**Genomics Proteomics Bioinformatics**

REVIEW

# Profiling Chromatin Accessibility at Single-cell Resolution

Sarthak Sinha [1,*], Ansuman T. Satpathy [2], Weiqiang Zhou [3], Hongkai Ji [3]
Jo A. Stratton [1,4,5], Arzina Jaffer [1], Nizar Bahlis [6], Sorana Morrissy [4,6,7]
Jeff A. Biernaskie [1,4,5,*]

[1] *Department of Comparative Biology & Experimental Medicine, Faculty of Veterinary Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada*
[2] *Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA*
[3] *Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA*
[4] *Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB T2N 4N1, Canada*
[5] *Hotchkiss Brain Institute, University of Calgary, Calgary, AB T2N 4N1, Canada*
[6] *Arnie Charbonneau Cancer Institute, University of Calgary, Calgary, AB T2N 4Z6, Canada*
[7] *Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada*

**Abstract** How distinct transcriptional programs are enacted to generate cellular heterogeneity and plasticity, and enable complex fate decisions are important open questions. One key regulator is the cell's epigenome state that drives distinct transcriptional programs by regulating chromatin accessibility. Genome-wide chromatin accessibility measurements can impart insights into regulatory sequences (in)accessible to DNA-binding proteins at a single-cell resolution. This review outlines molecular methods and bioinformatic tools for capturing cell-to-cell chromatin variation using single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) in a scalable fashion. It also covers joint profiling of chromatin with transcriptome/proteome measurements, computational strategies to integrate multi-omic measurements, and predictive bioinformatic tools to infer chromatin accessibility from single-cell transcriptomic datasets. Methodological refinements that increase power for cell discovery through robust chromatin coverage and integrate measurements from multiple modalities will further expand our understanding of **gene regulation** during homeostasis and disease.

* Corresponding authors.
    E-mail: sarthak.sinha1@ucalgary.ca (Sinha S), jeff.biernaskie@ucalgary.ca (Biernaskie JA).

## Introduction

Since 1665, when Robert Hooke first described that single cells form the building blocks of complex tissues, biologists have sought to delineate cellular taxonomies based on form and function. This task, however, has proven to be extraordinarily difficult due to the remarkable diversity in function, regulation, and morphology, even between cells grouped as one type. Single-cell sequencing has become a turning point for cell biologists as it enables profiling of single cells without experimental purifications and provides an unbiased view of cells, their different states, and their interactions with neighboring cells through molecular cross-talks within an intact tissue [1–3]. Leading the revolution in single-cell measurement technologies is single-cell RNA sequencing (scRNA-seq), which has imparted significant insights into gene expression. Owing to rapid developments in technology, such as droplet-based microfluidics capable of profiling thousands of single-cells by counting 3′-end of transcripts [4,5], organism-scale atlases have started to detangle transcriptional heterogeneity in cells comprising complex tissues [6–14].

Although powerful in resolving cell-specific gene programs, scRNA-seq fails to capture the diverse and dynamic chromatin landscape that regulates gene expression. Waddington famously conceptualized all cells as derivatives of a single totipotent cell canalized through different troughs in the sinuous epigenetic landscape, similar to a ball (cell) rolling down a hill [15]. Analysis of transcriptomes, detached from properties of chromosomes that form this epigenetic landscape, is akin to studying the features of the ball without appreciating its relationship and residence on this hill. Indeed, bulk assay for transposase-accessible chromatin (ATAC-seq) [16,17] and Deoxyribonuclease I digestion (DNase-seq) [18] has revealed that chromatin accessibility is a key component of the epigenetic landscape. By exposing different promoters, enhancers, and regulatory elements, cells enable transcription factors (TFs) to regulate gene expression and to navigate it through bifurcating valleys in the epigenetic landscape. For example, during development, specialized cells progressively diverge from their progenitors by gaining or losing chromatin accessibility [19,20]. At any given time point through this process, measurements of chromatin state can impart insights into a cell's past (lineage relationships) and provide a window into its future (terminal fate) that cannot be captured by gene expression alone [20]. In contrast to developmental trajectories, tumorigenesis is a reversion towards a de-differentiated state as transformed cells reactivate chromatin states reminiscent of embryonic stem cells [20,21]. However, their upward climb on the epigenetic landscape is haphazard as they also adopt chromatin profiles (and subsequently activate gene programs) of noncognate lineages [20].

Is there an example of epigenomic reversion where cells truly de-differentiate by climbing the same hill they previously rolled down from? This would imply retention and maintenance of information regarding prior fate decisions. A number of models have emerged to test this hypothesis in tissues like the skin [22,23], jaw [24], digit tips [25], and the intestine [26], where injury induces an embryonic phenotype, enabling the tissue to regenerate and restore function. To fully appreciate the fundamental mechanisms by which cells navigate the epigenetic landscape requires techniques that capture cell-to-cell chromatin variations in defined populations. To this end, epigenomic techniques such as ATAC-seq [27,28], ChIP-seq [29], Hi-C [30,31], and DNase-seq [32] have recently been developed to assess chromatin structure at single-cell resolution. While the sensitivity of ChIP-seq traditionally required large ($\sim 10^7$ cells) inputs, recent developments such as nano-ChIP [33], iChIP [34], MOWChIP [35], SurfaceChIP [36], and single-cell ChIP using droplet separation termed Drop-ChIP [37] have significantly reduced the input material. Impressively, Drop-ChIP profiled post-translational histone modifications that either enable (H3K4me3) or repress (H3K27me3) transcription at single-cell resolution with relatively high coverage ($\sim 1000$ unique reads per cell) [37]. In particular, this review focuses on single-cell ATAC-seq as it has demonstrated significant promise in capturing large cell numbers (up to 100,000 single cells from 13 rodent tissues profiled at 9000–23,000 unique reads per cell depth [38]) to create organism-scale atlases comparable to those created using scRNA-seq. These datasets are now being integrated with single-cell transcriptomics and proteomics to generate multi-omic measurements. This review catalogs and compares recently developed scATAC-seq technologies and analysis tools that can be exploited to delineate principles of gene regulation and features of regulatory sequences, as well as extrapolates on what may be possible by integrating chromatin accessibility with other single-cell measurements.

## scATAC-seq technologies

### scATAC-seq library construction

The key element employed in ATAC-seq for detecting nucleosome-free regions in the genome is Tn5 transposase, a dimer of two chemically identical monomers that typically catalyzes the movement of transposons to different parts of the genome through a "cut-and-paste" mechanism (**Figure 1**). Since endogenous transposase is relatively inactive (as transpositions are highly mutagenic), a hyperactive version of the Tn5 transposase [17,39] is first loaded with sequencing adapters to create a dimeric transposome complex (Figure 1A), and then introduced into intact nuclei where it simultaneously cuts exposed DNA and ligates the loaded sequences (Figure 1B). Since tightly packed heterochromatin has high steric hindrance, these sites remain inaccessible to transposase and makes fragmentation at these sites less probable. Following fragmentation, transposase inserts its adaptor payload to permanently mark active regulatory sites. Adapter-ligated fragments are isolated, and inserts that contain two different ends (s5 and s7) can be extended to add identifiers such as cell barcodes and sample indexes [16,40] (Figure 1B and C). Highly accessible regions return significantly more sequencing reads that form peaks when called with peak-calling algorithms like ZINBA [41] (Figure 1D and E). Interestingly, close inspection of bulk ATAC peaks also reveal subregions (usually a few bp in length) that had escaped tagmentation since they were actively occupied by DNA-binding proteins and are considered "footprints" of those binding protein [16]. Bulk ATAC became a useful method for profiling accessible chromatin since: 1) its measurements are highly correlated with single- and double-cut DNase-seq [16,42], 2) libraries can be generated from 500 cells (compared to >100,000 needed for
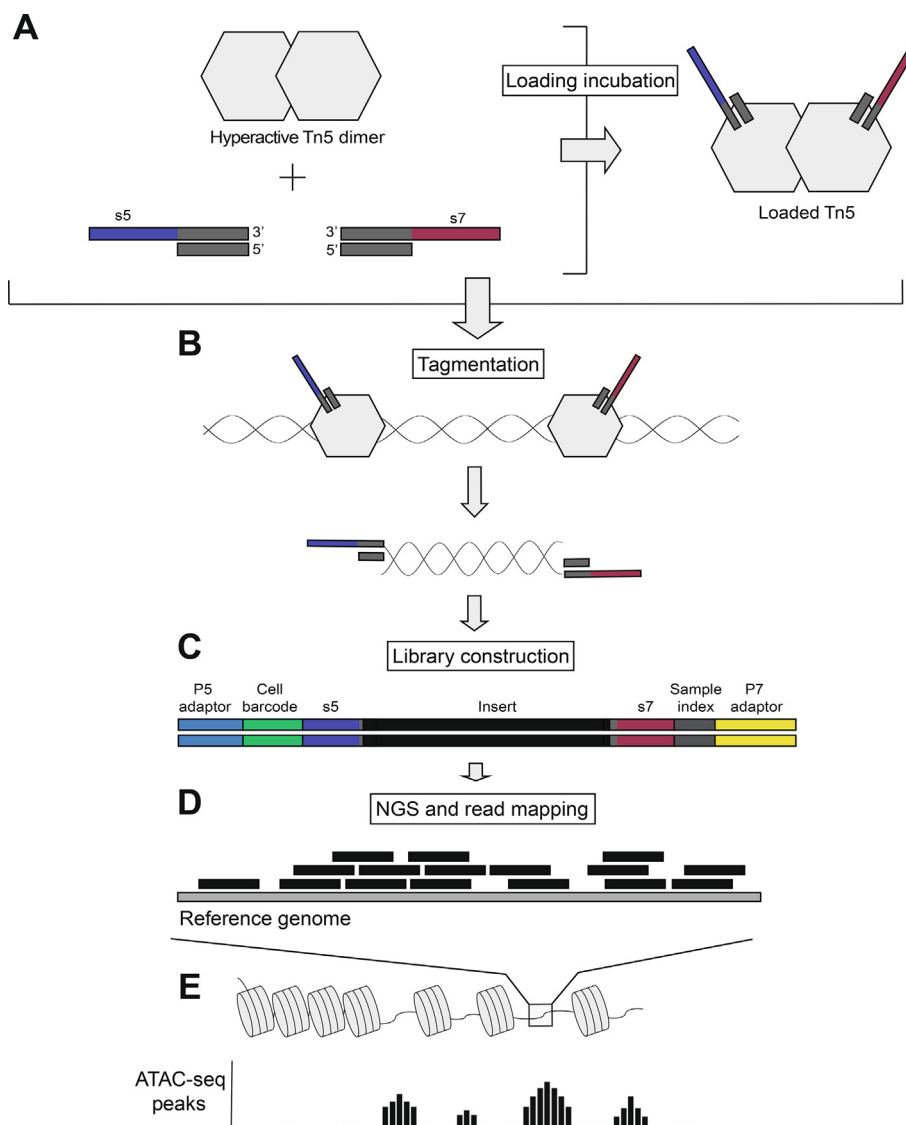
**Figure 1    ATAC-seq probes genome-wide chromatin accessibility using hyperactive Tn5 transposase**
**A.** Schematic illustrating hyperactive Tn5 being loaded with sequencing adapters by mixing equal amounts of two indexed oligos (s5 and s7) with Tn5 and incubating the mixture for approximately one hour. **B.** During Tn5 tagmentation (fragmentation and tagging), the transposase cleaves accessible DNA and attaches adaptor overhangs within intact nuclei. Since nuclei are not fragmented in this process, bulk Tn5 tagging can be performed in scATAC reactions prior to partitioning tagged nuclei. **C.** Tagmentation generates three different products: 1) sequence with s5 at both ends, 2) sequence with s7 at both ends, or ideally, 3) sequence with s5 and s7 at opposite ends (as shown in the diagram). Only the final product (containing different ends) is amplifiable. Final library is generated by appending additional identifiers such as cell/sample-specific barcodes using PCR. **D.** scATAC libraries are paired end sequenced and mapped to a reference genome. **E.** Peak-calling algorithms identify enriched (peak) regions which correspond to open chromatin states. ATAC-seq, assay for transposase-accessible chromatin using sequencing; scATAC, single-cell assay for transposase-accessible chromatin.

DNase-seq), and 3) library generation requires two hours (compared to multiday DNase-seq preparation) [16].

To resolve chromatin accessibility at a single-cell resolution, the Shendure and Chang/Greenleaf laboratories modified methods that were originally developed for single-cell transcriptomics and *de novo* genome assembly to tag transposed DNA fragments from the same cell with a unique molecular barcode (**Figure 2**) [27,28]. The two groups employed different molecular methods to achieve this. The study by Shendure's group profiled > 15,000 single cells using combinatorial cellular indexing, where nuclei were first tagged with barcoded Tn5 transposase in a 96 well plate, then pooled and diluted, and 15–25 nuclei were redistributed to another plate where a second barcode was added during PCR with primers targeting Tn5 adapters (**Figure 2**A) [28]. They reasoned that this "split-pool" approach would generate a large array of barcode combinations, and all fragments originating from the same cell can be identified by their shared combinatorial barcode. Although this approach overcame the need to individually compartmentalize single nuclei, since these libraries were prepared from thousands of single cells sequenced to an average depth of 2500 reads per barcode, the libraries had low
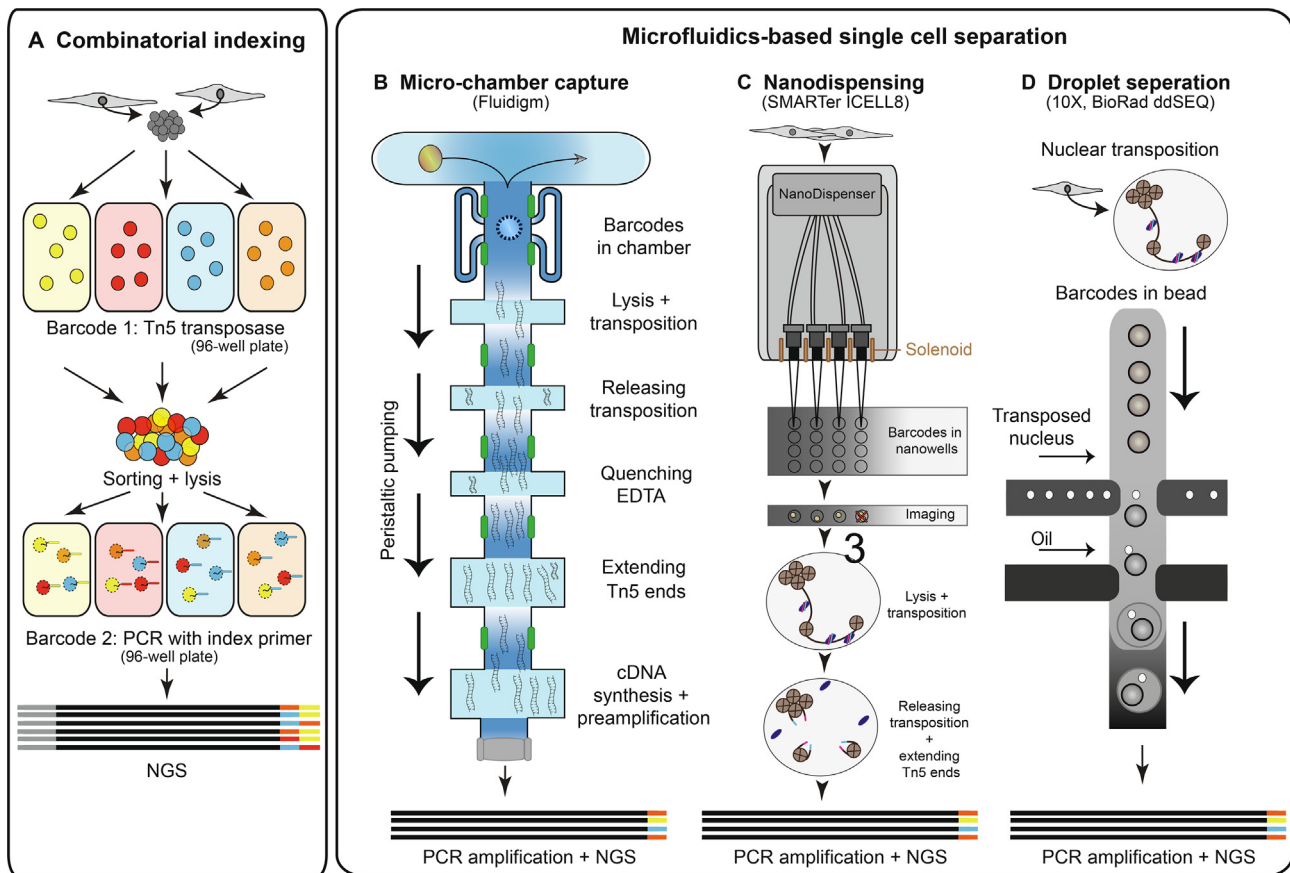
**Figure 2    Schematic diagrams showing methods for transposition, barcoding, and library preparation for scATAC-seq**
**A.** In combinatorial indexing, nuclei are first tagged in bulk via barcoded Tn5 transposase in a 96 well plate. Then, cells are pooled, and 15 to 25 nuclei are randomly sorted into another 96 well plate where a second barcode is added during PCR. The probability for two cells to share the same combination of barcode is between 6%−11%. **B.** Micro-chamber capture utilizes a plate with thousands of non-adherent, barcode-containing microwells with a central and a serpentine microfluid flow path. The first cell entering a microwell gets trapped in the central path and blocks entry of subsequent cells, forcing them to take the serpentine path and be captured by a downstream chamber. The cell trapped in the central path is subsequently subjected to lysis, transposition, and downstream library construction within the chamber. **C.** Nanodispensers use non-contact dispensing to place a single cell (stained with live/dead Hoechst stain) into a nanowell containing preprinted barcodes. Only wells containing a dispensed cell (approximately one third of the 5184 nanowells) are transposed to generate sequencing libraries. **D.** In Drop-seq, transposed nuclei are compartmentalized into nanoliter-sized aqueous droplets containing unique barcodes that are carried in a continuous oil phase.

complexity (number of unique fragments recovered). The Chang/Greenleaf laboratories programmed Fluidigm's microfluidics chip to capture, transpose, and amplify DNA fragments from a single nucleus captured onto a microchamber (Figure 2B) [27]. This approach sampled fewer cells (~ 1600), but each cell received 30 times the sequencing coverage for an average of 73,000 reads per cell, generating at least a two-fold increase in library complexity. However, even with this approach, only 10% of all ubiquitously open regions were recovered from a single cell [43]. Notwithstanding limitations of sparse coverage, combinatorial indexing has been employed to profile accessibility from 13 different tissues (~ 100,000 cells) [38], developing *Drosophila* embryos (23,085 cells) [44], mouse forebrains (~ 15,000 cells) [45], and cortices of mice with Down Syndrome (13,766 cells) [46].

Recently, two additional microfluidic capture strategies have been developed. First, Takara Bio's SMARTer ICELL8 platform employs individually indexed nanoliter-scale wells that can theoretically enable 5184 ATAC reactions to run in parallel (Figure 2C) [47]. A particular advantage of this platform is the integration of multi-color fluorescence imaging to determine wells that received single cells to selectively add reagents, thereby reducing library preparation/sequencing cost to ~ 81¢ per cell. Integration of imaging may also foreshadow methods that enable simultaneous imaging of accessible chromatin to reveal its positional identity within the native configuration of the nucleus, similar to its bulk counterpart ATAC-see [48]. Second, Bio-Rad's SureCell® ATAC kit (http://www.bio-rad.com/en-ca/life-science-research/news/bio-rad-launches-scatac-seq-solution-for-early-access-customers?vertical = LSR&ID = Bio-Rad-Launches-scA_1537378901) [49] and 10X Genomics's scATAC solution [40] employ microfluidics to partition up to 10,000 transposed nuclei into nanosized droplets, where content from each droplet is distinguished by one of ~ 750,000 unique barcodes (Figure 2D). Using the 10x technology, Satpathy and colleagues demonstrated that up to 6000 single cells can be individually barcoded whilst maintaining a reasonable multiplet rate (~ 1%), an approach that costs ~ 25¢

per cell [40]. Impressively, Lareau and colleagues further extend Bio-Rad's microfluidics platform by combining droplet capture with combinatorial indexing (termed droplet-based sciATAC-seq) to enable simultaneous profiling of ~ 100,000 cells per microfluidic run. This conjugate approach profiled 502,207 cells from diverse mammalian and human tissues, including the rodent brain (~ 39 K cells), human blood (~ 60 K cells), and experimentally perturbed cell populations (~ 76 K cells) [49]. Since both commercial kits leverage their widely adopted barcoding platforms, ddSEQ Single-Cell Isolator and the Chromium System, an advantage for existing customers is that the scATAC-seq assay can be easily integrated into existing infrastructure.

Finally, it is worth recognizing that scATAC is still in its infancy and its evolution is constrained by what is computationally and technologically feasible. One example is whether single cells or single nuclei should be the starting point for ATAC assays. Currently, scATAC-seq technologies have been optimized to isolate single nuclei to reduce mitochondrial reads and downstream sequencing costs (https://www.10xgenomics.com/solutions/single-cell-atac/) [42,50]. However, two recent reports reveal that the 10- to 100-fold higher mutation rate in mtDNA (compared to nuclear DNA) can be employed as a natural barcode to reconstruct clonal architecture and lineage relationships in complex organisms [51,52]. Impressively, both studies demonstrate mitochondrial mutation tracking in human tissues and tumors, a system where there is currently no method for fate mapping and studying clonal dynamics. Indeed, how scATAC will be further customized to impart novel single-cell measurements and go beyond deciphering causality between gene expression and chromatin accessibility are important open questions. In this regard, we highlight recent developments (Table 1) and extrapolate on what may be possible.

## Customizations to scATAC-seq library construction

To corroborate chromatin accessibility with a cell's function and to quantitate the impact of genetic perturbations on the epigenome, the Chang and Khavari laboratories repurposed a candidate reverse transcription (RT)-based approach previously paired with single-cell transcriptomics (Figure 3) [53–55]. Briefly, this approach, referred to as "Indexed Single-cell Seq," seeks to enrich specific amplicons during PCR steps of library construction to provide an orthogonal measurement of the cell's state or of the perturbations introduced.

By recovering the single guide RNA (sgRNA) through RT of an index called "guide barcode" during a microfluidics-based scATAC-seq run, Rubin and colleagues identified regions of the genome deleted in individual Cas9-expressing cells, and then analyzed the impact of those deletions on the regulatory dynamics broadly (Figure 3A) [56]. Since gRNA can target any region of the genome, this combinatorial deletion approach, termed perturbation-indexed single-cell ATAC-seq (or Perturb-ATAC), can unearth new insights into how coding (*i.e.*, TFs, chromatin regulators) and noncoding genomes enact distinct chromatin regulatory networks in each cell. Already, a knockout screen on differentiating keratinocytes *in vitro* has revealed positive and negative regulation amongst interacting TFs and delineated how differentiation trajectories can be rerouted as a result of TF perturbation.

As well, these analyses also reveal conceptual models of TF interactions that can only be observed by perturbing them in pairs. For example, it identified that, while the majority of TF interactions were *additive* (no interaction between the pair perturbed), a number of them were *synergistic* (positive interaction) or *buffering* (negative interaction). Further measurements supported the notion that genomic co-localization and co-binding of TF pairs may regulate synergistic TF activation, a mechanism that may be especially important for activating key regulatory genes that in turn, enact a distinct cell state.

Combining Perturb-ATAC with other candidate index or mitochondrial lineage tracing methods can unearth unprecedented insights. For example, it's theoretically possible to combine Satpathy and colleagues' TCR sequencing with ATAC approach (termed 'T-ATAC-seq', Figure 3B) that enables studying chromatin landscapes in clonal T cells [57] with Perturb-ATAC [56] and mitochondrial lineage tracing [51]. Since T cells exhibit particularly high clonal dynamics, and since understanding these dynamics are critical in states like tumor infiltrating T lymphocytes to boost anti-tumor response, simultaneous assessments of perturbations that enhance T cell efficacy alongside TCR and epigenome/transcriptome sequencing would be critical for identifying the repertoire and regulatory drivers that can be targeted. Indeed, Yost, Satpathy, and colleagues forecast the potential of such multiomic T cell profiling approaches, as their combination of TCR and transcriptome sequencing revealed novel repertoires of T cell clones that may be suitable candidates for modulation in skin cancers [58].

## Methods integrating scATAC-seq with other omic approaches

Deciphering regulatory associations between different genomic layers can impart novel perspectives on how cellular information flows from one layer to the next [59]. This approach, applied to single cells, can further reveal the layering of regulatory controls (and its evolution over time) that underlie complex processes such as fate choice during development or disease. Combining single-cell chromatin accessibility measurements with transcriptomics is of particular interest, as their joint analysis can reveal novel *cis*-regulatory sequences (noncoding DNA regions which modulate TF activity) and help catalog its impact on gene expression (Figure 4). This is particularly important as these regulatory sequences may play a more important role than their gene-coding counterpart [60,61] and their purposeful correction/modulation may be crucial for regulating cell function [62].

To this end, two methods have been developed to simultaneously profile chromatin and transcriptome from single cells. The first method, called single-cell Chromatin Accessibility and Transcriptome Sequencing (scCAT-seq), sorts single cells into a 96-well plate, where the cell membrane is lysed and the nucleus separated from the cytoplasm due to physical dissociation (Figure 4A) [63]. Physical separation of the nucleus and the cytoplasm enables simultaneous transposition of the nucleus and full-length transcript capture from the cytoplasm using Smart-seq2. This generates high-confidence regulatory interactions for a small number of cells. The second method, called single-cell combinatorial indexing-based profiling of chromatin accessibility and mRNA (sci-CAR), trades the high

**Table 1    Summary of molecular methods for single-cell isolation and ATAC library preparation**

| No. | Approach | Overview | Advantage | Disadvantage | Commercial platform | Status | Ref. |
|---|---|---|---|---|---|---|---|
| **Methods for single-cell chromatin accessibility profiling** | | | | | | | |
| 1 | sciATAC-seq | Single nuclei are tagged using transposase, pooled, and redistributed into a new set of wells where a second barcode is added during PCR. | Doesn't require compartmentalization of single cells, high throughput and easily scalable by increasing the split-pool indexing rounds. | Improper barcode collision rate can result in more than one nucleus receiving the same index combination. | Illumina scATAC-seq (Cell Index) | Published (May 2015) | [28] |
| 2 | Microfluidic chip | Bulk transposed nuclei are encapsulated into nanoliter-sized droplets containing unique barcodes. Sequencing reads from the same cell are identified by this shared barcode. | High-throughput nuclei profiling. 500–10,000 nuclei per channel for up to 8 channels with 65% capture efficiency. | Multiplet (doublet) rate increases with the number of nuclei loaded. Sparse coverage of chromatin accessibility. | 10X Genomics, Bio-Rad, Illumina scATAC-seq (Microfluidics) | Preprints (Apr 2019) Trial Dataset: support.10x genomics.co m/single-cell-atac/datasets | [40] |
| 3 | μATAC-seq | Single cells stained with dyes are loaded onto nanowell chips. Fluorescence imaging identifies wells containing only one cell. Cell lysis, transposition, and PCR amplification are only implemented in wells containing a single live cell. | High throughput due to massively parallel nanowell arrays on a chip. Low library generation cost (~ 81¢ /cell) as reagents are only added to wells containing single cells. | No nuclei extraction through cell lysis. Preparation can include mitochondrial contamination leading to high mitochondrial reads. | SMARTer ICELL8 Single-Cell System | Published (Sept 2018) | [47] |
| 4 | Perturb-ATAC | Detects CRISPR sgRNA and accessible chromatin from single cells. sgRNA are detected by reverse transcription using targeted primers. | Enables studying the impact of CRISPR perturbations on chromatin accessibility using high-throughput single-cell screens. 40 sgRNA genotypes detected. | ~ 4300 single cells with 63 unique sgRNA genotypes analyzed. Optimization of cells per sgRNA genotype will be challenging, particularly when scaling up the number of sgRNAs included in a library. | Fluidigm C1 | Published (Jan 2018) | [56] |
| 5 | T-ATAC-seq | Simultaneous profiling of chromatin accessibility and T cell antigen receptors. TRA and TRB sequences are reversely transcribed and amplified along with accessible chromatin to study epigenetic landscapes of specific T cell clonotypes. | Links TCR identity to chromatin accessibility and identifies cis/trans regulators of T cell subsets. Can be further modified to recover specific targeted transcripts such as B cell receptors. | Low throughput (96 cells per chip). | Fluidigm C1 | Published (Apr 2018) | [57] |
| **Joint multi-omic measurement techniques** | | | | | | | |
| 6 | scPi-ATAC-seq | Cells are fixed, permeabilized, and subjected to antibody staining and bulk transposition. They are then sorted into barcoded wells where protein measurements are taken from fluorescence emitted by antibodies. Transcripts from each cell share well-specific barcodes. | Can be used on fixed cells or tissue. | Number of proteins screened is limited by the number of fluorophores resolved by an epifluorescence microscope. | N/A | Published (Nov 2018) | [65] |
| 7 | scCAT-seq | Nucleus and cytoplasm are separated. Cytoplasm is subjected to full-length transcript capture and nucleus is subjected to transposition, both marked by well-specific barcodes. | Works well when starting with low input material. | Number of cells with transcriptomes captured are not in the order achieved by microfluidics partitioning. Nuclear RNA is not captured either. | Illumina Smart-seq2 | Published (Jan 2019) | [63] |
| 8 | sciCAR-seq | Split-pool barcoding to co-assay chromatin accessibility and gene expression from the same cell. Can correlate regulatory elements to target genes by calculating covariance of accessibility and expression. | The number of split-pool indexing rounds can be increased to co-assay millions of single cells. | Uneven data sparsity between gene expression and chromatin accessibility. The ATAC co-assay was leveraged to recover the chromatin landscape but can't be used autonomously to discover the full diversity of cell states/types. | Illumina scATAC-seq (Cell Index) | Published (Aug 2018) | [64] |

*Note*: PCR, polymerase chain reaction; CRISPR, clustered regularly interspaced short palindromic repeats; sgRNA, single guide  RNA; TCR, T cell receptor; TRA, T cell receptor alpha; TRB, T cell receptor beta; sciATAC-seq, single-cell combinatorial indexing; μATAC-seq, nano-well scATAC-seq; SMARTer, switching mechanism at the 5′ end of the RNA transcript; Perturb-ATAC, perturbation-indexed scATAC-seq; T-ATAC-seq, transcript-indexed ATAC-seq; scPi-ATAC-seq, single-cell protein-indexed ATAC; scCAT-seq, single-cell chromatin accessibility and transcriptome sequencing; sciCAR-seq, single-cell combinatorial indexing-based chromatin accessibility and mRNA.
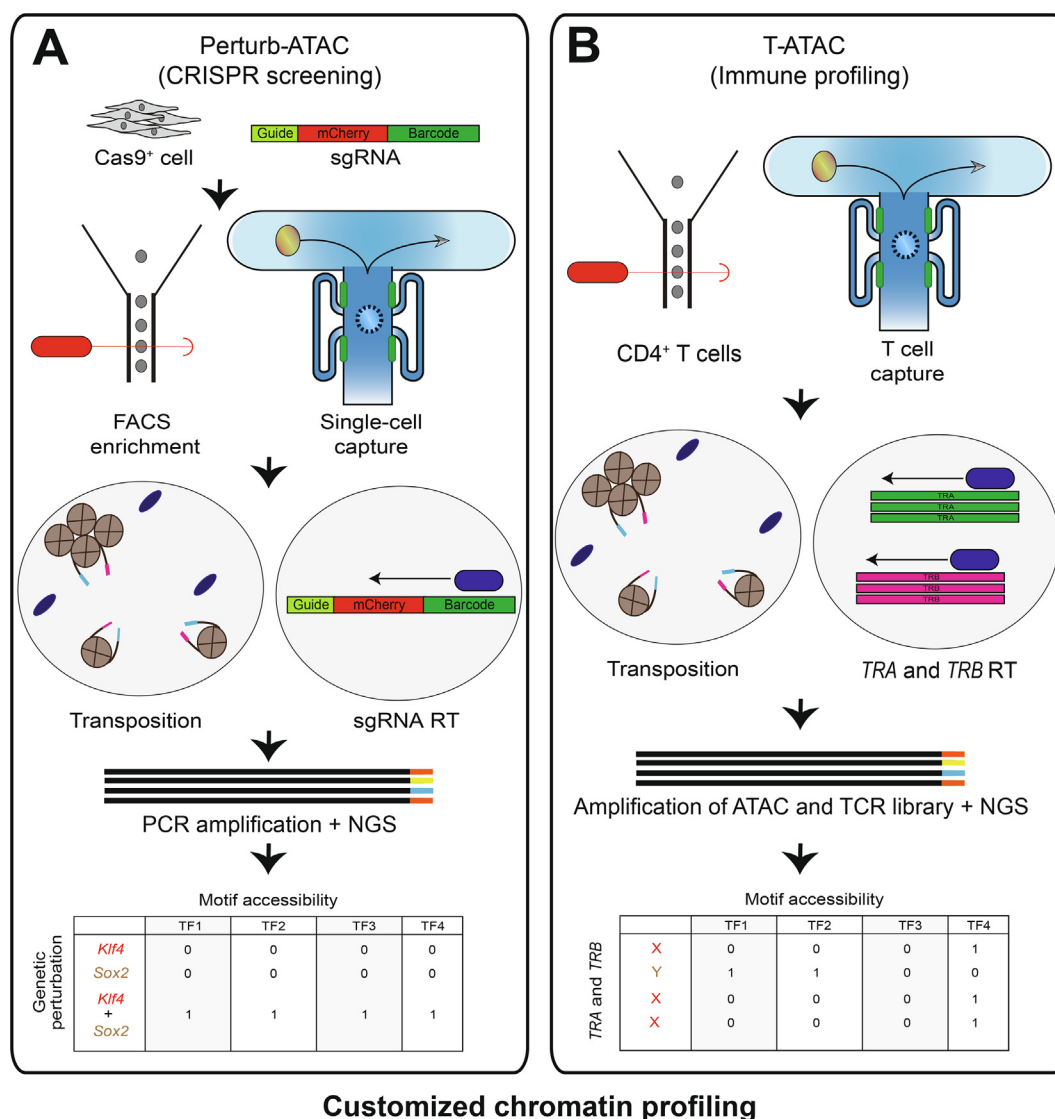
**Figure 3    Customizations to scATAC-seq enables high-throughput CRISPR screening and T cell clonotyping**

**A.** Perturb-ATAC maps the impact of CRISPR perturbation on chromatin accessibility in single-cells. First, cells are transduced by sgRNA vectors containing a reporter sequence. FACS enriched cells are captured on microchambers (Figure 1C) and transposed with Tn5 enzyme. Following transposition, CRISPR sgRNAs are reversely transcribed using primers targeting the common 3′ end of sgRNA vectors. sgRNA and ATAC amplicons are amplified, pooled, sequenced, and analyzed for changes in TF features following genetic perturbations. **B.** T-ATAC-seq simultaneously profiles chromatin accessibility and TCRs in single T cells. Single CD4$^+$ T cells are captured on microchambers (Figure 1C) where they are lysed, and their accessible chromatin transposed with Tn5 enzyme. TRα and TRβ transcripts (*TRA* and *TRB*) are reversely transcribed with primers targeting *TRA* and *TRB*, and ATAC amplicons are PCR amplified with well-specific barcodes, pooled, and sequenced. TF, transcription factor; sgRNA, single guide RNA; CRISPR, clustered regularly interspaced short palindromic repeats; FACS, fluorescence-activated cell sorting; Perturb-ATAC, perturbation-indexed scATAC-seq; T-ATAC-seq, transcript-indexed ATAC-seq; TCR, T cell receptor; TRA, T cell receptor alpha; TRB, T cell receptor beta; RT, reverse transcription; CDR3, complementarity-determining region 3.

sensitivity of Smart-seq2 for high-throughput offered with combinatorial indexing [64]. The key modification to the pool-sort strategy used for ATAC alone [28] is the addition of an RNA-seq index through RT that contains: 1) poly(T) sequence to capture transcripts, 2) a well/pooling round-specific barcode to trace the origin of transcripts, and 3) a unique molecular identifier (UMI) to unambiguously quantitate the number of transcript copies recovered (Figure 4B). Finally, although there is a paucity of methods that combine chromatin with comprehensive proteomics, the recently developed Protein-indexed ATAC (Pi-ATAC) seeks to correlate chromatin accessibility to phenotypes that are delineated based on protein markers [65]. By measuring fluorescence from antibodies raised against epitopes of interest in each well and using well-specific barcodes for ATAC fragments, this approach can generate phenotype-resolved accessibility signatures (Figure 4C). While such a method is limited by the number of resolvable fluorophores, it is well-suited for a system where
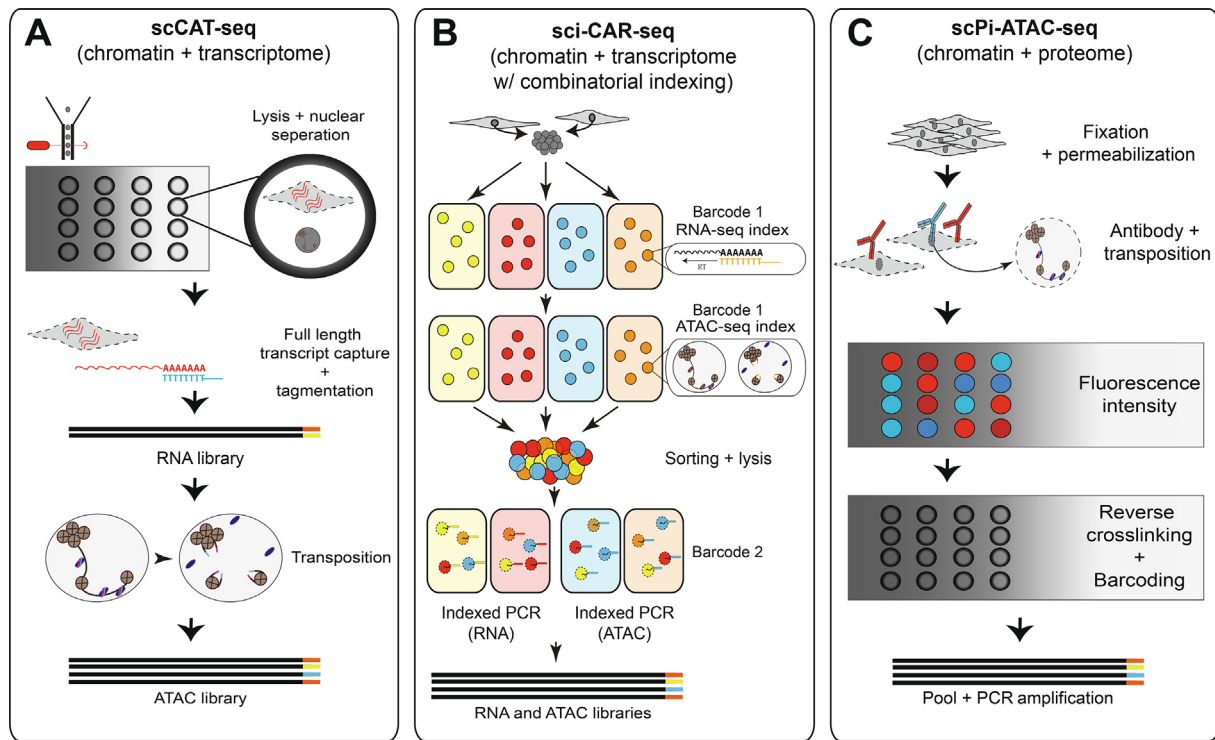
**Single-cell multi-omics**

**Figure 4    Methods for single-cell multi-omics that integrate chromatin accessibility with proteomics and transcriptomics**
**A.** scCAT-seq separates the nucleus and the cytoplasm from single cells sorted in a 96-well plate. The cytoplasm is subjected to full-length transcript capture using Smart-seq2 and the nucleus to transposition, and both are marked by a barcode unique to each well. **B.** sci-CAR-seq profiling starts with nuclei distributed in a 96-well plate. First, nuclear RNA is indexed by reversely transcribing poly(A) mRNA with a poly(T) primer carrying a well-specific barcode and a UMI. Then, accessible chromatin is indexed with transposase carrying a well-specific barcode. All nuclei are pooled, and 15 to 25 are randomly sorted into another 96 well plate where a second barcode is added during indexed PCR for RNA-seq or for ATAC-seq. Amplicons from both libraries are pooled and sequenced. **C.** scPi-ATAC-seq starts with fixed and permeabilized cells that are subjected to antibody staining and bulk transposition. Cells are then sorted into a 96 well plate where fluorescence emitted by antibodies are quantified, proteins are reverse crosslinked, and barcodes are added by indexing PCR. scCAT-seq, single-cell chromatin accessibility and transcriptome sequencing; sci-CAR-seq, single-cell combinatorial indexing-based chromatin accessibility and mRNA; scPi-ATAC-seq, single-cell protein-indexed ATAC-seq; UMI, unique molecular identifier.

(relative) expressions of key markers robustly identify cell states, which can then be correlated to its chromatin status. Albeit in its infancy, co-assaying the 'input' (chromatin accessibility) alongside gene or protein 'outputs' will reveal the combinatorial logic that generates diversity and dynamics in biological systems and eventually inform *in silico* models that predict functional impacts of perturbations to (non)coding regions of the genome [66,67].

## scATAC-seq analysis

### Computational challenges in scATAC-seq analysis

In comparison to bulk chromatin accessibility, where cell ensembles exhibit a range of signal intensities across accessible regions, data from single cells is essentially binary. Although it is possible to have multiple Tn5-transposase insertions at one site in a single cell, since such events are rare [27]. The binarization of read count matrix is often the first processing step. By

the same account, if a site receives no read, it is difficult to discern if that is truly an inaccessible chromatin region or a site that was missed by transposase or lost during amplification [68]. scATAC-seq data analysis is also more challenging than scRNA-seq due to the former's considerably larger data matrices that have much higher sparsity. Since scATAC-seq matrices can contain counts from over hundreds of thousands of regulatory sites (compared to 20,000 protein coding genes typically assessed in RNA-seq) across thousands of single cells, the result is large datasets containing sparsely mapped reads. Similar to tackling the dropout problem in scRNA-seq data, where only a small percentage of the transcriptome is retrieved for each cell, resulting in a sparse gene-barcode matrix, an antidote to scATAC-seq sparsity is aggregating signals across similar cells and across sites that share a common feature [27,28]. However, aggregating cells results in a loss of the single-cell nature of scATAC-seq. Moreover, grouping sites by a common feature (*i.e.*, TF motif) requires significant *a priori* knowledge. Despite these challenges, scATAC-seq tools can: 1) perform discrete and continuous cell groupings [69–

**Table 2    Summary of open source bioinformatic tools for scATAC-seq analysis**

| No. | Tool | Algorithm | Key feature | Limitation | Platform | Availability | Status | Refs. |
|---|---|---|---|---|---|---|---|---|
| **Single-cell ATAC-seq toolkits** | | | | | | | | |
| 1 | chromVAR | Deviation in chromatin accessibility for each annotation from a null model (generated by averaging accessibility across all cells) using k-mer analysis. | Genome-wide assessment of features (*i.e.*, Transcription factor motifs) in different cell types. | 1. Restricts analysis to pre-defined genomic feature sets. Doesn't consider the full list of chromosomal locations. 2. Cells with similar total accessibility for a genomic feature can have different accessibility at different regulatory elements comprising that feature. | R | GitHub, Bioconductor | Published (Aug 2017) | [69] |
| 2 | cisTopic (similar to SCENIC for scRNA-seq) | Bayesian topic modelling using Latent Dirichlet Allocation to discover cis-topics and cell types. Data is binarized and subsampled (Gibbs sampler) for efficiency. | Identification of epigenome-based cis-regulatory topics. Topics and transcription factors identified corroborated with gene regulatory networks independently identified by scRNA-seq. | Discovery of new cis-regulatory topics requires a topics database to enable overarching biological interpretations of regulatory topics. | R | GitHub | Preprint (Jul 2018) | [83] |
| 3 | BROCKMAN | Chromatin accessibility is represented as word frequencies. This generates a matrix of k-mer frequencies to find hidden structure in the data. | K-mer landscape can resolve cell types, batch effects, active and co-varying TFs, and score cycling cells. | Restricts analysis to pre-defined genomic feature sets. Doesn't consider the full list of chromosomal locations. | R | GitHub | Published (Jul 2018) | [70] |
| 4 | SCRAT | One or multiple features can be used for clustering (once reads are aggregated by feature). Cell regulome can be compared to ENCODE DNase-seq profiles to infer cell type. Parametric and non-parametric statistical tests can be applied for differential feature analysis. | Graphical User Interface to identify regulatory sites according to features in existing databases (*i.e.*, transcription factor binding sites, gene annotations). | Restricts analysis to pre-defined genomic feature sets. Doesn't consider the full list of chromosomal locations. | R | Web (https://zhiji.shinyapps.io/scrat) and GitHub | Published (Sept 2017) | [75] |
| 5 | Cicero | Cell grouping performed by clustering or pseudotemporal ordering. Covariance calculations to compute co-accessibility scores between regulatory sites using graphical LASSO. | Identifies co-accessible DNA elements, connecting regulatory elements to target genes. Can also identify "chromatin hubs" and infer change across pseudotime. | Co-accessibility of DNA elements does not necessitate regulatory interactions. | R | GitHub, Bioconductor | Published (Sept 2018) | [76] |
| 6 | scABC | Weighted K-medoids clustering where cells with higher sequencing coverage are up-weighted. Cells are assigned to cluster landmarks (which depict a prototypical cell within that cluster). | *De novo* cell clustering by chromatin accessibility signatures without reliance on predefined features. | 1. Ineffective when a population can't be clustered (*i.e.*, it lies in a continuum). 2. Discounting cells with low sequencing depth may miss rare cell types/states. | R | GitHub | Published (Jun 2018) | [77] |
| 7 | g-chromVAR | Causal genetic variants identified using fine mapped variant posterior probabilities for Genome-Wide Association Study (GWAS) traits. | Score single cells for GWAS enrichment. This can be used to identify regulatory elements forming the basis for disease-associated genetic variants at a single-cell resolution. | Genome-wide association studies is restricted to people from European descent. Insights limited to hematopoiesis and blood cells. | R | GitHub | Preprint (Jan 2018) | [78] |

**Table 2**    **Summary of open source bioinformatic tools for scATAC-seq analysis**

| No. | Tool | Algorithm | Key feature | Limitation | Platform | Availability | Status | Refs. |
|---|---|---|---|---|---|---|---|---|
| 8 | Destin | Cell clustering using weighted principal component analysis where distance to transcription factor start site and accessibility peaks from reference experiments are factored in. | One stop package for quality control, cell clustering, differential chromatin accessibility analysis, and mapping cluster-specific GWAS features. | 1. Discounts regulatory elements located far away from transcription factor start sites. 2. Restricts analysis to pre-defined genomic feature sets. Doesn't consider the full list of chromosomal locations. | R | GitHub | Preprint (Nov 2018) | [79] |
| 9 | Scasat | K-medoid-based cell clustering. Fisher exact test and information gain for differential accessibility analysis. | One stop package for quality control and downstream analysis (*i.e.*, clustering, differential accessibility analysis) in a Jupyter notebook. | User-specified K limits its utility to cases where the true number of cell types is known. | Python (initial processing), R (data analysis) | GitHub | Published (Oct 2018) | [80] |
| 10 | PRISM | A matrix of angular cosine-based distances between cells and the group norm (spatial centroid) is used to measure variability in chromatin accessibility for a defined set of genomic features. | Predicting cell-to-cell variability with high fidelity (as demonstrated by ignoring CTCF binding sites). | Restricts analysis to pre-defined genomic feature sets. Doesn't consider the full list of chromosomal locations. | R | GitHub | Published (Aug 2018) | [81] |
| 11 | Scitools | The package is a collection of wrappers that use classic tools such as bedtools and samtools. | Streamlined analysis package for combinatorial indexing sc-ATAC experiments. | No new algorithm included in the package. | Perl | GitHub | Preprint (Sept 2018) | [72] |
| 12 | Dr.seq2 | The package is a collection of wrappers that use classic tools such as Macs14 for peak calling. | One stop package for quality control and basic downstream analysis (*i.e.*, clustering) for sc-RNA-seq and sc-ATAC-seq. | No new algorithm included in the package. | Python | Web (https://zhang lab.tongji.edu .cn/softwares/ Drseq2/index. html) and GitHub | Published (Jul 2017) | [82] |
| 13 | STREAM | Each cell is scored with chromVAR to calculate k-mer z-scores. Pseudotemporal tree structure is learnt by ElPiGraph. | Constructing and visualizing developmental trajectories and identifying key regulators in a user-friendly fashion. | Same as chromVAR as k-mer z-scores are calculated using its *matchKmers* function. | Docker/Python/R | Web (stream.pinell olab.org) and GitHub | Preprint (Apr 2018) | [71] |
| 14 | ChromA | De-noising sparse sampling biases by employing a Bayesian statistical model to impute chromatin accessibility. | Improves power for cell state identification, enables transcription factor footprinting, and recovers less prominent peaks. Overall improved signal-to-noise ratio. | Consensus annotations are only as strong as the training replicates. Probabilistic models can be further refined by pooling chromatin accessibility information from more sources as they become available. | Python | GitHub | Preprint (Mar 2019) | [73] |
| 15 | Garnett | Uses markup language (that use tags to define gene expression features) for classifying cell types based on signature markers. | Systematic annotation of cell types. The use of hierarchical markup language enables cell subtype delineation as well. | The efficiency of pre-trained classifiers is limited by molecularly defined cell types it catalogs. | R | GitHub | Preprint (Feb 2019) | [74] |
| 16 | SnapATAC | Clusters cells based on genome-wide chromatin accessibility profile and therefore does not rely on pre-defined, population-based peaks annotations. | Outperformed other methods for separating heterogenous cell populations and for analyzing large (> 80,000 cells) datasets. As well, can identify cis-regulatory sequences active in rare cell types. | None reported. Highly recommended for separating cell populations and for examining large (*i.e.*, sci-ATAC-seq mouse dataset) without down sampling within reasonable time frames (~ 700 minutes for sci-ATAC-seq). | R | GitHub | Preprint (May 2019) | [38,100,101] |

**Table 2    Summary of open source bioinformatic tools for scATAC-seq analysis**

| No. | Tool | Algorithm | Key feature | Limitation | Platform | Availability | Status | Refs. |
|---|---|---|---|---|---|---|---|---|
| 17 | Signac | Extends scRNA-seq package Seurat to identify cell type-specific chromatin accessibility following scRNA/ATAC integration. | Enables visualization of 'pseudo-bulk' genome tracks and enriched DNA sequence motif. As well, has a wrapper to run chromVAR[72] to calculate and visualize per-cell motif activity. | Currently under active development. | R | Web (satijalab.org/signac/) | Unpublished | N/A |
| **Integrated multi-omic analysis toolkits** | | | | | | | | |
| 18 | Coupled NMF | Gene expression-based clustering and chromatin accessibility-based clustering is optimized iteratively according to a pre-defined coupling matrix. | Integrated cell clustering by combining single-cell RNA and ATAC-seq measurements. Integrated clustering was found to be superior to clustering from just one data type. | The coupling matrix provided is fixed to a finite number of predicted genes based on prior bulk datasets. | Python | GitHub | Published (Jul 2018) | [86] |
| 19 | SOM | Self-organizing map, a type of artificial neural network, that uses unsupervised learning to perform dimensionality reduction. Here, SOMs are independently created from sc-ATAC-seq and sc-RNA-seq data and then linked using a linking function. | Linked SOMs can reveal corroborating changes in gene expression and chromatin accessibility that jointly change, providing a more thorough regulatory map. | Since SOMs are calculated separately and then linked, this assumes both measurements have a similar power for independently detecting cell types/states. | C++ and R | GitHub | Preprint (Oct 2018) | [85] |
| 20 | Seurat (v3) | Measurements from different modalities are compared using canonical correlation analysis to identify shared biological features termed "anchors". Weighted anchors are used to calculate correction vectors for each cell to enable joint dimensionality reduction. | Single-cell data from diverse technologies can be integrated and compared to generate harmonized atlases. | Cell populations that are uniquely present in one dataset can form incorrect anchors with closely resembling cell states in the reference dataset. | R | CRAN, GitHub | Preprint (Nov 2018) | [84] |
| 21 | MATCHER | Employs manifold alignment to calculate correlations between scRNA and scATAC-seq data. Manifold representations of both datasets are corrected such that they can be plotted on a comparable pseudotime scale. | MATCHER can reveal sequential transcriptomic and epigenetic changes underlying cell trajectories. As well, it can probe correlations between different genomic layers (*i.e.*, transcriptome, chromatin accessibility, and histone modification). | A key assumption is that the cell trajectory assessed is linear and monotonic (progresses unidirectionally). | Python | GitHub | Published (July 2017) | [107] |
| 22 | LIGER | Employs integrative matrix factorization to identify shared and non-shared metagenes between two single-cell datasets. | This platform can be adopted to integrate scATAC-seq datasets with scRNA-seq, spatial transcriptomics, *etc.* | User-specified parameters in join factorizations can alter the weight assigned to dataset-specific metagenes. While this can enable integration of divergent datasets, it can also lead to spurious integration where cell types don't corroborate across datasets. | R | GitHub | Preprint (November 2018) | [105] |
| **Chromatin accessibility inference from single-cell transcriptomics** | | | | | | | | |
| 23 | BIRD | Predicts chromatin accessibility and TFBSs using small-cell-number and/or pooled single-cell RNA-seq datasets. | BIRD-derived chromatin accessibility predictions based on pooling scRNA-seq profiles from 10 or more cells were superior to those obtained from pooling single-cell ATAC/DH profiles. | Predictive potential is only as robust as the number and quality of training datasets.<br><br>Currently limited to human and rodent datasets. | C++ and R | GitHub | Preprint (Jan 2016) | [114] |

*Note*: CRAN, comprehensive R archive network; scABC, single cell Accessibility Based Clustering; SOM, self-organizing maps; coupled NMF, coupled nonnegative matrix factorizations; STREAM, single-cell trajectories reconstruction, exploration and mapping; Scasat, single-cell ATAC-seq analysis tool; Destin, detection of cell-type specific difference in chromatin accessibility; SCRAT, single cell R analysis toolkit; BROCKMAN, brockman representation of chromatin by K-mers in mark-associated nucleotides; SCENIC, single-cell rEgulatory network inference and clustering; TFBS, transcription factor binding site; MATCHER, manifold alignment to CHaracterize experimental relationship; BIRD, Big data regression for predicting DNase I hypersensitivity.

82], 2) identify *cis*-regulatory regions and their target genes [76,83], and impressively, 3) integrate single-cell chromatin dataset with other single-cell measurements, such as scRNA-seq [84–86].

## Bioinformatics workflow for scATAC-seq analysis

A full comparative overview of bioinformatic tools, either published through a peer-reviewed stream (9/19) or uploaded on preprint servers (10/19) that are accessible through open code repositories are described in Table 2. Briefly, scATAC-seq processing constitutes the following five sequential steps.

### Processing raw sequencing data

This includes trimming sequencing adapters, eliminating poor quality reads, mapping paired-end reads, eliminating cells with a library size that falls below a chosen threshold, and aggregating cell barcodes for downstream processing. For users employing a commercial platform like 10X Genomics, accompanying analysis pipelines such as CellRanger-ATAC (https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac) may readily perform these steps.

### Pre-defined feature selection

Typically, the complete list of chromosomal locations is not inspected for differential accessibility. Instead, the overall variance in a dataset is reduced by restricting the analysis to pre-defined motifs or a list of genomic/gene set annotations. For example, chromVAR, one of the first scATAC-seq tools developed by the Greenleaf lab, aggregated reads based on known TF motifs. This aggregation converted sparse accessibility by cell barcode matrix to a more stable, bias-corrected deviation per TF motif matrix, enabling cell-to-cell similarity measurements by analyzing gain or loss of accessibility within defined genomic features [69].

### Heterogeneity calculation of single cells

Using the filtered accessibility–barcode matrix, a dissimilarity measurement that quantitates the extent of divergence between two cells is calculated based on differential peak accessibility. For example, Scasat employs Jaccard distance as a dissimilarity index [87] by identifying shared and distinct peaks. SCRAT first aggregates co-activated sites to derive pathway level accessibility. It then uses pathway-level accessibility as a feature to compute cell dissimilarity. It is shown that aggregation can increase signal-to-noise ratio and lead to improved dissimilarity measures that better separate cells by cell type [75].

### Dimensionality reduction

Since the accessibility matrix is a high-dimensional dataset, reducing its dimensionality, such that biological variance is retained but random variables contributing to noise are eliminated, is critical for downstream analysis and visualization. Multidimensional scaling (MDS) [88], t-distributed stochastic neighbor embedding (tSNE) [89], and Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) [90] are common dimensionality reduction techniques where the number of dimensions included is user-specified and depends on the overall complexity of the dataset.

### Cell clustering and differential accessibility analysis

Based on heterogeneity in chromatin accessibility, clustering algorithms modified for scATAC-seq (*i.e.*, k-medoid giving higher weight to cells with higher sequencing coverage [77]) hierarchically cluster cells into different groups. A number of different statistical tests (*i.e.*, Fisher exact, Information gain) can then be employed to calculate differential accessibility signatures between two clusters. A more targeted two-sided t-test can be used to determine whether a particular motif is differentially accessible between two clusters.

Extending differential accessibility measurements, a number of tools have started to impart novel perspectives on the chromatin regulatory landscape. Below, we highlight three particularly impressive approaches that enable: 1) mapping co-accessible DNA elements to connect regulatory sequences with its targets, 2) predicting cell-type specific transcription factor activity, and 3) integrating scATAC-seq data with scRNA-seq, spatial transcriptomics, and other single-cell measurements. We conclude by highlighting that methods capable of predicting chromatin status from scRNA-seq datasets will further improve the power of both techniques and enable superior integration across single-cell modalities.

## Predicting DNA interactions by mapping chromatin co-accessibility

The mechanisms by which *cis*-regulatory sequences engage in long-range interactions to recruit transcriptional machinery and control gene expression is highly cell type-, state-, and context-specific [91,92]. Indeed, such widespread and context-dependent enhancer–promoter associations pose a significant challenge to generating computational models of gene expression and the way they change as a result of perturbation [66]. Although long-range contacts can be quantified in cell ensembles using structural genomic approaches, such as ChIA-PET [93], Hi-C [94], and HiChIP [95,96], there are no methods to infer this in single cells. Cicero, a software developed in the Shendure/Trapnell labs, seeks to infer these interactions by mapping the co-accessibility of promoters with other regulatory sites. This infers that the *cis*-regulatory landscapes that enact a given homeostatic gene program enable tracking changes in this landscape over pseudotime by modifying trajectory inference algorithms developed for Monocle2 [97,98]. As a proof-of-concept, Cicero's application to myoblast differentiation revealed that while accessible promoter analysis explained 17% of the variance, the addition of promoter–distal site linkages explained 2.27-fold higher variance in expression. Extending this idea further, sequences that are in close physical proximity, share common TF targets, and undergo synchronized changes in accessibility over pseudotime may qualify as "chromatin hubs" [99], and their modulation could preferentially bias a cell's fate. Tracking chromatin co-accessibility enables interrogating fundamental biological questions in a new light, including the fate reversion question posed in the beginning of this review. Knowing the extent to which reacquisition of embryonic-exclusive TFs are driven by interactions with the same distal regulatory sequences after injury, and

their accessibility/co-accessibility status in the uninjured adult tissue, will be critical for assessing whether a cell remembers its lineage history and harbors a latent capacity to navigate Waddington's valley in a reverse direction.

**Predicting cell type-specific TF function**

An emerging concept in single-cell regulomics is that expression, activation, and function of any particular TF is determined by the complete composition of co-factors active at a given time. As such, methods that use differences in features for cell grouping fail to appreciate how the presence or absence of other factors may impact the function of each feature used. To circumvent this and to improve prediction of the cell type-specific impact of a TF, Aerts group's cisTopic [83] employs topic modelling, where co-accessible regions are assigned to one or multiple *regulatory topics* with varying strengths, and these topics (and not its constituent features) are then used for cell grouping. This approach revealed an interesting feature about the *SOXE* TFs: while a core *SOXE* topic (*SOX9/10*) was robustly present in a variety of cell types (*i.e.*, melanocytes, melanoma, oligodendrocytes, astrocytes), its predicted motif and co-regulatory factors were cell-type and cell-state dependent. These observations were explained by a model where *SOXE* proteins function as pioneering TFs that unmask chromatin domains and recruit cell type-specific co-factors (*i.e.*, *OLIG* in oligodendrocytes) to enforce somatic cellular programs. A recent assessment comparing ten scATAC tools on a standardized benchmarking platform revealed that, while *cis*Topic was slightly more sensitive to noise, along with SnapATAC [100] and *Cusanovich2018* [38], it outperformed other methods used for separating cell types/states [101]. Revisiting the discussion on the epigenetic basis of injury-induced fate reversions, it can be concluded that it is possible, at least in theory, to carry out an assessment of somatic cells' propensity to activate pioneering TFs (perhaps by alleviating repression that prevents their function during adulthood) that subsequently recruit lineage-specific TFs to reenact development.

## Leveraging strengths of different single-cell modalities

Understanding biochemical mechanisms that endow cell type-specific functions requires comprehensive integration of modalities, measuring distinct cell states and purposeful use of one measurement to inform interpretations of another. For example, out of all the single-cell measurement assays, transcriptomics is the most powered for detecting cell types, yet it is devoid of positional context of each cell in a three-dimensional tissue. While spatial transcriptomics can resolve the distribution of transcripts, limitations in how closely positional barcodes (or barcoded beads) can be placed precludes single-cell [102,103] or subcellular resolution [104]. Similarly, while scATAC-seq profiles chromatin accessibility, and with inclusion of mitochondrial reads enables robust lineage reconstruction [51,52], its sparse and binarized nature derives statistical strength from cluster (not single-cell) analysis. As such, algorithms that integrate each modality to generate harmonized atlases add significant value.

To this end, algorithms have been devised to perform join dimensionality reduction to compare two single-cell datasets. In Seurat Version 3 [84], Linked Inference of Genomic Exper-

imental Relationships (LIGER) [105], a concept similar to mutual nearest neighbors is applied, where each cell in the first dataset (say scATAC-seq) is queried against all cells in a second dataset (say scRNA-seq) to identify its closest neighbor across datasets [106]. This is subsequently repeated, but now starting from the second (scRNA-seq) dataset where each cell is queried against cells in the first (scATAC-seq) dataset. If both analyses identify a mutual pair, these two separate measurements are thought to be taken from the same underlying cell state and the pairs are integrated. MATCHER is another tool that integrates data types by performing the alignment of pseudotime from transcriptome and epigenome data, although it works well on trajectories that are linear rather than branched [107]. A similar but more generalizable approach utilizes self-organizing maps (SOMs) [85], a type of artificial neural network that is trained to generate low-dimensional representations of either scRNA-seq or scATAC data. Resulting SOMs for each data type can then be used to identify groups of co-expressed genes and open chromatin elements in order to identify those that jointly change across datasets. Advantages of this approach include the ability to analyze time series and a lack of reliance on inferred pseudotime. While benchmarking studies have compared scRNA-seq analysis methods [108] (*i.e.*, tools for Gene Regulatory Network construction [109,110]) and a recent study extended this approach to compare ten scATAC tools [101], none have systematically compared data integration algorithms. Parameters that will need to be assessed using standardized benchmarking frameworks include joint dimensionality reduction/clustering, label transfer, resistance to variable sequencing coverage, impact of changing signal-to-noise ratios, as well as practical considerations such as running times and CPU/memory requirements especially when considering integration of large (*i.e.*, > 80,000 single-cell in sci-ATAC-seq mouse atlas [38]) datasets across multiple modalities.

An essential prerequisite to integration is sufficient batch correction of all datasets being integrated [111], especially if the variance introduced due to batch is comparable to variance across cell types/states. An ever-evolving challenge continues to be the need to improve molecular methods to recover greater percentage of the data for each cell and to develop imputation and predictive methods, both of which can increase statistical power for integration. Together, robust integrative analyses will enable the transfer of biological and computational phenotypes from one dataset to another. For example, a lineage trajectory predicted based on RNA splicing kinetics from transcriptomics [112] can be directly compared with one delineated based on mtDNA mutations from accessible chromatin [51,52]. Such analyses will provide unprecedented insights and inform both the biology and the bioinformatic tools employed to decipher it.

## Predicting chromatin accessibility from transcriptomics

Single-cell RNA-seq is currently the most widely used single cell functional genomic technology. Most scRNA-seq datasets are generated without accompanying scATAC-seq data. Conventionally, scRNA-seq is only used for analyzing transcriptome but not chromatin accessibility. However, a recent study from the Ji lab has shown that chromatin accessibility can be predicted using scRNA-seq (**Figure 5**) [113]. They used

DNase-seq and RNA-seq samples from matching cell types generated by the ENCODE and Roadmap Epigenomics projects to train prediction models, leveraging the recently developed Big Data Regression for Predicting DNase I hypersensitivity (BIRD) algorithm [114]. The trained prediction models are then applied to new bulk or single-cell RNA-seq samples to predict genome-wide chromatin accessibility (Figure 5A). Interestingly, using bulk DNase-seq data from the corresponding cell type [115] as 'ground truth', BIRD prediction using low-input RNA-seq data from 30 and 100 cells outcompeted ATAC-seq signals from 500 cells. Moreover, compared to scATAC-seq performed using the same number of cells, chromatin accessibility predicted by BIRD from scRNA-seq is more continuous, less sparse, and can more accurately reconstruct gold standard bulk DNase-seq and bulk ATAC-seq profiles in the same cell type (Figure 5B).

This prediction approach has at least three potential applications. First, when an investigator only has scRNA-seq data but not scATAC-seq data, one can use scRNA-seq to predict the chromatin accessibility landscape and *cis*-regulatory element activities. Because the vast majority of single-cell genomic data are scRNA-seq, prediction may substantially increase the value of both existing and new scRNA-seq experiments. Second, when scATAC-seq data are available, chromatin accessibility predicted by scRNA-seq can be used as pseudo-replicates to improve signal quality. It has been shown that combining these pseudo-replicates with experimental scATAC-seq data can tremendously increase the signal-to-noise ratio and the accuracy for inferring *cis*-regulatory element activities [113]. Third, when both scRNA-seq and scATAC-seq data are available, prediction may be used as a bridge to align and integrate these two distinct data types. Thus, predicting
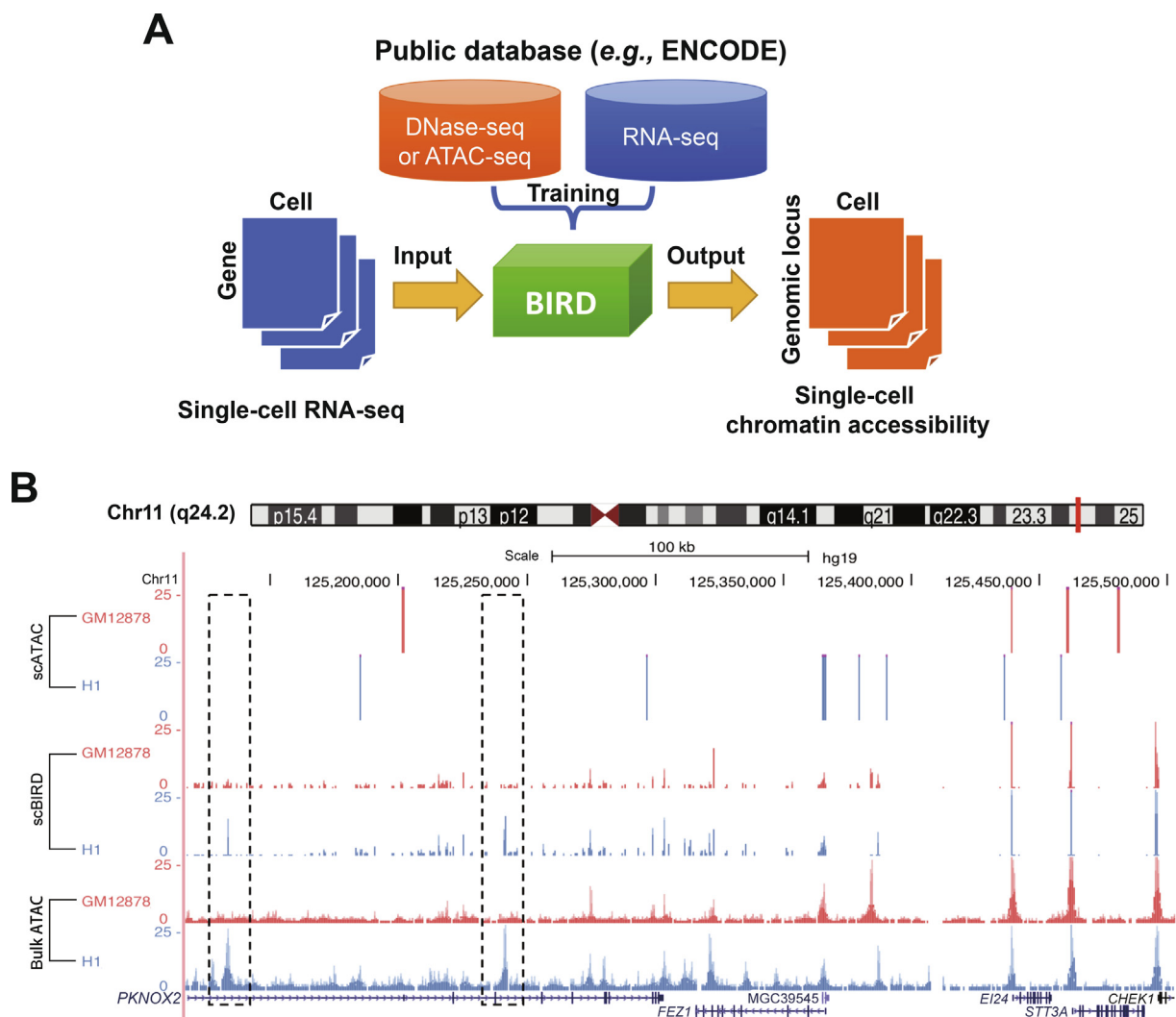


**Figure 5  Predicting chromatin accessibility from single-cell transcriptomics**
**A.** Overview of the prediction approach. **B.** BIRD-predicted chromatin accessibility, experimental scATAC-seq data, and bulk ATAC-seq data for two cell types (GM12878 and H1) are compared in a sample genomic region. The scATAC-seq signals are sparse and discrete, while BIRD-predicted signals are more continuous and correlate better with the bulk ATAC-seq signals. BIRD, Big Data Regression for predicting DNase I hypersensitivity.

chromatin accessibility using scRNA-seq offers a new tool for single-cell regulome analysis. It is useful both when scATAC-seq data is unavailable and when it is available.

One limitation of prediction is that it requires training data. Currently, only human and mouse have matched chromatin accessibility and transcriptome data from a broad spectrum of cell types that may support training prediction models. Most other species do not have enough training data. However, as more data become available, this limitation will eventually be resolved in the future. Moreover, most of today's scRNA-seq data (94% of scRNA-seq samples in SRA) are from human and mouse. Thus, prediction in these two species alone would be tremendously useful. Overall, predicting chromatin accessibility from scRNA-seq is a new area that has yet to be fully explored. There is plenty of room to improve prediction accuracy. New methods that can better handle technical biases in single-cell data and more effectively train prediction models for ultra-high-dimensional responses (*i.e.*, genome-wide chromatin accessibility) using ultra-high-dimension predictors (*i.e.*, transcriptome) could make this approach more powerful in the future.

## Conclusion

How organisms achieve a division of labor amongst trillions of cells that enable emergent properties like consciousness to arise has intrigued philosophers for centuries. Scientists have since appreciated that specialized cell function, which forms the basis of division in labor, is the final product of many layers of combinatorial control. Single-cell genomics has started to impart a detailed glimpse of the epigenetic landscapes that maintain this specialized cell function and has started to correlate these landscapes with distinct gene and protein expression outputs. Methods that assess chromatin accessibility at a single cell resolution are becoming powerful tools for inferring this epigenetic landscape and for asking how it changes following perturbation, injury, disease, or more gradually over micro/macro-evolutionary timescales. We predict that its applications will impart novel insights into important questions, like: Do somatic cells retain an epigenetic memory of their lineage? If so, how do they recall the trajectories they traversed to adopt a final fate? Such memories may explain their latent capacity to revert back to an embryonic stage and reactivate tissue-specific developmental programs after sustaining an injury [22–26]. Understanding how specialized cells maintain such memories can inform drug treatments that purposefully activate regenerative programs to improve the quality of healing and restore native tissue function.

Single-cell chromatin measurements can also unearth novel dynamics in non-coding regions that underlie changes over evolutionary or disease processes. Comparative epigenomic studies have implicated changes in chromatin accessibility in driving *cis*-regulatory evolution across different species [116,117]. The ability to profile cell-to-cell divergences in chromatin landscapes can enable "cellular anthropology" with unprecedented resolution and track how noncoding regulatory elements change with the emergence of new cell types and species. Indeed, integration of single-cell transcriptomic and chromatin accessibility data from rodents has revealed that enhancers may facilitate modifications to existing gene regulatory networks by integrating *de novo* genes transcribed from regions of the genome that were previously non-coding. Such wiring and rewiring of gene regulatory networks may explain phenomena like the birth of *de novo* gene sets across eumetazoan evolution [118]. In the context of human diseases, the majority of pathogenic variants lie within non-coding functional elements that lead to gene dysregulation. Since multiple layers of the epigenome are dysfunctional in multiple cell types, leading to tumor initiation and progression [119,120], integration of scATAC-seq with low-input or single-cell ChIP-seq for selected TFs/histone modifications may be an attractive approach. Simultaneous measurements of TF binding and chromatin accessibility enabled by 'ChIPmentation' [121] or 'CUT&Tag' [122] can reveal definitive binding sites of key TFs (eliminating reliance on motif inferences or deciphering ATAC TF footprints). Since 'CUT&Tag' relies on tethering pA-Tn5 transposomes to antibodies bounded to a target chromatin protein, it is well suited for low cell inputs and can be configured with commercial single-cell platforms, such as 10X Genomics droplet separation system (Figure 2B) or SMARTer ICELL8 nano-dispensation system (Figure 2D) [122]. Alternatively, orthogonal single-cell ChIP and ATAC measurements assaying the same underlying cell types/states may be bioinformatically integrated by identifying shared biological states, such that information captured by one technique may inform the interpretation of another. Notwithstanding these expectations, single-cell chromatin accessibility profiling will lead to many unexpected discoveries, each revealing mechanisms of gene regulation that underlie the remarkable cell diversity and plasticity seen in complex tissues.

## CRediT author statement

**Sarthak Sinha:** Conceptualization, Writing - original draft, Visualization, Writing - review & editing. **Ansuman T. Satpathy:** Writing - review & editing. **Weiqiang Zhou:** Writing - original draft, Writing - review & editing, Visualization. **Hongkai Ji:** Writing - original draft, Writing - review & editing, Visualization. **Jo A. Stratton:** Writing - review & editing. **Arzina Jaffer:** Visualization, Writing - review & editing. **Nizar Bahlis:** Writing - review & editing, Supervision. **Sorana Morrissy:** Writing - review & editing, Supervision. **Jeff A. Biernaskie:** Writing - review & editing, Supervision, Funding acquisition. All authors read and approved the final manuscript.

## Competing interests

ATS is an advisor to Immunai. There are no other conflicts of interest to be declared.

## Acknowledgments

## ORCID

0000-0002-0770-3150 (Sarthak Sinha)
0000-0002-5167-537X (Ansuman T. Satpathy)
0000-0001-9396-9571 (Weiqiang Zhou)
0000-0002-6480-0141 (Hongkai Ji)
0000-0002-1205-1353 (Jo A. Stratton)
0000-0002-9197-9060 (Arzina Jaffer)
0000-0001-7353-7034 (Nizar Bahlis)
0000-0001-8526-562X (Sorana Morrissy)
0000-0002-0633-0698 (Jeff A. Biernaskie)

## References

[1] Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 2014;343:776–9.

[2] Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, et al. Single-cell reconstruction of the early maternal–fetal interface in humans. Nature 2018;563:347–53.

[3] Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol 2018;18:35.

[4] Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 2015;161:1202–14.

[5] Stratton JA, Sinha S, Shin W, Labit E, Chu TH, Shah PT, et al. Droplet barcoding-based single cell transcriptomics of adult mammalian tissues. J Vis Exp 2019;143:e58709.

[6] Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. Cell 2016;166:1308–23.

[7] Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, et al. Molecular architecture of the mouse nervous system. Cell 2018;174:999–1014.

[8] Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature 2018;562:367.

[9] Plass M, Solana J, Alexander Wolf F, Ayoub S, Misios A, Glažar P, et al. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. Science 2018;360: eaaq1723.

[10] Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, et al. The *Drosophila* embryo at single-cell transcriptome resolution. Science 2017;358:194–9.

[11] Fincher CT, Wurtzel O, de Hoog T, Kravarik KM, Reddien PW. Cell type transcriptome atlas for the planarian Schmidtea mediterranea. Science 2018;360:eaaq1736.

[12] Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 2017;357:661–7.

[13] Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science 2018;360:eaar3131.

[14] Hammond TR, Dufort C, Dissing-Olesen L, Giera S, Young A, Wysoker A, et al. Single-cell RNA sequencing of microglia throughout the mouse lifespan and in the injured brain reveals complex cell-state changes. Immunity 2019;50:253–71.

[15] Waddington CH. The strategy of the genes. A discussion of some aspects of theoretical biology. London: Allen & Unwin; 1957.

[16] Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10:1213.

[17] Adey A, Morrison HG, Asan XX, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. Genome Biol 2010;11:R119.

[18] Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell 2008;132:311–22.

[19] Chen T, Dent SYR. Chromatin modifiers and remodellers: regulators of cellular differentiation. Nat Rev Genet 2014;15:93–106.

[20] Stergachis AB, Neph S, Reynolds A, Humbert R, Miller B, Paige SL, et al. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell 2013;154:888–903.

[21] Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. Science 2018;362:eaav1898.

[22] Ito M, Yang Z, Andl T, Cui C, Kim N, Millar SE, et al. Wnt-dependent *de novo* hair follicle regeneration in adult mouse skin after wounding. Nature 2007;447:316–20.

[23] Wang X, Hsi TC, Guerrero-Juarez CF, Pham K, Cho K, McCusker CD, et al. Principles and mechanisms of regeneration in the mouse model for wound-induced hair follicle neogenesis. Regeneration 2015;2:169–81.

[24] Ransom RC, Carter AC, Salhotra A, Leavitt T, Marecic O, Murphy MP, et al. Mechanoresponsive stem cells acquire neural crest fate in jaw regeneration. Nature 2018;563:514.

[25] Carr MJ, Toma JS, Johnston APW, Steadman PE, Yuzwa SA, Mahmud N, et al. Mesenchymal precursor cells in adult nerves contribute to mammalian tissue repair and regeneration. Cell Stem Cell 2019;24:240–56.

[26] Nusse YM, Savage AK, Marangoni P, Rosendahl-Huber AKM, Landman TA, De Sauvage FJ, et al. Parasitic helminths induce fetal-like reversion in the intestinal stem cell niche. Nature 2018;559:109–13.

[27] Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 2015;523:486–90.

[28] Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 2015;348:910–4.

[29] Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, Weitz DA, et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat Biotechnol 2015;33:1165–72.

[30] Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 2013;502:59–64.

[31] Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, et al. Massively multiplex single-cell Hi-C. Nat Methods 2017;14:263–6.

[32] Jin W, Tang Q, Wan M, Cui K, Zhang Y, Ren G, et al. Genome-wide detection of DNase i hypersensitive sites in single cells and FFPE tissue samples. Nature 2015;528:142–6.

[33] Adli M, Zhu J, Bernstein BE. Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. Nat Methods 2010;7:615.

[34] Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, et al. Chromatin state dynamics during blood formation. Science 2014;345:943–9.

[35] Cao Z, Chen C, He B, Tan K, Lu C. A microfluidic device for epigenomic profiling using 100 cells. Nat Methods 2015;12:959.

[36] Ma S, Hsieh YP, Ma J, Lu C. Low-input and multiplexed microfluidic assay reveals epigenomic variation across cerebellum and prefrontal cortex. Sci Adv 2018;4:eaar8187.

[37] Grosselin K, Durand A, Marsolier J, Poitou A, Marangoni E, Nemati F, et al. High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. Nat Genet 2019;51:1060.

[38] Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, et al. A single-cell atlas of *in vivo* mammalian chromatin accessibility. Cell 2018;174:1309–24.

[39] Goryshin IY, Reznikoff WS. Tn5 *in vitro* transposition. J Biol Chem 1998;273:7367–74.

[40] Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol 2019;37:925–36.

[41] Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol 2011;12:R67.

[42] Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 2017;14:959.

[43] Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet 2019;20:207–20.

[44] Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, et al. The *cis*-regulatory dynamics of embryonic development at single-cell resolution. Nature 2018;555:538–42.

[45] Preissl S, Fang R, Huang H, Zhao Y, Raviram R, Gorkin DU, et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat Neurosci 2018;21:432–9.

[46] Spektor R, Yang JW, Lee S, Soloway PD. Single cell ATAC-seq identifies broad changes in neuronal abundance and chromatin accessibility in Down Syndrome. bioRxiv 2019;561191.

[47] Mezger A, Klemm S, Mann I, Brower K, Mir A, Bostick M, et al. High-throughput chromatin accessibility profiling at single-cell resolution. Nat Commun 2018;9:3647.

[48] Chen X, Shen Y, Draper W, Buenrostro JD, Litzenburger U, Cho SW, et al. ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. Nat Methods 2016;13:1013.

[49] Lareau CA, Duarte FM, Chew JG, Kartha VK, Burkett ZD, Kohlway AS, et al. Droplet-based combinatorial indexing for massive scale single-cell epigenomics. Nat Biotechnol 2019;37:916–24.

[50] Montefiori L, Hernandez L, Zhang Z, Gilad Y, Ober C, Crawford G, et al. Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. Sci Rep 2017;7:2451.

[51] Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. Cell 2019;176:1325–39.

[52] Xu J, Nuno K, Litzenburger UM, Qi Y, Corces MR, Majeti R, et al. Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. Elife 2019;8:e45105.

[53] Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell 2016;167:1853–66.

[54] Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. Cell 2016;167:1867–82.

[55] Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al. Pooled CRISPR screening with single-cell transcriptome readout. Nat Methods 2017;14:297.

[56] Rubin AJ, Parker KR, Satpathy AT, Qi Y, Wu B, Ong AJ, et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. Cell 2019;176:361–76.

[57] Satpathy AT, Saligrama N, Buenrostro JD, Wei Y, Wu B, Rubin AJ, et al. Transcript-indexed ATAC-seq for precision immune profiling. Nat Med 2018;24:580.

[58] Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat Med 2019;25:1251–9.

[59] Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol 2017;18:83.

[60] Niu M, Tabari E, Ni P, Su Z. Towards a map of *cis*-regulatory sequences in the human genome. Nucleic Acids Res 2018;46:5395–409.

[61] Mathelier A, Shi W, Wasserman WW. Identification of altered *cis*-regulatory elements in human disease. Trends Genet 2015;31:67–76.

[62] Luizon MR, Ahituv N. Uncovering drug-responsive regulatory elements. Pharmacogenomics 2015;16:1829–41.

[63] Liu L, Liu C, Quintero A, Wu L, Yuan Y, Wang M, et al. Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. Nat Commun 2019;10:1410.

[64] Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science 2018;361:1380–5.

[65] Chen X, Litzenburger UM, Wei Y, Schep AN, LaGory EL, Choudhry H, et al. Joint single-cell DNA accessibility and protein epitope profiling reveals environmental regulation of epigenomic heterogeneity. Nat Commun 2018;9:4590.

[66] Packer J, Trapnell C. Single-cell multi-omics: an engine for new quantitative models of gene regulation. Trends Genet 2018;34:653–65.

[67] Kelley DR, Snoek J, Rinn JL. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. Genome Res 2016;26:990–9.

[68] Pott S, Lieb JD. Single-cell ATAC-seq: strength in numbers. Genome Biol 2015;16:172.

[69] Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 2017;14:975–8.

[70] de Boer CG, Regev A. BROCKMAN: deciphering variance in epigenomic regulators by k-mer factorization. BMC Bioinformatics 2018;19:253.

[71] Chen H, Albergante L, Hsu JY, Lareau CA, Bosco G Lo, Guan J, et al. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat Commun 2019;10:1903.

[72] Sinnamon JR, Torkenczy KA, Linhoff MW, Vitak SA, Pliner HA, Trapnell C, et al. The accessible chromatin landscape of the murine hippocampus at single-cell resolution. Genome Res 2019;29:857–69.

[73] Gabitto MI, Rasmussen A, Wapinski O, Allaway K, Carriero N, Fishell GJ, et al. Characterizing the epigenetic landscape of cellular populations from bulk and single-cell ATAC-seq information. bioRxiv 2019;567669.

[74] Pliner HA, Shendure J, Trapnell C. Supervised classification enables rapid annotation of cell atlases. Nat Methods 2019;16:983–6.

[75] Ji Z, Zhou W, Ji H. Single-cell regulome data analysis by SCRAT. Bioinformatics 2017;33:2930–2.

[76] Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, et al. Cicero predicts *cis*-regulatory DNA interactions from single-cell chromatin accessibility data. Mol Cell 2018;71:858–71.

[77] Zamanighomi M, Lin Z, Daley T, Chen X, Duren Z, Schep A, et al. Unsupervised clustering and epigenetic classification of single cells. Nat Commun 2018;9:1–8.

[78] Ulirsch JC, Lareau CA, Bao EL, Ludwig LS, Guo MH, Benner C, et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. Nat Genet 2019;51:683–93.

[79] Urrutia E, Chen L, Zhou H, Jiang Y. Destin: toolkit for single-cell analysis of chromatin accessibility. bioRxiv 2018;461905.

[80] Baker SM, Rogerson C, Hayes A, Sharrocks AD, Rattray M. Classifying cells with Scasat, a single-cell ATAC-seq analysis tool. Nucleic Acids Res 2019;47:e10.

[81] Cai S, Georgakilas GK, Johnson JL, Vahedi G. A cosine similarity-based method to infer variability of chromatin accessibility at the single-cell level. Front Genet 2018;9:319.

[82] Zhao C, Hu S, Huo X, Zhang Y. Dr.seq2: a quality control and analysis pipeline for parallel single cell transcriptome and epigenome data. PLoS One 2017;12:e0180583.

[83] González-Blas CB, Minnoye L, Papasokrati D, Aibar S, Hulselmans G, Christiaens V, et al. cisTopic: *cis*-regulatory topic modeling on single-cell ATAC-seq data. Nat Methods 2019;16:397–400.

[84] Stuart T, Butler A, Hoffman P, Hafemeister C, Iii WMM, Stoeckius M, et al. Comprehensive integration of single-cell data. Cell 2019;177:1888–902.

[85] Jansen C, Ramirez R, El-Ali N, Gomez-Cabrero D, Tegner J, Merkenschlager M, et al. Building gene regulatory networks from single-cell ATAC-seq and RNA-seq using Linked Self-Organizing Maps. PLoS Comput Biol 2019;15:e1006555.

[86] Duren Z, Chen X, Zamanighomi M, Zeng W, Satpathy AT, Chang HY, et al. Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations. Proc Natl Acad Sci U S A 2018;115:7723–8.

[87] Jaccard P. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. Bull Soc Vaudoise Sci Nat 1901;37:241–72.

[88] Borg I, Groenen P. Modern multidimensional scaling: Theory and applications. J Educ Meas 2003;40:277–80.

[89] van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9:2579–605.

[90] McInnes L, Healy J, Melville J. Umap: uniform manifold approximation and projection for dimension reduction. arXiv 2018;1802.03426.

[91] Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 2012;148:84–98.

[92] Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, et al. Chromatin connectivity maps reveal dynamic promoter–enhancer long-range associations. Nature 2013;504:306.

[93] Li X, Luo OJ, Wang P, Zheng M, Wang D, Piecuch E, et al. Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. Nat Protoc 2017;12:899.

[94] Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 2014;159:1665–80.

[95] Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods 2016;13:919.

[96] Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, Cho SW, et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat Genet 2017;49:1602.

[97] Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 2017;14:979.

[98] Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mRNA quantification and differential analysis with Census. Nat Methods 2017;14:309.

[99] de Laat W, Grosveld F. Spatial organization of gene expression: the active chromatin hub. Chromosom Res 2003;11:447–59.

[100] Fang R, Preissl S, Hou X, Lucero J, Wang X, Motamedi A, et al. Fast and accurate clustering of single cell epigenomes reveals *cis*-regulatory elements in rare cell types. bioRxiv 2019;615179.

[101] Chen H, Lareau C, Andreani T, Vinyard ME, Garcia SP, Clement K, et al. Assessment of computational methods for the analysis of single-cell ATAC-seq data. bioRxiv 2019;739011.

[102] Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science 2016;353:78–82.

[103] Moor AE, Itzkovitz S. Spatial transcriptomics: paving the way for tissue-level systems biology. Curr Opin Biotechnol 2017;46:126–33.

[104] Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. Science 2019;363:1463–7.

[105] Welch J, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko E. Integrative inference of brain cell similarities and differences from single-cell genomics. bioRxiv 2018;459891.

[106] Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 2018;36:421.

[107] Welch JD, Hartemink AJ, Prins JF. MATCHER: manifold alignment reveals correspondence between single cell transcriptome and epigenome dynamics. Genome Biol 2017;18:138.

[108] Tian L, Dong X, Freytag S, Lê Cao KA, Su S, JalalAbadi A, et al. Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. Nat Methods 2019;16:479–87.

[109] Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods 2020;17:147–54.

[110] Jackson CA, Castro DM, Saldi G-A, Bonneau R, Gresham D. Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. bioRxiv 2019;581678.

[111] Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018;36:411.

[112] La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature 2018;560:494.

[113] Zhou W, Ji Z, Fang W, Ji H. Global prediction of chromatin accessibility using small-cell-number and single-cell RNA-seq. Nucleic Acids Res 2019;47:e121.

[114] Zhou W, Sherwood B, Ji Z, Xue Y, Du F, Bai J, et al. Genome-wide prediction of dnase i hypersensitivity using gene expression. Nat Commun 2017;8:1038.

[115] Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57.

[116] Peng P-C, Khoueiry P, Girardot C, Reddington JP, Garfield DA, Furlong EEM, et al. The role of chromatin accessibility in *cis*-regulatory evolution. Genome Biol Evol 2019;11:1813–28.

[117] Prescott SL, Srinivasan R, Marchetto MC, Grishina I, Narvaiza I, Selleri L, et al. Enhancer divergence and *cis*-regulatory evolution in the human and chimp neural crest. Cell 2015;163:68–83.

[118] Majic P, Payne J. Enhancers facilitate the birth of *de novo* genes and their functional integration into regulatory networks. bioRxiv 2019;616581.

[119] Ting AH, McGarvey KM, Baylin SB. The cancer epigenome— components and functional correlates. Genes Dev 2006;20:3215–31.

[120] Yau HL, Ettayebi I, De Carvalho DD. The cancer epigenome: exploiting its vulnerabilities for immunotherapy. Trends Cell Biol 2019;29:31–43.

[121] Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods 2015;12:963.

[122] Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun 2019;10:1930.