenome and transcriptome with spatial-ATAC-RNA-seq and spatial-CUT&tag-RNA-seq. *Nat Protoc* 2025;**20**:2383–417. <https://doi.org/10.1038/s41596-025-01145-9>
102. Zhang K, Zemke NR, Armand EJ *et al*. A fast, scalable and versatile tool for analysis of single-cell omics data. *Nat Methods* 2024;**21**:217–27. <https://doi.org/10.1038/s41592-023-02139-9>
103. Chen X, Li K, Cui X *et al*. EpiAgent: foundation model for single-cell epigenomics. *Nat Methods* 2025;**22**:2316–27. <https://doi.org/10.1038/s41592-025-02822-z>
104. Wu J, Wan C, Ji Z *et al*. EpiFoundation: a foundation model for single-cell ATAC-seq via peak-to-gene alignment. *bioRxiv* 2025; 2025.02.05.636688. <https://doi.org/10.1101/2025.02.05.636688>
105. LeRoy NJ, Zheng G, Khoroshevskiy O *et al*. Atacformer: a transformer-based foundation model for analysis and interpretation of ATAC-seq data. *bioRxiv* 2025; 2025.11.03.685753. <https://doi.org/10.1101/2025.11.03.685753>
106. Javed N, Weingarten T, Sehanobish A *et al*. A multi-modal transformer for cell type-agnostic regulatory predictions. *Cell Genom* 2025;**5**:100762. <https://doi.org/10.1016/j.xgen.2025.100762>